THE FRAMEWORK TO SUPPORT MULTI-SCALE ATMOSPHERIC ENVIRONMENTAL CAPACITY ESTIMATION WITH SATELLITE BIG DATA

Xiran Zhou (1)(2), Yong Xue (1)(2), Kai Qin (1)(2), Sheng Zhang (1), Chunlin Jin (1)

[1] School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou, Jiangsu Province, 221116, China
[2] Artificial Intelligence Research Institute, China University of Mining and Technology, Xuzhou, Jiangsu Province, 221116, China
Email: xrzhou@cumt.edu.cn; yxue@cumt.edu.cn; qinkai@cumt.edu.cn; hellozs@qq.com; TS19160057A31LD@cumt.edu.cn

KEY WORDS: atmospheric environment capacity, remote sensing, deep learning, multi-model data

ABSTRACT: A series of big socio-economic development changes in the new era of China lead to new challenges that the methodology of air pollution control is increasingly complex and the marginal of air pollution control becomes gradually diminishing. These challenges call for a pressing demand for understanding the sensitivity of a region's atmospheric environment to pollutants and the potentials of atmospheric resources. This paper exploits an indicator called atmospheric environment capacity to measure the maximal allowed emissions of air pollutants in a region while meeting the air quality objectives. To enhance the accuracy of atmospheric environment capacity, this paper integrates remote sensing techniques and other dataset to generate a result without breaks over spatial and time dimension. The framework of our proposed methodology includes multi-modal remote sensing data processing, critical parameters mining from remote sensing data, and atmospheric environment capacity simulation.

1. INTRODUCTION

A series of big socio-economic development changes in the new era of China lead to new challenges that the methodology of air pollution control is increasingly complex and the marginal of air pollution control becomes gradually diminishing. These challenges call for a pressing demand for understanding the sensitivity of a region's atmospheric environment to pollution and the potential of atmospheric resources in this region. The capacity of atmospheric environment measures the maximal allowed emissions of air pollutants in a region while meeting the air quality objectives (Guo et al., 2018; Xu, Wang and Zhu, 2018). Thus, atmospheric environment capacity estimation becomes a significant technology for efficient and effective air pollution control in a region.

China initially carried out urban-scale atmospheric environmental capacity estimation work in 2003. The national work plan regarding urban environmental capacity verification defined the target and requirements, the technical methods and the main contents of the atmospheric environmental capacity estimation. Besides, a number of approaches have been proposed to study the characteristics of atmospheric environmental capacity by three types: objective (e.g. meteorological conditions and topography), subjective (air quality standards, pollutant emissions, and output from external sources), and the scarcity of resources. These conventional approaches relied on the massive point data acquired from monitoring stations. However, atmospheric conditions could change anytime in everywhere, making the spatial density and spatial distribution of point data insufficient to support a fine-detailed, accurate and complete estimation on atmospheric environment capacity.

Remote sensing is the technique that spatially seamless monitor the characteristics of a place, ensuring no breaks occurs in the study area . Previous researches regarding atmospheric remote sensing have proved that multi-modal remote sensing data could enhance the completeness of atmospheric research over spatial and time dimension (Gupta et al., 2006; Patino and Duque, 2013; Xue, et al., 2017).

To our knowledge, the efforts on atmospheric environment capacity with remote sensing have been rarely reported yet. This paper reports our work on using remote sensing data to enhance the estimation of atmospheric environment capacity. The framework of our proposed methodology includes multi-modal remote sensing data processing, critical parameters mining from remote sensing data, and atmospheric environment capacity simulation.

2. RELATED WORKS

2.1 Atmospheric Environment Stress, Atmospheric Environment Carrying Capacity and Atmospheric Environment Capacity

Atmospheric Environment Stress (AES). AES (Zhou and Zhou, 2017) refers to the total amount of various atmospheric pollutants produced by natural causes and human activities. The commonly seen AES might include a variety of emissions of air pollutants.

Atmospheric Environment Carrying Capacity (AECC). AECC (Guo et al., 2018; Li et al., 2019) refers to the atmospheric environment's capability of decreasing or removing air pollutants. Considering that the atmospheric pollution is a dynamic processing, to distinguish the pollution sources people define the "implicit" AECC and "explicit" AECC. The explicit AECC includes the reduction amount or reduction plan for atmospheric pollutants. Otherwise, the "implicit" AECC includes some indirect actions, such as environmental investment, industrial structure, energy consumption proportion, etc.

Atmospheric Environment Capacity (AEC). AEC (Guo et al., 2018; Xu, Wang and Zhu, 2018) refers to the maximum cutoff for pollutant emissions that could prevent a region from suffering from atmospheric pollution. Since the role of each atmospheric pollutant varies, the AEC should be given for each air pollutant. AEC is significance of the appropriate air pollutant emissions. When a type of air pollutant exceeds its AEC, it will cause atmospheric pollution. Therefore, atmospheric environment could be clear unless the AEC for all possible atmospheric pollutants meet the criteria.

The interaction of AES, AECC and AEC determine the conditions of atmospheric environment. When AES>AECC, a severe atmospheric pollution could be occurred. Otherwise, when AES<AECC and (AECC-AES)<AEC, a slight atmospheric pollution might be happened. In comparison, (AECC-AES)>AEC means atmospheric environment is in good condition. I general, AES and AECC are available from the inventories of air pollutant emission, and the related statistics, respectively. Therefore, AEC is the critical indicator for the accurate prediction on atmospheric environmental conditions in a region.

2.2 Site-based AEC estimation and satellite-based AEC estimation

The monitoring sites and systems of atmospheric environment enrich the knowledge about atmospheric pollution and environment. In the era of big data and Internet of Things, the rapid improvement of the ground-based atmospheric monitoring network has improved the ability to sense air pollution in the surrounding area. The atmospheric monitoring sites such as AeroNeT, CARSNet have been commonly used in a great number of investigations regarding atmospheric pollutions.

However, the increasing of monitoring sites still poses a challenge for assessing the spatial and temporal changes of atmospheric environment in a region. Although many interpolations and reconstruction methods have been proposed to generate a data that no break occurs in spatial dimension, the point-based data might not be appropriate to represent the complex changes in a spatial space (Martin, 2008; She et al., 2018). Moreover, the vertical and long-range transportations of atmospheric pollutants are impossible to be represented by the data derived from ground-based observations (Mao et al., 2010).

Thus, an accurate and precise AEC estimation relies on the site-based observations and satellite-based earth observation data. Site-based observations provides timely-update and accurate data, and satellite-based earth observation data offer the imagery product that covers a large-scale region without spatial breaks.

3. FRAMEWORK OF AEC ESTIMATION

3.1 Data layer

Table 1 lists the data necessary for AEC estimation. Based on the achievements of previous research, we take population density distribution and transportation network into the consideration for AEC estimation. Moreover, besides the dataset used by the-state-of-the-art approaches for AEC, this paper integrates remote sensing and GIS data including digital elevation models, high-resolution satellite images and multispectral satellite images to support the calculation of AEC.

Table 1 Atmospheric capacity calculation parameters and data support.

Parameters	**Data Type**
Atmospheric pollution elements	Multispectral satellite image
Atmospheric pollution degree	Multispectral satellite image
Weather condition	Surface weather observation
Atmospheric pollution condition	Multispectral satellite image
Sources of air pollution emission	Multispectral satellite image
Atmospheric pollution delivery	Multispectral satellite image
AECC	Empirical data
AES	Empirical data
Landform and terrain	Digital elevation models
Land cover/land use	High-spatial resolution satellite image
Environmental protection investment	Economic statistics
Energy consumption structure	Economic statistics
Industrial structure	Economic statistics
Air quality standards	Text data
Emission standards for air pollutants	Text data
GDP	Text data
Population density and distribution	GIS data
Traffic network	GIS data
Sources of air pollution emission	GIS data

The nature of these data listed in Table 1 includes multiple data sources, multiple data modalities, and multiple data domain.

The data listed in Table 1 are from a wide range of sources. For example, satellite imagery might be available from different satellite-based, airborne or unmanned aerial vehicle (UAV) -based sensors; meteorological observations can come from different types of environmental or air pollution monitoring stations; and digital elevation models can come from LiDAR point cloud data or mapping data. Studies of big data have shown that data from a wide range of sources can affect the Value, Veracity and Validity of the data. In addition, data from different sources, even if they contain the same content, can have different structures for the same content due to different data production methods.

The data in Table 1 includes several aspects of modality: (1) Multimodal data structure: the data structure in Table 1 includes matrix structure, raster structure, vector structure, and so on. (2) Multimodal of data types: the data types in Table 1 include data optical image data, multispectral data cubes, text data, spatial data, and so on. (3) Multimodal of data manipulation: data manipulation depends on data structure and data type, and due to the diversity of data structure, there is diversity of data manipulation. For example, convolution operation is mainly applied to optical image data, On-line Analytical Processing (OLAP) is mainly applied to multispectral or hyperspectral data, buffer operation is mainly applied to spatial data, and so on.

Multi-domain data are data that contain information from different domains. For example, the data in Table 1 includes different domains such as atmospheric environment, geographic information, geological survey, market economy, and population distribution. With the accumulation of research on deep learning for big data integration, multi-domain data is beginning to be valued. Compared to the more obvious and direct heterogeneity of data in multi-source and multimodal data, the heterogeneity of multi-domain data is more implicit and exists in the different standards and semantics in the process of data production and use. For example, data on the atmospheric environment tend to be point data based on monitoring stations, while population distribution data tend to be facet data. Economic statistics and textual data do not have coordinates, whereas coordinates of GIS data, remotely sensed image data, etc., are basic requirements.

3.2 High performance computing (HPC) layer

Compared with the traditional single-CPU computing model, heterogeneous computing builds a platform that includes processors with different architectures. In recent years, research on geological big data computing has demonstrated that heterogeneous computing can effectively enhance the "4Ps" of high-performance computing for geological data - performance, productivity, power, and economy (Liu et al., 2019). Therefore, it is necessary to design a high-performance heterogeneous computing framework for remote sensing-based capacity calculation of the atmospheric environment, and construct a high-performance computing model that can meet the needs of calculation and analysis of atmospheric environmental data in terms of both performance and efficiency, as well as in terms of economy. Considering the temporal and spatial consistency of remote sensing image data, the high-performance heterogeneous computing model for remote sensing-based atmospheric capacity calculation can be divided into data space-based heterogeneous parallel, data time-based heterogeneous parallel, and data content-based heterogeneous parallel.

The data space-based heterogeneous parallelism divides the raw data in the same time range

into different data regions based on the spatial scope, and then uses the same set of "CPU+GPU" hybrid algorithm for the different data regions. The spatial range can be divided according to quadratic tree, octagonal tree and other tree structures with equal-area rule, or irregularly according to hierarchical enclosing tree, BSP tree, k-d tree and other tree structures. Heterogeneous parallelism based on data time will cover the same spatial scope of the original data based on time distribution combined into different data streams, and then the different data streams using the same set of "CPU + GPU" hybrid algorithm.
Data content-based heterogeneous parallelism is more complex and is based on the content of the original data. I

3.3 Information fusion layer

Considering the data for AEC estimation includes large-scale various dataset, information fusion layer is designed to fuse those data into a enhanced dataset. In general, the modality of these data varies in spatial coverage and temporal trend.

Different in spatial coverage modalities. The continuous release of the atmospheric environment in space is basically in accordance with the first law of geography, that is, the atmospheric environment of each region is related, and regions closer to a certain location are more susceptible to the influence of the atmospheric environment of that location than regions farther away. Therefore, the estimation of the atmospheric capacity of a region is not only considering the atmospheric environment of the region, but also needs to include the atmospheric environment of the surrounding area in the analysis.

Different temporal trend modalities. The change of the atmospheric environment in time is a dynamic process, and the atmospheric environmental conditions at one moment are correlated with the atmospheric environmental conditions at the previous moment or even the previous time period. Therefore, the estimation of the atmospheric capacity of a region at a given moment needs to take into account the atmospheric conditions of the region in the previous time period.

Moreover, according to the current classification of data fusion, data fusion can be strategically divided into: platform-level fusion, data-level fusion, feature-level fusion, semantic-level fusion, and decision-level fusion. Platform-level fusion mainly refers to the data are fused in the data sources. Data-level fusion mainly refers to data fusion are processed based on data properties and data metadata. Feature-level fusion mainly refers to the fusion is processed based on features derived from each data. A great number of classical methods regarding data fusion employ this strategy. Semantic-level fusion is less researched in the early stages and mainly means that some classical methods such as all can be used to support the semantic-level fusion strategy.

3.4 Estimation modeling layer

Based on the foundation of previous research, this paper incorporates population density distribution and transportation network into the calculation of atmospheric environmental capacity, and combines the tasks that can be provided by digital elevation models, high-resolution remote sensing images and multispectral remote sensing images produced by satellite remote sensing technology to propose the data set required for the calculation of atmospheric environmental capacity based on remote sensing and GIS data, as shown in Figure 1. The required data are multi-source, multi-modal and multi-domain in nature. In particular, the red part of Figure 1 shows the computational categories specific to remote sensing and GIS data-based data.

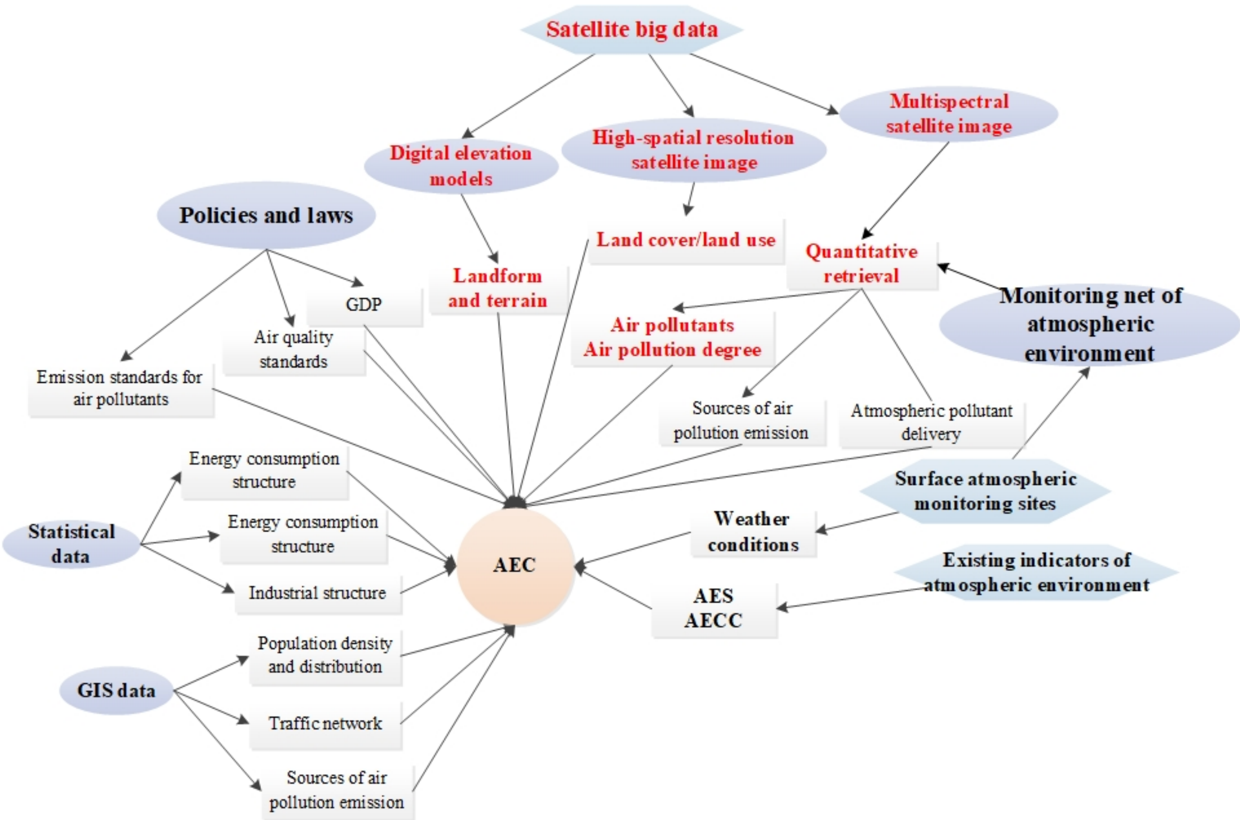

Figure 1. The framework to support multi-scale AEC estimation.

Then, the methods corresponding to the parameters are summarized as follows,
Based on the multispectral satellite image, quantitative retrieval approaches could be useful for atmospheric pollution elements, atmospheric pollution degree, atmospheric pollution condition, atmospheric pollutant delivery.

- Based on surface weather observation, statistical analysis such as linear regression could be useful for weather condition modeling.
- Based on empirical data, AECC and AEC could be available.
- Based on high-spatial resolution satellite image, the conditions of land cover/land use could be obtained by data classification.
- Based on digital elevation models, the 3-D landform and terrain features could be detected by terrain analysis.
- Based on economic statistics, a variety of statistical computing algorithms could generate the results of environmental protection investment, energy consumption structure, and industrial structure.
- Text mining and statistical analysis could support to generate the numerical information from air quality standards, emission standards for air pollutants, and GDP.
- Geo-spatial analysis could derive the spatial pattern for population density and

distribution, traffic network, and sources of air pollution emission from GIS data.

4. Conclusion

Efficient atmospheric pollution control and management acts as an important role of the construction for ecological civilization and the project “Building a beautiful China.” In the new era, the gradual changes of industrial structure, energy consumption patterns, urbanization and other socio-economic developments in China has progressing a new routine of atmospheric environment. In details, air pollution control has become increasingly challenging, the marginal effects of air pollution control has gradually diminishing, and the room for significant improvement of air quality has become emerging. All of these new challenges pose a pressing a demand for strategies that support an accurate understanding on the roles of human and natural causing air pollution. Moreover, to enhance an accurate atmospheric emission reduction and control, the capacity of the atmospheric environment is necessary to help scientists and governments assess the sensitivity of a regional atmospheric environment for pollutants and the potentials of atmospheric resources.

References from Journals:

Guo, Q., Wang, J., Yin, H. and Zhang, G., 2018. A comprehensive evaluation model of regional atmospheric environment carrying capacity: Model development and a case study in China. Ecological Indicators, 91, pp.259-267.

Gupta, P., Christopher, S.A., Wang, J., Gehrig, R., Lee, Y.C. and Kumar, N., 2006. Satellite remote sensing of particulate matter and air quality assessment over global cities. Atmospheric Environment, 40(30), pp.5880-5892.

Mao, J., Ren, X., Chen, S., Brune, W.H., Chen, Z., Martinez, M., Harder, H., Lefer, B., Rappenglück, B., Flynn, J. and Leuchner, M., 2010. Atmospheric oxidation capacity in the summer of Houston 2006: Comparison with summer measurements in other metropolitan studies. Atmospheric Environment, 44(33), pp.4107-4115.

Martin, R.V., 2008. Satellite remote sensing of surface air quality. Atmospheric environment, 42(34), pp.7823-7843.

Li, C., Li, H., Feng, S., Liu, X. and Guo, S., 2019. A study on the spatiotemporal characteristics and change trend of the atmospheric environmental carrying capacity in the Jing-Jin-Ji region, China. Journal of Cleaner Production, 211, pp.27-35.

Liu, J., Xue, Y., Ren, K., Song, J., Windmill, C. and Merritt, P., 2019. High-performance time-series quantitative retrieval from satellite images on a gpu cluster. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(8), pp.2810-2821.

Patino, J.E. and Duque, J.C., 2013. A review of regional science applications of satellite remote sensing in urban settings. Computers, Environment and Urban Systems, 37, pp.1-17.

She, L., Xue, Y., Yang, X., Leys, J., Guang, J., Che, Y., Fan, C., Xie, Y. and Li, Y., 2018. Joint retrieval of aerosol optical depth and surface reflectance over land using geostationary satellite data. IEEE Transactions on Geoscience and Remote Sensing, 57(3), pp.1489-1501.

Xu, D., Wang, Y. and Zhu, R., 2018. Atmospheric environmental capacity and urban atmospheric load in mainland China. Science China Earth Sciences, 61(1), pp.33-46.

Xue, Y., He, X., de Leeuw, G., Mei, L., Che, Y., Rippin, W., ... & Hu, Y. (2017). Long-time series aerosol optical depth retrieval from AVHRR data over land in North China and Central Europe. Remote Sensing of Environment, 198, 471-489.

Zhou, Y. and Zhou, J., 2017. Urban atmospheric environmental capacity and atmospheric environmental carrying capacity constrained by GDP–PM2.5. Ecological indicators, 73, pp.637-652.