MEIAT-CMAQ v1.0: A Modular Emission Inventory Allocation Tool for Community Multiscale Air Quality Model Version 1.0

Haofan Wang[1,2,3], Jiaxin Qiu[4], Yiming Liu[1,2,3,*], Qi Fan[1,2,3,*], Xiao Lu[1,2,3], Yang Zhang[5], Kai Wu[6] Ao Shen[1,2,3], Yifei Xu[1,2,3], Yinbao Jin[1,2,3], Yuqi Zhu[1,2,3], Jiayin Sun[1,2,3], Haolin Wang[1,2,3]

5   [1]School of Atmospheric Sciences, Sun Yat-sen University, and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, 519082, PR China
[2]Guangdong Provincial Field Observation and Research Station for Climate Environment and Air Quality Change in the Pearl River Estuary, Guangzhou, 510275, PR China
[3]Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, Sun Yat-sen University, Zhuhai, PR
10   China
[4]College of New Energy and Environment, Jilin University, Changchun 130012, China
[5]College of Resources and Environment, Chengdu University of Information Technology, Chengdu, 610225, China
[6]Department of Civil and Environmental Engineering, University of California, Irvine, CA 92697, USA

15   *Correspondence to*: Yiming Liu (liuym88@mail.sysu.edu.cn), Qi Fan (eesfq@mail.sysu.edu.cn)

Abstract. The Modular Emission Inventory Allocation Tool for Community Multiscale Air Quality Model (MEIAT-CMAQ) v1.0 is a resource that enables the refinement of coarse emission inventories by delivering complete temporal, species, and vertical allocations. This tool generates model-ready emission files for CMAQ, and it effectively addresses the challenges concerning the pinpointing of grid information and the spatial allocation for spatial surrogates with specific shapes. These
20   features significantly enhance the accuracy of the allocation of emissions from transportation sectors. Additionally, MEIAT-CMAQv1.0 features an efficient operational algorithm and a modular design, thus conferring greater flexibility and making it suitable for both gridded and tabulated emissions inventories. By inputting pre-assessment and post-assessment emissions separately into the CMAQ model, we observe that post-allocation inventory has a significant positive effect on both O_3 and $PM_{2.5}$ simulations. The development of MEIAT-CMAQv1.0 provides valuable insights into the automated operation of air
25   quality models and the development of emission inventory allocation tools.

1 Introduction

High-resolution emission inventories of air pollutants are essential for formulating effective atmospheric environmental mitigation strategies (Lu et al., 2023; Qiu et al., 2023; H. Wang et al., 2022), improving ambient air quality forecasts (Wu et al., 2014), and promoting air pollution research (Wang et al., 2021; Yang et al., 2020). Determining pollutant concentrations
30   necessitates employing an air quality model (AQM) (Liu and Wang, 2020a, 2020b), such as the Weather Research and Forecast (WRF)–Community Multiscale Air Quality (CMAQ) model. For an effective analysis of regional air quality using regional numerical models like WRF-CMAQ, emission inventories with a 1-5 km resolution are required. However, both "top-down" and "bottom-up" emission inventory approaches generally yield resolutions above 5 km, as exemplified by the

Multi-resolution Emission Inventory for China (MEIC, http://meicmodel.org.cn, last accessed: May 7, 2023) with a 0.25°
35   resolution (Li et al., 2017; Zheng et al., 2018), developed using the "bottom-up" methodology, and the satellite-derived NO_x
emissions (Yang et al., 2021) with a 9 km resolution, developed using the "top-down" approach.

Moreover, due to limitations in observational methods and economic constraints, emission inventory production typically
focuses on monthly or annual scales for conventional species such as NO_x, NH_3, SO_2, VOC, $PM_{2.5}$, and PM_{10}. Yet, AQM
40   demand hourly scale inventory emissions with more detailed species descriptions based on chemical mechanisms, such as
CB06 (Luecken et al., 2019) or SAPRC07 (Hutzell et al., 2012; Xie et al., 2013). This discrepancy necessitates significant
effort in adapting original emission inventories for AQM use. Consequently, there is an urgent need for a tool to facilitate
spatial, temporal, and species allocation in emission inventory processing.

45   In the recently published study by Wang et al. (2023), the authors introduced Inventory Spatial Allocation Tools version 2.0
(ISATv2.0), offering a pioneering investigation into the spatial, temporal, and species distribution of emissions. Boasting a
WRF nested domain design and advanced inventory downscaling capabilities, ISATv2.0 enables streamlined data
preparation for air quality modeling. Moreover, this novel approach employs a "sub-grid" method to expedite processing
efficiency, a crucial advancement in the field. This simplification greatly streamlines the process of preparing model-ready
50   emission files for air quality model users. However, the inherent trade-off in utilizing the sub-grid method is the reduction in
the accuracy of the tool in pinpointing grid information. Consequently, this limitation induces biases in the resulting
emission inventory, which will input into AQM. Addressing this problem, the development of a method that simultaneously
upholds distribution efficiency and accuracy, a breakthrough that holds substantial implications for advancing air quality
research.

55

In the term of spatial allocation, ISATv2.0 employs spatial surrogates to apportion a fraction of total national and regional
emissions to a target grid. Proxies such as population density, land use, and road maps have been developed as spatial
surrogates to represent the intensity of real-world activities in sectors including residential, agriculture, and transportation.
However, it is essential to emphasize that ISATv2.0 solely supports gridded spatial surrogates, necessitating the conversion
60   of spatial surrogates with specific shapes, such as roads, into gridded data. This conversion process may introduce distortions
in spatial surrogates, and the substantial computational effort required for the transformation constrains the attainable
resolution. Consequently, these areas that are open to further refinement may contribute to deviations in the resulting
emission inventory. As a result, the implementation of a method adept at efficiently managing spatial surrogates with
specific shapes carries significant ramifications for the distribution of emission inventories in the transportation sector.
65   Furthermore, MEIAT-CMAQ includes a module to enable users to allocate emissions vertically for inventories that only
cover one layer. This enhancement enables more accurate representation of upper atmospheric emissions, including those
from aviation and a portion of power emissions.

In a novel approach, we have developed a modular emission inventory allocation tool, dubbed MEIAT-CMAQ. This
innovative solution not only boasts an efficient operational algorithm, but also addresses challenges related to pinpointing
grid information and spatial allocation for spatial surrogates with specific shapes. Furthermore, the modular design of
MEIAT-CMAQ confers greater flexibility compared to an integrated tool, rendering it suitable not only for gridded emission
inventories but also for tabulated emissions inventories.

This report undertakes a thorough introduction of the algorithmic principles at the foundation of MEIAT-CMAQ and
establishes its accuracy through the use of the WRF-CMAQ model. Apart from the introduction in Section 1, we offer an in-
depth introduction of the workflow and algorithmic principles underlying MEIAT-CMAQ in Section 2. Section 3 outlines
the model configuration and provides an overview of the data utilization. Section 4 evaluates the performance of tool using
the WRF-CMAQ model, and Section 5 presents our conclusions. Additionally, a step-by-step guide for MEIAT-CMAQ has
been furnished in the appendix.

2 Model Mechanism

MEIAT-CMAQ has been specifically designed for the Windows operating system and is developed using Python 3 and GIS
(Geographic Information System) interfaces. To optimize processing efficiency, the tool is configured to utilize 50% of the
CPU (Central Processing Unit) by default, enabling parallel processing of the entire allocation process. Figure 1 provides an
overview of MEIAT-CMAQ, which consists of three main parts that transform the original emissions inventory into model-
ready emission data.

Part 2 comprises two scripts which encompasses spatial allocation, temporal allocation, and species allocation. The first
script, 'coarse_emission_2_fine_emission.py,' performs spatial allocation by converting coarse-resolution emission
inventories to fine-resolution emission inventories using spatial surrogates. The second script, 'Create-CMAQ-Emission-
File.py', handles temporal and species allocation, converting the assigned results into model-ready emission data. The
temporal allocation process supports only monthly scale emission inventories and assume that there are four weeks in a
month with emissions remaining constant across each week and the module can be modified to account for weekly and
diurnal variations in weekly and hourly profiles, facilitating allocation of monthly emissions to hourly emissions.

Therefore, Part 1 serves as a pre-processing module that includes components enabling the conversion of annual-scale
emission inventories into monthly-scale emission inventories based on monthly profiles. Additionally, the pre-processing
module features scripts for processing commonly-used emission inventories into GeoTIFF format, as is required by MEIAT-

CMAQ. Part 3 is an optional component that handles vertical allocation. This part is omitted for emissions occurring solely
100   near the surface, while it is essential for sectors where emissions may occur at aloft.

2.1 Spatial Allocation

Spatial allocation constitutes a critical component of MEIAT-CMAQ, with temporal and species allocation relying on its
output for subsequent processing. Figure 2 displays a conceptual representation of spatial allocation, which aims to transform
emissions from coarse to fine resolution using user-specified spatial surrogates. For instance, in Figure 2, a coarse emission
105   grid with a value of 40 Mg is converted into a 3×3 grid configuration. Spatial surrogates are then employed to represent the
intensity of activity within the coarse grid at a finer resolution, ultimately enabling the calculation of fine grid emissions (Eq.
1).

$$E_{fine} = E_{coarse} \times \frac{ss_{fine}}{ss_{coarse}} \tag{1}$$

110   In Eq.1, E_{fine} is the fine grid emission, E_{coarse} denotes the coarse grid emission, ss_{fine} corresponds to the value of the
spatial surrogate grid associated with the fine grid, and ss_{coarse} is the sum value of the spatial surrogate grid within the
coarse grid range. Thus, the sum value of ss_{fine} within the same coarse grid equals 1, allowing the coarse grid to be
considered as a computational cell.

115   Figure 3 elucidates the process of spatial distribution within MEIAT-CMAQ, exhibiting the constituents of this allocation,
which encompass modules such as "create fined grid", "fine grid information", "calculate allocation factor", "calculate coarse
grid emission", and "calculating fine grid emissions". The GRIDDESC file, an output from the Meteorology-Chemistry
Interface Processor (MCIP), is employed in the fabrication of a shapefile delineating the simulation grid. Moreover, spatial
surrogates in raster or line shapefile format are utilized in the computation of allocation factors. The initial coarse emission
120   inventory is segregated and quantified by a coarse grid. Eventually, the final refined grid emission is calculated in
accordance with the allocation factor.

2.1.1 Create Fine Grid

The "create fine grid" module is designed to transform GRIDDESC into a shapefile with Lambert Conformal Conic (LCC)
projection, a projection utilized in regional numerical models, including WRF and CMAQ. This shapefile fundamentally
125   comprises multiple small rectangles, with each representing a grid within the CMAQ simulation domain. To locate each
small rectangle, four points are required, which can be computed using Eq. 2-5. The points correspond to the left bottom
position $(x_{\min}, y_{\min})$, the right bottom position $(x_{\max}, y_{\min})$, the left upper position $(x_{\min}, y_{\max})$, and the right upper
position $(x_{\max}, y_{\max})$. Additionally, x_{org} and y_{org} represent the coordinates in the horizontal and vertical directions,
respectively, while dx and dy denote the resolutions in these respective directions.

130 $$x_{min} = x_{org} + dx \times i \tag{2}$$

$$x_{max} = x_{org} + dx \times (i + 1) \tag{3}$$

$$y_{min} = y_{org} + dy \times j \tag{4}$$

$$y_{max} = y_{org} + dy \times (j + 1) \tag{5}$$

2.1.2 Fine Grid Information

135 Figure 4 (a) depicts an idealized conceptual model, assuming that the boundary of each coarse grid emission seamlessly overlaps with multiple fine grids. However, gridded emission inventories often face challenges in precisely aligning with multiple fine grids due to variations in coordinate systems. Additionally, the boundaries of some tabulated inventories might exhibit irregular shapes. These circumstances can result in the smaller grid being divided into one or more sub-grids along the boundary of the coarse grid. Consequently, addressing these split sub-grids in a reasonable manner becomes the primary

140 challenge for MEIAT-CMAQ.

Figure 4 (b) presents a conceptual illustration of the "fine grid information" module, demonstrating how the coarse grid subdivides the fine grid. As shown in Figure 3 (b), the fine grid is partitioned into four sub-grids (F1-F4), each belonging to distinct coarse grids (C1-C4). The primary objective of the "fine grid information" module is to identify the sub-grid with the

145 largest area and determine the corresponding coarse grid as the current attribution for the fine grid. This approach addresses the issue of insufficient accuracy in pinpointing grid information and accelerates computation through the implementation of parallel computing techniques.

2.1.3 Calculate Allocation Factor

The allocation factor is a critical parameter for spatial allocation, and the "calculate allocation factor" module employs high-

150 resolution spatial surrogates to represent the intensity of activity in each sector. Notably, this module supports both raster and line shapefile formats as spatial surrogates, with the inclusion of line shapefiles marking a significant breakthrough in transportation sector allocation. Furthermore, the "calculate allocation factor" module integrates the output file containing fine grid information about coarse grid labels from the "fine grid information" module to compute ss_{coarse} and ss_{fine}.

155 Spatial surrogates in raster format exhibit distinct boundaries, which can divide coarse or fine grids. Consequently, the "calculate allocation factor" module employs the "nearest" method for these values. Eq. 6 illustrates the calculation of the allocation factor, where v_{fine} represents the sum value within the fine grid and v_{coarse} denotes the sum value within the coarse grid.

$$A_{fine} = \frac{v_{fine}}{v_{coarse}} \tag{6}$$

160

Line shapefile format spatial surrogates commonly represent roads, aircraft flight paths, and shipping channels, acting as essential indicators of activity intensity within the transportation sector. Nevertheless, existing inventory assignment tools typically handle roads by converting them into raster format to serve as spatial surrogates. However, due to the significant computational costs and format limitations associated with spatial surrogates, these road-based raster often lack sufficient
165   resolution, resulting in distortions in the final emissions within the fine grid.

Figure 5 displays the fine grid NO_x emissions for the transportation sector in January, allocated from the 2017 Multiscale Emissions Inventory of China (MEIC) developed by Tsinghua University, using both a 3 km resolution road raster and line shapefile spatial surrogates. The 3 km road raster in Figure 5 (a) is converted from the four distinct roads depicted in Figure
170   5 (b), with values in the raster representing the sum of the four different road lengths, each multiplied by their respective coefficients indicating activity intensity. Figure 5 (c) reveals that the fine grid emissions allocated by raster spatial surrogates exhibit pronounced striping patterns due to the low resolution of the road raster. In contrast, Figure 5 (d) demonstrates that employing line shapefile spatial surrogates effectively addresses this issue.

175   Calculating the allocation factor for line shapefile spatial surrogates is similar to that for raster spatial surrogates. The length in the coarse grid is treated as V_{coarse} and the length in the fine grid as V_{fine}. The allocation factor (A_{fine}) is computed following Eq. 6. As most spatial surrogates for the transportation sector are provided in the form of one or more-line shapefiles, converting these line shapefiles into raster format with a specific factor is highly computationally intensive. MEIAT-CMAQ enables users to input an unlimited number of shapefile files and easily specify coefficients for each line
180   shapefile, offering a crucial function for allocating emission inventories in the transportation sector.

2.1.4 Calculate Coarse and Fine Grid Emissions

Based on the boundaries of each coarse grid, the sum value within the grid represents the coarse grid emission. Thus, the "calculate coarse grid emission" module functions as an independent processor, requiring only the input of the coarse grid and original coarse grid emission. Additionally, users can create a fishnet shapefile as the coarse grid for gridded emission
185   inventories, ensuring complete overlap of the boundaries, or they can bypass this step to directly generate the output file of the "calculate coarse grid emission" module for tabulated emission inventories representing countries or cities within each coarse grid. The "calculate fine grid emission" module combines the coarse emission grid table, grid information table, and allocation factor table to compute fine grid emissions.

2.2 Temporal and Species Allocation

190   The "CMAQ emission file" module incorporates both temporal and species allocation, conforming to the stringent stipulations of CMAQ emission files. These specifications dictate that the unit for gaseous species must be reported in "mol/s" and particulate species in "g/s." While annual emission inventories are more accessible than their monthly

counterparts, the MEIAT-CMAQ model solely accommodates monthly-scale emission inventories across all modules due to

the higher temporal resolution in monthly-scale emission inventories.

195

To address this limitation, MEIAT-CMAQ introduces an efficient alternative procedure, "year2month.py", which effectively

converts annual emission inventories into monthly emission inventories. This method harnesses monthly-scale temporal

profiles, enabling the use of transformed data within the MEIAT-CMAQ framework. The "year2month.py" script is

governed by a monthly temporal profile ("monthly.csv") found in the "temporal" directory.

200

Alongside the monthly-scale temporal profile, the module "CMAQ emission file" also employs weekly and hourly-scale

temporal profiles arranged by emission sectors, resulting in a highly organized and efficient system. Thus, the temporal

allocation process aims to assigns monthly emissions inventory data to hourly emissions. Nonetheless, species allocation is

also necessary to fulfil the requirements of CMAQ emission files. Therefore, the complete calculation for both temporal and

205   species allocation adheres to Eq. 7.

$$E = E_{m,sector} \times F_{w,sector} \times F_{h,sector} \begin{cases} \times 10^6 \times \frac{1}{3600} \times S_f \ (U_o = Mmol \ \& \ U_t = mol/s) \\ \times 10^6 \times \frac{1}{3600} \times S_f \ (U_o = Mg \ \& \ U_t = g/s) \\ \times 10^6 \times \frac{1}{3600 \times MW} \times S_f \ (U_o = Mg \ \& \ U_t = mol/s) \end{cases} \tag{7}$$

In Eq. 7, E represents the final emission value within the CMAQ emission file, while $E_{m,sector}$ denotes the monthly emission

for each sector. $F_{w,sector}$ and $F_{h,sector}$ correspond to the weekly and hourly temporal allocation factors for each sector, as

210   provided by "weekly.csv" and "hourly.csv" files, respectively. Furthermore, S_f, MW, U_o, and U_t symbolize the split factor

for the same species, molecular weight, unit of original emission inventories, and unit of CMAQ emission files, respectively.

These parameters can be defined in the species files located within the "species" directory.

2.3 Vertical Allocation

The input for the "vertical allocation" module is derived from the output file of the "CMAQ emission file" module, which

215   has undergone spatial, temporal, and species allocation. However, current emission files primarily characterize near-surface

sources and are not yet suitable for high-altitude emissions. Consequently, accurately characterizing emissions at high

altitudes, such as stack emissions from the industrial and power sectors, proves challenging. The "vertical allocation" module

can utilize user-provided vertical profiles to complete the vertical allocation according to Eq. 8.

$$E_{layer} = F_{layer} \times E_0 \tag{8}$$

220

In Eq. 8, E_{layer} denotes the emission value for each respective layer, whereas E_0 signifies the emission value of the initial

layer. F_{layer} represents the vertical allocation factor provided by the vertical profile.

3 Materials and methods

3.1 Model configuration and observation data

3.1.1 WRF-CMAQ configuration

The WRFv4.4.2 (Skamarock et al., 2019) and CMAQv5.4 (https://zenodo.org/record/7218076, last accessed: June 3, 2023) models were compiled and executed on a server within a Linux environment. The WRFv4.4.2 model simulated meteorological conditions, with initial and boundary conditions obtained from the NCEP 1°×1° Final (FNL) reanalysis dataset (National Centers for Environmental Prediction, National Weather Service, NOAA, U.S. Department of Commerce, 2000). Figure 6 illustrates the adoption of three nested domains with horizontal resolutions of 27, 9, and 3 km, respectively. The outermost domain encompasses China and its surrounding countries, while the innermost domain concentrates on the Pearl River Delta (PRD). Table 1 details the physical parameterizations employed within the WRF model.

Table 1: the physical parameterizations employed within the WRF.

Model attribution	Configuration
Microphysics	Purdue Lin (Chen and Sun, 2002)
PBL physics scheme	MYJ (Janjić, 1994)
Shortwave radiation	Goddard (Chou et al., 2001)
Longwave radiation	Rapid Radiative Transfer Model (RRTM) (Mlawer et al., 1997)
Land surface model	Noah land surface model (LSM) (Ek et al., 2003)
Urban scheme	Single-layer urban canopy model (UCM) (Kusaka and Kimura, 2004)

The initial and boundary conditions of CMAQv5.4 were derived from default profiles representing a clean atmosphere. The Model of Emissions of Gases and Aerosols from Nature version 3.2 (MEGANv3.2) calculated BVOC emissions and was integrated into CMAQv5.4 for in-line computations. The CMAQv5.4 model was configured with the CB06 and AERO7. To represent winter, spring, summer, and autumn in PRD, the model simulated January, April, July, and October, respectively. In addition, to mitigate meteorological drift and clean atmosphere bias, an additional 10-day sign-up time was incorporated for each month, which was excluded from the analysis period.

3.1.2 Scenario configuration

In order to assess the influence of post-assignment emission inventories compared to pre-assignment inventories on the performance of CMAQ simulation, a comprehensive study was devised. This study incorporates two distinct scenarios to examine their impact. The first scenario, named BASE, employs emission allocations without the utilization of spatial surrogates. Conversely, the second scenario, known as EXPR, utilizes emission allocations with spatial surrogates to quantify the influence of MEIAT-CMAQ on air quality modeling performance. Furthermore, an additional scenario, EXPR-V, was established to evaluate the impact of vertical allocation. This scenario encompasses both the assignment processing and the vertical allocation, providing insights into the effects of vertical allocation on the overall evaluation.

250

Table 2 shown the details of the all-scenario configuration and present that all scenarios employ the identical species (Shi et al., 2015; Yuan et al., 2010; Zeng et al., 2021) (Table S1-S5) and temporal profiles (Cai et al., 2018; Zhang et al., 2018) (Table S6-S7), which allows for the attribution of pollutant concentration disparities across scenarios to the distinct spatial allocation algorithms employed by various emissions. In the BASE scenario, coarse emissions are uniformly allocated to the

255 fine grid, resulting in fine emissions. Conversely, EXPR utilizes population gridded data in 2017, sourced from the LandScan Population Data Explorer (https://landscan.ornl.gov/, last accessed: May 7, 2023), to allocate residential sector emissions. Road data in line shapefile format, obtained from Open Street Map (https://www.openstreetmap.org/, last accessed: May 7, 2023), is for the allocation of transportation sector emissions. In addition, the China Land Cover Dataset (CLCD) was used for the allocation of agriculture, industry, and power sectors (Yang and Huang, 2021). Furthermore,

260 emissions from other sectors, including agriculture, power, and industry, are evenly distributed to the fine grid as fine emissions and the emissions from power and industry adopt the vertical profile (Terrenoire et al., 2015) (Table S8) for vertical allocation.

Table 2: Scenario configuration

Scenario Names	Spatial Surrogates	Vertical Profile
BASE	NO	NO
EXPR	YES	NO
EXPR-V	YES	YES

3.1.3 Ambient Monitoring Data

265 In this study, hourly ground-level meteorological observations, such as 2-m temperature (T2), 10-m wind speed (WS10), and 2-m relative humid (RH2), are acquired from 27 national basic meteorological stations provided by China Meteorological Data Service Centre (CMDC). Additionally, hourly ambient levels of air pollutants, including $PM_{2.5}$ and O_3, are obtained from 56 monitoring stations managed by the China National Environmental Monitoring Centre (CNEMC) and the locations of these stations are depicted in Figure 6. The collected data undergo stringent quality control procedures and are

270 subsequently utilized to evaluate the performance of the WRFv4.4.2-CMAQv5.4 model.

3.2 Statistical Analysis

In the study conducted by Emery et al. (2017), several statistical metrics were employed to evaluate the performance of CMAQ. These metrics include the correlation coefficient (R), normalized mean bias (NMB), and normalized mean error (NME). The recommended metrics for $PM_{2.5}$ and O_3, as suggested by Emery et al. (2017), are listed in Table 3. It is worth

275 highlighting that both the goal category and criteria category are referenced again in the subsequent evaluation of the impact of various emission inventories on CMAQ.

In addition to the indicators mentioned above, we employ a range of metrics, such as mean bias (MB), mean error (ME), root mean square error (RMSE), mean fractional error (MFE), and mean fractional bias (MFB) for WRF evaluation, which are
280 derived using Eq.S1 to Eq.S5.

Table 3: Recommended benchmarks for CMAQ performance statics and the formulas for metrics.

Species	NMB		NME		R	
	Goal	Criteria	Goal	Criteria	Goal	Criteria
Hourly O3	$< \pm 5\%$	$< \pm 15\%$	$< 15\%$	$< 25\%$	> 0.75	> 0.50
Daily PM2.5	$< \pm 10\%$	$< \pm 30\%$	$< 35\%$	$< 50\%$	> 0.70	> 0.40

R	$\dfrac{\sum_{i=1}^{n}(M_i - \bar{M}_i)(o_i - \bar{O}_i)}{\sqrt{\sum_{i=1}^{n}(M_i - \bar{M}_i)^2}\sqrt{\sum_{i=1}^{n}(O_i - \bar{O}_i)^2}}$
NMB	$\dfrac{\sum_{i=1}^{n}(M_i - O_i)}{\sum_{i=1}^{n} O_i}$
NME	$\dfrac{\sum_{i=1}^{n}|M_i - O_i|}{\sum_{i=1}^{n} O_i}$

Where M_i and O_i are the simulated and observed data, respectively. $\bar{M}_i$ and $\bar{O}$ are the average of the simulated and observed data, respectively. n is the number of samples.

4 Results and Discussion

4.1 Evaluation for WRF

285 Table 4 presents a comprehensive assessment of meteorological verification metrics for RH2, T2, and WS10 across all simulation periods. Regarding RH2, underestimations occurred in all simulation periods, with MB of -13.09%, -14.41%, -21.40%, and -12.11%, respectively. Nevertheless, WRF model simulations successfully captured RH2 trends in all periods, exhibiting R of 0.80, 0.76, 0.71, and 0.73, respectively. For T2, all simulations accurately captured both trends and values, with R values of 0.93, 0.89, 0.82, and 0.93 for January, April, July, and October, respectively, and MB values of 1.33℃,
290 1.61℃, 1.59℃, and 0.67℃ for the same months. Although the verification metrics for WS10 were not as robust as those for RH2 and T2, R values of 0.52, 0.55, 0.53, and 0.61 were achieved in January, April, July, and October, respectively, while MB values were maintained within 0.5 m/s. In summary, the meteorological variables from these simulations effectively reproduced actual meteorological conditions and can be employed as meteorological fields for CMAQ. Table S9-S20 presents the comprehensive statistical indicators corresponding to each meteorological station.

295

Table 4: Comprehensive Meteorological Verification Metrics

Month	Variable	Ave. Sim.	Ave. Obs.	R	MB	ME	MFB	MFE	NMB	NME
January	RH2	65.90	79.03	0.80	-13.09	13.72	-11.69%	12.37%	-16.58%	17.36%
	T2	17.84	16.51	0.93	1.33	1.61	5.46%	6.76%	8.10%	9.83%
	WS10	2.55	2.35	0.52	0.19	0.99	12.70%	33.71%	9.19%	43.74%
April	RH2	67.43	81.83	0.76	-14.41	15.17	-12.86%	13.58%	-17.66%	18.58%
	T2	24.05	22.44	0.89	1.61	1.97	4.86%	5.90%	7.20%	8.80%
	WS10	3.03	2.58	0.55	0.45	1.25	17.46%	36.05%	19.51%	49.89%
July	RH2	63.08	84.48	0.71	-21.40	21.79	-19.13%	19.45%	-25.35%	25.82%
	T2	29.93	28.34	0.82	1.59	2.05	3.64%	4.75%	5.61%	7.23%
	WS10	2.98	2.58	0.53	0.40	1.32	17.02%	37.26%	17.85%	52.70%
October	RH2	60.32	72.43	0.73	-12.11	13.44	-11.61%	13.05%	-16.81%	18.61%
	T2	25.12	24.45	0.93	0.67	1.41	1.71%	3.95%	2.71%	5.76%
	WS10	3.43	3.11	0.61	0.33	1.35	14.89%	34.33%	11.13%	44.74%

4.2 Comparison of the pre-assessment and post-assessment emissions

300 Figure 7 illustrates the emissions after undergoing spatial allocation. Specifically, (a)-(d) and (i)-(l) represent the emissions in the transportation and agriculture sectors, respectively, for the BASE scenario. On the other hand, (e)-(h) and (m)-(p) depict the emissions in the transportation and agriculture sectors, respectively, for the EXPR scenario. Notably, the transportation sector is associated with the pollutant NO_2 while NH_3 is linked to the agriculture sector.

305 It is worth noting that the emission inventories in the BASE scenarios exhibit distinct grid boundaries for both the transportation and agricultural sectors. However, their coarse resolution fails to accurately capture the detailed spatial patterns of emissions from these sources. In the case of transportation sources, the emission inventories in the BASE scenario provide only a general overview of high emission values at the junction of Guangzhou and Foshan, as well as in the Shenzhen area. Conversely, the post-assessment emission inventories of transportation sources offer a more detailed

310 portrayal of the spatial distribution along roads, thus providing a finer view of the emission patterns.

Similarly, for agricultural source inventories assigned using grid data, the post-assessment emission inventories offer a more detailed representation of the relationship between agricultural source emissions and agricultural land compared to the pre-assessment inventories.

315

Figure 8 illustrates the vertical spatial pattern of the industrial emission inventory, while Figure S1 showcases the vertical spatial pattern of the power emission inventory, both demonstrating the performance of the "vertical allocation" module. It is crucial to consider that industrial emissions are predominantly released at higher altitudes, a factor that is often overlooked in certain studies (Y. Wang et al., 2022; Wu et al., 2023). Neglecting this aspect can significantly impact the performance of

320 AQM simulations.

In Figure 8, the description reveals that industrial emissions are primarily concentrated at heights of 995 hPa and 990 hPa. Conversely, emissions at 1000 hPa are relatively lower in comparison.

4.3 Various emissions performance in CMAQ

325 Figure 9 presents the overall validation metrics for all stations. In this analysis, the mean values of the simulated and observed results for all stations at corresponding time points are utilized as the new observed and simulated values. The data reveals that the simulations of O_3 and $PM_{2.5}$ based on the assigned emission inventories do not exhibit significant performance improvements in the overall metrics. However, both pollutants display better simulation indicators in addition to other factors.

330

However, significant improvements in the performance of the simulation for O_3 in CMAQ can be observed through the statistics for each site indicator, as depicted in Figure 10. In the month of January, the proportion of the goal category increases from 3.6% in the BASE scenario to 7.1% in the EXPR scenario. Similarly, the proportion of the criteria category rises from 69.6% in the BASE scenario to 76.8% in the EXPR scenario. Furthermore, the EXPR-V scenario demonstrates

335 further enhancements compared to the EXPR scenario, with the proportion of the goal category increasing to 12.5% and the proportion of the other category decreasing to 14.3%.

Among the four months evaluated, the most substantial increase in the proportion of the goal category is observed in April. In this month, the proportion of the goal category climbs from 3.6% in the BASE scenario to 18.0% in the EXPR-V scenario.

340 In July, the proportion of the goal category remains steady at 51.8%, while the proportion of the criteria category increases from 35.7% in the BASE scenario to 46.4% in the EXPR-V scenario. Additionally, in October, the proportion of the goal category rises from 16.1% in the BASE scenario to 23.2% in the EXPR-V scenario.

Figure 11 displays the classification percentages by stations for $PM_{2.5}$, highlighting the significant improvements in the

345 simulation of $PM_{2.5}$ due to both pre-assignment and post-assignment emission inventories. In the month of January, the proportion of the criteria category increases from 14.3% in the BASE scenario to 41.1% in the EXPR-V scenario, representing the highest increase among the four months evaluated. Furthermore, the proportion of the goal category also rises from 0% in both the BASE and EXPR scenarios to 3.6% in the EXPR-V scenario. In April, the proportion of the goal and criteria categories increases from 1.8% and 10.7% in the BASE scenario to 10.7% and 14.3% in the EXPR-V scenario,

350 respectively. Although the proportion of the goal category decreases from 1.8% in the BASE scenario to 0% in the EXPR scenario, the proportion of the criteria category increases from 10.7% to 23.2%. Additionally, the proportion of the other category decreases from 87.5% in the BASE scenario to 76.8% in the EXPR scenario.

The most significant increase in the proportion of the goal category occurs in July, rising from 3.6% in the BASE scenario to
16.1% in the EXPR-V scenario, while the proportion of the criteria category remains stable. In October, the proportion of the
goal category increases from 7.1% in the BASE scenario to 17.9% in the EXPR-V scenario, while the proportion of the other
category decreases from 62.5% in the BASE scenario to 35.7% in the EXPR-V scenario.

Indeed, all scenarios successfully capture the spatial patterns in both $PM_{2.5}$ and O_3 concentrations, as depicted in Figure 12,
which displays the monthly averages of CMAQ simulation results alongside the observation data. It is noteworthy that not
only does the simulation of $PM_{2.5}$ concentrations effectively capture the spatial pattern across various scenarios, but the
simulation of O_3 concentration also demonstrates a notable level of spatial consistency with the observed data.

Although post-assignment emission inventories may not exhibit significant superiority in the overall assessment with the
observation stations, both the EXPR and EXPR-V scenarios are capable of delineating hotspot areas in greater detail
compared to the BASE scenario. This enhanced ability to identify and represent hotspot areas is a notable advantage of the
EXPR and EXPR-V scenarios.

5 Conclusion

We have developed a user-friendly modular tool called MEIAT-CMAQ for preparing emission files specifically designed for
the CMAQ. MEIAT-CMAQ encompasses spatial, species, and temporal allocations for the original emission inventory.
While we primarily use gridded emission inventories as input to generate model-ready emission files in this paper, the
modular structure of MEIAT-CMAQ allows for flexibility in handling various formats, including tabulated emission
inventories, which are commonly used for local emissions in a statistical format.

Furthermore, we have included pre-packaged spatial surrogates in the program to facilitate the spatial allocation process.
Additionally, we have prepared temporal allocation profiles for the five main sectors (residential, transportation, power,
industry, and agriculture), species allocation profiles for the CB06 mechanism and the SAPRC07 mechanism, as well as
vertical allocation files for the industry and power sectors. These pre-packaged files are readily available for users to directly
utilize. However, if these files do not fully meet the specific requirements of users, they can easily customize them by
utilizing the pre-packaged files as templates and making necessary modifications as per their needs.

Additionally, we conducted a comparison between the pre-assessment emissions and the post-assessment emissions to
evaluate their impact on the performance of the CMAQ model. The post-assessment emissions, which are derived with the
aid of spatial surrogates, enable the CMAQ model to identify hotspot areas in a more detailed manner during the simulation
of pollutants. This refined spatial allocation leads to notable improvements in the proportion of the goal category for both O3

and PM2.5 simulations in CMAQ, particularly in the EXPR and EXPR-V scenarios, across different months of the year. These findings suggest that employing a high-resolution emission inventory can enhance the accuracy and scientific validity of the model, especially when conducting higher-resolution regional air quality simulations.

390   MEIAT-CMAQ offers an efficient approach for assigning inventories, simplifying the process of transforming coarse-resolution emission inventories into high-resolution emission files that can be directly inputted into the model. Furthermore, MEIAT-CMAQ supports user flexibility by allowing the selection of any desired period for generating the model-ready emission file, providing valuable technical support for automating operations within the business department.

395

400

405

410

415

Appendix A: MEIAT-CMAQ step-by-step guide

Taking the nested domains in Figure 5 with a spatial resolution of 3 km as an example, and the MEIC in 2017 as the original emission inventories.

420 **Step 1: namelist.input configuration**

(a) namelist.input attribution introduction

Table 5: the namelist.input attribution.

Attribution	Description	Process
Global		
griddesc_file	The path of GRIDDESC file.	A
griddesc_name	The grid name of domain.	A
big_grid_file	The path of coarse grid shapefile.	S
geotiff_dir	The path of original emission directory (GeoTIFF format).	S
inventory_label	The prefix in the original emission inventory file name.	S
inventory_year	The year identifier in the original emission inventory file name.	S
sectors	The sectors identifier in the original emission inventory file name.	A
allocator	The file name of spatial surrogates in the "allocator" directory.	S
allocator_type	The type of spatial surrogates (raster or line). If this option is line, the allocator will read the "line" part.	S
inventory_mechanism	The chemical mechanism of original emission inventories.	Sp. & T
target_mechanism	The chemical mechanism of model-ready emission file.	Sp. & T
start_date	The start date of model-ready emission file.	Sp. & T
end_date	The end date of model-ready emission file.	Sp. & T
cores	Number of parallel cores.	A
Line		
line_files	The file name of line shapefile in the "allocator" directory.	S
line_factors	The weight of each road mentioned in line_files.	S
Control		
create_grid	The control of 'Create fine grid' module.	S
grid_info	The control of 'Fine grid information' module.	S
create_factor	The control of 'Calculate allocation factor' module.	S
coarse_emission	The control of 'Calculate coarse grid emission' module.	S
create_source	The control of 'Calculate fine grid emission' module.	S

A = All processes; S = Spatial allocation process; Sp. = Species allocation process; T = Temporal allocation process.

425

(b) namelist.input configuration in this example

Table 6: Contents of case namelist.input

```
&global
    griddesc_file = "input/GRIDDESC.PRD274x181"
    griddesc_name = "PRD274x181"
    big_grid_file = "shapefile/MEIC-0P25.shp"
    geotiff_dir = "input/MEIC"
    inventory_label = "MEIC"
    inventory_year = "2017"
    sectors = 'transportation', 'residential', 'power', 'agriculture', 'industry'
    allocator = 'line', 'landscan-global-2017_nodata.tif', 'power.tif', 'agriculture.tif', 'industry.tif',
    allocator_type = "line", "raster", "raster", "raster", "raster"
    inventory_mechanism = "MEIC-CB05"
    target_mechanism = "CB06"
    start_date = "2017-01-01"
    end_date = "2017-01-02"
    cores = 4
/

&line
    line_files = "motorway.shp", "primary.shp", "residential.shp", "secondary.shp"
    line_factors = 0.435798, 0.326848, 0.081712, 0.155642,
/

&control
    create_grid = 1,
    grid_info = 1,
    create_factor = 1,
    coarse_emission = 1,
    create_source = 1,
/
```

Step 2: spatial allocation

430   Enter the command "python coarse_emission_2_fine_emission.py" into the terminal.

Step 3: create model-ready emission file (species and temporal allocation)

(a) Species allocation profile introduction

Table 7 displays a portion of the transportation source species allocation profile with the "pollutant" column representing the species name identifier in the original emission inventory file name, while the "emission_species" column denotes species

435   names recognized by CMAQ that will serve as variable names in the model-ready emission file. For the same species, the sum of "split_factor" should be theoretically equal to 1, which represents the weight of splitting the same species into species that can be identified by CMAQ. Typically, original emission inventories are provided in units of "Mg" or "Mmol",

which, even if not directly available, can be easily converted to "Mg" or "Mmol". Hence, the "inv_unit" column offers the user a choice between the two options. However, for use in CMAQ, emission files must be in "mol/s" (gaseous species) or 440 "g/s" (solid species) units. Therefore, only these units are supported in the "emi_unit" column. The "divisor" column denotes the molecular mass, which serves as the unit conversion factor.

Table 7: Portions of the transportation species profile

pollutant	emission_species	split_factor	divisor	inv_unit	emi_unit
SO2	SO2	1	64	Mg	mol/s
SO2	SULF	0	98	Mg	mol/s
NH3	NH3	1	17	Mg	mol/s
CO	CO	1	28	Mg	mol/s
PMC	PMC	1	1	Mg	g/s
NOx	NO	0.9	30	Mg	mol/s
NOx	NO2	0.092	46	Mg	mol/s
NOx	HONO	0.008	47	Mg	mol/s
OC	POC	1	1	Mg	g/s

(b) temporal allocation profile introduction

It is important to highlight that when allocating monthly-scale emission inventories to individual weeks, the model assumes 445 four weeks per month, with equal total emissions distributed across each week. Subsequently, the module apportions emissions for each day based on the weekly temporal profile. In this profile, values ranging from 0 to 6 in the "weekly" column (Table 8) correspond to Sunday through Saturday, respectively. Similarly, in the hourly temporal file, values from 0 to 23 in the "hourly" column (Table 9) signify 0 to 23 UTC, respectively. Except for the first column, the column names of the other columns are the emission sectors, which will keep consistent with the sector identification in the original emission 450 inventory file name.

Table 8: The content of weekly temporal profile.

weekly	power	industry	residential	transportation	agriculture
0	0.13	0.078	0.143	0.108	0.143
1	0.147	0.162	0.143	0.155	0.143
2	0.147	0.162	0.143	0.155	0.143
3	0.147	0.162	0.143	0.155	0.143
4	0.147	0.162	0.143	0.155	0.143
5	0.147	0.162	0.143	0.155	0.143
6	0.135	0.112	0.143	0.117	0.143

Table 9: The content of hourly temporal profile.

hourly	power	industry	residential	transportation	agriculture
0	0.032	0.026	0.038	0.017	0.026
1	0.03	0.007	0.038	0.013	0.019
2	0.029	0.007	0.03	0.014	0.019
3	0.028	0.007	0.045	0.015	0.018
4	0.029	0.007	0.045	0.016	0.019
5	0.032	0.007	0.038	0.016	0.021
6	0.035	0.007	0.03	0.029	0.029
7	0.04	0.029	0.03	0.056	0.033
8	0.0433	0.045	0.038	0.0599	0.0473
9	0.0457	0.068	0.038	0.059	0.0576
10	0.0479	0.068	0.03	0.0594	0.07
11	0.0495	0.068	0.045	0.0501	0.0885
12	0.0495	0.068	0.045	0.0501	0.0885
13	0.0497	0.068	0.038	0.0588	0.0823
14	0.0501	0.068	0.03	0.06	0.0803
15	0.05	0.068	0.03	0.062	0.07
16	0.0497	0.068	0.038	0.0594	0.0597
17	0.0489	0.066	0.075	0.0574	0.0453
18	0.0477	0.063	0.075	0.0557	0.0309
19	0.0473	0.037	0.075	0.049	0.0268
20	0.0466	0.037	0.075	0.0454	0.0226
21	0.044	0.037	0.054	0.0417	0.0206
22	0.0397	0.037	0.018	0.0308	0.0206
23	0.0352	0.037	0.018	0.0216	0.0206

455 **(c) complete species and temporal allocation**

Enter the command " python Create-CMAQ-Emission-File.py" into the terminal.

Step 4: vertical allocation

(a) Vertical allocation profile introduction

Table 10 presents the vertical allocation profile, listing "vglytop" as sigma coordinates. These values can be obtained by
460 querying the METCRO3D file, a MCIP output where any layers not involved are considered as unemitted for this layer. The
"fraction" column denotes the relevant distribution factor.

Table 10: vertical allocation profile

Vglvltop	fraction
1.000	0.06
0.995	0.16
0.990	0.75
0.985	0.03

(b) Complete vertical allocation

465 **Code and data availability**

The source code of MEIAT-CMAQv1.0 and input data used to produce the results used in this paper are archived on Zenodo at https://zenodo.org/record/8001532 (H. Wang, 2023).

Competing interests

The contact author has declared that none of the authors has any competing interests.

470 **Acknowledgments**

This work was supported by the Guangdong Major Project of Basic and Applied Basic Research (Grant No. 2020B0301030004), Science and Technology Program of Guangdong Province (Science and Technology Innovation Platform Category) (No. 2019B121201002), Key-Area Research and Development Program of Guangdong Province (No. 2020B1111360003), and the National Natural Science Foundation of China (No. 42075181, 42105097). This work was also
475 supported by the high-performance grid-computing platform of Sun Yat-sen University.

Author contribution

Haofan Wang performed the design of the tool, manuscript writing, and model execution, with Jiaxin Qiu, Yiming Liu, and Qi Fan contributing to the tool design. The manuscript was commented upon by Xiao Lu, Yang Zhang, and Kai Wu. Ao Shen, Yifei Xu, Yinbao Jin, Yuqi Zhu, Jiayin Sun, and Haolin Wang prepared the user's guide for the tool.

480 **References**

Cai, S., Li, Q., Wang, S., Chen, J., Ding, D., Zhao, B., Yang, D., Hao, J., 2018. Pollutant emissions from residential combustion and reduction strategies estimated via a village-based emission inventory in Beijing. Environmental Pollution 238, 230–237. https://doi.org/10.1016/j.envpol.2018.03.036

Chen, S.-H., Sun, W.-Y., 2002. A One-dimensional Time Dependent Cloud Model. Journal of the Meteorological Society of
485   Japan 80, 99–118. https://doi.org/10.2151/jmsj.80.99

Chou, M.-D., Suarez, M.J., Liang, X.-Z., Yan, M.M.-H., Cote, C., 2001. A thermal infrared radiation parameterization for atmospheric studies.

Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., Tarpley, J.D., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta
490   model. J. Geophys. Res. 108, 2002JD003296. https://doi.org/10.1029/2002JD003296

Emery, C., Liu, Z., Russell, A.G., Odman, M.T., Yarwood, G., Kumar, N., 2017. Recommendations on statistics and benchmarks to assess photochemical model performance. Journal of the Air & Waste Management Association 67, 582–598. https://doi.org/10.1080/10962247.2016.1265027

Hutzell, W.T., Luecken, D.J., Appel, K.W., Carter, W.P.L., 2012. Interpreting predictions from the SAPRC07 mechanism
495   based on regional and continental simulations. Atmospheric Environment 46, 417–429. https://doi.org/10.1016/j.atmosenv.2011.09.030

Janjić, Z.I., 1994. The Step-Mountain Eta Coordinate Model: Further Developments of the Convection, Viscous Sublayer, and Turbulence Closure Schemes. Mon. Wea. Rev. 122, 927–945. https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2

500 Kusaka, H., Kimura, F., 2004. Coupling a Single-Layer Urban Canopy Model with a Simple Atmospheric Model: Impact on Urban Heat Island Simulation for an Idealized Case. Journal of the Meteorological Society of Japan 82, 67–80. https://doi.org/10.2151/jmsj.82.67

Li, M., Liu, H., Geng, G., Hong, C., Liu, F., Song, Y., Tong, D., Zheng, B., Cui, H., Man, H., Zhang, Q., He, K., 2017. Anthropogenic emission inventories in China: a review. National Science Review 4, 834–866.
505   https://doi.org/10.1093/nsr/nwx150

Liu, Y., Wang, T., 2020a. Worsening urban ozone pollution in China from 2013 to 2017 – Part 1: The complex and varying roles of meteorology. Atmos. Chem. Phys. 20, 6305–6321. https://doi.org/10.5194/acp-20-6305-2020

Liu, Y., Wang, T., 2020b. Worsening urban ozone pollution in China from 2013 to 2017 – Part 2: The effects of emission changes and implications for multi-pollutant control. Atmos. Chem. Phys. 20, 6323–6337.
510   https://doi.org/10.5194/acp-20-6323-2020

Lu, Y., Yang, X., Wang, H., Jiang, M., Wen, X., Zhang, X., Meng, L., 2023. Exploring the effects of land use and land cover changes on meteorology and air quality over Sichuan Basin, southwestern China. Front. Ecol. Evol. 11, 1131389. https://doi.org/10.3389/fevo.2023.1131389

Luecken, D.J., Yarwood, G., Hutzell, W.T., 2019. Multipollutant modeling of ozone, reactive nitrogen and HAPs across the
515   continental US with CMAQ-CB6. Atmospheric Environment 201, 62–72. https://doi.org/10.1016/j.atmosenv.2018.11.060

Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J., Clough, S.A., 1997. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J. Geophys. Res. 102, 16663–16682. https://doi.org/10.1029/97JD00237

520 National Centers for Environmental Prediction, National Weather Service, NOAA, U.S. Department of Commerce, 2000. NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999.

Qiu, J., Fang, C., Tian, N., Wang, H., Wang, J., 2023. Impacts of land use and land cover changes on local meteorology and PM2.5 concentrations in Changchun, Northeast China. Atmospheric Research 289, 106759. https://doi.org/10.1016/j.atmosres.2023.106759

525 Shi, J., Deng, H., Bai, Z., Kong, S., Wang, X., Hao, J., Han, X., Ning, P., 2015. Emission and profile characteristic of volatile organic compounds emitted from coke production, iron smelt, heating station and power plant in Liaoning Province, China. Science of The Total Environment 515–516, 101–108. https://doi.org/10.1016/j.scitotenv.2015.02.034

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Liu, Z., Berner, J., Wang, W., Powers, J.G., Duda, M.G., Barker,
530   D.M., Huang, X.-Y., 2019. A Description of the Advanced Research WRF Model Version 4. UCAR/NCAR.
        https://doi.org/10.5065/1DFH-6P97
Terrenoire, E., Bessagnet, B., Rouïl, L., Tognet, F., Pirovano, G., Létinois, L., Beauchamp, M., Colette, A., Thunis, P.,
        Amann, M., Menut, L., 2015. High-resolution air quality simulation over Europe with the chemistry transport
        model CHIMERE. Geosci. Model Dev. 8, 21–42. https://doi.org/10.5194/gmd-8-21-2015
535 Wang, H., Liu, Z., Wu, K., Qiu, J., Zhang, Y., Ye, B., He, M., 2022. Impact of Urbanization on Meteorology and Air Quality
        in   Chengdu,   a   Basin   City   of   Southwestern   China.   Front.   Ecol.   Evol.   10,   845801.
        https://doi.org/10.3389/fevo.2022.845801
Wang, H., Liu, Z., Zhang, Y., Yu, Z., Chen, C., 2021. Impact of different urban canopy models on air quality simulation in
        Chengdu,   southwestern   China.   Atmospheric   Environment   267,   118775.
540   https://doi.org/10.1016/j.atmosenv.2021.118775
Wang, H., 2023. MEIAT-CMAQ v1.0: A Modular Emission Inventory Allocation Tool for Community Multiscale Air
        Quality Model Version 1.0, Zenodo [code and data set], 10.5281/zenodo.8001532,
Wang, K., Gao, C., Wu, K., Liu, K., Wang, H., Dan, M., Ji, X., Tong, Q., 2023. ISAT v2.0: an integrated tool for nested-
        domain configurations and model-ready emission inventories for WRF-AQM. Geosci. Model Dev. 16, 1961–1973.
545   https://doi.org/10.5194/gmd-16-1961-2023
Wang, Y., Yang, X., Wu, K., Mei, H., De Smedt, I., Wang, S., Fan, J., Lyu, S., He, C., 2022. Long-term trends of ozone and
        precursors from 2013 to 2020 in a megacity (Chengdu), China: Evidence of changing emissions and chemistry.
        Atmospheric Research 278, 106309. https://doi.org/10.1016/j.atmosres.2022.106309
Wu, K., Zhu, S., Mac Kinnon, M., Samuelsen, S., 2023. Unexpected deterioration of O3 pollution in the South Coast Air
550   Basin of California: The role of meteorology and emissions. Environmental Pollution 330, 121728.
        https://doi.org/10.1016/j.envpol.2023.121728
Wu, Q.Z., Xu, W.S., Shi, A.J., Li, Y.T., Zhao, X.J., Wang, Z.F., Li, J.X., Wang, L.N., 2014. Air quality forecast of
        PM₁₀ in Beijing with Community Multi-scale Air Quality Modeling (CMAQ) system:
        emission and improvement. Geosci. Model Dev. 7, 2243–2259. https://doi.org/10.5194/gmd-7-2243-2014
555 Xie, Y., Paulot, F., Carter, W.P.L., Nolte, C.G., Luecken, D.J., Hutzell, W.T., Wennberg, P.O., Cohen, R.C., Pinder, R.W.,
        2013. Understanding the impact of recent advances in isoprene photooxidation on simulations of regional air quality.
        Atmos. Chem. Phys. 13, 8439–8455. https://doi.org/10.5194/acp-13-8439-2013
Yang, J., Huang, X., 2021. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. Earth Syst. Sci.
        Data 13, 3907–3925. https://doi.org/10.5194/essd-13-3907-2021
560 Yang, X., Wu, K., Wang, H., Liu, Y., Gu, S., Lu, Y., Zhang, X., Hu, Y., Ou, Y., Wang, S., Wang, Z., 2020. Summertime
        ozone pollution in Sichuan Basin, China: Meteorological conditions, sources and process analysis. Atmospheric
        Environment 226, 117392. https://doi.org/10.1016/j.atmosenv.2020.117392
Yang, Y., Zhao, Y., Zhang, L., Zhang, J., Huang, X., Zhao, X., Zhang, Y., Xi, M., Lu, Y., 2021. Improvement of the
        satellite-derived NOx emissions on air quality modeling and its effect on ozone and secondary inorganic aerosol
565   formation in the Yangtze River Delta, China. Atmos. Chem. Phys. 21, 1191–1209. https://doi.org/10.5194/acp-21-
        1191-2021
Yuan, B., Shao, M., Lu, S., Wang, B., 2010. Source profiles of volatile organic compounds associated with solvent use in
        Beijing, China. Atmospheric Environment 44, 1919–1926. https://doi.org/10.1016/j.atmosenv.2010.02.014
Zeng, X., Kong, S., Zhang, Q., Ren, H., Liu, J., Feng, Y., Yan, Q., Qin, S., Zheng, S., Yao, L., Fan, Z., Zhang, Y., Liu, X.,
570   Yan, Y., Zhu, K., Ding, F., Liu, W., Liu, D., Qi, S., Fu, P., 2021. Source profiles and emission factors of organic
        and inorganic species in fine particles emitted from the ultra-low emission power plant and typical industries.
        Science of The Total Environment 789, 147966. https://doi.org/10.1016/j.scitotenv.2021.147966
Zhang, L., Chen, Y., Zhao, Y., Henze, D.K., Zhu, L., Song, Y., Paulot, F., Liu, X., Pan, Y., Lin, Y., Huang, B., 2018.
        Agricultural ammonia emissions in China: reconciling bottom-up and top-down estimates. Atmos. Chem. Phys. 18,
575   339–355. https://doi.org/10.5194/acp-18-339-2018
Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., Li, H., Li, X., Peng, L., Qi, J., Yan, L., Zhang, Y., Zhao, H., Zheng,
        Y., He, K., Zhang, Q., 2018. Trends in China's anthropogenic emissions since 2010 as the consequence of clean air
        actions. Atmos. Chem. Phys. 18, 14095–14111. https://doi.org/10.5194/acp-18-14095-2018

580

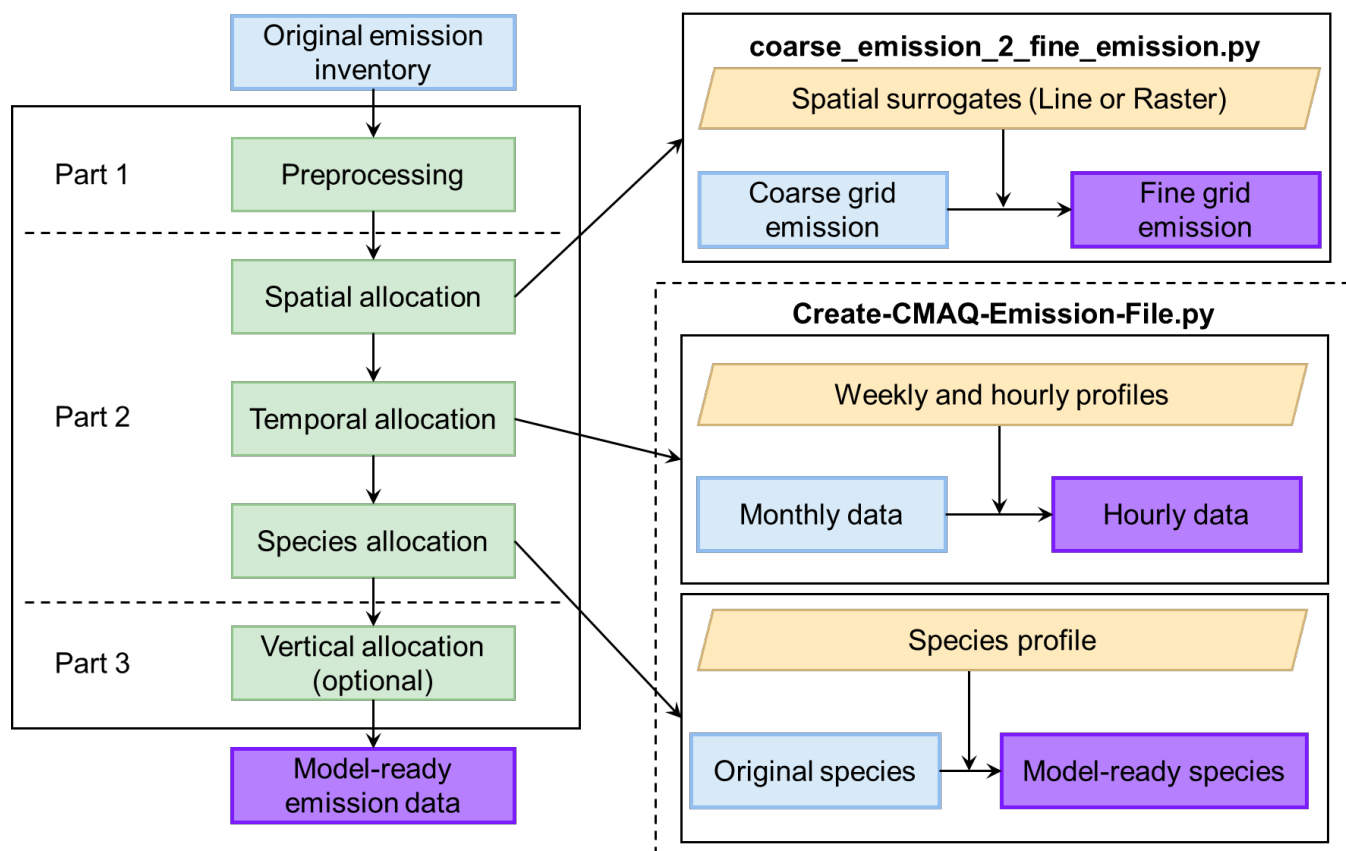



Figure 1: Operational workflow for MEIAT-CMAQ.

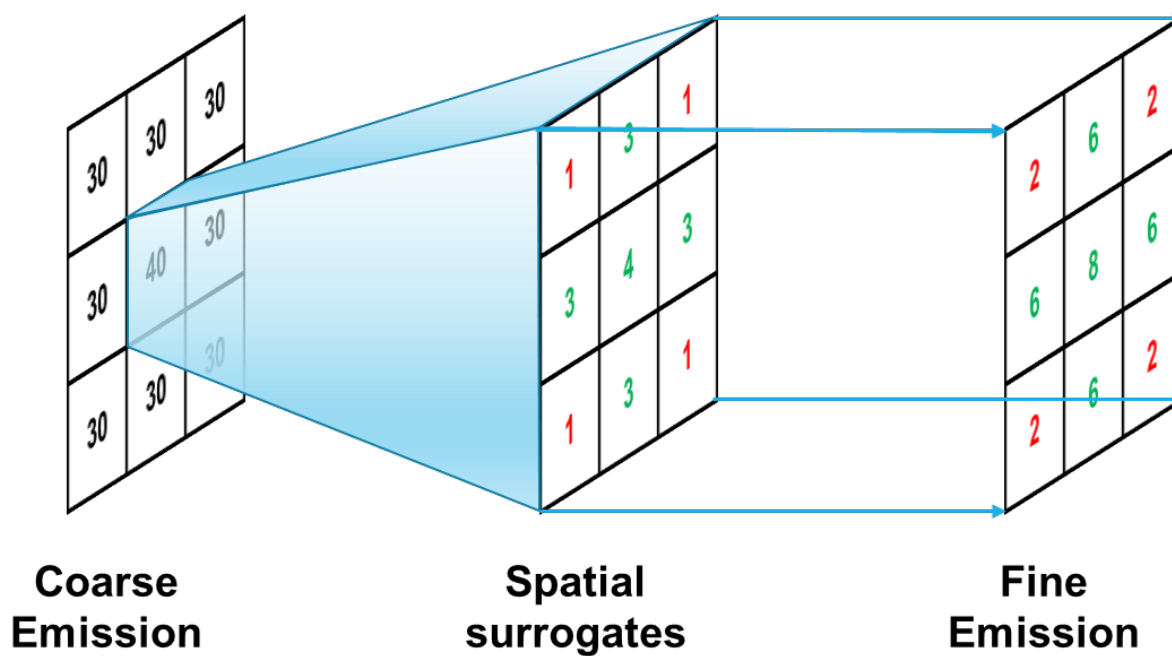
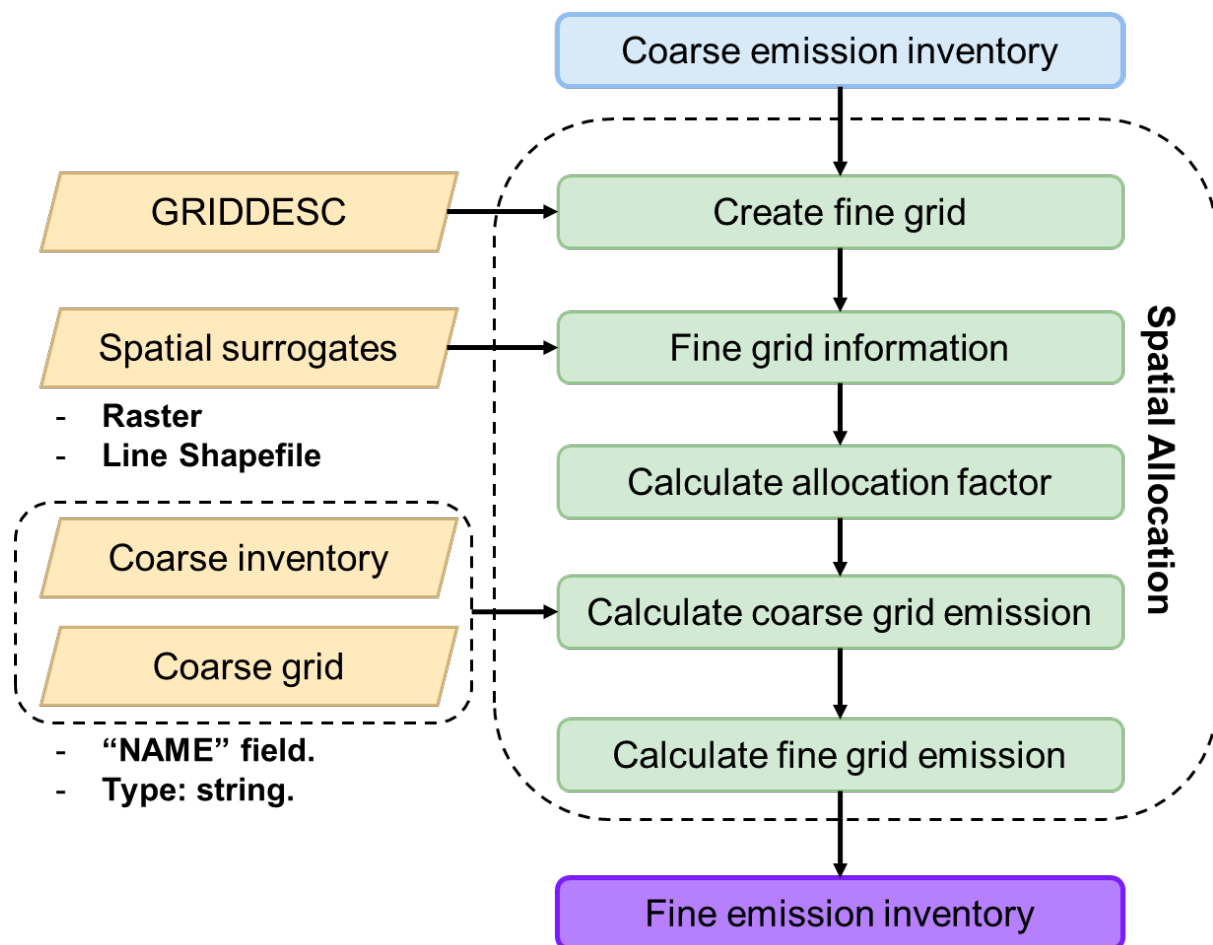



Figure 2: Conceptual representation of spatial allocation.

590   **Figure 3: Operational workflow for spatial allocation in MEIAT-CMAQ**

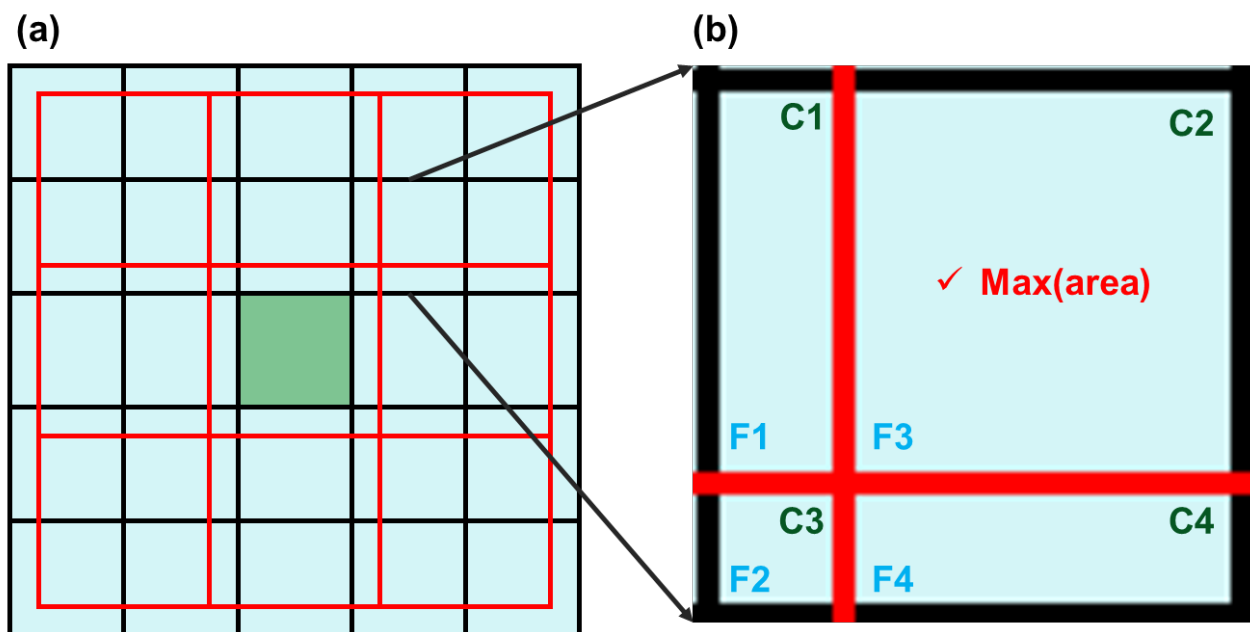



Figure 4: Conceptual illustration of the "Fine Grid Information" module. In panel (a), red grids represent coarse grids, while black grids denote fine grids. C1-C4 serve as labels for each coarse grid, and F1-F4 designate labels for each sub-grid originating from the fine grid.

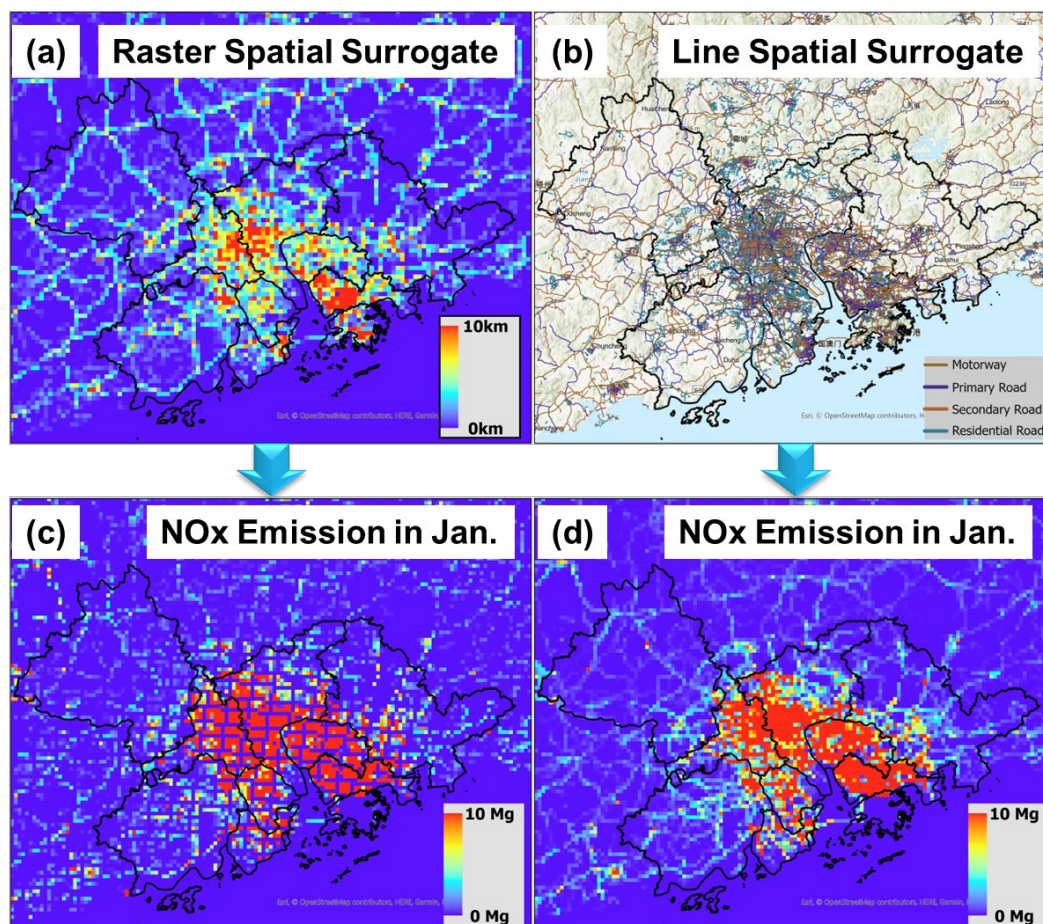



Figure 5: (a) Road spatial surrogate in raster format, **(b)** road spatial surrogate in line shapefile format, **(c)** NO$_x$ fine grid emissions for January using MEIAT-CMAQ with spatial surrogate from panel (a), and **(d)** same as (c) but from panel (b).

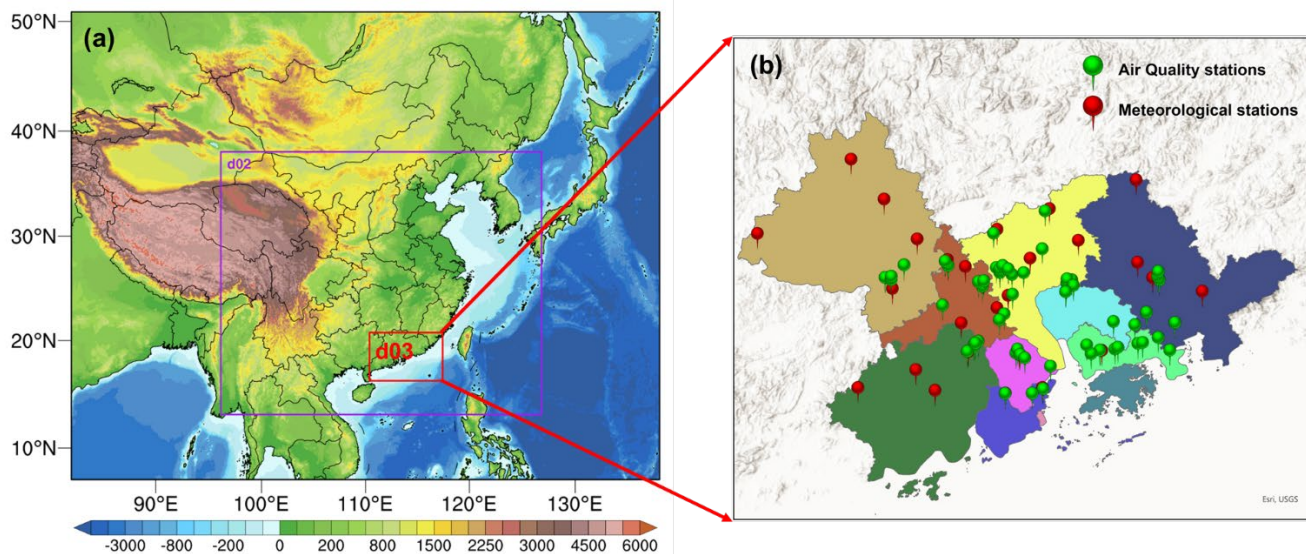



Figure 6: Depiction of the simulation domains and observation station locations. (a) Illustrates the spatial extent of the simulation domains, and (b) highlights the positions of the observation stations within the simulation domains.

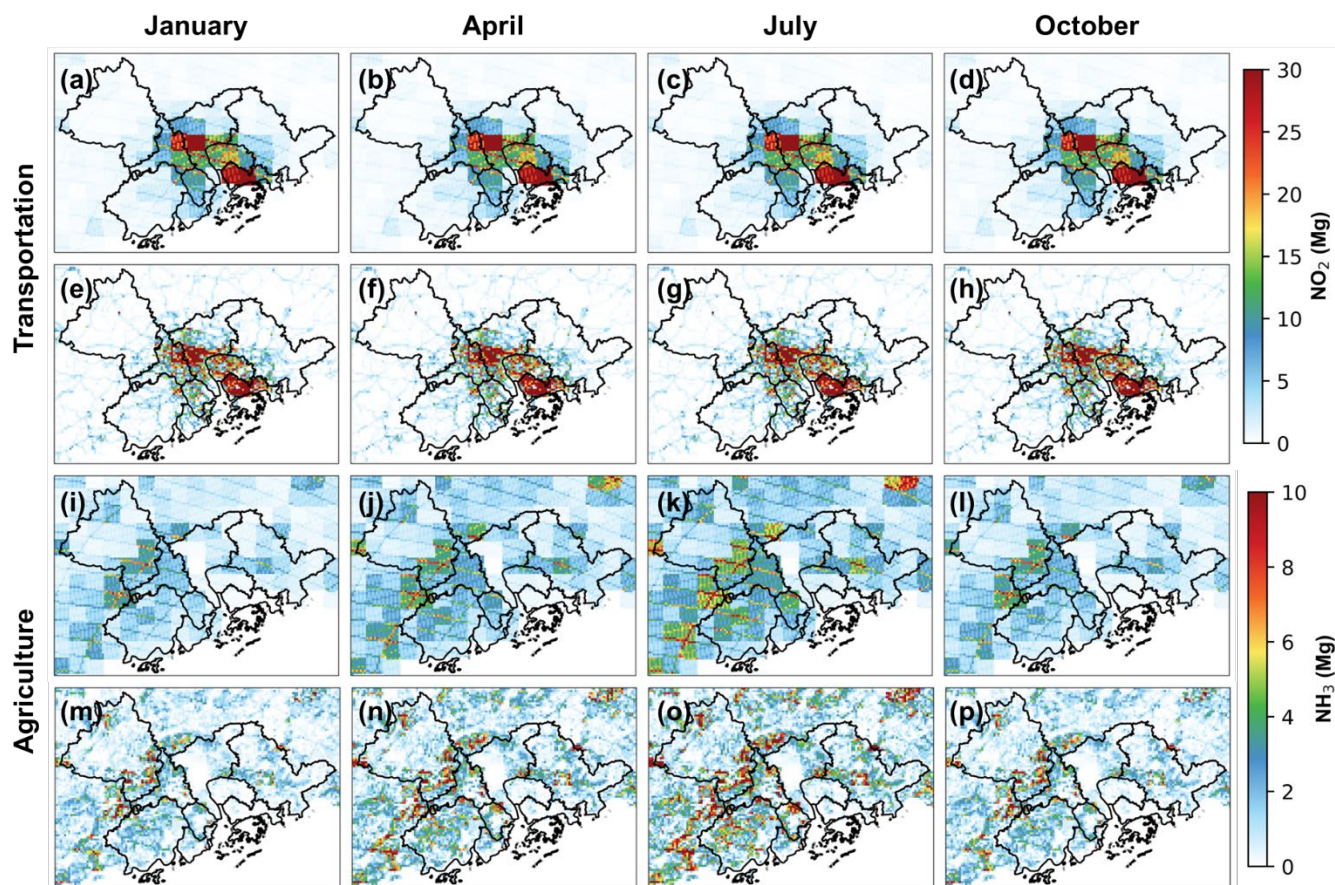



Figure 7: The emissions after spatial allocation. (a)-(d) and (i)-(l) are the emissions of BASE scenario in transportation and agriculture sectors, respectively. (e)-(h) and (m)-(p) are the emissions of EXPR scenario in transportation and agriculture, respectively. The pollutants of the transportation and agriculture sectors are NO₂ and NH₃.

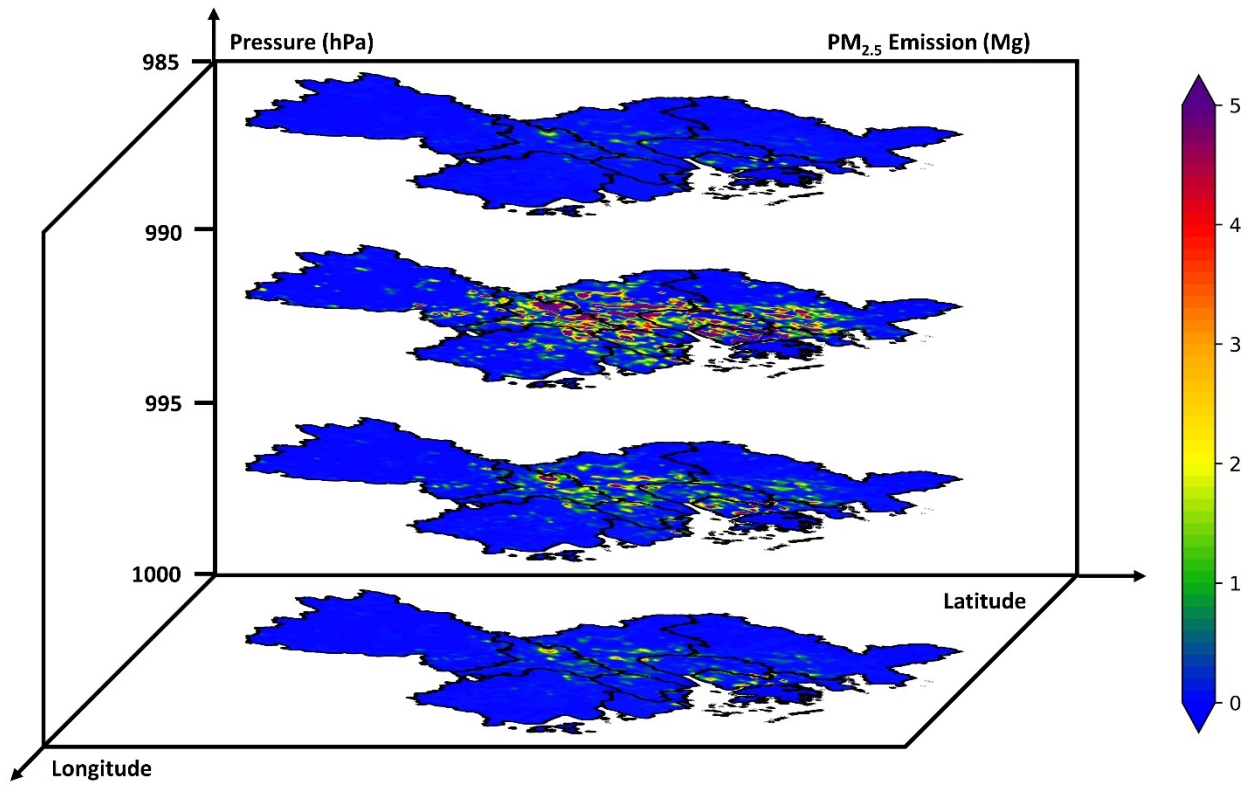



Figure 8: vertical spatial pattern of the industrial emission inventory.

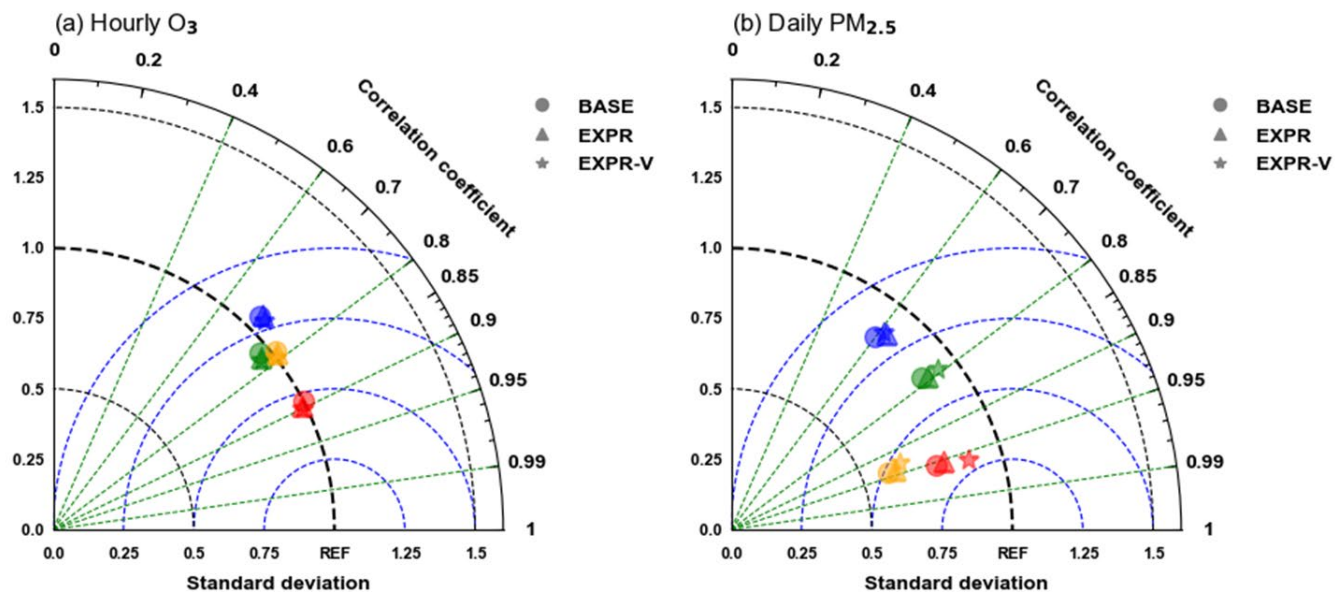



Figure 9: Taylor images comparing overall hourly O₃ concentrations (a) and overall daily PM₂.₅

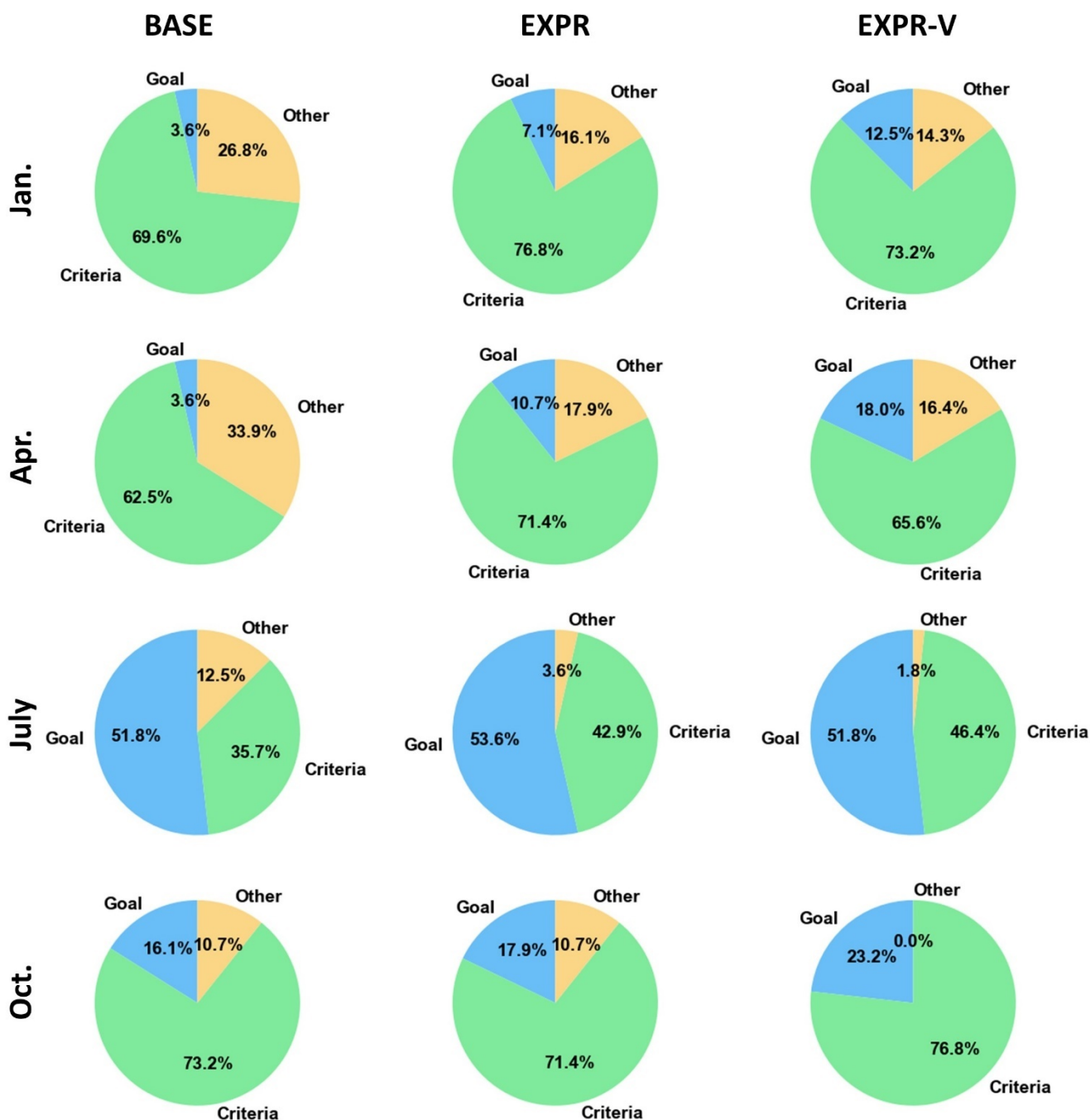



Figure 10: Classification percentage by stations for O₃. Classification categories, including Goal and Criteria classifications (as described in Table 3), along with the 'Other' category representing that do not meet the Goal and Criteria indicators.

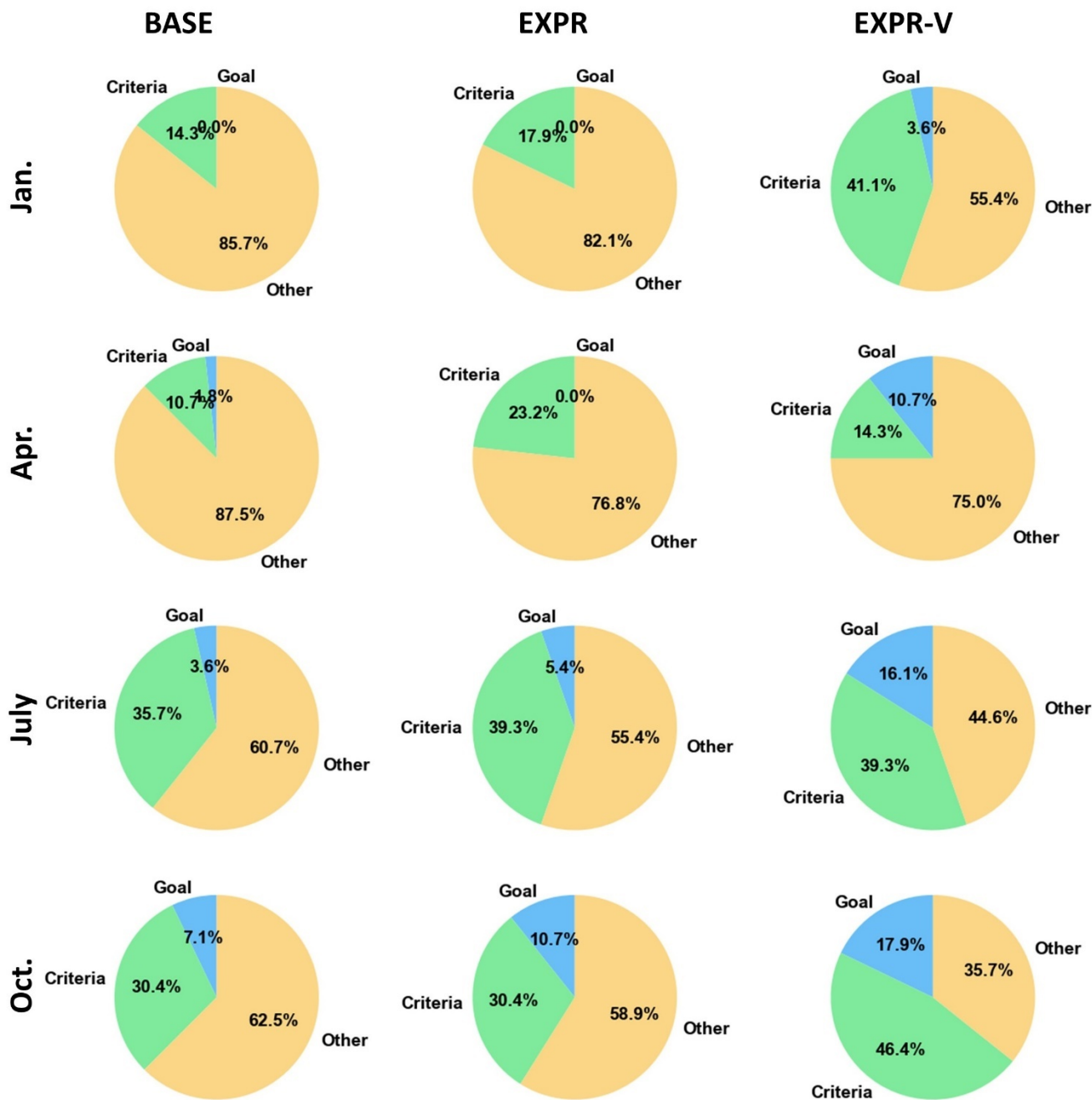



Figure 11: Classification Percentage by Stations for PM2.5. Classification categories, including Goal and Criteria classifications (as described in Table 3), along with the 'Other' category representing that do not meet the Goal and Criteria indicators.

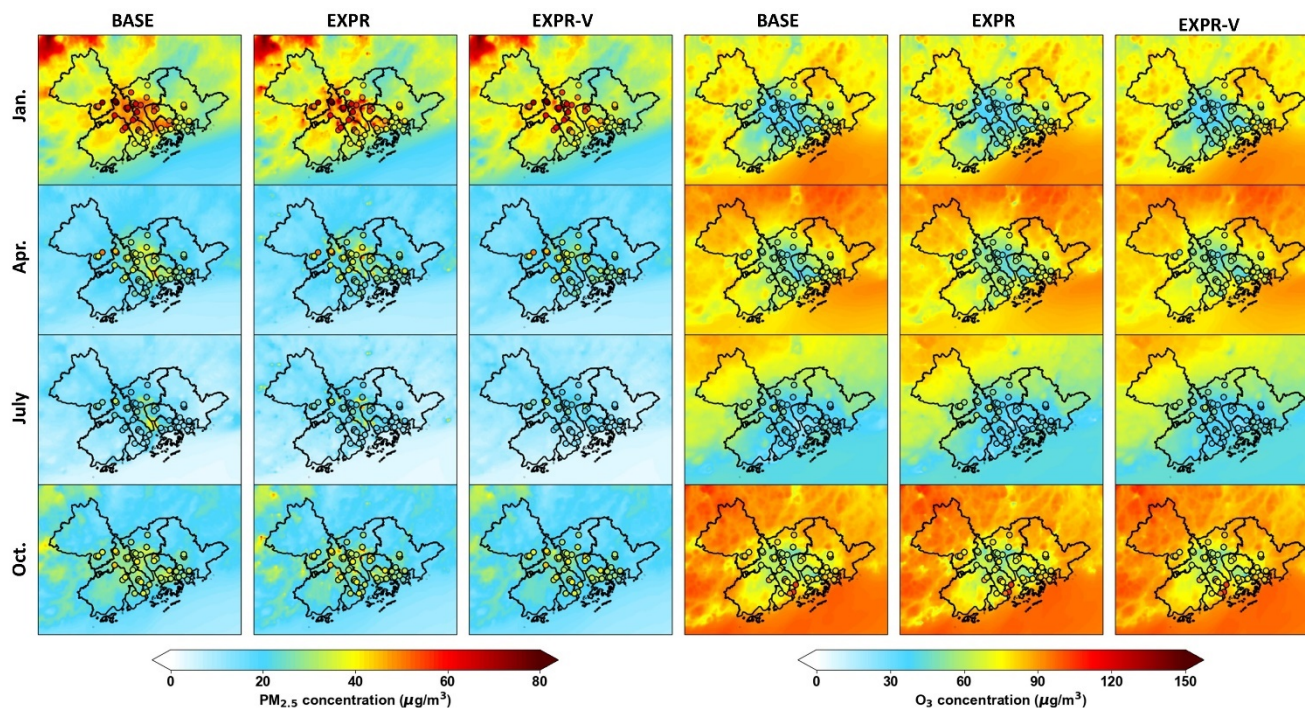



Figure 12: Spatial pattern of the simulation for monthly average PM$_{2.5}$ and O$_3$ concentrations via various emission inventories. The point is the monthly average values of each station, and the contour is the monthly average value of CMAQ simulation results.