LAPORAN PENELITIAN

"Atmospheric Modeling with Focus on Management of Input/Output Data and Potential of Cloud Computing Applications"

Didin Agustian P. Ph.D

INSTITUT TEKNOLOGI NASIONAL
BANDUNG - 2020

4

Atmospheric Modeling with Focus on Management of Input/Output Data and Potential of Cloud Computing Applications

Atmospheric modeling has been widely used to understand the physical and chemical processes in the atmosphere that involve the climate, weather and air quality research. Several air quality dispersion models exist which are applied worldwide, including CAMx, CMAQ, GEOS-Chem and CHIMERE, among others. These air quality models are used to establish quantitative relationships between the emissions and the ambient air concentrations driven by meteorological variables commonly produced by meteorological models. This chapter provides an introduction to the atmospheric modeling with emphasis on the conceptual modeling structure. Specific features of several commonly applied air quality models are provided together with their online information. Further, this chapter describes the model application practices and highlights the issue of big input and output datasets involved where applications of cloud computing would hold potential. Several dispersion modeling case studies covering different domain scales in Southeast Asia (urban, national and regional) conducted by the air quality research group at the Asian Institute of Technology (Thailand) are provided to facilitate the comprehension of modeling application process and the involved big input and output datasets for processing and handling in the modeling system.

Chapter written by Thi Kim Oanh NGUYEN, Nhat Ha Chi NGUYEN, Nguyen Huy LAI and Didin Agustian PERMADI.

TORUS 3 – Toward an Open Resource Using Services: Cloud Computing for Environmental Data, First Edition. Edited by Dominique Laffly.

4.1. Introduction

4.1.1. *Atmospheric modeling*

The compositions of the Earth's atmosphere have been vastly changing more rapidly in recent decades due to the increased emissions from human-made activities. As a result, there is a significant increase in the levels of toxic constituents that are harmful to human health, ecosystems and the Earth's climate system. Several atmospheric issues such as global warming, ozone layer depletion and air pollution have been intensively studied in the last few decades (Ramanathan and Feng 2008, and references therein).

The atmospheric processes are of a wide range of scale and far too complex to be reproduced physically in the laboratory (Schneider *et al*. 1989). To some extent, instead of building a physical analog of the whole atmospheric system, these processes can be simulated mathematically using some computer model systems known as the atmospheric modeling systems. Such a modeling system uses a system of mathematical equations to describe physical and chemical processes in the atmosphere, and this modeling approach is applied for climate, weather forecast and air quality studies. Although these models are not able to reproduce the full complexity of reality, they can reveal the logical consequences of assumptions about how the atmospheric system changes. Atmospheric models include at least four major categories: (1) atmospheric general circulation model (AGCM) for simulating the global climate, (2) regional climate model (RCM) to simulate the climate in different regions, (3) mesoscale meteorological model for simulating the details of regional and local meteorological phenomena and (4) chemistry transport model (CTM) to simulate changes in chemistry in the atmosphere that are applied in air quality studies (Dawson *et al*. 2008).

AGCM simulates the global climate with a grid resolution at about one degree (~100 km), while RCM simulates the regional scale climate and can use the input supplied by AGCM. Air quality models (or CTM) are equipped with various chemistry mechanisms which use complex chemical reactions of species involved to simulate, for example, the formation of ground-level ozone (O_3), secondary particulate matters (PM) and acidic compounds in the atmosphere. These models can treat multiple air quality issues and simulate the consequences of past and future emission scenarios to examine the effectiveness of abatement strategies (Daly and Zannetti 2007). Meteorological models provide important input data of meteorological variables/fields to CTM. These meteorological fields affect the dispersion, chemical reaction rates and formation of secondary air pollutants in the atmosphere. Atmospheric models have different details of processes simulated, and they can run with different grid resolutions and time scales, which in turn require specific input and output data management. In this chapter, we focus on mesoscale meteorological models and the air quality models which are commonly used worldwide to provide scientific information to develop and analyze air quality management (AQM) strategies.

Specifically, this chapter describes the common model types that are available and their typical features. Several case studies of the model applications conducted by the air quality research group at the Asian Institute of Technology (AIT) are provided for with illustrations. The detailed input and output data for PM and O_3 air quality modeling in various domain scales are described in the case studies.

4.1.2. *Roles of modeling in air quality management*

High levels of air pollution are observed in many developing countries that cause harmful effects on human health and ecosystems (Kim Oanh *et al.* 2013a and references therein). The high emission strength, mainly from human-made activities, and insufficient management capacities are among the important factors leading to the serious pollution situations. Effective AQM requires an in-depth understanding of the causal relationship between the emissions and meteorology on one side, and the resulting air pollution concentrations, deposition and consequent effects on the other side. Air pollution modeling is a technical tool used in the AQM to establish such a quantitative relationship. Particularly, air quality models reveal the deterministic relationship between emissions and concentrations/depositions that can be used to analyze past and future emission scenarios in order to assess the efficacy of various abatement strategies (Daly and Zannetti 2007). Three-dimensional (3D) air quality models are normally driven by 3D mesoscale meteorological models or, in other words, the meteorological models provide required meteorological input data to run the air quality models. Applications of air quality models for AQM have attracted attention among researchers around the world. Especially in Asian developing countries where the air quality is worsening, the applications of air quality models become increasingly popular.

Air quality models can be used for historical simulations (hindcast) and provide comprehensive spatial coverage for air pollution data that are especially useful for the places where the monitoring data coverage is insufficient. The models can be used to simulate the present and future air quality (nowcast and forecast) using the corresponding emissions and meteorological input data. The hindcast simulation can be applied for the investigation of past air pollution episodes, assessment of associated health, environmental and climate effects. The nowcast and forecast, in addition to the aforementioned applications, also provide the current and future air quality information to assess the associated health and ecosystem impacts. Interactions between air quality and climate change can also be investigated using the modeling approach, and this helps to analyze the potential co-benefits (to air quality and climate) resulting from different interventions (Permadi *et al.* 2018 and references therein).

4.1.3. *Existing modeling systems*

4.1.3.1. *Air quality dispersion model*

The air quality dispersion model has been evolving from the simple Gaussian plume models to the complex one-atmosphere models that are commonly used at present. The simple Gaussian plume model was initially applied for estimating the maximum ground-level impact of the plumes and the distance of maximum impact from the source, especially for non-reactive pollutants (Bosanquet 1936). However, the air pollution problem is a complex phenomenon that involves chemical reactions, dispersion and deposition processes at various scales, hence cannot be handled by this type of simple model. More atmospheric processes need to be simulated, especially for the secondary pollutants which are formed in the atmosphere, such as ground-level O_3 and secondary inorganic (such as sulfate and nitrate particles) and organic particles. The model domains need to be enlarged to simulate the large-scale phenomena at regional and global scales. Long-range transport models were then developed following the Lagrangian approach, mainly based in Europe (Rohde 1972). At the urban scale, the first-generation Eulerian photochemical smog models were developed in the United States (Reynolds 1973). These became a basis for the development of new one-atmosphere models, which simulate the changes of pollutant concentrations in the atmosphere at multiple scales. These CTM models have been equipped with chemistry mechanisms with complex reactions and species involved, such as the carbon bond (CB-IV, CB-V), Regional Acid Deposition Model (RADM) and State-wide Air Pollution Research Centre (SAPRC) mechanism (Stockwell *et al.* 1990; Carter 2000). The inclusion of these complex chemistry mechanisms added requirements to higher computational resources and the need for big data management.

Presently, the advanced CTMs are most commonly used. They are believed to be the most powerful to involve the least restrictive assumptions, and to be the most computationally intensive. The CTM development has been mainly initiated in the United States, but similar efforts have been conducted in other regions lately. A summary of commonly applied photochemical smog models is presented in Table 4.1. In the United States, the 3D variable-grid Urban Air shed Model (UAM-V) was developed and maintained by the System Applications International (SAI), which was formerly the most widely applied photochemical air quality model. Later on, more advance models have been developed, such as the Models-3/Community Multi-scale Air Quality (CMAQ) and the Comprehensive Air Quality Model with extensions (CAMx), which are well maintained and updated by the community. The above-mentioned models belong to the Eulerian photochemical dispersion model category that allows a "one-atmosphere" assessment of gaseous and particulate air pollution at different scales ranging from sub-urban to continental. These models simulate the emissions, dispersion, chemical reactions and removal of pollutants in the troposphere on a system of nested 3D grids.

Model		Full name	Model type	Chemistry mechanisms	Grid size and features	Website
US-based models	CAMx	Comprehensive Air Quality Model with Extensions	Eulerian, multiscale, two-way nested, multilayer	CB-IV, RADM2, SAPRC, CB-V	4–36 km	http://www.camx.com/home.aspx
	MODEL-3/ CMAQ	Models-3/Community Multi-Scale Air Quality	Eulerian, multiscale, two-way nested, multilayer	CB-IV, RADM2, SAPRC, CB-V	4–36 km	https://www.cmascenter.org/cmaq/
	GEOS-Chem		3D chemical transport model (CTM)	FlexChem; GEOS-Chem Chemical Mechanism	Global and regional scale, 0.25°–5°	http://acmg.seas.harvard.edu/geos/
	WRF-Chem	Weather Research and Forecasting (WRF) model coupled with Chemistry	Eulerian, multiscale, two-way nested, multilayer	CB-IV, CB-V, RADM2, SAPRC99	1–100 km	https://ruc.noaa.gov/wrf/wrf-chem/
European-based models	CHIMERE		Eulerian, multiscale, one-way nested, multilayer	MELCHIOR	2–100 km	www.lmd.polytechnique.fr/chimere/

Model		Full name	Model type	Chemistry mechanisms	Grid size and features	Website
	MATCH	Multiple-Scale Atmospheric Transport and Chemistry Model	Eulerian, multiscale, one-way nested, multilayer	EMEP (Simpson 1993)	2–100 km	https://www.smhi.se/en/research/research-departments/air-quality/match-transport-and-chemistry-model-1.6831
	EURAD	European Air Pollution Dispersion	Eulerian, multilayer, nested grid, one-way nested	RADM2, RACM	Variable grid (about 80 km)	http://www.uni-koeln.de/math-nat-fak/geomet/eurad/index_e.html?/math-nat-fak/geomet/eurad/modell/eurad_descr_e.html
Japanese-based model	SPRINTARS	Spectral Radiation-Transport Model for Aerosol Species	Eulerian with aerosol module	Gas phase – aerosol interaction (Takemura *et al*. 2000)	30–100 km	https://sprintars.riam.kyushu-u.ac.jp/indexe.html

Table 4.1. *Commonly applied air quality dispersion models (modified from Kim Oanh and Permadi 2009)*

Several CTMs have also been developed in Europe and applied worldwide. CHIMERE is a 3D multi-scale Eulerian model, which is applied for real-time forecasts and long-term simulations (http://www.lmd.polytechnique.fr/chimere/). It consists of a European-scale model with a horizontal resolution of 0.5° × 0.5° and different regional models (Schmidt *et al.* 2001). The Multiscale Atmospheric Transport and Chemistry Model (MATCH), another European model, is a 3D one-way nesting Eulerian model that can be configured for different geographical areas with different resolutions. The model is driven by meteorological input data taken from operational numerical weather prediction models, such as the High-Resolution Limited Area Model (HIRLAM) or the European Centre for Medium-Range Weather Forecasts (ECMWF), or objective analyses such as Mesoscale Analysis (MESAN) or from combinations of these. The Spectral Radiation-Transport Model for Aerosol Species (SPRINTARS) is a global CTM developed in Japan which is mainly used to simulate aerosol concentrations over different scales (Table 4.1).

Beside the models listed in Table 4.1, other models have been developed by various projects, such as MOZART (Model for OZone and Related chemical Tracers), LOTOS (Long-Term Ozone Simulation) or REM3 (Regional Eulerian Model with three different chemistry schemes), which also gained applications in several studies. These and other models being further developed may gain popular applications in the future.

4.1.3.2. *Mesoscale meteorological model*

Several mesoscale meteorological models exist. Among the first to be mentioned are the Systems Applications International Meteorological Model (SAIMM), which used a terrain-following vertical coordinate system with variable vertical grid spacing and the 4D data assimilation (SAI 1995), the Fifth-Generation Pennsylvania State University/National Centre for Atmospheric Research (PSU/NCAR) Mesoscale Meteorological Model (MM5) (http://www.mmm.ucar.edu/mm5/) and the ECMWF model. More recently, the community support for MM5 has been diverted to a new model, namely the Weather Research Forecast (WRF) meteorological model (http://www.wrf-model.org/).

Key input data for those mesoscale meteorological models can be derived from local meteorological observations (e.g. SAIMM) or from National Centres for Environmental Prediction (NCEP) final (FNL) (global NCEP FNL) and reanalysis data (e.g. MM5 and WRF). The most recent models used lateral boundaries and initial meteorological conditions taken from the NCEP FNL global analyses that are available at 1° × 1° grid resolution for every six hours (http://rda.ucar.edu/datasets /ds083.2/). Each model is normally equipped with a module to process the geographical input data (i.e. land use, vegetation index, soil type, albedo, etc.) which can be obtained from the global dataset. Physics options are provided in the models to represent physical processes that occurred in the atmosphere: microphysics, land-surface scheme, Rapid Radiative Transfer Model (RRTM), planetary boundary layer (PBL) parameterization scheme and convection scheme. These schemes should be scrutinized to select the most suitable for a specific region of interest (Jankov *et al.* 2005; Osuri *et al.* 2012).

4.2. Model architecture of chemistry transport model

A more detailed description on structure, architecture and data management is provided in this section for the chemistry transport model (CTM) with the mesoscale meteorological modeling considered as a supporting tool to provide meteorological input data for CTM operations.

4.2.1. *Conceptual framework and structure*

CTMs require input data including emissions, topography and meteorology that determine the processes of formation of atmospheric pollutants and their accumulation in a domain. Two main types of CTMs exist, namely the Eulerian and Lagrangian. However, most of the currently used models belong to the Eulerian model type, which uses a fixed reference/observer with respect to the earth and calculates the pollutant concentration in each cell by using equation [4.1]:

$$\boxed{\text{Change C with time in a cell}} = \boxed{\text{Advection by average motion}} + \boxed{\text{Turbulent diffusion}} + \boxed{\text{Molecular diffusion}} + \boxed{\text{Source/sink (E + R + D)}} \qquad [4.1]$$

where "Change C" is the change of the average level of an air pollutant over the time period ΔT, and this is defined as the sum of several processes involving the pollutants listed on the right-hand side of equation [4.1], for example advection described by the net amount of pollutants carried in and out by the 3D wind field, turbulent diffusion caused by 3D wind fluctuations, molecular diffusion and emission source/sink parameters (E – emission, D – deposition, R – reactions). All these processes are considered in every grid cell (differential control volume). Note that in the atmosphere, the molecular diffusion is considered negligible as it is several orders of magnitude smaller than the turbulence diffusion. The atmospheric turbulent diffusion of the pollutants through the control volume is caused by wind fluctuations in x, y and z directions respectively. More details of each term are provided in Kim Oanh and Permadi (2009).

Equation [4.1] is then expressed in a mass continuity differential equation [4.2] that relates the changes in the pollutant concentration with time due to dispersion (i.e. advection and turbulent diffusion), chemical reactions, deposition and emissions.

$$\frac{\partial f}{\partial t} + \nabla(uf) = \nabla(k\nabla f) + P - L \qquad [4.2]$$

where f is the vector containing concentrations of all species in every grid, u is the 3D wind vector, k is the eddy diffusivity, P is the formation due to chemical reactions and emissions, and L is the decay due to deposition and chemical reactions.

The CTMs attempt to solve equation [4.2], but different models have their own software architecture to solve the numerical equations. The general structures of CHIMERE and CAMx are given as examples in Figure 4.1.

Overall, each CTM consists of at least five major modules: emission, meteorology, boundary and initial condition, geographical data and chemistry mechanisms. CAMx has a core model, which is supported by the pre-processors interface that processes the output of the five major modules before getting into the core process. CMAQ, another USEPA model, has a similar structure to CAMx but with an additional emission processing module of SMOKE.

In CHIMERE, the whole system is driven by the top calling script *chimere.sh*, which reads parameters of the simulation, links all necessary files into a temporary directory where all programs are executed, compiles all necessary code, runs the three dynamic pre-processors (meteorology, emission and boundary conditions) and runs the model itself. It has the following sub-directories:

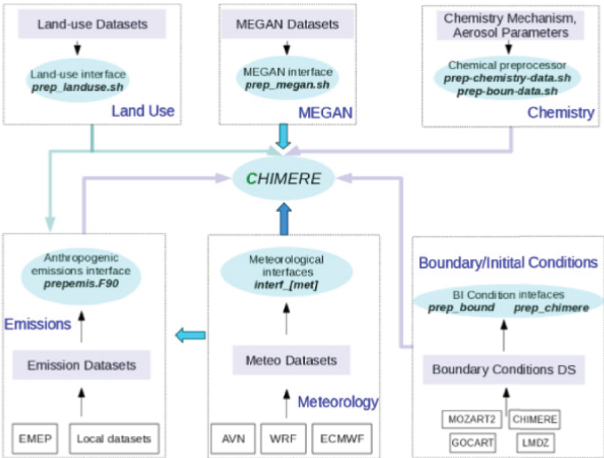
– *Chempreps*: data files for gas, aerosols and chemistry;

– *Domain*: set coordinates of the domain;

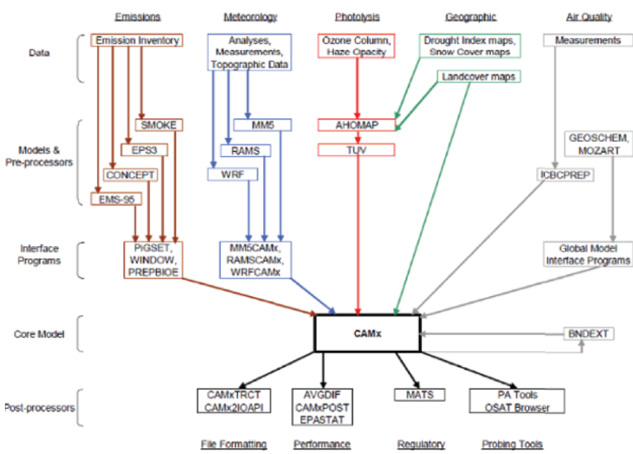
– *src/*: source code of CHIMERE;

– *Util/*: some utility programs.

One important feature in CTM is the chemistry mechanism. Many models now have several choices of chemistry mechanisms incorporated. CTM ideally requires a numerical simulation of all involved chemical reactions, but this would be practically impossible due to a huge number of reactions involved. To avoid excessive simulation time and cost, several chemistry mechanisms have been developed which commonly use condensed computationally affordable kinetic schemes. Accordingly, the chemistry mechanism can use surrogate species where a few volatile organic compounds (VOCs) are selected as the representative of the respective VOC class, or lumped molecule where the VOC from the same class are lumped together and a hypothetical species is then defined to represent the entire class. An example of the surrogate species type is the SAPRC mechanism (Byun and Ching 1999), while an example of lumped molecule type is the RADM version 2 (RADM-2) (Stockwell *et al.* 1990). The most popular mechanism is the lumped structure in which the VOC classification is based on their hydrocarbon (HC) chemical bonds and reactivity, for example CB Mechanism versions 4 and 5 (CB-IV and CB-V respectively) (Dodge 2000). CAMx, WRF/Chem and CMAQ have all the mentioned mechanisms incorporated. However, the CHIMERE model has its own chemistry mechanism, namely the MELCHIOR mechanism (Middleton *et al*. 1990), which consists of 33 species including trace gases and aerosol.



a) CHIMERE



b) CAMx

Figure 4.1. *General model structures of CTMs: (a) CHIMERE (www.lmd.polytechnique .fr/chimere/) and (b) CAMx (http://www.camx.com/home.aspx). For a color version of this figure, see www.iste.co.uk/laffly/torus3.zip*

4.2.2. *Data flow and processing*

Figure 4.2 presents the main framework on how the data flow in most of the CTMs.

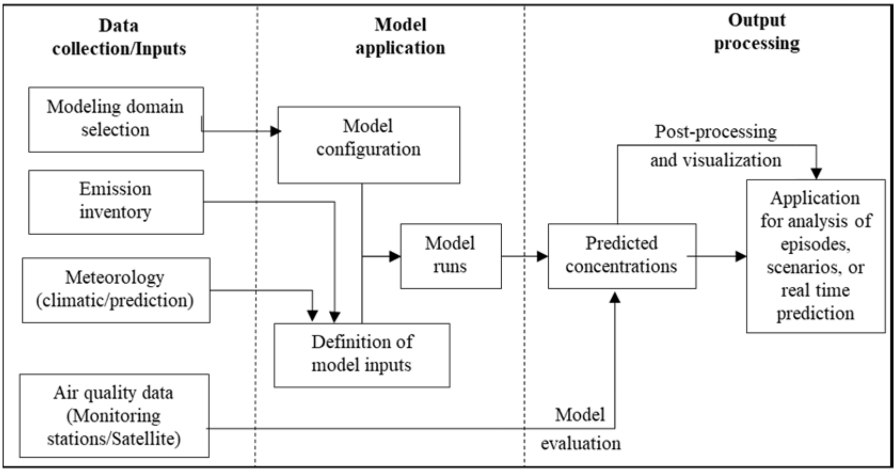


Figure 4.2. *Framework of air quality dispersion model application with data flows (adapted from Kim Oanh and Permadi 2009)*

4.2.2.1. *Input/output (I/O)*

CTMs require 3D meteorological fields, which are supplied by a meteorological model (e.g. MM5, WRF, etc., as described above), and include wind speed and direction, temperature, pressure, water vapor, cloud cover, rainfall and vertical diffusivity. Advection and diffusion processes in the atmosphere are largely affected by the meteorological fields. These meteorological fields further interact with the chemistry mechanisms to treat sources and sinks in the main module of CTMs. Thus, there is a transfer of big 3D meteorological input data, and it is normally managed using a CTM-meteorological module. For example, CMAQ uses a meteorology chemistry interface processor (MCIP), while CAMx uses a MM5-CAMx or WRF-CAMx module as an interface program to handle various meteorological parameters and to only screen the necessary ones to be used for further calculations. These interface programs work in an "offline" mode, which means that 3D meteorological parameters need to be prescribed before transferring the data to be processed by the main module of CTMs. The stand-alone simulation by the mesoscale meteorological models is used in this case. The advantage is that the meteorological data can be evaluated prior to the CTM run so that necessary improvements can be done to reduce potential bias caused by the meteorological input. However, this creates a disadvantage involving massive data transfer, which in turn would increase computational power and time for processing.

A recent CTM model of WRF/Chem can generate meteorological fields simultaneously with the simulation of the chemistry; hence, the total simulation time can be reduced. However, this "online" mode could not isolate potential bias caused by meteorological input parameters alone, but rather the results of simulation are the interaction between the meteorology and chemistry. The later application is however useful to study the interaction of air pollution and climate change, for example.

Other required input data include emission, boundary and initial conditions of air quality, land use and albedo, O_3 column and photolysis rate. Each mentioned dataset should be pre-treated by a stand-alone module, and often users need to develop their own program to prepare emission input data while others are provided in the model source code repository. Model-ready input data should be stored in the provided folders when running the main process, which will call all the data from the respective pathway.

The typical output of the CTM is 3D time-varying average concentration files for user-selected species for surface layer and all other layers of the master grid file and fine grid file (if nesting is applied). An instantaneous concentration file is also generated, for example the output of the last two hours of simulation for model restart. The *I/O* file types are model-specific, i.e. CAMx requires binary input and produces binary output, while CMAQ requires *netcdf* and binary input and produces binary output. CHIMERE uses a *homogen* data format in *netcdf* that requires a specific tool for visualization and manipulation. The common tools are CHIMPLOT for visualization and NCO/CDO for *netcdf* data processing. The Fortran program is commonly used to process the binary output from CAMx and CMAQ.

4.2.2.2. *Main process*

The main processing of CTM can only be done after all the input data are ready in the provided places (mainly located in a working folder). A "*run*" folder contains a job script that will compile the executable file and call the data prepared earlier in the working folder. Model configuration such as domain definition (vertical and horizontal), time for simulation and atmospheric mechanisms should be declared in this job script. The pathway for *I/O* should also be defined in this file.

The main process simulates the codes that are compiled, including dynamical, physical and chemical processes of the model represented by equation [4.2]. There is considerable time required to process input data, computation, as well as to write the output to the output file. However, 85–95% of the total time is required to finish the computational tasks while others are considered to be minor. Depending on the model configuration and computer architecture, one model can perform better than others. To illustrate this, here is an example: a recent version of CHIMERE requires a total simulation time of seven hours to simulate 10 days of air quality in the Southeast Asia (SEA) domain (30 km × 30 km) with a total number of grids of almost 40,000 using an HPC with 32 processors in an application at AIT, Thailand (Permadi *et al.* 2018). CHIMERE processes biogenic emissions online using the global land-cover and meteorological data; hence, more time is required. For a smaller domain, a metropolitan city in SEA, with 400 grids, CAMx required 4 hours for 30 days air quality simulation using the cluster computing with 8 processors available at AIT (Ha Chi 2018). Note that for the CAMx case, all emission data need to be *a priori* prepared and no emission calculation is incorporated in the main process.

4.3. Output data processing

4.3.1. *Output data processing*

As mentioned above, the output format of CTMs varies depending on the model. Therefore, the output data need to be processed/manipulated for various purposes such as visualization, performance evaluation or other analyses such as trend, status and emission scenarios. For example, CAMx and CMAQ generate output files with the binary format that cannot be directly extracted. Therefore, to manipulate the file, a Fortran program can be prepared to convert the file to a text file. The most common tool to visualize the binary output is the open-source tool of the Package for Analysis and Visualization of Environmental data (PAVE) or the Visualization Environment for Rich Data Interpretation (VERDI). However, the support for PAVE was stopped in 2008 and the community continues the online support for VERDI. The software is also capable of converting the binary files to the text file directly. Note that the software has size bandwidth, which means the too large files should be split into several smaller files. The data transfer is also required for this process; hence, it may take time and spaces. CHIMERE and WRF/Chem generate output files in *netcdf* format that are easier to view since there are some readily available software to visualize this file format, such as NCVIEW or NCAR Command Language (NCL). There are also some ready package tools to manipulate/process this file format, such as *netcdf* operators (NCO) or climate data operator (CDO), which can process the data directly.

By using these tools, the output data which are normally generated in a large size (hundreds of megabytes to a maximum of two gigabytes) can be handled and used for various data analyses and presentation. However, this process requires data transfer and additional spaces are required to store the intermediate and output files that can be double in size. It is worth mentioning that the whole process deals with big datasets and thus needs to be well managed. For example, if the data are to be publicized for public awareness of air quality forecast during a forest fire event, they may need to be transferred to a mobile application. Hence, the data should first be transferred to Internet of Things (IOT) format. To manage this, CC is a promising interface which can also deal with big data management. The above-mentioned post-processing tools should be prerequisitely used, for example, to enable data streaming in a cloud-compatible environment which in turn can be publicized.

4.3.2. *Model performance evaluation*

Model performance evaluation is normally done using both statistical and graphical methods to assess the modeled pollutant concentrations against the observed concentrations. Model performance evaluation is a must step which assesses if the model is adequately representing the real situation and analyzes the accuracy of the prepared input data (emission and meteorological fields). The air quality model performance can be statistically evaluated using a set of criteria, for example the USEPA (1991) guidelines for ground-level O_3 (Table 4.2) and that proposed by Boylan and Russell (2006) for PM (Table 4.3). Besides the ground-level observations, the satellite measurements such as the aerosol optical depth (for PM) are also useful for the model performance evaluation.

Statistical measure	**Equation**	**Suggested criteria**
Mean normalized bias (MNB)	$MNB = \frac{1}{n}\sum_1^n \left[\frac{(C_p - C_o)}{C_o}\right]$	±15%
Mean normalized error (MNE)	$MNE = \frac{1}{n}\sum_1^n \left[\frac{\lvert C_p - C_o \rvert}{C_o}\right]$	35%
Unpaired peak prediction accuracy (UPA)	$\frac{C_{p\max} - C_{o\max}}{C_{o\max}}$	±20%

NOTE.– n is the number of monitoring stations or number of measurement times. C_o is the observed values. C_p is the predicted value at monitoring station i for hour t. $C_{o\,max}$ is the maximum hourly observed concentration over all hours or all monitoring stations. $C_{p\,max}$ is the maximum hourly predicted concentration over all hours or all surface grid squares.

Table 4.2. *USEPA (1991) recommended statistical measures for air quality model for O_3*

Parameters	Formula	Suggested criteria
Mean fractional bias (MFB)	$MFB = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(M_i - O_i)}{(M_i/2 + O_i)}\right] \times 100\%$	≤ ±30% (goal) ≤ ±60% (criteria)
Mean fractional error (MFE)	$MFE = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\lvert Mi - Oi\rvert}{(M_i/2 + O_i)}\right] \times 100\%$	≤ 50% (goal) ≤ 75% (criteria)

NOTE.– *O* is the observation value, *M* is the model prediction value and *N* is the number of observations.

Table 4.3. *Evaluation criteria for model performance for PM air quality simulation (Boylan and Russell 2006)*

Using a separate dedicated tool for the statistical analyses mentioned above would be resourceful in terms of data transfer and processing. For example, using simple Microsoft EXCEL or IBM SPSS® would require a huge data transfer, especially to work on two different operating systems (e.g. from Linux to Windows). It would be useful to have cloud-based statistical processing to reduce data transfer and time for analysis.

4.4. Potential applications of cloud computing in atmospheric modeling

4.4.1. *Current status of cloud computing applications in atmospheric modeling*

CC is defined as a computing environment where the computing needs of one party can be outsourced to another party and when the computing power is needed, it can be accessed via the Internet (Jadeja and Modi 2012). Thanks to the large-scale proliferation of the Internet around the world, applications can now be delivered as services over the website, and this reduces the overall cost.

Atmospheric models generally require a large amount of central processing unit (CPU) power, which increases drastically for a big geographical domain with a high resolution (Goga *et al*. 2017). A large number of computing resources are required to satisfy the computational demand through infrastructures such as clusters in grid or cloud. The traditional solution using a high-performance computing (HPC) system requires a considerable cost to build the cluster infrastructure, and its core module should be installed and operated inside the HPC. Recently, all major information technology (IT) players (Amazon, Google, Microsoft, IBM, etc.) have provided public cloud service for users to use their powerful computing resources which is known as Infrastructure as a Service (IaaS). In addition, the most common service recently is pay-as-you-go CC service tailored for atmospheric models including mesoscale meteorological and air quality models, such as WRF and WRF/Chem (http://www.sabalcore.com/vertical-markets/weather-environment/wrf/) and other air quality models (e.g. CALPUFF, AERMOD and CMAQ). These are known as Software as a Service (SaaS). Lakes Environment Co. Ltd. developed the Air Quality Management Information System (AQMIS Cloud), which provides a state-of-the-science cloud solution for emissions management and air quality modeling (https://www.weblakes.com/products/aqmis/index.html). This allows users to manage emissions from thousands of sources, run models, issue permits, forecast air quality impacts and much more, all through the Internet.

CC solution has not been widely applied in developing countries, especially regarding the air quality modeling purposes. Perhaps the public cloud-based services mentioned above do not always provide enough simplicity in the practical applications and some cost is still applied. It is generally known that air quality models are complex, especially CTMs, and the scientific communities in the developing countries still prefer the more traditional HPC solution. IaaS solution can be opted as an appropriate alternative solution to the HPC, and the CC-based solution for air quality modeling is expected to be more widely applied in the future.

4.4.2. *Potential applications of cloud computing in air quality modeling*

From the preceding content of this chapter, it is shown that different types of air quality models often require, to a certain extent, different input data, and generate different output datasets. The purposes of the model applications also determine the types of input data required and the types of output data generated. In addition, the data transfer and data management would be model-type-specific. With the increase in the spatial scale, coupled with more refined grid resolution, and an increase in the temporal scales of model applications, the size of input data and model generated output datasets also increases.

In summary, the following parts of the whole modeling process involve the big data management and the potential CC roles:

– input: meteorological pre-processors, geographical static data processing and emission processing. These are processed separately and generate big data; hence, CC could play an important role;

– main process: compilation of input data, calculation and output data writing. The best solution for this remains with HPC, but future development of HPC in cloud can be a breakthrough;

– output: processing (analysis) and storage. A cloud-based solution would be appropriate for both the processing and the storage of the output.

4.5. Case studies of air pollution modeling in Southeast Asia

Selected case studies of air pollution modeling applications in SEA conducted at AIT are presented in this section to highlight the modeling approach for understanding air pollution problems in the region. The case studies also highlight the difference in geographical and time scale, the purpose of model applications and the size of the input and output datasets associated.

4.5.1. *Modeling air quality in Vietnam*

4.5.1.1. *Domain configuration and input data*

Vietnam, located on the Indochinese Peninsula, is bordered by China, Laos and Cambodia to the north, northwest and southwest respectively. On the east side, Vietnam is bordered by the East Sea. The country is influenced by its tropical monsoon climate, with typical Southwest (SW) monsoon season prevalent from May to October and Northeast (NE) monsoon season from November to April. The CTM modeling system of WRF/CAMx was applied for the simulation of O_3 and PM air quality in Vietnam for one month of August to represent the rainy season and one month of December to represent the dry season of 2010. The CAMx coarse domain covered Vietnam with a resolution of 12 km × 12 km and a fine domain of the Hanoi Metropolitan Region (HMR) (4 km × 4 km) for PM and the Eastern region of Southern of Vietnam (ERS) surrounding Ho Chi Minh City (HCMC) for O_3 air quality, as presented in Figure 4.3. The emission inventory (EI) database for the year 2010 was prepared for major emission anthropogenic sources (on-road mobile, industry, thermal power plants, residential cooking, biomass open burning and gasoline stations) and biogenic (natural) sources (Huy 2015; Huy and Kim Oanh 2017). Meteorological data at the major airports and meteorological stations in Vietnam and the synoptic weather charts were used to evaluate the WRF performance. Available air quality monitoring data (PM and O_3) in Vietnam were used to evaluate performance of the CAMx/WRF modeling system. Both time series and spatial distribution patterns of the pollutants were considered for the model performance evaluation.

The WRF simulations were done for three domains as shown in Figure 4.3. The vertical structure of the WRF domain consisted of 30 sigma layers, ranging from the ground surface level to the top of 15.797 km. The CAMx domain consisted of 15 vertical layers matching the layer interface of WRF layers. The emission data of 10 different air pollutants were aggregated appropriately to the model species using the CB-IV (Fu *et al*. 2004). The profile speciation for VOC and PM was applied to speciate the original 10 species into 30 species. Then, they were also segregated spatially into 81 × 162 grids (for the CAMx coarse domain of Vietnam), and temporally first into the monthly profile for the two simulated months and then hourly for the model input. This results in a dataset of 23 million points of daily emission data input for the coarse domain alone for two months. Likewise, the processing and computing had to deal with more than 1.3 and 5.9 million of emission data points daily for the fine domain of the HMR and ERS respectively. The models were run on a PC/Linux platform installed at AIT. The initial and boundary conditions for CAMx domain were extracted from the study of Permadi (2013) and Permadi *et al*. (2018) who simulated air quality for the entire SEA domain using CHIMERE/WRF.

The WRF-CAMx output data included hourly values of 12 PM species (e.g. fine and coarse fractions: CCRS, CPRM, FPRM; sulfate-PSO_4; nitrate-PNO_3, ammonium-PNH_4; primary organic aerosol-POA; secondary organic aerosol-SOA_1-SOA_5) and one O_3 to analyze the PM and O_3 concentrations respectively. To get the hourly average concentrations of PM and O_3, more than 245, 14.1 and 61.8 million of hourly data points for Vietnam, HMR and ERS domains respectively were processed for two months. Besides the 12 PM species and one O_3 output, there were 19 other species that made a total of 32 species of the model output dataset. In total, there were more than 604 million hourly data points for the coarse domain, and 34.7 and 152.2 million for the HMR and ERS domain respectively for the two-month simulation.

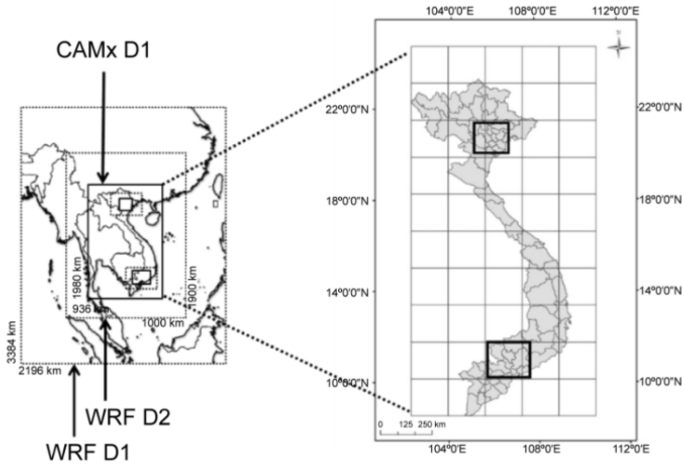


Figure 4.3. *The modeling domain, Vietnam (12 km grid size), Hanoi Metropolitan Region (4 km grid size) and Eastern Region of Southern Vietnam (4 km grid size) (Huy 2015). For a color version of this figure, see www.iste.co.uk/laffly/torus3.zip*

4.5.1.2. *Result highlights*

Monthly gridded average ground-level O_3 concentrations in the Vietnam domain for August and December 2010 were analyzed, which showed lower O_3 concentrations in the center of HCMC and Hanoi, and this was attributed to the titration effect by NO freshly emitted from the traffic.

In principle, the precursors of O_3 were transported from urban areas to downwind locations, and when passing the area with rich biogenic emissions such as forest, more O_3 can be formed. Accordingly, in August, high ground-level O_3 concentrations were shown in the downwind regions (north and northeast of HCMC; and north and northwest of Hanoi) due to the dominant SW monsoon during the month. For example, the SW monsoon directed the plume of O_3 in HCMC to the provinces of Binh Duong and Dong Nai.

In December, Vietnam was affected by the NE monsoon. A high concentration of O_3 was also seen downwind of Hanoi (Nam Dinh, Ninh Binh and northern central coastal regions). The long-range transport of emissions from China would contribute to the high O_3 levels in northern Vietnam. The O_3 plume from HCMC moved south-westerly and showed high concentrations at Ben Tre, Tra Vinh and south coastal regions. As a way of example, the spatial distribution of O_3 in the Eastern region of Southern Vietnam is presented in Figure 4.4.

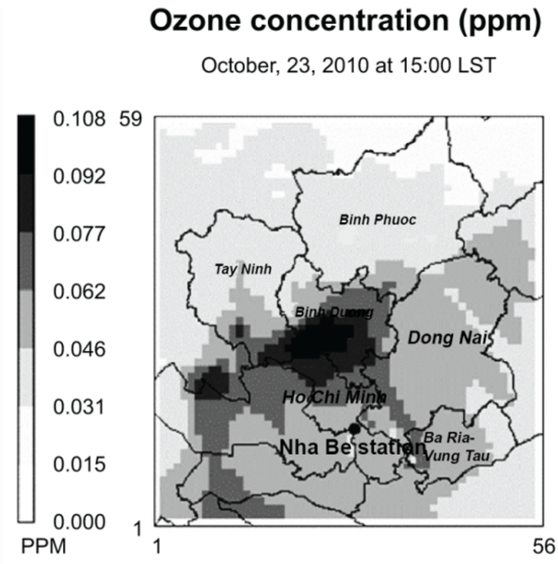


Figure 4.4. *Simulated hourly ground-level O_3 in the ERS domain, Vietnam (Danh* et al. *2016) at 15:00 LST on October 23, 2010*

The highest hourly concentrations of O_3 in the domain of Vietnam were 143 ppb in August and 162 ppb in December, i.e. they exceeded the National Ambient Air Quality Standard (NAAQS) of one-hour average 100 ppb for O_3. The eight-hour O_3 had the maximum, in August, of 97 ppb (Hanoi), and in December, of 122 ppb (HCMC), both exceeding the eight-hour standard for O_3 of 60 ppb (MONRE 2013). High emissions from local sources such as traffic, open burning and industry in these large urban areas would significantly influence the O_3 concentrations.

Monthly average concentrations of $PM_{2.5}$ (fine particles with an aerodynamic diameter less than 2.5 µm) in both months are presented in Figure 4.5. High levels of PM occurred in the surrounding areas of Hanoi and HCMC respectively that show the influence of the intensive emissions from transportation, industrial and construction activities and the biomass open burning (rice straw field burning) in the Red River and Mekong River Deltas. Cooking activities from urban regions with high population density also considerably contributed to the PM pollution (Huy 2015). As expected, higher PM levels were simulated in the dry month of December as compared to those in the rainy month of August. The PM plumes in August and December were shown to follow the main wind directions of the months.

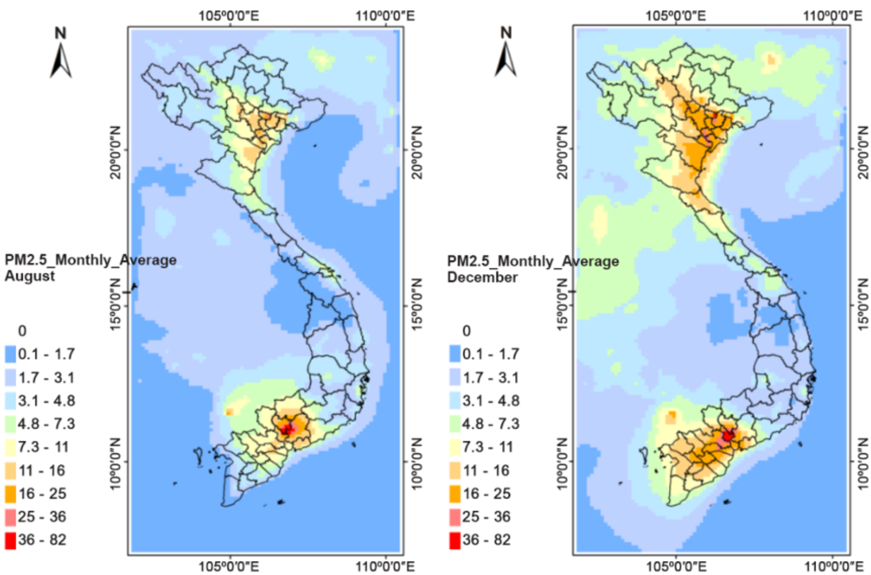


Figure 4.5. *Monthly average of $PM_{2.5}$ for August and December in Vietnam by model (Huy 2015). For a color version of this figure, see www.iste.co.uk/laffly/torus3.zip*

4.5.1.3. *Model evaluation*

Only fragmented observation data were available in Vietnam for the model performance evaluations. The model performance in ozone simulation was done using the time series and scatter plot of the modeled versus observed O_3 concentrations in August and December available at three stations (Da Nang, Nha Be in HCMC and Nguyen Van Cu in Hanoi) in August and two stations (Nha Be and Nguyen Van Cu) in December. The analysis showed that the model underestimated O_3 concentration at Da Nang and Nha Be stations (both months) and overestimated O_3 concentration at Nguyen Van Cu station (a traffic site). The statistical criteria were examined for these stations using the USEPA-suggested parameters of MNB, MNE and UPA, but only some criteria were satisfied, for example MNB at Nha Be and UPA at Nguyen Van Cu in August, while in December none was satisfied. More observation data were still required for in-depth model performance evaluation.

For PM, a statistical analysis of WRF/CAMx model performance for August and December was conducted, which showed that both MFB and MFE were satisfactorily met for the Da Nang station, but none was met for Nha Be and Nguyen Van Cu stations. It is further noted that the modeled values represent the grid-based average while the monitoring data are point-based; therefore, the differences are expected. Lack of the spatially distributed observed data is a key obstacle to evaluate the model performance for both O_3 and PM. There are also uncertainties in observed data and model results that lead to the discrepancy.

4.5.2. *Modeling air quality in the Bangkok Metropolitan Region*

4.5.2.1. *Domain configuration and input data*

The Bangkok Metropolitan Region (BMR) administratively includes Bangkok, the capital city of Thailand, and its five surrounding provinces – Samut Prakan, Nonthaburi, Pathum Thani, Nakhon Pathom and Samut Sakhon. The BMR covers an area of about 7,762 km^2. The WRF-CAMx was used to simulate PM and O_3 air quality in the BMR. Two-way nesting was applied to simulate the interaction between the inner domain and the outer domain. The outer domain thus covered the central part of Thailand and some parts of the Gulf of Thailand (Figure 4.6) and was named as the CENTHAI domain. It had an area of 300×300 km^2 and consisted of 50×50 horizontal grid cells at a 6 km grid resolution. The inner BMR domain had an area of 100×70 km^2 and consisted of 50×35 grid cells at 2 km grid resolution.

Meteorological input data were generated by the WRF simulation at AIT using the Computing on Kepler Architectures (COKA) cluster, a multicore high HPC system installed at the University of Ferrara, Italy, as detailed in Chapter 18 of

Volume 1. The emissions data were prepared from the updated EI for key anthropogenic sources and a natural (biogenic) emission source in the inner BMR domain for 2016. Similar to the case of the simulation for the Vietnam domain above, there were 30 species in the model emission input file to CAMx. The emission data were segregated spatially into 50 × 35 grid cells of the BMR domain and temporally monthly and then hourly for the simulation period of one year (2016). The emission dataset thus consisted of about 15 million data points (24 hours × 12 months × 30 species × 50 × 35 grid cells). The WRF-CAMx output data included hourly values of 12 PM species and O_3, i.e. similar to the case of the Vietnam domain above. To get the daily average concentration of PM and O_3, about 500,000 data points (24 hours × 13 species × 35 × 50 grid cells) needed to be processed; hence, about 16 and 200 million data points were used to compute the monthly and annual average concentration respectively.

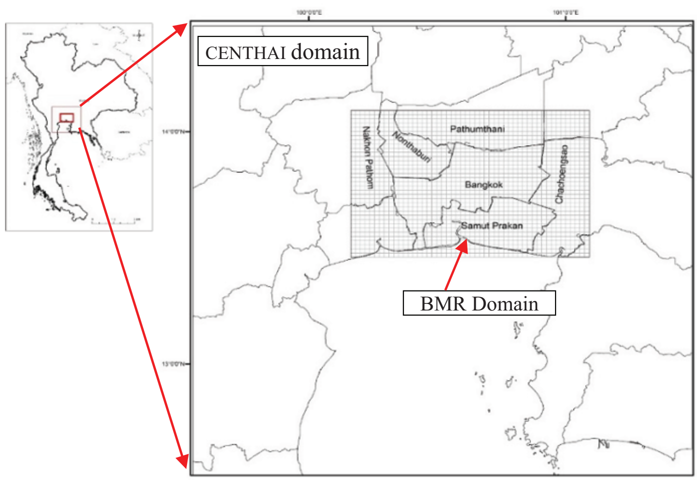
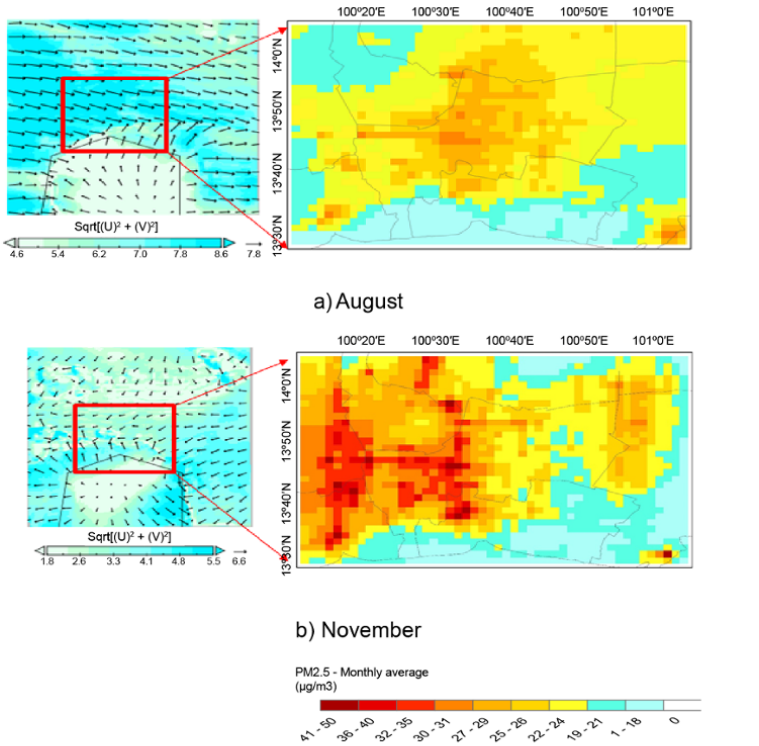


Figure 4.6. *Location and coverage of CENTHAI and the BMR domain*

4.5.2.2. *Result highlights*

The simulated results for PM by WRF-CAMx for 2016, extracted for every month for $PM_{2.5}$ and PM_{10} (particles with an aerodynamic diameter less than 10 µm), showed higher levels in the dry season and lower levels in the wet season, i.e. opposite to the precipitation trend. The monthly average $PM_{2.5}$ in the domain was 11 µg/m^3 in March, 7 µg/m^3 in August and 22 µg/m^3 in November. Spatially, the higher values were found in the center of Bangkok city every month, for example 7–24 µg/m^3 in August and 20–51 µg/m^3 in November. The dispersion plume of $PM_{2.5}$ was influenced by the wind field that in turn reflected the monsoon regime (Figure 4.7).



NOTE.–The values are the monthly average per grid (2 km); wind speed is in m/s

Figure 4.7. *Monthly wind fields and spatial distribution of grid average $PM_{2.5}$ (Ha Chi 2018). For a color version of this figure, see www.iste.co.uk/laffly/torus3.zip*

4.5.2.3. *Model performance evaluation*

The WRF-CAMx performance was evaluated using the statistical criteria. The modeled $PM_{2.5}$ and PM_{10} concentrations of hourly (1 hour), daily (24 hour) average, and weekly averages were compared with the observed data collected from the Pollution Control Department (PCD) stations and from a research project on “A Study on Urban Air Pollution Improvement in Asia” conducted by AIT in collaboration with the Japan International Cooperation Agency (JICA) and the Asia Pacific Clean Air Partnership (APCAP) of Japan.

The model performance was shown to be better for some stations than others. The evaluation for hourly PM showed that all the criteria were satisfied for station ST59 (representing the general area of the BMR) for both $PM_{2.5}$ and PM_{10}, while for station ST27 (industrial estate), the MFB and MFE for PM_{10} were satisfactorily met. The weekly simulated $PM_{2.5}$ and compositions (nitrate-PNO_3, sulfate-PSO_4, organic carbon and elemental carbon) were evaluated against the AIT-JICA project monitoring data and showed that at the PCD site (Bangkok city center), the model can better simulate PNO_3 than PSO_4, whereas at the AIT site (sub-urban area), PSO_4 was better simulated than PNO_3. Modeled elemental carbon (EC) and organic carbon (OC) levels in both sites were lower than observed (Ha Chi 2018). The sparsely available monitoring data would not allow a more comprehensive model performance evaluation. In addition, the observed data are point-based, hence may not be fully comparable with the grid-averaged modeled values.

4.5.3. *Modeling air quality in the Southeast Asia domain*

SEA is a region with a large population and fast-growing economy and is an important contributor to the emissions of air pollution and greenhouse gases in Asia (Zhang *et al*. 2009). The simulation domain covered the Southeast Asia (SEA) region in this case study with the purpose to understand the impacts of emissions, both local and transboundary, on air pollution levels, climate forcing and human health.

4.5.3.1. *Configuration and input data*

CHIMERE was used to simulate 3D aerosol concentrations in the domain using the meteorological fields generated by the WRF model. The domain extended from southern China (24°N, 95°E) to eastern parts of Indonesia (9°S, 137°E) consisting of 169 × 133 grids with eight vertical layers (from 20 to 5,500 m) and a horizontal grid resolution of 0.25° × 0.25° (~30 × 30 km^2). CHIMERE is equipped with the MELCHIOR 2 chemistry mechanism, which consists of about 120 reactions and 40 chemical species. One-year simulation (January 1–December 31, 2007) was performed with a spin-up period of one week prior to the main simulation period. The details of model configuration are given in Permadi *et al*. (2018).

The emissions input was prepared for three countries (Indonesia, Thailand and Cambodia), while for the rest of the domain, the emissions were taken from the available online gridded EI databases (grid size of 0.5°~50 km) compiled by the Centre for Global and Regional Environmental Research (CGRER) (Zhang *et al*. 2009). The hourly speciated emissions of each species were gridded into 0.25° × 0.25° using the Geographic Information System tool. In total, the emission input included 24 hour × 33 species × 169 × 133 (grids) data points. Other model input data included 3D hourly gridded meteorological parameters for the entire one-year simulation period plus 10 spin-up days and the associated gridded geographical data of the domain, i.e. land cover, albedo, etc., as detailed in Permadi *et al.* (2018). The output data included 3D hourly gridded concentrations of 33 chemical species (i.e. trace gases and aerosol) for one year.

4.5.3.2. *Result highlights*

The hourly concentrations of 33 species (i.e. NO, NO_2, HONO, CO, SO_2, NH_3, HCL, SULF, CH_4, C_2H_6, NC_4H_{10}, C_2H_4, C_3H_6, C_5H_8, APINEN, OXYL, HCHO, CH_3CHO, CH_3COE, CH_3OH, C_2H_5OH, GLYOX, MGLYOX, MVK, TOL, TMB, PPM_big, PPM_coa, PPM_fin, OCAR_fin, BCAR_fin, PNO_3 and H_2SO_4_fin) were produced by the model for each grid of the domain. The data were then extracted and processed. For example, the spatial distributions of the modeled monthly average PM_{10}, $PM_{2.5}$ and black carbon (BC) are presented in Figure 4.8 for January and August 2007.

The highest monthly average concentrations of PM_{10} in January and August over the domain were 69 and 58 $\mu g/m^3$ respectively, while the corresponding values of $PM_{2.5}$ were 40 and 37 $\mu g/m^3$. The simulated highest hourly PM_{10} in January and August 2007 were 325 and 245 $\mu g/m^3$ respectively, while the corresponding values of $PM_{2.5}$ were 188 and 150 $\mu g/m^3$. In January, in the Northern Hemisphere, NE monsoon transported pollutants from the source regions to the southwest direction, while in the Southern Hemisphere part (Indonesia), the plume moved to the northeast/east direction. The opposite is seen in August where, in the Southern Hemisphere part of the modeling domain, the plumes of PM moved north-westerly and turned into north-easterly after reaching the equator line. In August, the dry months in the southern domain, the PM_{10} and $PM_{2.5}$ plumes, were shown to reflect the effects of biomass open burning (forest fire and crop residue) emissions in Indonesia that originated in the Riau province (Sumatera Island) and the western and southern parts of the Borneo island, and were seen clearly moving northeast-ward. In January, a dry season month in the northern part of the domain, the plumes of PM_{10} and $PM_{2.5}$ were intensified by biomass open burning in the central and northern areas of Thailand, which were seen moving southwest-ward. The simulated maximum monthly average BC concentration in the domain was somewhat higher in January (8.2 $\mu g/m^3$) as compared to August (7.8 $\mu g/m^3$). The BC plumes were generally seen originating from large cities in the domain that confirms the significant influence of the fossil fuel combustion emission, specifically traffic, and other urban activities for all months of the year. During the dry period, BC plumes from the areas were also seen being intensified by biomass open burning emissions but were not as clearly shown as the PM plumes. This may be because biomass open burning contributed more to organic carbon (OC) than to BC emissions.

4.5.3.3. *Model evaluation*

The daily (24-hour) modeled PM_{10} concentrations were calculated using the hourly output data, and the results were compared with the 24-hour data gathered from the governmental monitoring networks available in three large cities of SEA (i.e. one station in Kuala Lumpur, two stations in Bangkok and one station in Surabaya). For BC, the 24-hour BC measured by the optical method available at several SEA sites under the AIRPET project (Kim Oanh *et al*. 2013b) was used. The statistical evaluation of PM_{10} showed that in all three cities, the MFB and MFE values for 24-hour PM_{10} and BC satisfactorily met the suggested criteria (Permadi *et al*. 2018). The model system was used to analyze the effects of emission reduction strategies in the key sectors in Indonesia and Thailand on BC levels in 2030 and potential co-benefits on air quality (and health) and climate forcing in SEA region (Permadi *et al*. 2018).

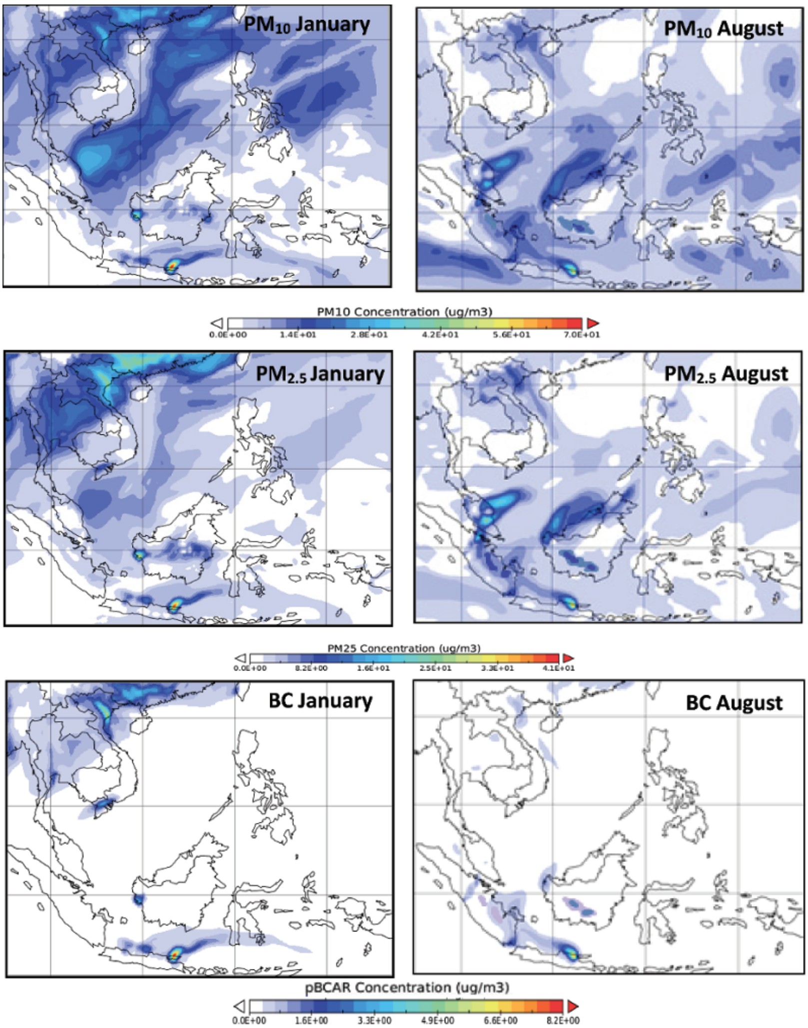


Figure 4.8. *Spatial distribution of monthly average PM_{10}, $PM_{2.5}$ and BC in the selected months, 2007. For a color version of this figure, see www.iste.co.uk/laffly/torus3.zip*

4.6. Summary and conclusion

Atmospheric dispersion modeling is widely used to simulate air quality in various domains. The modeling tool is essential to link the causes (emissions, meteorology) to the effects (air quality and associated effects on health and ecosystems, and climate forcing), hence can be applied to support air quality management activities by various stakeholders, i.e. scientific communities, policy makers and the general public, among others. Many types of atmospheric dispersion modeling systems are available worldwide. Such systems included meteorological models (i.e. MM5, WRF, etc.) that generate meteorological fields to drive the air quality models (i.e. CAMx, CMAQ, GEOS-Chem, CHIMERE, etc.). Several case studies conducted by the AIT air quality research group are presented in this chapter to illustrate the modeling applications and magnitude of input–output data size. These include the urban scale (i.e. BMR Thailand by using WRF/CAMx), the country scale (i.e. Vietnam by using WRF/CAMx) and the regional scale (SEA by using WRF/CHIMERE).

Two important types of input data, i.e. meteorological data and emission data, are required for most of the atmospheric modeling applications. The input data are required for every grid in the domain with a grid size range of 0.10–0.5° for the large regional domain to as small as 2 km or lower for urban and local scale domains. Temporally, the air quality models often require hourly input data for the required pollution species. The number of input species varies with the model system and purpose of the applications; for example, for O_3 and PM, simulations using CTMs normally require more than 20 species, including gases and aerosol. The output datasets normally consist of hourly levels of modeling species for every grid in the domain, not only on the surface layer but also in every vertical layer. This results in the handling and transferring of big datasets where CC could play important roles. The model output quality cannot be better than the input quality; hence, the uncertainty related to meteorology and emission input data should be scrutinized and reduced. Evaluation of the model performance is done both quantitatively and qualitatively to ensure the accuracy of the model output.

Overall, the big datasets of both inputs (meteorology and emissions) and output (simulated air quality) require large storage resources that may be potentially handled by CC. The cloud computing for big dataset management, especially for air quality data has emerged as one of the potential tools for storing, processing and publicizing data in an efficient way.

4.7. References

Bosanquet, C.H. (1936). The spread of smoke and gas from chimneys. *Trans. Faraday Soc.*, 32, 1249.

Boylan, J.W. and Russell, A.G. (2006). PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models. *Atmos. Environ.*, 40(26), 4946–4959.

Byun, D.W. and Ching, J.K.S. (1999). Science algorithms of the EPA models-3 community multi-scale air quality (CMAQ) modelling system. *NERL*, Research Triangle Park, North Carolina, USA.

Carter, W.P.L. (2000). Implementation of the SAPRC-99 chemical mechanism into the Models-3 framework. Report, US Environmental Agency, USA.

Daly, A. and Zannetti, P. (2007). Ambient Air Pollution. In *Air Pollution Modelling – An Overview*, Zannetti, P., Al-Ajmi, D., and Al-Rashied, S. (eds). The Arab School for Science and Technology and the EnviroComp Institute, Fremont, CA, USA.

Danh, N.T., Huy, L.N., and Kim Oanh, N.T. (2016). Assessment of rice yield loss due to exposure to ozone pollution in Southern Vietnam. *Sci. Tot. Environ.*, 566–567, 1069–1079.

Dawson, J.P., Racherla, P.N., Lynn, B.H., Adams, P.J., Adams, J., and Pandis, S.N. (2008). Simulating present-day and future air quality as climate changes: Model evaluation, *Atmos. Environ.*, 42, 4551–4566.

Dodge, M.C. (2000). Chemical oxidant mechanisms for air quality modelling: Critical review. *Atmos. Environ.*, 34, 2103–2130.

Emmons, L.K., Walters, S., Hess, P.G., Lamarque, J.F., Pfister, G.G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S.L., and Kloster, S. (2010). Description and evaluation of the model for ozone and related chemical tracers, version 4 (MOZART-4). *Geosci. Model Dev.*, 3, 43–67.

Fu, J.S., Jang, C.J., Chen, C., and He, K. (2004). Attachment 8: Project for the MCNC/US: ICAP 2 Project-Task 8 pilot project of national/regional air quality modelling assessment in China (MCNC contract no. 10176.003). Environmental Protection Agency, USA.

Goga, K., Parodi, A., Ruiu, P., and Terzo, O. (2017). Performance analysis of WRF simulations in a public cloud and HPC environment. In *Complex, Intelligent, and Software Intensive Services*, Barolli L., and Olivier T. (eds). Springer, Turin, Italy.

Ha Chi, N.N. (2018). Assessment of particulate matter, surface ozone air quality and associated health burdens in the Bangkok Metropolitan Region using a photochemical grid model. Master Thesis, Asian Institute of Technology, Bangkok, Thailand.

Huy, L.N. (2015). Evaluation of performance of photochemical smog modelling system for air quality management in Vietnam. Master Thesis, Asian Institute of Technology, Bangkok, Thailand.

Huy, L.N. and Kim Oanh, N.T. (2017). Assessment of national emissions of air pollutants and climate forcers from thermal power plants and industrial activities in Vietnam. *Atmos. Pollut. Res.*, 8(3), 503–513.

Jadeja, Y. and Modi, K. (2012). Cloud computing – Concepts, architecture, and challenges. *International Conference on Computing, Electronics and Electrical Technologies*, ICCEET, Tamil Nadu, India.

Jankov, I., Gallus Jr., W.A., Segal, M., Shaw, B., and Koch. S.E. (2005). The impact of different WRF model physical parameterizations and their interactions on warm season MCS rainfall. *Weather Forecast.*, 20(6), 1048.

Kim Oanh, N.T. and Permadi, D.A. (2009). Photochemical smog modelling for ozone air quality management. In *Modelling of Pollutants in Complex Environmental Systems*, vol. I, Hanrahan, G. (ed.). ILM Publications, Hertfordshire, UK.

Kim Oanh, N.T., Permadi, D.A., Zhang, B.N., Huy, T.N.Q., Phuong, N.L., Kanabkaew, T., and Iqbal, A. (2013a). Applications of photochemical smog models for assessment of ozone, particulate matter air quality, and acid deposition in Asian cities. In *Integrated Air Quality Management: Asian Case Studies*, Kim Oanh, N.T. (ed.). CRC Press, Florida, USA.

Kim Oanh, N.T., Pongkiatkul, P., Templonuevo Cruz, M., Nghiem, T.D., Philip, L., Zhuang, G., and Lestari, P. (2013b). Monitoring and source apportionment for particulate matter pollution in six Asian Cities. In *Integrated Air Quality Management: Asian Case Studies*, Kim Oanh, N.T. (ed.). CRC Press, Florida, USA.

Middleton, P., Stockwell, W.R., and Carter, W.P. (1990). Aggregation and analysis of volatile organic compound emissions for regional modelling. *Atmos. Environ.*, 24, 1107–1133.

MONRE (2013). National technical regulation on ambient air quality. Ministry of Natural Resources and Environment, Vietnam [Online]. Available: http://www.monre.gov.vn/ [Accessed 31 May 2018].

Osuri, K.K., Mohanty, U.C., Routray, A., Kulkarni, M.A., and Mohapatra, M. (2012). Customization of WRF-ARW model with physical parameterization schemes for the simulation of tropical cyclones over North Indian Ocean. *Nat. Hazards*, 63, 1337–1359.

Permadi, D.A. (2013). Assessment of efficiency of black carbon emission reduction measures and co-benefit on climate forcing mitigation and air quality improvement in Southeast Asia. PhD Thesis, Asian Institute of Technology, Bangkok, Thailand.

Permadi, D.A., Kim Oanh, N.T., and Vautard, R. (2018). Integrated emission inventory and modelling to assess distribution of particulate matter mass and black carbon composition in Southeast Asia. *Chem. Phys.*, 18, 2725–2747.

Ramanathan, V. and Feng, Y. (2008). Air pollution, greenhouse gases and climate change: Global and regional perspectives. *Atmos. Environ.*, 43, 37–50.

Reynolds, S., Roth, P., and Seinfeld, J. (1973). Mathematical modelling of photochemical air pollution. *Atmos. Environ.*, 7, 1033–1061.

Rohde, H. (1972). A study of the sulphur budget for the atmosphere over northern Europe. *Tellus*, 24, 128.

SAI (1995). User's Guide to the Systems Applications International Mesoscale Model. Systems Applications International, California, USA.

Schmidt, H., Derognat, C., Vautard, R., and Beekmann, M. (2001). A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in Western Europe. *Atmos. Environ.*, 35, 6277–6297.

Schneider, S.H., Krenz, J.H., Glantz, M.H., Salby, L.M., Niiler, P.P., Dickinson, R.E., Aber, J.D., Turco, R.P. (1989). *Climate System Modelling*, Cambridge University Press, London, UK.

Simpson, D. (1993). Photochemical model calculations over Europe for two extended summer periods: 1985 and 1989 model results and comparison with observations. *Atmos. Environ.*, 27A, 912–943.

Stockwell, W.R., Middleton, P., Chang, J.S., and Tang, X. (1990). The second generation regional acid deposition model chemical mechanism for regional air quality modelling. *J. Geophys. Res.*, 95, 16343–16367.

Takemura, T., Okamoto, H., Maruyama, Y., Numaguti, A., Higurashi, A., and Nakajima, T. (2000). Global three-dimensional simulation of aerosol optical thickness distribution of various origins. *J. Geophys. Res.*, 105, 17853–17873.

USEPA (1991). Guideline for the regulatory application of the Urban Airshed Model. Report, United States Environmental Protection Agency, Research Triangle Park, North Carolina, USA.

Zhang, Q., Streets, D.G., Carmichael, G.R., He, K.B., Huo, H., Kannari, A., Klimont, Z., Park, S., Reddy, S., Fu, J.S., Chen, D., Duan, L., Lei, Y., Wang, L.T., and Yao, Z.L. (2009). Asian emissions in 2006 for the NASA INTEX-B mission. *Atmos. Chem. Phys.*, 9, 5131–5153.