

# A Data-Driven Simulation Model for China Haze Monitor and Governance

# Xiaoyan Lu, Hong Chen, Miao Wang, Zhengying Cai\*

China Three Gorges University, Yichang, China Email: \*master cai@163.com

Received 29 January 2016; accepted 28 May 2016; published 31 May 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY). <u>http://creativecommons.org/licenses/by/4.0/</u> Open Access

# Abstract

Recently, the China haze becomes more and more serious, but it is very difficult to model and control it. Here, a data-driven model is introduced for the simulation and monitoring of China haze. First, a multi-dimensional evaluation system is built to evaluate the government performance of China haze. Second, a data-driven model is employed to reveal the operation mechanism of China's haze and is described as a multi input and multi output system. Third, a prototype system is set up to verify the proposed scheme, and the result provides us with a graphical tool to monitor different haze control strategies.

# **Keywords**

Data-Driven, Haze Monitor, MIMO, Simulation Model

# **1. Introduction**

Today China's economy and social life are increasingly suffered from haze and air pollution [1]. Evaluation and governance of pollution sources are so complex that involves all kinds of factors [2], including economic production, legislation and law enforcement, cultural and social life, etc. In performance evaluation, PM2.5 index is often used [3]. Recent researches on the modelling and governance of China haze can be classified as four aspects as follows.

Above all, by view of weather conditions, it is important to strengthen and optimize urban planning and to reduce the construction of large and high buildings, increase green space and improve the local ecological environment [1] [2]. Hence weather law plays an indispensable role in haze control, and a comprehensive analysis of the relationship between urban planning and local climate characteristics is needed. Then corresponding governance strategies include planning and designing an air control system to ventilate the pollutant dispersion.

<sup>\*</sup>Corresponding author.

How to cite this paper: Lu, X.Y., Chen, H., Wang, M. and Cai, Z.Y. (2016) A Data-Driven Simulation Model for China Haze Monitor and Governance. *World Journal of Engineering and Technology*, **4**, 374-381. http://dx.doi.org/10.4236/wjet.2016.42036

In the next, by view of enterprise production, it is important to strengthen the cooperation between enterprises to implement energy conservation policies [3]. This kind of strategy is helpful to effectively reduce aerosol emissions of industrial gases, agricultural pollution, and traffic emission. General strategies include diminishing the thermal power enterprises, promoting the clean energy transformation, limiting the volatile organic compounds in petrochemical companies, and actively application and development of renewable, efficient, clean, and low-emission energy sources.

Moreover, the explosive growth of private car in China also increases the exhaust emissions of motor vehicle [4]. Therefore it is very necessary to improve exhaust emission standards, actively promote hybrid vehicles and new energy vehicles, and greatly develop subway, light rail, electric buses and other green public transport system to effectively reduce vehicle exhaust pollution.

Lastly, legislation is also important to strengthen emission regulatory policies to direct economic development [5]. Based on the technical countermeasures of consumption habit, some laws and economic policies are out of date and should be adjusted as more effective long-term control measures. Generally these strengthen measures may be more effective to deal with the haze in the long run.

Additionally haze pollution assessment is a complex decision-making process influenced by many factors, including complex monitor network and a large number of data [6] [7]. Hence it is very meaningful to establish and improve the haze dada management system for the prevention and treatment decisions of haze pollution control [8] [9]. By visualization in Variable Intensity Computational Environment currently [10], a big data management system can be established with data visualization to support our haze monitor and governance system.

However, these data in haze control are nonlinear and dynamic, can not be easily integrated or described by traditional mathematics model. Today data-driven mode-free control is popular in control engineering with good theoretical foundation, including sensor networks optimization [11], complex process control [12], and the automatic performance improvement [13]. It is apparent that data-driven model is an effectively approach to solve this kind of nonlinear system with a high dynamic adaptation, so it may be helpful to employ data-driven model to control China haze.

Here, this paper proposes a data-driven simulation model to monitor the development and government of China haze, and the whole paper is organized as follows. Section 2 introduces a multi-dimensional performance evaluation index system of China haze, and the data-driven model. Then Section 3 illustrates the prototype and experiments. Lastly, some interesting conclusions and future work are indicated in Section 4.

# 2. Data-Driven Model for China's Haze

# 2.1. Evaluation of China Haze

In haze governance, there involves all sorts of conflicted factors. But everyone expects a green China, and different roles also need to undertake the responsibility to governance haze together. Apparently, among all roles, the government should bear the most important responsibility because of its dominant role of social manager and regulatory maker for the industrial production and social life. Next, enterprises, especially large enterprises in China can also not be absent in haze governance because they the main emitters of air pollutions, and have to do some good things without doubt. Additionally every citizen is not only a victim in China haze, but also a responsible person to reduce haze, because each citizen is not only the recipient of the environment, but also the resource consumer and polluter.

Then, according to the characteristics of China haze [1]-[6], a six dimensional evaluation index system of haze control is proposed here, including air quality, haze days, pollution source, greening, public health, GDP, and so on. Obviously, the evaluation performance of China haze governance is a multi-input multi-output (MIMO) data-driven model, including government, enterprise, citizen, and six performance indexes.

It is assumed that k is time variable, input vector is  $\mu(k)$  ( $k = 1, 2, 3, \cdots$ ), output vector is y(k) ( $k = 1, 2, 3, \cdots$ ), and transfer function is  $\varphi(k)$  ( $k = 1, 2, 3, \cdots$ ), then the transfer function of China haze control can be written as a vector

$$y(k+1) - y(k) = \varphi(k) \left[ \hat{\mu}(k) - \hat{u}(k-1) \right]$$

$$\tag{1}$$

where

$$\varphi(k) = \begin{bmatrix} \varphi_{1}(k)^{r} \\ \varphi_{2}(k)^{r} \\ \vdots \\ \varphi_{n}(k)^{r} \end{bmatrix} = \begin{bmatrix} \varphi_{11}(k) & \varphi_{12}(k) & \cdots & \varphi_{1m}(k) \\ \varphi_{21}(k) & \varphi_{22}(k) & \cdots & \varphi_{2m}(k) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{n1}(k) & \varphi_{n2}(k) & \cdots & \varphi_{nm}(k) \end{bmatrix}$$
(2)

where  $\varphi(k)$  is called as quasi gradient matrix.

Hence the evaluation of China haze is a multi-objective evaluation problem, involving all kinds of nonlinear input and output data. There are many practical systems, which are multi-input multi-output, where some systems, such as China haze control system, their input or output variables are strongly coupled and decoupled. This kind of model can not be easily solved by traditional algorithm, so it is necessary to seek a non-decoupling algorithm for haze model of multi-input multi-output nonlinear system.

#### 2.2. A Data-Driven Model

As China haze control system is a multi-output, the gradient vector can be presented as a gradient matrix with several single output vectors. Similar to the case of single-output, all haze control data generated by the gradient matrix will form control law of the data-driven model. Then the input-output relationship of China haze can be given as the equation of these vectors, greatly reducing the solving difficulty of control law.

Assuming

$$r(k) = rank \left\{ \varphi(k) \right\} \tag{3}$$

where  $r(k) \le n$ . Let  $D_t(k)$  be a full rank sub phalanx  $\varphi(k)$  with rank r(k),  $t = 1, 2, \dots, N$ , where N is the number of full rank matrices  $\varphi(k)$  with rank r(k), and  $||D_t(k)||$  is defined as the norm of  $D_t(k)$ , so there are r(k) s sub phalanxes  $D_t(k)$  with full rank. That is

$$D(k) = \begin{bmatrix} \phi_{i_{1}j_{1}}(k) & \phi_{i_{1}j_{2}}(k) & \cdots & \phi_{i_{1}j_{r}}(k) \\ \phi_{i_{2}j_{1}}(k) & \phi_{i_{2}j_{2}}(k) & \cdots & \phi_{i_{2}j_{r}}(k) \\ \vdots & \vdots & & \vdots \\ \phi_{i_{r}j_{1}}(k) & \phi_{i_{r}j_{1}}(k) & \cdots & \phi_{i_{r}j_{r}}(k) \end{bmatrix}$$
(4)

There is

$$\left|D(k)\right| = \max_{1 \le t \le N} \left\|D_t(k)\right\| \tag{5}$$

where D(k) is often called an advantage sub-square of  $\varphi(k)$  with rank r(k). Correspondingly,

 $u_{j_1}(k), u_{j_2}(k), \dots, u_{j_r}(k)$  are intended advantage control variables of the system, and it is obvious that, at different times, intended advantage control variables of the system are different.

Assuming  $r(k) = rank \{\hat{\varphi}(k)\}\$ , the sub-order advantage phalanx  $\hat{D}(k)$  of  $\hat{\varphi}(k)$  and r(k) can all be determined.

Thus the advantaged controlled variables  $u_{j_1}(k), u_{j_2}(k), \dots, u_{j_r}(k)$ , and output variables  $y_{j_1}(k), y_{j_2}(k), \dots, y_{j_r}(k)$  in accord with row  $\hat{D}(k)$  can also be gotten.

$$\hat{y}(k+1) = \left(y_{i_1}(k+1), y_{i_2}(k+1), \cdots, y_{i_r}(k+1)\right)^t$$
(6)

$$u(k) = \left(u_{j_1}(k), u_{j_2}(k), \cdots, u_{j_r}(k)\right)^r$$
(7)

After the removal of  $\hat{y}(k+1)$  from vector y(k+1), the rest component is noted as  $y^{-}(k+1)$ . Similarly

the rest component of the vector u(k) is named as  $u^{-}(k)$  after the removal of  $\hat{u}^{*}(k)$ . After the rearrangement of the order of components y(k+1) and u(k), there is

$$y(k+1) = (\hat{y}(k+1)^{\tau}, y^{-}(k+1)^{\tau})^{\tau}$$
(8)

$$\hat{u}(k) = \left(\hat{u}^{*}(k), \hat{u}^{-}(k)^{r}\right)^{r}$$
(9)

Since every vector in the vector group  $\hat{\varphi}_1(k), \hat{\varphi}_2(k), \dots, \hat{\varphi}_n(k)$  can be expressed as a linear independent vectors  $\hat{\varphi}i_{i_1}(k), \hat{\varphi}_{i_2}(k), \dots, \hat{\varphi}_{i_r}(k)$ , for the expected output at time k + 1, only component

 $y_{i_1}(k+1), \dots, y_{i_r}(k+1)$  can be given independently. Assuming it is set as

$$\hat{\varphi}^{*}(k) = \begin{bmatrix} \hat{\varphi}i_{i_{1}}(k)^{r} \\ \hat{\varphi}_{i_{2}}(k)^{r} \\ \vdots \\ \hat{\varphi}_{i_{r}}(k)^{r} \end{bmatrix}$$
(10)

The Equation (1) can be further simplified as follows

$$y^{*}(k+1) - y^{*}(k) = \hat{\varphi}^{*}(k) \Big[ \mu(k) - \hat{u}(k-1) \Big]$$
(11)

In the formula (11), making  $\hat{\varphi}^{-}(k) = \hat{u}^{-}(k-1)$ , the formula (11) can be simplified as

$$y^{*}(k+1) - y^{*}(k) = \hat{D}(k) \Big[ \mu^{*}(k) - \hat{u}^{*}(k-1) \Big]$$
(12)

So,  $\hat{\mu}^*(k)$ , the estimated value of  $\mu^*(k)$  can be gotten easily

$$\hat{u}^{*}(k) = \hat{u}^{*}(k-1) + \hat{D}^{-1}(k) \left[ y^{*}(k+1) - y^{*}(k) \right]$$
(13)

where  $\hat{D}^*(k)$  denotes the adjoined matrix of  $\hat{D}(k)$ , and  $|\hat{D}(k)|$  represents the determinant of  $\hat{D}(k)$ .

In summary, the required control law of data-driven model can be described as follows

$$\begin{cases} \hat{u}^{*}(k) = \hat{u}^{*}(k-1) + \frac{\lambda_{k}}{\left|\hat{D}(k)\right|} \hat{D}^{*}(k) \left[ y^{*}(k+1) - y^{*}(k) \right] \\ \hat{u}^{-}(k) = \hat{u}^{-}(k-1) \end{cases}$$
(14)

$$\hat{u}(k) = \left(\hat{u}^{*}(k)^{\tau}, \hat{u}^{-}(k)^{\tau}\right)^{\tau}$$
(15)

In which,  $y^*(k+1)$  represents the desired value of an available separated component of y(k+1) at time k + 1, and

$$\lambda_{k} = \begin{bmatrix} \lambda_{1}(k) & & 0 \\ & \lambda_{2}(k) & & \\ & & \ddots & \\ 0 & & & \lambda_{r}(k) \end{bmatrix}$$
(16)

where  $\lambda_1(k), \lambda_2(k), \dots, \lambda_r(k)$  is the appropriate parameters, and  $\lambda_k$  is the control parameter matrix.

# 3. Experiment and Analysis

### **3.1. Simulation Parameter Settings**

To verify the proposed model, a prototype is built. The initial simulation data comes from the following sources.

1) According to *Bulletin China Environment* in World Environment Day (June 5, 2014) and the latest version of *China Ambient Air Quality Standard*, among 74 main cities in China, there are only three cities of Haikou, Zhoushan, and Lhasa whose air qualities reach the national standard, so the compliance rate is only 4.1%. In the whole of 2014, the average haze days is 35.9 days, exceeding previous year by 18.3 days, and is the worst since 1961.

2) According to *China Index of Air Quality Monitoring City Rankings* in April 2014, the initial simulated values are selected from 100 China cities before rank 100 in air quality ranking, with higher than  $SO_2 = 38 \mu g/m^3$ ,  $NO_2 = 56 \mu g/m^3$ ,  $PM2.5 = 63 \mu g/m^3$ , and  $PM10 = 100 \mu g/m^3$ , respectively.

3) According to 2013 *Chinese Greening State Gazette*, up to March 12 in 2014, the national green area is 1.812 million hectares, 93,000 hectares more than in 2012; urban green is 12.26 square meters per capita, 0.46 square meters more than in 2012; national urban green coverage and greening levels are 39.59% and 35.72% respectively.

4) GDP is 568,845 trillion at the end of 2013, with an increasing of 7.7% over the previous year.

Based on the actual data above, basic simulation parameters are generated. To facilitate the analysis, all kinds of data are normalized in the interval [0,1], and it does not affect the strategy simulation and decision-making.

The simulation environment includes hardware configuration, such as Core i3 M 350@2.27GHz dual-core processor, and 2GB Memory, and software as Matlab7.11.0.584 (R2010b), 32-bit. Each simulation takes about six minutes.

## 3.2. Results and Analysis

According to the referred data above, the initial values of simulation parameters are set as follows. Government profit = 0.8, regulatory cost = 0.2, regulatory probability = 0.8; enterprise profit = 0.3, penalty cost = 0.5, treatment cost = 0.2; citizen profit = 0.6, participation cost = 0.3, not-participated loss = 0.5. The initial simulation stage of China haze control is shown in Figure 1.

As we can see, in **Figure 1**, at early stage of simulation, the SO<sub>2</sub> is 0.42, NO<sub>2</sub> is 0.51, PM = 0.45; industrial emission is 0.41, traffic emission = 0.32, coal combustion is 0.33; haze days is 0.443, civic health equals to 0.086, GDP was 0.564. The average haze days in the map is 44.38 days, and there are many red circles in the China map meaning bad haze situation. The radar chart in the lower right area can present the six-dimensional evaluation indexes for decision support. In the whole simulation process, all kinds of parameters will change according to the variation of multi input of data-driven model.

When simulation is finished, the results are shown in **Figure 2**.

From Figure 2, as can be seen, air quality is apparently improved, such as  $SO_2$  is decreased to 0.15,  $NO_2$  is down to 0.15, PM is lower than 0.02, and air quality in radar chart is lessened from 0.347 in Figure 1 to 0.144 in Figure 2. At the same time, several pollutant sources are also diminished, for example, industrial emissions = 0.45, traffic exhaust = 0.21, coal combustion = 0.03, and haze source in radar chart is decreased from 0.232 in Figure 1 to 0.095 in Figure 2. In haze distribution map the average haze days is lessened to 4.473 days, and red area is decreased evidently. Additionally radar chart verifies that the haze control strategy seems to be effective, for example, number of haze days is cut down to 0.044, civic health rises to 0.231, GDP grows to 0.642.

Therefore, it is apparent that to strengthen government governance is effective way to diminish China haze. To comparison, two strategies to strengthen government governance are employed to simulation, where the one is to strengthen legislation and lay enforce, the other is to promote the green energy, as shown in **Figure 3** and **Figure 4**, respectively.

As we can see from **Figure 3**, legislation strategy may be helpful to haze control, and may produce a certain difference. The average haze source is 0.107, and haze days is 0.017, air quality is 0.149, GDP = 0.670, health = 0.350, and greening = 0.165. From **Figure 3**, legislation strategy does not affect the industrial production and is helpful to get higher GDP and better greening, though air quality is relative stable with higher GDP according to **Figure 2**. Whereas green energy strategy in **Figure 4** may be different from legislation strategy in **Figure 3**. The average haze source is 0.100, and haze days is 0.058, air quality is 0.139, GDP = 0.641, health = 0.273, and











Figure 3. Experimental results with legislation strategy.



Figure 4. Experimental results with green energy strategy.

greening = 0.161. From Figure 4, green energy strategy may be different in different areas and different industrial distribution, and is helpful to get higher health, but is not helpful to higher GDP.

In practice, because China is vast, and industrial development in China is deeply uneven, environmental department and district government in China can select different strategies according to the local characteristics and the difference between different areas and industries. Based on this simulation system, different districts can input different data by its owner and try different governance strategies before its practice to avoid risk and optimize cost.

# **4.** Conclusion

This paper proposes a data-driven model to describe the China haze governance problem, and analyses the impact of government, enterprises and citizens on China haze control. By data visualization technology, the proposed model can help us select reasonable haze governance strategy to coordinate the complex relationship between government, enterprise and citizen. After modifying the governance parameters, the simulation system can try different haze control strategies and manage more complex data. In future research work, it is important to further analyse the interaction action between different governance processes, dig data value in haze big data with more effective tools, and build intelligent decision support system with artificial intelligence technology in data-driven method, etc.

# Acknowledgements

This research was supported by the National Natural Science Foundation of China (No. 71471102), and Science and Technology Research Program, Hubei Provincial Department of Education in China (Grant No. D20101203).

# **References**

- Ho, R.C., Zhang, M.W. and Ho, C.S. (2014) Impact of 2013 South Asian Haze Crisis: Study of Physical and Psychological Symptoms and Perceived Dangerousness of Pollution Level. *BMC Psychiatry*, 4, 12. http://dx.doi.org/10.1186/1471-244x-14-81
- [2] Hoerst, S.M. and Tolbert, M.A. (2014) The Effect of Carbon Monoxide on Planetary Haze Formation. *Strophysical Journal*, 781.
- [3] Park, S.-S., Jung, S.-A. and Gong, B.-J. (2013) Characteristics of PM2.5 Haze Episodes Revealed by Highly Time-Resolved Measurements at an Air Pollution Monitoring Supersite in Korea. *Aerosol And Air Quality Research*, 13, 957-976.
- Yu, S.C. (2014) Water Spray Geoengineering to Clean Air Pollution for Mitigating Haze in China's Cities. *Environmental Chemistry Letters*, 12, 109-116. <u>http://dx.doi.org/10.1007/s10311-013-0444-0</u>
- [5] Othman, K.A., Li, N. and Abdullah, E.H. (2013) Haze Monitoring System in City of Kuala Lumpur Using Zigbee Wireless Technology Implementation. *World Congress on Engineering*, 1448-1452.
- [6] Wruck, W., Peuker, M. and Regenbrecht, C.R.A. (2014) Data Management Strategies for Multinational Large-Scale Systems Biology Projects. *Briefings in Bioinformatics*, 15, 65-78. <u>http://dx.doi.org/10.1093/bib/bbs064</u>
- [7] Stanzel, S., Weimer, M. and Kopp-Schneider, A. (2013) Data Management in Large-Scale Collaborative Toxicity Studies: How to File Experimental Data for Automated Statistical Analysis. *Toxicology in Vitro*, 27, 1402-1409. http://dx.doi.org/10.1016/j.tiv.2012.12.009
- [8] Martinez-de Dios, J.R., Lferd, K. and de San Bernabe, A. (2013) Cooperation between UAS and Wireless Sensor Networks for Efficient Data Collection in Large Environments. *Journal of Intelligent & Robotic Systems*, 70, 491-508.
- [9] Sepper, R., Ross, P. and Tiik, M. (2011) Nationwide Health Data Management System: A Novel Approach for Integrating Biomarker Measurements with Comprehensive Health Records in Large Populations Studies. *Journal of Proteome Research*, 10, 97-100. <u>http://dx.doi.org/10.1021/pr1007784</u>
- [10] Yilmaz, E. and Aliabadi, S. (2013) Surface Conformed Linear Mesh and Data Subdivision Technique for Large-Scale Flow Simulation and Visualization in Variable Intensity Computational Environment. *Computers & Fluids*, 80, 388-402. <u>http://dx.doi.org/10.1016/j.compfluid.2012.01.017</u>
- [11] Jian, Y., Liu, E., Wang, Y., Zhang, Z. and Lin, C. (2013) Scale-Free Model for Wireless Sensor Networks. Proceedings of the IEEE Wireless Communications and Networking Conference, 1-5.
- [12] Lee, S.W., Shimojo, S. and O'Doherty, J.P. (2014) Neural Computations Underlying Arbitration between Model-Based and Model-Free Learning. *Neuron*, 81, 687-699. <u>http://dx.doi.org/10.1016/j.neuron.2013.11.028</u>
- [13] Marden, J.R., Ruben, S.D. and Pao, L.Y. (2013) A Model-Free Approach to Wind Farm Control Using Game Theoretic Methods. *IEEE Transactions on Control Systems Technology*, 21, 1207-1214. http://dx.doi.org/10.1109/TCST.2013.2257780