

Article

# Design of a Model Base Framework for Model Environment Construction in a Virtual Geographic Environment (VGE)

Chunxiao Zhang, Gang Ai \*, Xinqi Zheng \*, Kun Fang and Jiayang Li

School of Information Engineering, China University of Geosciences in Beijing, No. 29, Xueyuan Road, Haidian District, Beijing 100083, China; zhangcx@cugb.edu.cn (C.Z.); 2009010022@cugb.edu.cn (K.F.); frankie1100@126.com (J.L.)

\* Correspondence: aigang@cugb.edu.cn (G.A.); zhengxq@cugb.edu.cn (X.Z.); Tel.: +86-10-8232-1744 (G.A.); +86-10-8232-2116 (X.Z.)

Academic Editors: Milan Konecny and Wolfgang Kainz

Received: 20 February 2017; Accepted: 27 April 2017; Published: 4 May 2017

**Abstract:** The model environment is a key component that enables a virtual geographic environment (VGE) to meet the scientific requirements for simulating dynamic phenomena and performing analyses. Considering the comprehensiveness of geographic processes and the requirements for the replication of model-based research, this paper proposes a model base framework for a model environment of a VGE that supports both model construction and modelling management, resulting in improved reproducibility. In this framework, model management includes model metadata, creation, deposition, encapsulation, integration, and adaptation; while modelling management focuses on invoking the model, model computation, and runtime control of the model. Based on this framework, to consider the problem of ever-worsening air quality, we applied the Linux-Apache-MySQL-Perl stack plus Supervisor to implement the model base to support a VGE prototype using professional meteorological and air quality models. Using this VGE prototype, we simulated a typical air pollution case for January 2010. The prototype not only illustrates how a VGE application can be built on the proposed model base, but also facilitates air quality simulations and emergency management.

**Keywords:** model environment; model base; virtual geographic environment (VGE); air quality management

---

## 1. Introduction

An important priority of geographic analysis has emerged as researchers have shifted from identifying static snapshots to exploring dynamic phenomena and supporting decision- and policy-making, providing early warnings, crisis management, and so on [1,2]. Dynamic phenomena usually include a series of physical or human geographic processes that change across both spatial and temporal dimensions. Such diverse and complex geographic dynamics lead to challenges in connecting the relevant information and understanding the underlying mechanisms [3]. Focusing on dynamic geographic research, computer-aided geographic experiments are becoming an increasingly common strategy for simulating and analysing dynamic phenomena with process models—mainly mathematical abstractions representing the essential features of complex systems and often utilized as basic tools [4,5]. Hartmann has argued that large portions of current research in the natural and social sciences can no longer be imagined without simulations. In particular, simulations implemented on a computer are highly effective methodological tools [6]. In this context, taking the model environment as the most important component, the virtual geographic environment (VGE), as an extension of a

GISystem, has been proposed as the new generation of “geographic language” for studying dynamic geography and a fast-growing trend in geographic research [7–11].

By including both geographic processes and analytical models, a VGE can be used to perform geographic analysis, simulate geographic phenomena, represent and predict changes in the geographic environment, and evaluate human influences on the environment [9]. In this context, a VGE is generally built with two cores—a geo-database and a geographic process model base. In contrast to the spatial analysis (e.g., network analysis) and spatial statistical analysis algorithms employed by previous GISystems, such as ArcGIS, QGIS, etc., geographic process models refer to models that are used for computing dynamic geographic processes. Combining such geographic process models with GISystem-based analyses can improve the understanding of gradual geographic evolution and allow new quantitative and systematic examinations of spatial and temporal patterns of dynamic geographic processes to occur [4,12,13]. Therefore, featuring a model environment along with data, visualization, and interactive and collaborative components, a VGE is considered to be a suitable potential tool for advanced geographic analysis and problem-solving [9,13,14].

With the aim of studying the complex mechanisms of dynamic phenomena, since 2000, extensive research has been published regarding VGEs that has focused on model integration, computing, sharing, etc. Xu et al. developed a platform that integrated a computational grid and a VGE system with a meteorological model to improve model computational efficiency [15]. To enable researchers in various fields to collaboratively produce computer-aided geographic environments, researchers have developed collaborative model environments for conducting virtual flood experiments and air pollution simulations [16–18]. Concerning human geography, Song et al. combined pathogen diffusion and crowd simulation models in a VGE to simulate crowd evacuations for urban management [19]. In addition to the aforementioned case studies, several researchers have focused on the theoretical underpinnings of model environments to solve dynamic geographic problems. Lin et al. suggested that technologies related to model management, reuse, and coupling should be explored to support collective modelling and simulation in VGEs [4]. To assist modellers and model users in choosing an appropriate modelling approach for their integrated modelling, Kelly et al. compared five common approaches that are able to integrate different models regarding model development, spatiotemporal details and treatment of uncertainty [20]. With regard to model sharing, different modelling and model reuse frameworks have been developed as open environments to take advantage of cloud computing [21].

Although these case-based research and theoretical achievements are valuable in enhancing the simulation capacity of VGEs to reproduce or predict geographic processes, they have mostly focused on model integration, sharing and reuse. However, reproducibility, which is a key problem in model-based research, is rarely mentioned because researchers have argued that in geography or ecology, where models tend to be more complex and more closely linked to real systems, modelling starting with replication is virtually absent [22]. Axelrod stated that “Replication is one of the hallmarks of cumulative science. It is needed to confirm whether the claimed results of a given simulation are reliable in the sense that they can be reproduced by someone starting from scratch” [23]. To replicate an existing simulation, sharing modelling information is required. Such sharing includes not only the model itself but also the modelling information regarding how to run it and how to understand its output files so that the whole simulation process can be traced correctly. To offset this replication limitation in VGEs, researchers developed a model base that includes both model sharing and model management [9,13]. For example, Zhang et al. developed a VGE prototype designed to share both data and modelling parameters so that different users could reproduce modelling scenarios by retrieving the database [24]. Given that modelling management technologies have contributed to the reproducibility in VGEs, researchers have increasingly realized the potential contribution of the model base theory and modelling management for improving the model reuse and modelling reproducibility of VGE models [13,25].

This paper aims to design a framework employing model base theory for model environments of VGEs that supports model and modelling management to improve replication by enhancing model sharing, reuse and modelling traceability. The remainder of this paper is organized into four sections. We first discuss related works concerning (1) the research background and scope of reproducibility in model-based research and (2) the development of a model base and model management for geographic process modelling in Section 2. In Section 3, we propose a conceptual and theoretical framework employing model base theory for designing model environments in VGEs that considers model creation, decomposition, encapsulation, sharing, modelling management, and so on. Section 4 describes the implementation of a VGE prototype built using the Linux-Apache-MySQL-Perl (LAMP) stack and Supervisor to realize the framework design in a model environment, using a case study of air quality management in the Pearl River Delta (PRD), China. This paper culminates with conclusions and a discussion of the challenges in VGE model-based research and development.

## 2. Related Works

### 2.1. Reproducing Model-Based Research

Recognizing that the number of research projects using computational models has steadily increased, the resulting publications often lack a model code and documentation, which hinders replication, the verification of results, and the accumulation of knowledge. A key criterion in science is transparency—the potential for others to replicate a set of research procedures and simulation results. Transparency becomes ever more important as the systems we analyse become more complex [26]. Balaban emphasized a need to increase trust in models, a concept related to validity, replicability and tractability [27]. In contrast to the definition of replication as “experiment replication,” Wilensky and Rand defined replication “as the implementation (replicated model) by one scientist or group of scientists (model replicators) of a conceptual model described and already implemented (original model) by a scientist or group of scientists at a previous time (model builders)” [28]. This conception is known as “model replication” and is tested by comparing the output of the re-implementation with the output of the original implementation [22]. Model replication can lead to several benefits. Zhong and Kim concluded that replication can improve agent-based model reliability [29]. Thus, the model environment should have multiple levels by which not only the information regarding the model itself can be managed but also model execution, with respect to the data used and parameterizations (e.g., the spatial resolution of the air quality model used to compute different air quality indices) [30].

Based on the studies listed above, it can be concluded that, to replicate model-based research, it is necessary to have a sufficiently detailed description of the original computational model, its operation and objectives, as well as an accessible runtime environment (or sufficient information to recreate such an environment) in which to reproduce model-based research procedures [31]. The use of software engineering best practices to increase model transparency and improve research reporting practices, disclosing all the necessary data, improves comparisons of simulation results. With such detailed information of a simulation, Seagren conducted a successful replication study of an agent-based computational model and noted that the detailed information permitted them to additionally implement a richer and more complex version of the original model, leading to improved insights regarding the studied phenomena [32].

### 2.2. Model Base and Model Management

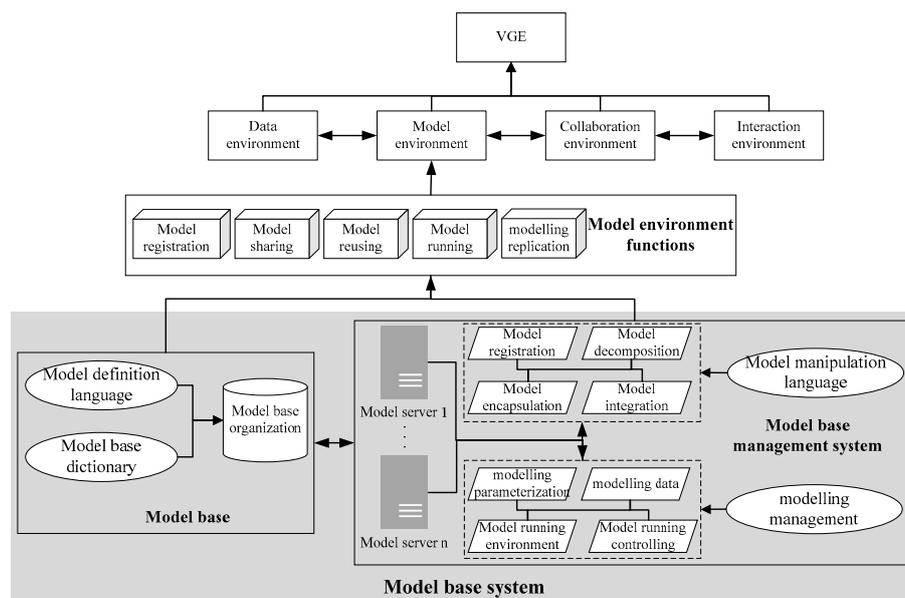
The terms model library [31], the Network for Computational Modelling in the Social and Ecological Sciences (CoMSES Net) Computational Model Library (CML) [33], and model bank [34] are treated as having the same meaning in this study and are referred to as “model base,” which is an organizational collection of models contained in an electronic storage medium accessible to users and programs [35]. The models in the model base are expected to be reused as building blocks for more complex systems, provided that they can be linked together in a meaningful way (matching the

variables, scales, etc.) [26]. As a valuable supplementation of the archived models in a model base, which is intended for sharing source code, documentation, and so on, managing the organizational models in a model base involves a wide variety of functions. These may include model creation and editing, querying and updating models, executing models, and generating reports so that others can replicate a set of modelling procedures [31,36,37]. The term model management was coined in the early-to-mid-seventies in the context of managing models for a decision support system, aiding in the creation, storage, retrieval, manipulation, and utilization of models for decision-makers, so they could make better use of the models [5,38,39]. These functions are analogous to those of a database management system, where the objects of interest are data records instead of models. Thus, a model base management system (MBMS) should be able to store and manipulate models in ways similar to the methods database management systems (DBMS) store and manipulate data [34]. Therefore, it is not surprising to see the strong influence of database technology in the creation of model management systems [35,40].

The analysis above shows that one of the key assumptions underlying an MBMS is that models, like data, are important organizational resources and should be managed with as much rigor and attention as data [5]. Thus, an MBMS should provide functional counterparts of a DBMS, including model description, model manipulation, and model control, although an MBMS is expected to be much more complex than a DBMS [35]. When developing VGEs, researchers have also mentioned the need for a model base frequently in recent years, but investigations of model base theory and of MBMS for VGE models is still in progress [9,13]. Based on this foundational research, it is reasonable to investigate the model base and MBMS and design the framework for a VGE model to improve the replication capabilities of model-based research.

### 3. Model Base Framework for Model Environments

The model component is the main component of a VGE because the fundamental goal of a VGE is to simulate and investigate dynamic processes by evolving integrated models [41]. Meanwhile, the model component is extremely relevant to the remaining three components, as was thoroughly analysed by Zhang et al. [13]. In this context, the model environment in a VGE focuses on model registration, sharing, reusing and execution [9]. Considering these four aspects, as well as the replication of model-based research, the model base framework is designed as shown in Figure 1.



**Figure 1.** Model base framework for the model environment portion of virtual geographic environments (VGEs).

### 3.1. Model Base and Model Organization

#### 3.1.1. Model Definition Language

Models in the model base are defined using model definition language (MDL), yielding a model definition which describes the model information instead of model implementation. Model definitions are mostly relatively independent from model implementation and execution, which occur in model servers. Using MDL, model providers can describe models in a flexible and structured manner and model users can obtain this descriptive information. As a standardized language, MDL reduces the heterogeneity of model descriptions.

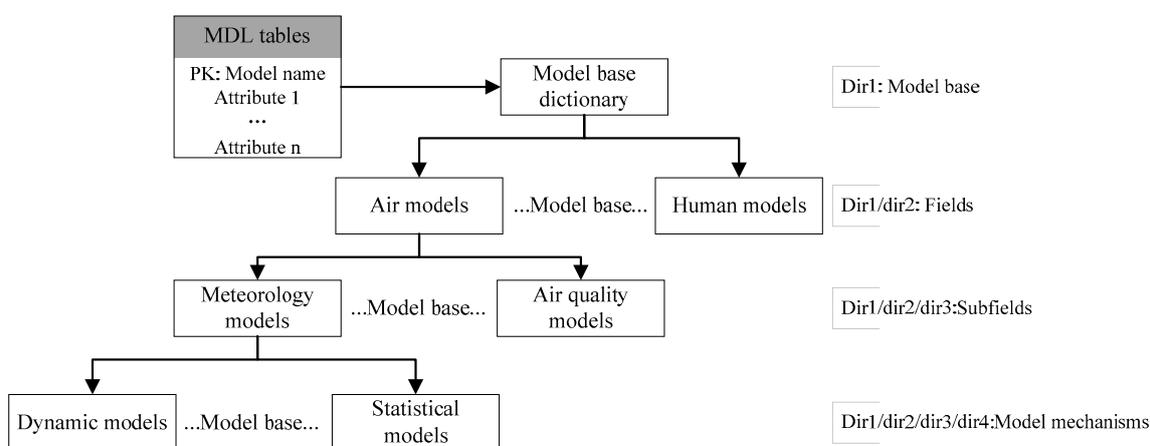
In the model base design, model base managers can define the MDL grammar. An example is given in Table 1. Implementing a model environment in a VGE, developers can define models according to MDL grammar accordingly in the formatting of tables.

**Table 1.** One instance of model definition language (MDL) grammar.

MDL Sections	Description
Model name	The model name
Description	A flexible expression of the model's design, problem focus, etc.
Key words	Key words in the description, used for model retrieval
Model base	The model base name containing this model
SubModel	The names of directly interactive submodels
Interface	Expression of the I/O capability of the model
Linkage	Model deployment and execution information to support the distributed model base
Reference	Typical resources and publications of models

#### 3.1.2. Model Organization

A large and diverse set of geographic models have been developed in recent years across multiple geographical fields that use various model mechanisms, including agent-based models, dynamic models, statistical models, and so on. When investigating a geographic problem, we first decompose the problem into relevant subfields and search for models that can solve these sub-problems well; then, we consider the model mechanism or its implementation. To assist researchers in searching for and managing models efficiently from a model base, based on the analysis above, a model base dictionary is one possible method for organizing the models. Supported by MDL, one instance of a model base dictionary is shown in Figure 2.



**Figure 2.** Model organization in model base.

### 3.2. MBMS

#### 3.2.1. Model Manipulation Language

Using a model manipulation language (MML) compiler in MBMS, advanced model users can add new models to the model base and deploy models for execution in model servers. In addition to such model-relevant manipulation, users also need to manage and maintain models through tasks including model registration, decomposition, encapsulation, integration, etc. (Figure 3).

Model registration enables the model base and model creation by interpreting MDL grammar; thus, new models can be added and stored in the model base. Meanwhile, users of VGEs can retrieve models using key words or search the model base to select desired models for geographic process simulations, using queries that consider modelling assumptions, run-time performance, etc. [42]. Sometimes, users may not be able to use a model directly from the model base. Therefore, to improve model reuse and sharing, MML provides capabilities for decomposing models into geographically meaningful submodels. Furthermore, model encapsulation equips models with standardized interfaces, contributes to forming model components in the model base and enables models to be transparently accessed [21]. Model integration, herein the same as model composition, involves searching the model base to generate a sequence of models to be executed, as well as identifying the linkage of each model for model execution [42].

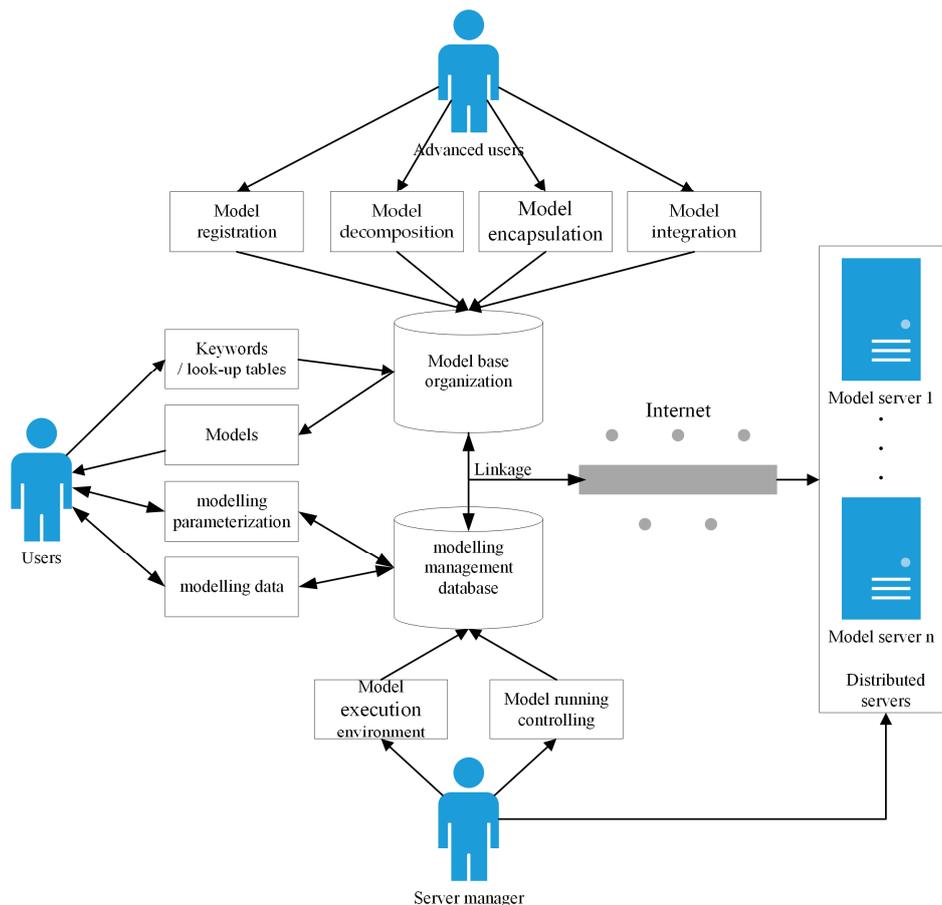


Figure 3. Conceptual design of a model base management system (MBMS).

#### 3.2.2. Modelling Management

Aiming at improving the ability to replicate model-based research, the goal of modelling management in such a framework is to manage model parameters and data, as well as model execution

and control. Model parameterization and model data functions are designed to manage both model settings and the relevant input as well as intermediate and output data as a whole in the MBMS. Furthermore, in comparison to the model definition in the model base, distributed models can be deployed on multiple platforms at geographically dispersed sites; they are activated and executed via the Internet. Thus, the model execution environment and model controls can be manipulated by server managers to provide a suitable modelling environment and give priority to or suspend the execution of a given model, etc.

Based on such a framework, when users begin to interact with a VGE through an interactive environment, they can conduct simulations through the following steps.

1. Users select a suitable model for simulation from the model base;
2. They provide required modelling parameters, including the model host, data and the modelling setting; these parameters are then uploaded to the MBMS;
3. Using the parameters in the database, the system can create a model control file and activate the corresponding model based on the linkage information defined by MDL in the model base via the Internet;
4. The activated model on a server conducts a simulation that adheres to the model execution function, which is controlled by server managers;
5. When the simulation is complete, the system uploads the simulation results to the MBMS;
6. Subsequently, users can retrieve the modelling parameters and the simulation results simultaneously.

Such a modelling management design is helpful in two ways. On one hand, different users can retrieve simulation results from the MBMS to check the data, the model settings, etc., to reach a consensus regarding the simulation result or to conveniently be able to reproduce the simulation. On the other hand, this design can also help new users learn from previous experience, as prior simulation parameters are stored in the database system.

#### **4. An Application of the Model Base Framework for Air Quality Management**

##### *4.1. Background*

The PRD region in southern China includes nine cities in Guangdong province with a population of approximately 65 million [43]. Rapid urbanization and high energy consumption have recently driven air pollution to harmful levels. Air pollution is a growing concern for both the PRD and its geographically adjacent region, the Hong Kong (HK) Special Administrative Region in China. It has long been thought that the air pollution situation in HK is significantly affected by the PRD region, particularly when secondary products such as O<sub>3</sub> and PM<sub>2.5</sub> are considered instead of classical primary pollutants such as SO<sub>2</sub> and the total suspended particulate matter [44–46].

Given the severe air quality problem and the regional characteristics of air quality management for HK and the PRD region, which are under the 'one country, two systems' principle, replication of policy-oriented simulations is urgent for both sides [24]. On one hand, to implement a scientific policy for managing such cross-boundary and regional air quality problems, different researchers or stakeholders typically simulate a set of scenarios that consider different potential policies [47]. However, because the simulations emanating from both HK and PRD have different starting points and targets, the simulation results from one group are difficult for other groups to accept [13]. In such a context, being able to replicate simulations would help the different participants in air quality management efforts reproduce the simulation procedures to achieve a consensus regarding the simulation results and even the potential policies, which are an important governing factor for air quality management [48]. On the other hand, air quality is such a complex geographic process, affected by geographic ground surface, meteorological conditions and the processing of emission inventory, that the simulation results could be very different due to the inappropriate utilization of datasets or

simulation parameters. This scenario reflects the need for modelling management and the capability to replicate each simulation so it can be traced and re-verified when necessary.

The current modelling systems for air quality management in the study area focus primarily on the mechanisms underlying the air quality problem or on collaborative simulations; they pay limited attention to the reproducibility of their modelling studies [17,49–52]. Considering this background analysis, we applied the framework of the model base to the model environment and developed one VGE prototype to solve the replication problem in air quality simulation for scientific regional management.

## 4.2. Technical Implementation

### 4.2.1. Meteorological and Air Quality Models

To develop the VGE for air quality simulation, a set of models were implemented in the model environment, including the Weather Research & Forecasting Model (WRF) for meteorological simulation, the Community Multiscale Air Quality Modelling System (CMAQ) for air quality simulation and the Sparse Matrix Operator Kernel Emissions Model (SMOKE) to prepare the emission inventory for CMAQ. In more detail, WRF was designed to be a flexible atmospheric simulation system in order to study numerically based weather prediction, regional climate, air quality, regional hydrology and water resources [53,54]. The model had numerous options for various physical processes, including a planetary boundary layer scheme, land-surface models, etc. [55].

The CMAQ modelling system was designed to simulate overall air quality, and included state-of-the-art capabilities for modelling multiple air quality issues, such as tropospheric ozone, fine particles, acid deposition, etc. Using the meteorological force field from the WRF, CMAQ simulated various chemical and physical processes that are considered important for understanding atmospheric trace gas transformations and distributions [56].

Air pollutant emission inventories are fundamental information for modelling air pollution formation and transport as well as for guiding air quality management. A localized SMOKE was applied by our research group to prepare the emission inventories for CMAQ that cover HK and the PRD region [57].

### 4.2.2. Model Base Design and Implementation

In this system, we applied Power Designer to design the conceptual model of the model base to store and manage model definition information (Table 2). Through MDL grammar, new models can be registered into the model base, and each model can be linked and executed on corresponding distributed servers through the linkage. Meanwhile, via an MML compiler, users can decompose, encapsulate and integrate models, and register them in the model base. For example, the definition of WRF in the model base is shown in Table 3; CMAQ and SMOKE have similar definitions. We deployed a MySQL database to store these model definitions and implement the model base.

**Table 2.** Model base design in an implementation with Model\_name as an identifier.

Item	Data Type
Model_name <pi>	Long characters (16)
Description	Text (124)
Key_words	Long characters (16)
Model_base	Long characters (16)
SubModel	Long characters (32)
Interface	Long characters (64)
Linkage	Long characters (32)
Reference	Text (124)

**Table 3.** Definition of the Weather Research & Forecasting (WRF) model in the model base.

Item	Value
Model_name	WRF
Description	A flexible atmospheric simulation system for the research of numerical weather prediction, regional climate, air quality, regional hydrology and water resources
Key_words	Meteorology, regional, dynamics
Model_base	Meteorological model base
SubModel	WRF Preprocessing System (WPS), Real, WRFrun
Interface	Earth surface data and boundary and initial condition data
Linkage	/121.194.93.67/Models/WRF
Reference	<a href="http://www.wrf-model.org/index.php">http://www.wrf-model.org/index.php</a>

#### 4.2.3. Implementation of Modelling Management

To improve the ability to replicate model-based research, we implemented our framework using LAMP architecture to build the modelling management features, including parameterizations and runtime control [24,58,59]. LAMP consists of several open source projects that provide a complete and robust solution for the operating system, the web server, the database, and a scripting language. The server is built on the Linux operating system. The MySQL database manages the modelling process information and is able to integrate well with different operating systems. Apache is used as the web server; Perl can access the MySQL database. Taking the WRF simulation as an example, Table 4 shows the parameters important to modelling and illustrates the database design. In this case, when users activate the WRF model and set parameters in the interface, this information will be uploaded to the database and the config files to execute the WRF model will be created. After the simulation, the results will be stored in the working directory. This approach forms the basic working procedure to implement modelling management and make such processes traceable and replicable, as explained in Section 3.2.2.

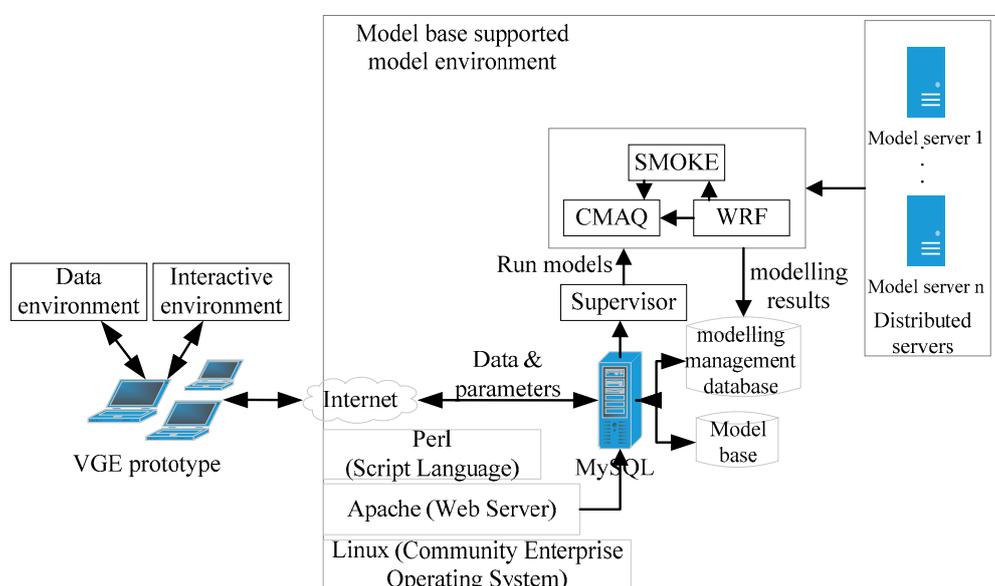
**Table 4.** Modelling management of WRF in an implementation with WRF\_casename as an identifier.

Item	Data Type
WRF_casename <pi>	Long characters (32)
D_host	Long characters (32)
Workin_dir	Long characters (64)
Geo_datapath	Long characters (64)
Geo_data res	Long characters (8)
BcandI_datapath	Long characters (64)
Parent_id	Short integer
Parent_gri_ratio	Short integer
I_parent_start	Integer
J_parent_start	Integer
Ref_lon	Float
Ref_lat	Float
X_resolution	Float
Y_resolution	Float
End_we	Integer
End_sn	Integer
Mp_physics	Long characters (32)
Ra_lw_physics	Long characters (32)
Ra_sw_physics	Long characters (32)
Sf_sfclay_physics	Long characters (32)
Sf_surface_physics	Long characters (32)
Bl_pbl_physics	Long characters (32)
Cu_physics	Long characters (32)

After launching a simulation, there will undoubtedly be a time when researchers want to stop or restart it, or see its status. The models we used are all deployed in the Linux operating system; therefore, we deployed Supervisor to provide control over the model execution. Supervisor is a client/server system that allows users to monitor and control multiple running processes (<http://supervisord.org/>).

#### 4.2.4. VGE Prototype Implementation

In addition to the model environment design and implementation, the data and interaction environments were also implemented in the VGE prototype (Figure 4). To facilitate optimal end user interaction, a Graphical User Interface (GUI) was developed using the C++ programming language. The compute-intensive dynamic models were deployed on the Community Enterprise Operating System (CentOS) on eight nodes, each of which has 24 cores. The system activates the dynamic models by calling Perl scripts using the Uniform Resource Locator (URL) to create the GUI via Supervisor. A box-based method was implemented for visualization based on the OpenSceneGraph 3D graphics toolkit to visualize the dynamic air quality process.

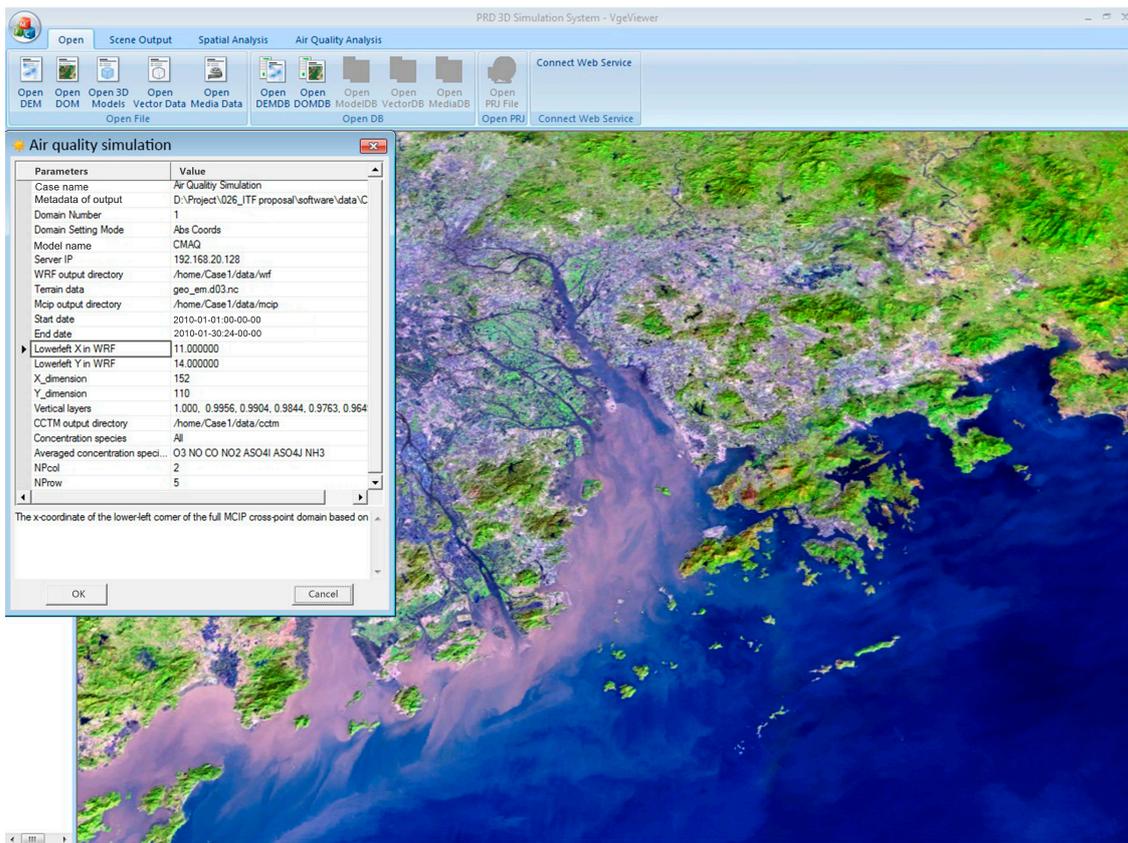


**Figure 4.** Implementation of a model-base-supported VGE.

### 4.3. Results and Discussion

#### 4.3.1. Application of the VGE Prototype

Taking the air quality of January 2010 as an example, we conducted a preliminary simulation to illustrate the replicative capabilities of our VGE prototype (Figure 5). Using the VGE GUI, users can select models from the model base and set the modelling parameters for different models or submodels. For example, we executed the CMAQ submodel, as shown in Figure 5a. After defining the modelling parameters and data, they were uploaded to the modelling management database (Figure 5b) and the system created a modelling config file, called Supervisor and activated the corresponding models for computation. When the simulation was completed, the modelling results were uploaded to the modelling management database and the input, modelling parameters and modelling output were linked together for further visual analysis (Figure 5c). Using this approach, different users can adopt the models from the model base and retrieve the modelling settings to reproduce a previously executed simulation conveniently. Therefore, it can be concluded that the framework has been implemented as a prototype, and that it is helpful for improving model reusability and for reproducing results.



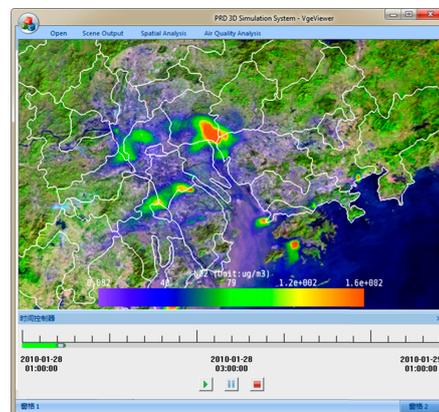
(a)

```
mysql> desc mcip_conf;
```

Field	Type	Null	key	Default	Extra
id	int(11)	NO		0	
server_ip	var char (32)	NO			
date_time	var char (128)	NO			
wrf_path	var char (128)	NO			
mcip_path	var char (128)	NO			
cctm_outdir	var char (128)	NO			
start_date	var char (128)	NO			
end_date	var char (128)	NO			
x	var char (32)	NO		0	
y	var char (32)	NO		0	
ncols	int(11)	NO		0	
nrows	int(11)	NO		0	
InterFile	var char (128)	NO			
cctm_lays	var char (128)	YES			
cctm_ncols	int(11)	NO		0	
cctm_nrows	int(11)	NO		0	
avg_conc_spcs	var char (128)	NO			
conc_spcs	var char (128)	NO			
upload_ip	var char (32)	NO			
user_name	var char (32)	NO			
comment	var char (512)	NO			

21 rows in set (0.00 sec)

(b)



(c)

**Figure 5.** Application of a VGE prototype for air quality simulation. (a) Graphical User Interface (GUI) for model selection and model setting; (b) Modelling parameters managed in MySQL database; (c) Visualization of simulated air quality process in the VGE.

#### 4.3.2. Discussion

The sample application illustrates the application of the prototype through a case study conducted solely by our research group. While this example is helpful in allowing users to understand the prototype workflow, a wider evaluation of the prototype conducted by other users would make the work of this paper more convincing. Because this research project is still under development, a wider application of the system and an assessment based on quantitative user experience will be conducted in future studies. Meanwhile, to expand the application and exchange of the system, we will also consider making our system Open Source based on the Berkeley Software Distribution (BSD) license.

## 5. Conclusions and Future Work

Models are widely acknowledged as important tools in dynamic geographical research; thus, models have become key components in VGEs. However, in developing a VGE, the ability to replicate model-based research is a challenge faced by geographers. To address this problem, employing model base theory, this paper designed a framework for VGE models that supports both model storage and model management with the goal of improving the ability to replicate models to enhance model sharing, reuse and modelling traceability. The design of such a framework involves designing a model base and an MBMS along with MDL and MML. Applying LAMP and Supervisor, we implemented a framework for the model environment and developed a VGE prototype, including the model data and an interactive environment based on C++ and openscenegraph technologies. Using a case study involving air quality simulation for January 2010, we demonstrated the prototype workflow and showed an example of the database tables that store and manage both data and parameters. Based on the records of model definitions and each executed simulation that is stored in the MBMS, various users can reuse models and reproduce simulations by retrieving existing models and parameters from the database. This prototype VGE demonstrated both modelling management and an improved ability to replicate model-based research.

Although the contribution of the model base framework is apparent for both model-based research and VGE development, many challenges remain. Model replication is complex for both model developers and model users. Thiele et al. even encouraged journal editors to create a new category called “Model replication and analysis” where modellers can publish their replication procedures when they led to new and important insights [22]. Concerning interactive geographic processes, interdisciplinary models that address various mechanisms are required, which makes model replication much more difficult to achieve. Furthermore, although designing a model base framework for VGEs is a valuable trial effort, a model base system is more complex than a typical database system. In this context, researchers are still a long way from designing a widely-recognized model base and MBMS with a more widely applicable efficiency assessment. Such a model base and MBMS would require a thorough review of models at suitable abstraction levels that must considering model domains, model details, and many other aspects.

**Acknowledgments:** We are grateful to Junyu Zheng and Zhijiong Huang for their data processing work in emission inventory, as well as to Hui Lin, Yulin Ding and Zhichao Zeng for their help in developing the VGE prototype. The paper was supported by the National Natural Science Foundation of China (No. 41601432), the Beijing Natural Science Foundation (No. 8174074), the National Key Research and Development Program of China (No. 2016YFA0602500), the China Postdoctoral Science Foundation (No. 2016M590117), the Fundamental Research Funds for the Central Universities (No. 2652015339) and the Key Laboratory of Spatial Data Mining & Information Sharing of Ministry of Education, Fuzhou University (No. 2017 LSDMIS06).

**Author Contributions:** Chunxiao Zhang designed the framework, led the implementation of the VGE prototype and wrote this paper. The co-authors participated in the framework design and system development.

**Conflicts of Interest:** The authors declare that there is no conflict of interests regarding the publication of this article.

## References

1. Goodchild, M.F. Twenty years of progress GIScience in 2010. *J. Spat. Inf. Sci.* **2010**, *1*, 3–20. [[CrossRef](#)]
2. Brunt, H.; Barnes, J.; Longhurst, J.W.S.; Scally, G.; Hayes, E. Local air quality management policy and practice in the UK: The case for greater public health integration and engagement. *Environ. Sci. Policy* **2016**, *58*, 52–60. [[CrossRef](#)]
3. Yuan, M. Geographic information systems (GIS) approaches for geographic dynamics understanding and event prediction—Art. No. 65781a. In Proceedings of the SPIE—Defense Transformation and Net-Centric Systems 2007, Orlando, FL, USA, 9–12 April 2007; Volume 6578, pp. A5781–A5808.
4. Lin, H.; Batty, M.; Jorgensen, S.E.; Fu, B.; Konecny, M.; Voinov, A.; Torrens, P.; Lu, G.; Zhu, A.X.; Wilson, J.P.; et al. Virtual environments begin to embrace process-based geographic analysis. *Trans. GIS* **2015**, *19*, 493–498. [[CrossRef](#)]

5. Iyer, B.; Shankaranarayanan, G.; Lenard, M.L. Model management decision environment: A web service prototype for spreadsheet models. *Decis. Support Syst.* **2005**, *40*, 283–304. [[CrossRef](#)]
6. Hartmann, S. The World as a Process: Simulations in the Natural and Social Sciences. Available online: <http://philsci-archive.pitt.edu/id/eprint/2412> (accessed on 26 August 2015).
7. Lu, G.; Yu, Z.; Zhou, L.; Wu, M.; Sheng, Y.; Yuan, L. Data environment construction for virtual geographic environment. *Environ. Earth Sci.* **2015**, *74*, 7003–7013. [[CrossRef](#)]
8. Yue, S.; Chen, M.; Wen, Y.; Lu, G. Service-oriented model-encapsulation strategy for sharing and integrating heterogeneous geo-analysis models in an open web environment. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 258–273. [[CrossRef](#)]
9. Lü, G. Geographic analysis-oriented virtual geographic environment: Framework, structure and functions. *Sci. China Earth Sci.* **2011**, *54*, 733–743. [[CrossRef](#)]
10. Chen, M.; Lin, H.; Hu, M.; He, L.; Zhang, C. Real geographic scenario based virtual social environment: Integrate geography with social research. *Environ. Plan. B* **2013**, *40*, 1103–1121. [[CrossRef](#)]
11. Chen, M.; Lin, H.; Kolditz, O.; Chen, C. Developing dynamic virtual geographic environments (VGEs) for geographic research. *Environ. Earth Sci.* **2015**, *74*, 6975–6980. [[CrossRef](#)]
12. Lin, H.; Chen, M.; Lu, G.; Zhu, Q.; Gong, J.; You, X.; Wen, Y.; Xu, B.; Hu, M. Virtual geographic environments (VGEs): A new generation of geographic analysis tool. *Earth Sci. Rev.* **2013**, *126*, 74–84. [[CrossRef](#)]
13. Zhang, C.; Chen, M.; Li, R.; Fang, C.; Lin, H. What's going on about geo-process modeling in virtual geographic environments (VGEs). *Ecol. Model.* **2016**, *319*, 147–154. [[CrossRef](#)]
14. Guo, F.; Zhang, Z.; Hu, D.; Yang, T.; Yue, S. Research on adaptive tide numerical simulation based on steering dynamic monitoring. *Environ. Earth Sci.* **2015**, *74*, 7029–7039. [[CrossRef](#)]
15. Xu, B.; Lin, H.; Chiu, L.S.; Tang, S.; Cheung, J.; Hu, Y.; Zeng, L. VGE-CUGrid: An integrated platform for efficient configuration, computation, and visualization of MM5. *Environ. Model. Softw.* **2010**, *25*, 1894–1896. [[CrossRef](#)]
16. Li, Y.; Gong, J.; Song, Y.; Liu, Z.; Ma, T.; Liu, H.; Shen, S.; Li, W.; Yu, Y. Design and key techniques of a collaborative virtual flood experiment that integrates cellular automata and dynamic observations. *Environ. Earth Sci.* **2015**, *74*, 7059–7067. [[CrossRef](#)]
17. Xu, B.; Lin, H.; Chiu, L.; Hu, Y.; Zhu, J.; Hu, M.; Cui, W. Collaborative virtual geographic environments: A case study of air pollution simulation. *Inf. Sci.* **2011**, *181*, 2231–2246. [[CrossRef](#)]
18. Zhu, J.; Yin, L.; Wang, J.; Zhang, H.; Hu, Y.; Liu, Z. Dam-break flood routing simulation and scale effect analysis based on virtual geographic environment. *IEEE JSTARS* **2015**, *8*, 105–113. [[CrossRef](#)]
19. Song, Y.; Gong, J.; Li, Y.; Cui, T.; Fang, L.; Cao, W. Crowd evacuation simulation for bioterrorism in micro-spatial environments based on virtual geographic environments. *Saf. Sci.* **2013**, *53*, 105–113. [[CrossRef](#)]
20. Kelly, R.A.; Jakeman, A.J.; Barreteau, O.; Borsuk, M.E.; ElSawah, S.; Hamilton, S.H.; Henriksen, H.J.; Kuikka, S.; Maier, H.R.; Rizzoli, A.E.; et al. Selecting among five common modelling approaches for integrated environmental assessment and management. *Environ. Model. Softw.* **2013**, *47*, 159–181. [[CrossRef](#)]
21. Wen, Y.; Chen, M.; Lu, G.; Lin, H.; He, L.; Yue, S. Prototyping an open environment for sharing geographical analysis models on cloud computing platform. *Int. J. Digit. Earth* **2013**, *6*, 356–382. [[CrossRef](#)]
22. Thiele, J.C.; Grimm, V. Replicating and breaking models: Good for you and good for ecology. *Oikos* **2015**, *124*, 691–696. [[CrossRef](#)]
23. Axelrod, R. Simulation in the social sciences. In *Handbook of Research on Nature Inspired Computing for Economy and Management*; Rennard, J.-P., Ed.; Idea Group Inc.: Hershey, PA, USA, 2007; pp. 90–100.
24. Zhang, C.; Lin, H.; Chen, M.; Zheng, X.; Li, R.; Ding, Y. A modelling system with adjustable emission inventories for cross-boundary air quality management in Hong Kong and the Pearl River Delta, China. *Comput. Environ. Urban* **2017**, *62*, 222–232. [[CrossRef](#)]
25. Blanning, R.W. Model management systems. *Decis. Support Syst.* **1993**, *9*, 9–18. [[CrossRef](#)]
26. Voinov, A.; Cerco, C. Model integration and the role of data. *Environ. Model. Softw.* **2010**, *25*, 965–969. [[CrossRef](#)]
27. Balaban, M. Credibility of modeling and simulation via triangulation. In Proceedings of the MODSIM World, Virginia Beach, VA, USA, 31 March–2 April 2015.
28. Wilensky, U.; Rand, W. Making models match: Replicating an agent-based model. *J. Artif. Soc. Soc. Simul.* **2007**, *10*, 2–28.

29. Zhong, W.; Kim, Y. Using model replication to improve the reliability of agent-based models. In Proceedings of the International Conference on Social Computing, Behavioral Modeling, and Prediction, Bethesda, MD, USA, 30–31 March 2010; Springer: Berlin/Heidelberg, Germany, 2010.
30. Pisoni, E.; Carnevale, C.; Volta, M. Sensitivity to spatial resolution of modeling systems designing air quality control policies. *Environ. Model. Softw.* **2010**, *25*, 66–73. [[CrossRef](#)]
31. Rollins, N.D.; Barton, C.M.; Bergin, S.; Janssen, M.A.; Lee, A. A computational model library for publishing model documentation and code. *Environ. Model. Softw.* **2014**, *61*, 59–65. [[CrossRef](#)]
32. Seagren, C.W. A replication and analysis of tiebout competition using an agent-based computational model. *Soc. Sci. Comput. Rev.* **2015**, *33*, 198–216. [[CrossRef](#)]
33. Barton, M.; Janssen, M.; Alessa, L.N.I. Openabm. Available online: <https://www.openabm.org/> (accessed on 8 January 2017).
34. Dolk, D.R. A generalized model management system for mathematical programming. *ACM Trans. Math. Softw.* **1986**, *12*, 92–126. [[CrossRef](#)]
35. Bharadwaj, A.; Choobineh, J.; Lo, A.; Shetty, B. Model management systems: A survey. *Ann. Oper. Res.* **1992**, *38*, 17–67. [[CrossRef](#)]
36. Wu, I.L. Model management system for IRT-based test construction decision support system. *Decis. Support Syst.* **2000**, *27*, 443–458. [[CrossRef](#)]
37. Yue, P.; Zhang, M.; Tan, Z. A geoprocessing workflow system for environmental monitoring and integrated modelling. *Environ. Model. Softw.* **2015**, *69*, 128–140. [[CrossRef](#)]
38. Will, H. J. Model management systems. In *Information Systems and Organization Structure*; Grochia, E., Szyperski, N., Eds.; Walter de Gruyter: Berlin, Germany, 1975; pp. 468–482.
39. Mayer, M.K. Future trends in model management systems: Parallel and distributed extensions. *Decis. Support Syst.* **1998**, *22*, 325–335. [[CrossRef](#)]
40. Li, J.; Sun, Y.; Liu, Z.; Cheung, T. Design and implementation of a model base management system. *J. Softw.* **1998**, *9*, 613–618. (In Chinese).
41. Lin, H.; Chen, M.; Lu, G.N. Virtual geographic environment: A workspace for computer-aided geographic experiments. *Ann. Assoc. Am. Geogr.* **2013**, *103*, 465–482. [[CrossRef](#)]
42. Krishnan, R.; Chari, K. Model management: Survey, Future Research Directions and a Bibliography. Available online: <http://www.informs.org/Pubs/ITORMS/Archive/Volume-3/No.-1-Krishnan-and-Chari> (accessed on 10 February 2017).
43. Wang, X.; Wu, Z.; Liang, G. WRF/CHEM modeling of impacts of weather conditions modified by urban expansion on secondary organic aerosol formation over Pearl River Delta. *Particuology* **2009**, *7*, 384–391. [[CrossRef](#)]
44. Zhang, Y.H.; Hu, M.; Zhong, L.J.; Wiedensohler, A.; Liu, S.C.; Andreae, M.O.; Wang, W.; Fan, S.J. Regional integrated experiments on air quality over Pearl River Delta 2004 (PRIDE-PRD2004): Overview. *Atmos. Environ.* **2008**, *42*, 6157–6173. [[CrossRef](#)]
45. Lam, K.S.; Wang, T.J.; Wu, C.L.; Li, Y.S. Study on an ozone episode in hot season in Hong Kong and transboundary air pollution over Pearl River Delta region of China. *Atmos. Environ.* **2005**, *39*, 1967–1977. [[CrossRef](#)]
46. Guo, H.; Wang, T.; Blake, D.R.; Simpson, I.J.; Kwok, Y.H.; Li, Y.S. Regional and local contributions to ambient non-methane volatile organic compounds at a polluted rural/coastal site in Pearl River Delta, China. *Atmos. Environ.* **2006**, *40*, 2345–2359. [[CrossRef](#)]
47. EPDHK. An Overview on Air Quality and Air Pollution Control in Hong Kong. Available online: [http://www.epd.gov.hk/epd/english/environmentinhk/air/air\\_maincontent.html](http://www.epd.gov.hk/epd/english/environmentinhk/air/air_maincontent.html) (accessed on 4 June 2014).
48. Environment Bureau; Transport & Housing Bureau; Food & Health Bureau; Development Bureau. A Clean Air Plan for Hong Kong. Available online: [http://www.enb.gov.hk/en/files/New\\_Air\\_Plan\\_en.pdf](http://www.enb.gov.hk/en/files/New_Air_Plan_en.pdf) (accessed on 4 June 2014).
49. Yin, X.; Huang, Z.; Zheng, J.; Yuan, Z.; Zhu, W.; Huang, X.; Chen, D. Source contributions to PM<sub>2.5</sub> in Guangdong Province, China by numerical modeling: Results and implications. *Atmos. Res.* **2017**, *186*, 63–71. [[CrossRef](#)]

50. Wang, X.; Situ, S.; Chen, W.; Zheng, J.; Guenther, A.; Fan, Q.; Chang, M. Numerical model to quantify biogenic volatile organic compound emissions: The Pearl River Delta region as a case study. *J. Environ. Sci.* **2016**, *46*, 72–82. [[CrossRef](#)] [[PubMed](#)]
51. Li, M.; Song, Y.; Mao, Z.; Liu, M.; Huang, X. Impacts of thermal circulations induced by urbanization on ozone formation in the Pearl River Delta region, China. *Atmos. Environ.* **2016**, *127*, 382–392. [[CrossRef](#)]
52. Wang, N.; Guo, H.; Jiang, F.; Ling, Z.H.; Wang, T. Simulation of ozone formation at different elevations in mountainous area of Hong Kong using WRF-CMAQ model. *Sci. Total Environ.* **2015**, *505*, 939–951. [[CrossRef](#)] [[PubMed](#)]
53. Chen, F.; Kusaka, H.; Bornstein, R.; Ching, J.; Grimmond, C.S.B.; Grossman-Clarke, S.; Loridan, T.; Manning, K.W.; Martilli, A.; Miao, S.G.; et al. The integrated WRF/urban modelling system: Development, evaluation, and applications to urban environmental problems. *Int. J. Climatol.* **2011**, *31*, 273–288. [[CrossRef](#)]
54. Akbar, M.; Aliabadi, S.; Patel, R.; Watts, M. A fully automated and integrated multi-scale forecasting scheme for emergency preparedness. *Environ. Model. Softw.* **2013**, *39*, 24–38. [[CrossRef](#)]
55. WRF. The Weather Research & Forecasting Model. Available online: <http://www.wrf-model.org/index.php> (accessed on 2 April 2012).
56. CMAQ. *Operational Guidance for the Community Multiscale Air Quality (CMAQ) Modeling System*, version 4.7.1; University of North Carolina: Chapel Hill, NC, USA, 2010.
57. Zheng, J.Y.; Zhang, L.J.; Che, W.W.; Zheng, Z.Y.; Yin, S.S. A highly resolved temporal and spatial air pollutant emission inventory for the Pearl River Delta region, China and its uncertainty assessment. *Atmos. Environ.* **2009**, *43*, 5112–5122. [[CrossRef](#)]
58. Tsai, S.J.S.; Luo, C.C. Synchronized power-quality measurement network with lamp. *IEEE Trans. Power Deliv.* **2009**, *24*, 484–485. [[CrossRef](#)]
59. Pasculescu, A.; Schoof, E.M.; Creixell, P.; Zheng, Y.; Olhovskiy, M.; Tian, R.; So, J.; Vanderlaan, R.D.; Pawson, T.; Linding, R.; et al. Coreflow: A computational platform for integration, analysis and modeling of complex biological data. *J. Proteom.* **2014**, *100*, 167–173. [[CrossRef](#)] [[PubMed](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).