The Research and Citation Landscape of PM2.5 in 1993-2017: A Retrospective Analysis with Bibliometric

Sun Meng
School of Finance
Yunnan University of Finance and Economics
Kunming, China
Mengsun_2004@163.com

Bingqing Ding
School of Finance
Yunnan University of Finance and Economics
Kunming, China
ingbingqing@qq.com

Jin Chen*
School of Finance
Yunnan University of Finance and Economics
Kunming, China
55658741@qq.com

Abstract—In recent years, some countries are facing the haze pollution because of industrialization. PM 2.5 is an important factor affecting the atmospheric hazy. There are many investigations about PM 2.5, but few of them are bibliometric analysis, including co-citation analysis, hot topics, burst detection, and emerging trends. Therefore, this paper adopts a bibliometric method to address the above issues and provide a general picture about the research field. By downloading 3792 articles from Web of Science and analyzing the results given by CiteSpace, the main conclusions of this paper are as follows: 1) The two hot topics in PM 2.5 area are “optical depth” and “chemical composition”. 2) The most productive countries, institutions, sources, authors, and interesting research directions in the PM 2.5 researches are visually demonstrated. 3) Atmospheric Environment, Science of the Total Environment, and Epidemiology are the major sources for PM 2.5 researches. 4) An emerging trend in PM 2.5 researches is “children”. This paper provides an overall analysis about PM 2.5 researches and gives some inspirations for researchers to do further investigations.

Keywords: PM 2.5, CiteSpace, optical depth, chemical composition, bibliometric method

I. INTRODUCTION

The haze pollution has become a serious environmental problem of the world because of its impact on human health, ecological environment and climate change. PM 2.5 is the main cause of haze pollution. As a result, the U.S. Environmental Protection Agency published the national ambient air quality standards which included a new particulate standard in 1997 (Venkataraman and Kao, 1999). Since then, related researches found that due to their ability to penetrate deeply and convey harmful chemical components, PM 2.5 has been related to various adverse health effects, such as the respiratory diseases (Trent et al., 2006; Rao et al., 2019), the cardiovascular disease (Polichtetti et al. 2009), the pulmonary inflammation (Zhang et al., 2011; Huang et al., 2019). The accurate forecasts of PM 2.5 have become significant due to the growth of epidemiological evidence which proved its negative impacts on human being's health. Therefore, some prediction methods, such as the GIS approaches (Liao et al., 2006), the hidden semi-Markov models (Dong et al., 2009), the spatiotemporal modeling with temporal-invariant variogram subgroups (Chen et al., 2012) and the chemical transport model (Di et al., 2016) have been developed to predict the concentration of PM 2.5. In addition to predicting the concentration of PM 2.5, some scholars also study how to reduce PM 2.5. In recent years, some measures were taken to reduce PM 2.5, such as changing the modes of electricity production (Markandya et al., 2009; Qu et al., 2019), reducing pollutant emission (Schleicher et al., 2012) and utilizing hot gas filters to capture PM 2.5 (Wang et al., 2016; Zhou et al., 2019a,b,c,d). It is found that PM 2.5 has been a hot issue.

With the increasingly concern about PM 2.5, some researchers have been analyzing the development of this research field. The bibliometric analysis is a scientific research technique and it is getting growing attentions (Alvarez-Betancourt and García-Silvante, 2014). Indeed, the field can be traced to prior quantitative studies in the early 20th century, which aimed to find out the rules of scientific activities, such as exploring the characteristics of the literatures on the production of bioenergy (Konur, 2012) and observing in macrology biomass as a source of biofuel feedstock (Coelho et al., 2014). These articles help researchers to find out the focuses and hot topics in the future. Meanwhile, many methods can be applied to further investigate PM 2.5 from the perspective of bibliometric analysis, such as CiteSpace, UCINET, VOSViewer and global maps of science. Among these bibliometric methods, CiteSpace is one of the most popular tools (Chen, 2006; Chen et al., 2010). It is particularly designed to support the analytic process of visualizing with the advantage of producing co-citation networks based on article citations to reveal the structure of a particular research field (Chen et al., 2012). Thus, the new method has been applied widely, such the bioinformatics (Song et al., 2014), the big data (Prathap, 2014), the renewable energy research progress in Mexico (Alemán-Nava et al., 2014), the geographic information systems (Wei et al., 2015), the carbon tax (Zhang et al., 2016), the hospitality research (Li, 2017), the low carbon development transformation (Wang et al., 2017), the carbon emissions trading (Yu and Xu, 2017), and the organic Rankine cycle technology (Imran et al., 2018). Therefore, CiteSpace has inspired scholars in various fields, telling them that there are still some new areas needed to be explored.

Due to the growing number of publications about PM 2.5, it is significant to make a good summary of the current status and the development trend in PM 2.5. Using CiteSpace to analyze PM 2.5 would help provide a clear picture on how knowledge and theories change as time goes by. Therefore, this paper analyzes all the related articles downloaded from Web of Science by CiteSpace. According to the statistics of keywords, authors, countries and cited references, this paper illustrates the
corresponding graphs and tables for the further researches based on modern bibliometrics to introduce the overall situation of the research on PM 2.5 and give some inspirations for scholars about their research directions.

To achieve the above aims, the paper is organized as follows: Section 2 mainly introduces the researches about PM 2.5 and citation analysis, in which the source of data and corresponding operational process are discussed. The current status, the research focuses and the hot topics are included in Section 3. Section 4 demonstrates the burst detection in the PM 2.5 research field and informs the potential research directions. Finally, Section 5 is the summary part based on the above results.

II. RESEARCH METHOD AND CITATION ANALYSIS OF THE PM 2.5 RESEARCH

Before using CiteSpace, analyzing the step of building database is must. Then, the productive countries & territories, institutions, hotspots, research trends, keywords and bursts of various papers in PM 2.5 research area are identified by CiteSpace. Several different visualization graphs can be generated.

A. Constructing database for CiteSpace

There are two steps in building a database for CiteSpace analysis. The first step is to choose the database. The database should contain the most complete literatures as suggested by Chen et al. (2010), co-citation networks as indicated by selection articles’ references should also be included in the database. Cited references by a given article provide valuable information regarding intellectual connections between various concepts and theories (Small, 1973). To fully capture the knowledge map and the evolution process of PM 2.5 discipline, WoS is selected as the data source for bibliometric analysis, because it is a huge platform that provides an access for readers to know specific information about the articles published in roughly 12,000 leading journals worldwide (Leeuwen, 2006). The Science Citation Index (SCI), the Social Sciences Citation Index (SSCI), and the Conference Proceedings Citation Index-Science (APCI-SCIENCE) databases are included in it. Second, choose appropriate keywords to select articles from database.

Researchers should pay more attention to the validity and representation of keyword and check performed to assess the relevance of each study returned. By researching the title “PM 2.5” or “PM2.5”, a total of 3792 articles (published from 1993 to 2017) were retained which exclude the conference proceeding citation index-science. There are two major reasons for this study to choose the research period: (1) The earlier publications were available for research were published in 1993. (2) The study focused on the recent research of PM 2.5, so the latest papers of this research field are taken into account.

B. Mapping knowledge using CiteSpace

After choosing the database, the mapping knowledge of PM 2.5 is introduced by using CiteSpace. First, in order to observe the trend of publications and citations over time, the number of publications and citations about PM 2.5 from 1993 to 2017 are analyzed. Second, CiteSpace is utilized to investigate the evolution of PM 2.5 researches. The analytical process consists of categorizing knowledge clusters in PM 2.5 research area, identifying major clusters in the PM 2.5 knowledge map, and conducting a timeline analysis on significant keywords in terms of origin, development and current status. Based on the frequency of keywords in articles, the hot research topics and emerging trends in the PM 2.5 field can be found. A further step is to identify the most productive countries & territories and institutions in PM 2.5 research development. For example, comparative analysis of knowledge evolution could be performed between 1993 and 2017. Then, contributing authors, sources and research direction were tabled. This can help identify major contributing factors for the evolution of the PM 2.5 knowledge map. Last but not least, the burst of certain references, authors and keywords can be used to analyze the evolution and the most recent research trends in PM 2.5 research. The burst refers to a significant change of a variable's values in a relatively short time. CiteSpace consider this type of change as a way to identify the front line of this research field. Next, a citation analysis of PM 2.5 studies will be investigated in detail.

C. Citation analysis of the PM 2.5 research

The literature data of PM 2.5 researches are downloaded from Social Sciences Citation Index (SSCI) and Science Citation Index Expanded (SCI-E) via Web of Science on Dec 31, 2017. Based on 3729 papers published from 1993 to 2017, the number of publications in each year of PM 2.5 researches can be seen in Figure 1.

The earliest fine particulate matter stems from the Maas Valley smoke event in Belgium in 1930. Moreover, there are a lot of typical cases in history, such as the USA Donora smoke incident (Bryson and Crhis, 1998) in 1948 and London smog incident (Laskin, 2006) in 1952. In order to analyze the current status and the development trends in PM 2.5 research field, the paper sorts out the publications and citations primarily in this area. Figure 1 illustrates the number of publications from 1993 to 2017. It is obvious that the literatures about travel motivation are increasing as time goes by. In order to analyze the change of publication amount, this paper divides the whole period into two phases: the development phase and the rapid growth phase. There are some fluctuations in the development phase. Furthermore, the number of published articles prior to 2004 was below 100 and increased slowly. Although some researchers claimed that PM 2.5 could significantly impact on people’s health, most articles presented an extensive analysis of the environment change rather than deep analysis on PM 2.5.

After 2008, many air-pollution incidents happened around the world, for example, Los Angeles had serious air pollution caused by mountain fire in 2008, Milan and Mexico's air pollution was serious in 2011 and Sao Paulo and Tehran were shrouded in haze in 2012, so the number of publications was...
up to 200 in 2013. Then, the publication number of PM 2.5 entered the rapid development period. Since 2014, PM 2.5 research issues have become one of the most significant concerns among related scientists, international environment organizations, and national organizations. Especially, the number of papers has researched to 416 in 2015, with a growth rate of 43.4% compared to the figure in 2014. This conclusion can be illustrated by Figure 1.

The development of new subject needs to be based on the knowledge accumulation of relevant subjects. Similarly, no single research paper can be developed solely on the basis of its own content. The paper uses past literatures and research results from its subjects or other related subjects for reference. In general, the research papers published in leading journals represent the frontiers of certain disciplines. This section illustrates the research focuses and hot topics of PM 2.5 through cluster analysis. With the assistance of computer software, the Document Co-citation Network (DCN) for the PM 2.5 shows the position of clusters and the association among clusters in co-citation network. It is easy to take a clear view of the whole picture in this field.

Co-citation analysis is more reliable to provide important insights into knowledge domains compared with citation-only analysis because citation-only analysis would risk excluding articles that hold promise. In addition, the half-life of an article is defined as the number of years that more than 50% of the total citations were made during these years since its publication. This is aimed to measure the frontier development of a research. The clustering function was performed by choosing 'T' as the labeling source and log-likelihood ratio as the method. These clusters are labeled by terms from their own cites. Therefore, a total of 12 significant co-citation clusters were divided by the cited references of the documents cited in each cluster and the log-likelihood ratio (LLR) algorithm. By analyzing the clusters, we can find out the major research areas about PM 2.5.

The second column, Size, in Table 1 demonstrates the number of the publications in the corresponding cluster. For example, the cluster (#0) has 108 members, which is one of the largest clusters in the PM 2.5 research field. The second cluster (#1) has 77 members and the third cluster has 75 in this field. The Silhouette value in Table 1 is an index that indicates the reliability of a cluster. If this score is higher than 0.5, then the corresponding result is reliable. In this paper the silhouette scores of 10 clusters ranged from 0.706 to 0.914, suggesting that the members of each cluster were consistent enough.

### TABLE 1 SUMMARY OF THE LARGEST 10 CLUSTERS IN PM 2.5 RESEARCH FIELD

| Cluster | Size | Silhouette | Label (LSI) | Label (LLR) | Mean (Year) |
|---------|------|------------|-------------|-------------|-------------|
| 0       | 108  | 0.818      | China       | Chemical Composition | 2009        |
| 1       | 77   | 0.834      | PM10        | Soot Content   | 1996        |
| 2       | 75   | 0.706      | China       | Hong Kong     | 2003        |
| 3       | 69   | 0.788      | Concentration| Outdoor Concentration | 2004        |
| 4       | 56   | 0.847      | Personal Exposure | Fine Particle | 1999        |
| 5       | 55   | 0.914      | China       | Optical Depth  | 2010        |
| 6       | 54   | 0.817      | Mass        | Semi-volatile | 1998        |
| 7       | 45   | 0.786      | Source apportionment | Organic Speciated PM2 | 2002        |
| 8       | 37   | 0.849      | Fine Particulate Matter | Air Pollution | 2004        |
| 9       | 34   | 0.903      | Characterization | Chemical Mass Balance Report | 1996        |

The LLR (Log-likelihood ratio) value means a kind of algorithm according to the probability density function to determine the maximum likelihood and find the most likely words. The index of Mean (Year) in Table 1 represents the mean cite year for each cluster, it implies whether it is formed...
by recent documents or old documents. Thus, cluster #1 is formed by older documents than other clusters. It can be learned that the cluster #5 “Optical Depth” is the newly emerging cited cluster and the average year of citation is 2010. Therefore, it can assume that the most recent research topics of PM 2.5 are “optical depth” and “chemical composition”. “Optical depth” is the fifth cluster, with 55 members. The clusters #0 “Chemical Composition” is the largest one, which indicates that the study of chemical composition is more popular. A total of 56 citing documents formed the fourth largest cluster (#4), it was marked as fine particle by LLR. This indicates a high quality cluster analysis of the PM 2.5 knowledge mapping.

In general, the results of the clusters are reliable according to the Silhouette values in Table 1. The yellow arc represents journals mainly include physics, chemistry, ecology, molecular, biology and etc. The green arc means medical aspects, such as dentistry, dermatology, surgery, medicine. It is clear to define the specialties in science mapping in terms of co-citation clusters. Moreover, the topic of “optical depth” is more potential for further research due to the size and mean year of this cluster.

Moreover, the results of double figures overlay in the field of PM 2.5. It indicates that the research on PM 2.5 are very extensive, the disciplines include science, biology, medical and clinical. Besides, PM 2.5 research became more interdisciplinary over time. The ellipse on the right of the chart shows the disciplines of cited papers, the most cited disciplines are environmental, health and medicine. However, the paper of biology introduces most in environment, toxicology and nutrition. There is a path starting from the left circle which can be considered as its original to the right circle which is the terminal. For example, the disciplines of cited papers about environmental, nutrition and toxicology are mostly cited from veterinary, molecular and biology. In recent years, researchers gave some new research methods and fields which can be used to solved the research problems.

In addition, analysis of keyword timeline is an effective way to show emerging trends and track topics of research over time because the timeline of keywords gives a succinctness and accurate high-level summarization of document. The changes in research topics have become an important research issue, which can help researchers to gain deeper insights into the development of a particular research field. The function of keywords to timeline analysis based on CiteSpace is to acquire keyword timeline matrix used for analysis of keyword timeline in this paper.

It is obvious that the more a country & territory or an institution publish its research findings, the more contributions it will make in certain research area. The results tell that the most productive counties & territories are at their prosperous stages of PM 2.5, and the most productive institutions are mainly from those countries & territories.

For the analysis of the characteristics of PM 2.5, Google earth was adopted to visualize information based on authors’ geographical position. The scholars who research on PM 2.5 are mainly from North America, Europe and Asia. The countries in North America, Europe and Asia form close academic networks with other countries. Additionally, the scientific collaboration network among countries & territories is analyzed by CiteSpace. Based on the results given by WoS and CiteSpace, further discussion can be seen below. Scientific collaboration has increasing in frequency and importance in recent years, many articles are collaborated by authors who from different continents (e.g., Dockery et al., 1993; Trenga et al., 2006; Chen et al., 2010). The collaboration has the potential to solve complex scientific problems, promote various agendas, and knowledge creation (Sonnenswald, 2007). As a result, there should be more international collaborations to help researchers consider research issues from multiple perspectives.

The development of the PM 2.5 research differs from countries to countries. The top 10 most productive countries & territories in this research field. The People’s Republic of China is the most productive country followed by USA, Taiwan, Italy, South Korea, Canada, India, Germany and England. Research teams in the China and USA are strong enough to conduct research independently or form cooperation systems.

At the institutional level, many institutions around the world are engaged in the PM 2.5 researches, and the top 10 productive institutions are presented in Table 2. Among these 10 institutions, there are six institutions in China, and the rest are from the USA. This could explain that why the China has become the most important contributor.

| Institutions                                    | Publication number | The percentage of total |
|-------------------------------------------------|--------------------|-------------------------|
| Chinese Academy of Sciences (China)             | 321                | 8.608%                  |
| US EPA (USA)                                    | 182                | 4.881%                  |
| Peking University (China)                       | 113                | 3.030%                  |
| Harvard University (USA)                        | 102                | 2.735%                  |
| Tsinghua University (China)                     | 99                 | 2.655%                  |
| Fudan University (China)                        | 71                 | 1.904%                  |
| University of Washington (USA)                  | 68                 | 1.824%                  |
| University of Chinese Academy of Sciences (China)| 65               | 1.743%                  |
| Chinese Research Academy of Environmental Sciences (China) | 64        | 1.716%                  |
| Desert Research Institute (USA)                 | 62                 | 1.6663%                 |

China is ranked at the top in terms of research outputs. Generally, the number of outputs is related to the number of research institutions, availability of research funding, and the proportion of those that have environment focus. Another significant reason for the high outputs is that some countries, such as China and USA, are highly influenced by environment change.

Those institutions presented in Table 2 made the majority of contributions to the total research development. According to Table 2, it can be found that among the top 10 productive institutions, more than half institutions are in China, such as Chinese Academy of Sciences, Peking University and Tsinghua University. Chinese Academy of Sciences takes the first place with 321 articles, followed by US EPA and Peking University. There is no doubt that the China is the most productive country in the PM 2.5 research field. The development and refinement of an academic research discipline depends greatly on the cooperation among institutions.
IV. BURST DETECTION IN THE PM 2.5 RESEARCH

Citation bursts, which refer to articles that have received sharp increases in citations, can in part reflect the dynamics of a field. Burst detection is a useful analytic method to find the articles that receive particular attention from the related scientific communities in a certain period of time. Therefore, several interesting points can be found by analyzing the results given by CiteSpace. In this section, the burst detection on the types of references, keywords, institutions and authors are illustrated to show articles that have received rapidly increasing citations and explore research directions intensively.

A. Emerging trends and new developments in PM 2.5 research field

The articles that have received rapidly increasing citations can be analyzed to reflect the emerging trends in the specific research field. The citation burst suggests that the scholars have paid special attention on the corresponding publications. In order to know the emerging trends of PM 2.5 researches, the reference burst detection is utilized in this study. As shown in Table 6, it is obvious to see that the earliest reference citation burst in this research field started from 1993. This paper investigates PM 2.5 on the aspects of the relationship of particulate air pollution with epidemiological of health (Pope et al., 1995a), daily mortality (Laden et al., 2000) and measurements (Lee, 2001). In other words, these topics are the research directions in 2003-2009. These studies presented in-depth analyses of PM 2.5 and provided theoretical basis for the further research. In recent years, PM 2.5 has become a vital element that affects people’s environment awareness due to the development of information technology and economic.

In consideration of the sudden detection of keywords, which are the fast growing topics in this research field. There are a number of keywords with bursts and the top 10 are listed on Table 7. “Aerosol” is firstly proposed in dust pollution. “Children” was proposed in 1999 which has attracted great attention from national government. Consequently, this keyword burst started from 1993 to 2013. Some topics show that the theme of the study changes quickly as time goes on.

B. Burst detection on institutions in the PM 2.5 research

The institution burst of United States Environmental Protection Agency started from 1993 and ended in 2004. This indicates that the researchers from United States Environmental Protection Agency paid a lot of attentions to PM 2.5 researches from 1993 to 2004. Among the above top 10 institutions shown in Table 8, seven of them are from the USA, China, Switzerland and Spain each has one. It can be seen that in the 1990s of the last century, United States Environmental Protection Agency started to study PM 2.5. While by 20th century, the research of PM 2.5 in Chinese Hong Kong University of Science and Technology was more active. Since 2000, the growth of environmental awareness has made many countries take measures to protect the environment which is the main reason why China has become active in the research field of PM 2.5.

C. Burst detection on authors in the PM 2.5 researches

By further studying the burst detection on authors in the PM 2.5 researches, we can identify the authors who have rapidly increasing publications. The top 10 authors with the strongest citation burst are listed on Table 9.

According to Table 10, most of authors are from China, the USA and Europe. The top ranked author by bursts is Chow with bursts strength of 6.2843, who is a leading expert on environment research. The second one is Kunzli with bursts strength of 5.9836. Some of the authors with the strongest citation bursts are also in the top 10 productive authors in PM 2.5 researches, such as Chow, Fang and Koutrakis and etc. In summary, the research area on PM 2.5 attracted many authors, such as Chow, Kunzli and Wilson and etc. Thus, the research about PM 2.5 is becoming mature in recent years.

V. CONCLUSIONS

In this study, we have made a bibliometric analysis on the evolution and development of the PM 2.5 research field by CiteSpace based on 3729 articles from Web of Science. According to the above analyses, some useful conclusions are derived as follows: In the past 24 years, there were 3729 journal articles related to PM 2.5. The published articles in 2013 are up to 207, which indicates people are paying increasing attention to this research field. Further, China is the biggest contributor with more publications compared with other countries. At the institutional level, Chinese Academy of Sciences takes a leading position. The research on PM 2.5 covers a quite wide range of topics, from environmental science to health, economy and medicine. In addition, the future research trends of PM 2.5 might be the environmental sciences ecology, the meteorology atmospheric sciences, the public environmental occupational health and etc.

According to the aforementioned visualization analysis and the review of PM 2.5 research papers, it is easy for readers to get a better understanding of its further development trends. Through different aspects of research, scholars are inspired to find more efficient and innovative methods to further investigate PM 2.5. In sum, this paper provides an overall analysis of the current state and emerging trends in PM 2.5 researches, which helps people to learn more about PM 2.5 and makes this research field step into a mature development stage.

ACKNOWLEDGMENT

This work was supported by the Social Science Youth foundation of Ministry of Education of China under Grant [18YJJC790118].

REFERENCES

[1] A. Markandya, B. G. Armstrong, S. Hales, A. Chiabai, P. Criqui, S. Mima, C. Tonne, and P. Wilkinson, Public health benefits of strategies to reduce greenhouse-gas emissions: Low-carbon electricity generation. Lancet, 2009, Vol. 373(9706), pp. 2006-2015.

[2] A. Seaton, W. Macnee, K. Donaldson, and D. Godden, Particulate air pollution and acute health effects. Lancet, 1995, Vol. 345(8943), pp. 176-178.

[3] B. Bryson, and C. Cthos, The Donora fluoride fog: A secret history of America’s worst air pollution disaster. Earth Island Journal, 1998, Vol. 13(4), pp. 36-37.

[4] C. A. Clayton, R. L. Perritt, E. D. Pellizzari, K. W. Thomas, R. W. Whitmore, L. A. Wallace, H. Ozkaynak, and J. D. Spengler, Particle Total Exposure Assessment Methodology (PTEAM) study: Distributions of aerosol and elemental concentrations in personal, indoor, and outdoor air samples in a southern California community. Journal of Exposure Analysis & Environmental Epidemiology, 1993, Vol. 3, pp. 227-250.

[5] C. A. Pope, D. W. Dockery, and J. Schwartz, Review of epidemiological evidence of health effects of particulate air pollution. Inhalation Toxicology, 1995a, Vol. 7(1), pp. 1-18.

[6] C. A. Pope, M. J. Thun, M. M. Namboodiri, D. W. Dockery, J. S. Evans, F. E. Speizer, and C. W. J. Heath, Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. American Journal of Respiratory & Critical Care Medicine, 1995b, VOL. 151(3 Pt 1), pp. 669-674.
