Towards efficient digital governance of city air pollution using technique of big atmospheric environmental data

Xuanzhengbo Ren\textsuperscript{1,3}, Yuqi Zhao\textsuperscript{2,3}, Lin Wu\textsuperscript{2}, Jinfang Jiang\textsuperscript{1}, Changbin Zhang\textsuperscript{4}, Xiao Tang\textsuperscript{2}, Lina Han\textsuperscript{3}, Mingming Zhu\textsuperscript{2,3}, Zifa Wang\textsuperscript{2}

1. Computer Network Information Center, Chinese Academy of Sciences
2. Institute of Atmospheric Physics, Chinese Academy of Sciences
3. University of Chinese Academy of Sciences
4. Huachang Technology (Dalian) Co., Ltd.

lin.wu@mail.iap.ac.cn

Abstract. The 11th goal of the United Nations Sustainable Development Goals (SDG) aims at making cities inclusive, safe, resilient and sustainable, to which the rising city air pollution poses a severe challenge. Achieving this goal necessitates governing city air quality at large scale. However, the lack of interconnection among air quality modelling, multi-source data, pollution diagnosis and control actions impedes scalable and efficient governance of city air pollution. Here we present a design of digital governance to make full use of science and information techniques, so that information and decisions can be well generated and processed to form scientific prevention and control against heavy city air pollution. We implement the desired interconnection based on open-architecture big atmospheric environmental data system, which supports digitizing the workflow and dataflow of governance practices. Through the network of digital governance, those cities with limited resources and capacities can henceforth access data and expertise, such as model-based air quality forecast and data-based mitigation actions, to realize regional joint prevention and control of heavy air pollution. With a demonstration case for the Wuhan city in China using the Nested Grid Air Quality Prediction Modelling System (NAQPMS) developed by the Institute of Atmospheric Physics of the Chinese Academy of Sciences, we show how our design of digital governance would empower scientific control of heavy city air pollution and help achieving the SDG goal 11.

1. Introduction

According to the Nobel laureate Stiglitz, urbanization in China would be one of the two keys to the mankind’s development in the 21\textsuperscript{st} century. At global scale, the UN World Urbanization Prospects projects an increase of the ratio of the world’s urban population from 54\% in 2016 to 68\% by 2050. This massive urban concentration of people and resources in about 3\% of the land area can lead to not only technological innovations and economic growth\cite{1}, but also disastrous effects on environment\cite{2}. Air pollution is one of the major adverse outcomes of urbanization. The World Health Organization estimated that about 3 million deaths a year are linked to exposure to ambient air pollution. Meanwhile air pollution can lower the happiness of urban residents\cite{3} and cause negative impacts on regional economy\cite{4}\cite{5}. In response, the United Nations propose the 11\textsuperscript{th} Sustainable Development Goal (SDG)\cite{6} calling for inclusive, safe, resilient and sustainable cities. In this SDG goal, air pollution is specifically
mentioned in order to promote vigorous mitigation actions which are urgent and vital for sustainable environment and citizens health.

The governance of air pollution is a joint effort among local governments, research and operational centers, and mitigation action executors. Various chemistry-transport models and monitoring networks have served air pollution control in developed economies such as Europe [7][8][9][10] and USA [11]. In developing economies [12][13], integrated modeling and observing systems also play an important role in curbing the emerging air pollution issues.

China, among other emerging economies, is experiencing city air pollution due to its rapid industrialization and urbanization, both at unprecedented rates and scale. This mix nature distinguishes China’s air quality issue as a regional compound air pollution problem, which is particularly difficult to tackle and demands both in-depth scientific understanding [14] and firm mitigation actions. For the 2013 to 2018 period, Air Pollution Action Plan has been conducted and achieved a 41.7% reduction in PM$_{2.5}$ annual ambient concentration for 74 monitored cities over China. Nevertheless, according to the China Ecology and Environment Bulletin in 2018, 64.2% of the Chinese 338 prefecture-level cities still have not met the national standard. As air pollution events in China are regional phenomena, successful abatements of air pollution also require joint prevention and control for all cities over the region. The issue is thus how governance of air quality can be extended to cities at large scale, efficiently.

The barrier for large-scale mitigations of air pollution is the absence of seamless access of diverse information needed for data- and science-based mitigation actions. Not all cities afford to operate governance systems for observational analysis and model-based forecast of air pollution due to their limited resources and capacities. Those governance systems for cities or provinces, if exist, are usual isolated and provide limited IT supports for regional prevention and control of air pollution such as video conferencing and data transfer and visualization.

Here we design a digital governance framework to fill the information gap arising from regional disparity. Advanced information techniques are adopted to build open-architecture systems enabling seamless interconnections among entities for air pollution control. Cities with limited resources can access analysis and forecasts of air pollution based on big atmospheric environmental data and cloud computing. We exemplify the design using Nested Air Quality Prediction Modeling System (NAQPMS) [15] for Wuhan city in central China. Hopefully, with this new paradigm of digital governance, data- and science-based prevention and control of city air pollution can be achieved efficiently at large scale, as helps fulfilling the SDG goal 11.

2. Design of digital governance framework

To break the information barrier, our digital governance framework adopts an open-architecture system design based on data protocol and computing interfaces. By this way, the complex governance system can be decoupled into components that can exchange information and develop independently. This decoupling is especially beneficial when bridging the information gap among existing local city systems. This open governance system should enable all cities in the region to access services which may be beyond their capacities due to heavy computational load and domain expertise. Such service can be provided via cloud computing by research or operational units from another cities or regions. As sharing of information, either from numerical model simulations or observational analysis, takes form of data transfer among research teams, environmental agencies and regulators, the key of the system design is the architecture design of the underlying data system. There are two important facets of the digital governance framework: the workflow indicating how each party reacts against heavy air pollution, and the dataflow describing how data are generated, transferred and transformed.

2.1. Big atmospheric environmental data

By big atmospheric environmental data, we mean the spatiotemporal data of 5V (Volume, Velocity, Variety, Veracity, and Value) features that can serve diverse otherwise difficult or impossible applications in air pollution prevention and control. Those atmospheric environmental data are from
various sources (such as numerical model simulations and observations from monitoring stations and satellites) and of various forms (such as NetCDF files, raster images, text files, and database records). The amounts of atmospheric environmental data are massive and their relationships are complex. For instance, the data generated by air quality model simulations can easily exceed terabytes. Big atmospheric environmental data is proposed to use big data technologies for the storage, transmission, management, visualization and analysis of those massive atmospheric environmental data.

2.2. Architecture design of the big atmospheric environmental data system

The big atmospheric environmental data system is characterized by its open architecture. Seamless access of information in the governance network is achieved by introducing data protocol for data sharing and computing interface for computing resources sharing. In fact, protocol and interfaces are used to encapsulate complexity and foster flexibility in different use cases.

![Architecture of big atmospheric environmental data system](image)

**Figure 1.** Architecture of big atmospheric environmental data system.

Internet is another open system dealing with immense complexity. It designs layers\cite{16} of protocols from which numerous applications are constructed. Inspired by the Internet’s layer architecture, we design a four-layer big atmospheric environmental data system (Fig.1). The key component of this system is the middleware layer that consists of protocol interface and computing interfaces. The adopted protocol is the Open-source Project for a Network Data Access Protocol (OPeNDAP), which extends from HTTP and is often used in geoscience\cite{17}. It has well-established software on both server and client side. This protocol implements a Common Data Model (CDM) that facilitates the managements of grided spatiotemporal data and supports efficient and flexible accesses of data. The computing interfaces encapsulate the complexity of computing, for instance, the simplification of usage of air quality models. Based upon the OPeNDAP protocol and computing interfaces, the big atmospheric environmental data system connects different data and computing resources, which enable data sharing, analysis and forecasts of air pollution as well as other data-fueled applications.
We classify two layers of applications. The applications layer contains operation-specific applications such as customized model computing, visualization of pollutants and feeds of information. By contrast, the basic services layer consists of the most general applications which means they are more or less used in the development of the operation-specific applications above. We have developed four basic services: system basic service, High Performance Computing (HPC) service, dataset service and data visualization service.

System basic service ensures that the whole system can work properly. It is responsible for system security and messaging. It also provides user managements and information indexing. The Dataset Service manages air pollution datasets. It interacts with the data center via the OPeNDAP protocol interface. Data can be uploaded and downloaded through this service. The HPC Service manages computing task submission and result queries. It receives key arguments from the user interface, then generates scripts from these arguments. By calling computing interface, task scripts are submitted to the high performance computers in the infrastructure layer. This service also maintains a task list, with each item of the list storing task information such as task ID and submission time. Users can thus quickly check the status of their submitted tasks. The computing results will be stored to data center and can be accessed via the dataset service.

The Data Visualization Service provides processed data (e.g. JSON files) for online data visualization applications. Similar to the dataset service, it transmits data via OPeNDAP protocol, but the data has less volume and more specific ranges (time, longitude and latitude) called data frame. The visualization applications request data frame from the data visualization service to plot the data on the web browser or other platforms.

2.3. Workflow of the digital governance framework

A typical workflow of digital governance on city air pollution proceeds as follows. Alerts at diverse levels can first be generated by diagnostics and forecast of the pollution processes based on atmospheric measurements and chemistry-transport modelling. Heavy air pollutions can then be alleviated or even avoided by alert-triggered regional and local interventions, such as staggering peak productions from large energy and industrial facilities as well as traffic restrictions by odd-even license plate policy.

Fig. 2 shows the flow chart that presents the workflow of our digital governance framework. Parties in this framework include environment agencies, superior regulators, city councils or local regulators and research teams. The environment agencies are responsible for monitoring air quality and sending alerts. Superior regulators can promote air pollution management strategy and policy. The city councils or local regulators are responsible for local mitigation actions. The research teams consist of scientists working on air quality modelling and analysis of pollution processes.

In the beginning of the workflow, environment agency monitors the air quality levels and originates alerts when air pollution episodes arise. Once the regulators receive the alert from environment agencies, the workflow goes to the early warning stage. In this stage, regulators will access the air quality monitoring platform to know how many areas are in pollution and how severe they are. On the basis of these information, regulators will decide what measures should take in the reaction stage and how many resources should be invested in these areas.

In the reaction stage, superior regulators will release air pollution information via electronic media and web platform and set up working groups. For local regulators or city councils, they will get notified about pollutant sources and levels of reactions, then execute planned mitigation actions such as shutdown of facilities and traffic control. After all above operations, if the pollutant concentration still cannot meet the targeted standard, the regulators will get notifications from the air quality monitoring platform and reevaluate the situation of pollution. This means that the operations will get back to the early warning stage and a new loop of governance will begin. This workflow continues until the pollutant concentrations meet the targeted standard. The regulators will then cancel alerts and summarize all works they have done to figure out where the original mitigation strategies can be
further improved. The research teams will also get feedbacks to see whether their air quality models or analysis need to be updated.

Figure 2. Workflow of digital governance framework on city air pollution.

2.4. Dataflow of the digital governance framework
The digital governance framework can be considered as an information system that consists of three steps of information processing. The information from multiple sources are first collected, then filtered according to the workflow of the digital governance, and finally dispatched to tailored receivers. As shown in Fig. 3, the big atmospheric environmental data can originate from air quality model simulations, data analysis and observing networks. During the filtering step, each process of the complex governance workflow will forward, transform or separate each collected information, and itself may generate new information. Those generated information would be properly recorded to form more comprehensive big atmospheric environmental data, so that more refined and efficient governance of air quality could be performed. The digital governance framework will operate the information flow and dispatch the information to proper receivers such as regulators or local executors.
The receivers are in fact fed by information in this big data system. There is no need for them to search for data with which they may be unfamiliar.

![Diagram of multi-source data flow](image)

**Figure 3.** Dataflow of digital governance framework on city air pollution.

### 3. Demonstration Case in Wuhan City

#### 3.1. Implementation of big atmospheric environmental data system

We choose the light-weighted Node.js and Express framework to ensure that the governance system can be agilely developed and rapidly iterated. Node.js is an asynchronous event-driven JavaScript runtime, with which it is easy to build scalable data-intensive system applications. Express is a minimal and flexible Node.js web application framework that provides a robust set of features for developing web applications. We implement the big atmospheric environment data system using the microservice architecture, which can help building decoupled modules for the complex workflow and dataflow of the governance system. We use the microservice toolkit Seneca for its ease of use and for its ability to foster clean and organized code that is easier to scale and deploy. For searching performance, MongoDB is used to maintain a dataset category. It caches meta information of air pollution datasets such as name, publication time, update frequency and other descriptions. Users can quickly acquire these information without querying them from data center, which may cost too much time.

The OPeNDAP middleware is implemented using Hyrax, which is a data server developed by the OPeNDAP community. It supports protocol parsing, data encoding and processing. It also supports dataset management and data access via protocol decoding. Back End Server (BES) and OPeNDAP Lightweight Front end Server (OLFS) are two main parts of Hyrax. Deployed on a Tomcat server, the
OLFS can parse protocol into the commands through its web interface. BES can then execute those commands to manage and extract data from dataset.

For the browser side development, we use Bootstrap and jQuery to construct web interfaces. As to data visualization, we adopt D3.js and ECharts to build our own visualization library for air pollution data.

3.2. The Wuhan demonstration case
To show how the digital governance framework works, we exemplify an application on heavy air pollution control for Wuhan city. Fig. 4 shows an operation interface of digital governance. The environmental agencies monitor the air quality and provide trends of the AQI sequences (left top plot in Fig. 4). It is provided by air quality model NAQPMS which can be installed either locally or in the cloud. More information are provided, such as the diagnosis of sources of pollutions (left middle plot in Fig. 4), dynamical analysis of the pollution processes (left bottom plot in Fig. 4), and the calculations of atmospheric carrying capacity (center plot and right middle plot in Fig. 4). These information helps regulators to pinpoint the pollution area where emissions need to be reduced as well as the sectors and the amount of emission reductions (right bottom plot in Fig. 4). Once the mitigation action plan is generated, air quality model simulation is performed to check whether the planned reduction of emissions can meet the targeted concentration level (right top plot in Fig. 4).

![Figure 4. Operation interface of digital governance for Wuhan city.](image)

4. Conclusion
A digital governance framework has been designed for city air pollution management based on big atmospheric environmental data. This new paradigm of digital governance fills the information gap due to regional disparity and paves the way for those cities with limited resources and capacities to join the regional joint prevention and control of heavy air pollution. Efficient governance of air pollution at large scale can help fulfilling the SDG goal 11 which aims at making cities inclusive, safe, resilient and sustainable.
We explained what the big atmospheric environmental data is and indicated that it is possible to process the massive data flow by building an open-architecture big data system. Data protocol and computing interfaces were used to implement this system that establishes interconnections among entities in the governance network such as cloud-based air quality models, multi-source data, pollution diagnosis and control actions. Based on the big atmospheric environmental data technique, the complex workflow and dataflow of practical governance can be digitized to supports efficient data- and science-based city air pollution mitigations.

Further in-depth developments are needed to make our digital governance framework performant and comprehensive. Our focus was to ensure the proper functioning of the governance system. Future thorough optimizations on performance will be essential for the operational use of the big data system. The implemented workflow and dataflow are for the moment preliminary. More complete and concrete procedures will be coded to deal with more complex situations for practical city air pollution mitigations. In addition, the real-time end-to-end broadcast using mobile internet techniques, for instance, WeChat mini program are under testing. Further efficiency would be obtained if blockchain technique\[18\] is added as an apparatus of our big data system to record immutable data and to trigger smart contracts for emission reduction trading.

References
[1] Bettencourt L and West G 2010 Nature 467 pp 912–3
[2] Seto KC, Guneralp B and Hutyra LR 2012 P. Natl. Acad. Sci. USA. 109 pp 16083-88
[3] Zheng S Q, J H Wang, C Sun, X N Zhang and M E Kahn 2019 Nat. Hum. Behav. 3 pp 237-43
[4] Han L J 2018 Prog. Geogr. 37 pp 1011-21
[5] Fan F Y, Wang H F and Fan L J 2019 Ecological Ec. 35 pp 157-63
[6] Elder M and Zusman E 2016 Strengthening the linkages between air pollution and the Sustainable Development Goals. IGES Policy Brief, no. 35(Kanagawa: Institute for Global Environmental Strategies)
[7] A Miranda, C Silveira, J Ferreira, A Monteiro, D Lopes, H Relvas, C Borrego and P Roebeling 2015 Atmospheric Pollutio. 6 pp 434-43
[8] J Brandt, J D Silver, L M Frohn, C Geels, A Gross, A B Hansen, K M Hansen, G B Hedegaard, C A Skjototh, H Villadsen, A Zare and J H Christensen 2012 Atmos. Environ. 53 pp 156-76
[9] Soret A, Jimenez–Guerrero P and Baldasano J M 2011 Atmos. Pollut. Res. 2 pp 255-66
[10] Brunt H, Barnes J, Longhurst J W S, Scally G and Hayes E 2016 Environ. Sci. Policy 58 pp 52-60
[11] Miller D J Sun K Zondlo M A Kanter D Dubovik O Welton E J Winker D M and Ginoux P 2011 J. Geophys. Res-Atmos. 116 p 19
[12] Sunil Gulia, S M Shiva Nagendra, Mukesh Khare and Isha Khanna 2015 Atmos. Pollut. Res. 6 pp 286-304
[13] Sunil Gulia, S M Shiva Nagendra, Jo Barnes and Mukesh Khare 2018 Sci. Total Environ. 619–620 pp 1308-18
[14] J M Hao, K B He, L Duan, J H Li and L T Wang 2007 Front. Environ. Sci. En. 1 p 129
[15] Z F Wang, Xie F Y, Wang X Q, An J L and Zhu J 2006 Chinese J. Atmospheri. 5 pp 778-90
[16] Kurose J F 2005 Computer networking: A top-down approach featuring the internet(Delhi: Pearson Education)
[17] Zeng L 2014 Inform. Communications 11 pp 35-7
[18] G Chapron 2017 Nature 545 pp 403-5

Acknowledgments
This work was carried out with financial support from the National Key R&D Program of China (No. 2016YFB0200800 & No. 2018YFC0213503 & 2017YFC0212604), the Knowledge Innovation Program of the Chinese Academy of Sciences (No. XXH13506-402, No. XXH13506-302), the
Strategic Priority Research Programme (XDC01040000), and the CAS Pioneer Hundred Talents Program.