Research on Polycyclic Aromatic Hydrocarbons in Environment from 2005 to 2020: A bibliometric analysis

Bingnan Ren1, 2, *

1Academician Workstation of Zhai Mingguo, Sanya University, Sanya, 572022, P. R. China.

2School of Health Industry Management, Sanya University, Sanya, 572022, P. R. China.

E-mail: rbnan@sanyau-edu.cn

Abstract. Polycyclic aromatic hydrocarbons (PAHs) are known to be toxic, mutagenic and carcinogenic. With extensive and persistent existence, polycyclic aromatic hydrocarbons can easily accumulate in the human body, posing a great risk to human health. With more and more studies on PAHs, it is essential to analyse hotspots and indicate emerging trends as well. The bibliometric study on PAHs in environment was analysed the papers by the method based on the CiteSpace, which published during the period of 2005-2020 in Web of Science database. The results indicated that the trend of this field is “pollution sources of polycyclic aromatic hydrocarbons”. And “PM2.5, health risk, impact, source identification” have become hot spots. The keywords with the strongest citation bursts are including toxicity, spatial distribution, health risk, pm2.5. There are many cross clusters links among organochlorine peptides, biodegradation, particle, positive matrix factorization.

1. Introduction

Polycyclic aromatic hydrocarbons (PAHs) are a class of toxic organic compounds composed of two or more benzene rings connected in linear, angular or clustered fused rings. Although PAHs are present in trace amounts in environmental media, the long–distance migration ability of PAHs makes its widespread and long–lasting existence in the environment possible. It then enters the human body via various routes such as respiratory, digestive and skin contact, accumulating and causing carcinogenic, teratogenic and mutagenic effects [1,2]. As a result, it has attracted wide attention from all over the world for a long time. The US Environmental Protection Agency and the European Union have listed 16 PAHs as precedence–controlled pollutants. In 1990, China also proposed seven PAHs for priority control in water [3] In addition to some natural activities that generate a certain number of PAHs, it is mainly derived from incomplete combustion of organic matter in human activities, such as coal/oil/biomass combustion in industry and daily life, waste incineration and vehicle exhaust emissions. Anthropogenic sources have gradually become major contributors to PAHs in the environment [4], and with the rapid development of urban construction and transportation, the exhaust of motor vehicles causes increasingly serious urban atmosphere pollution. Due to the lack of necessary research, little is known about the pollution status and impact of trace organic pollutants in the environment of many areas of China. Such situation has caused great difficulties in adopting reasonable management measures (such as developing regulations and standards) and engineering measures (such as developing pollution control and remediation technologies) [5,6].
In view of the harmfulness of PAHs on human beings, it is of great significance to research on PAHs in environment widely. Hence, many research papers have been published in this field. Bibliometric analysis is an interdisciplinary science that uses mathematical and statistical methods to quantitatively analyse the current situation and development trend of science and technology [7]. Bibliometric analysis has been adopted by different disciplines [8-11]. This study used bibliometric analysis to analyse the research on PAHs in environment based on online databases and with the use of scientific knowledge mapping software. It focuses on detecting and analysing the evolution trend of research frontiers, the relationship between research frontiers and their knowledge bases, and the internal relations among different research frontiers.

2. Literature sources and research methods
This study uses CiteSpace (version No. 5.6. R1se) to search the literatures in the web of science database. CiteSpace is a specialized in visual literature analysis software, in order to observe the trends of a certain research field. Through the analysis nodes of literature quantity, keywords, keyword emergence, keyword time and so on, it visually presents the hot topics in related research fields, and also shows the new research topics. Through visualization, CiteSpace can present the structure, regularity, and distribution of scientific knowledge [12].

The search terms for the data were set as “polycyclic aromatic hydrocarbon pollution” or “PAHs”. The Node Types was “keyword”, including title, author keyword and keyword plus. On this basis, the results were refined by year (2005–2020). This paper intends to do keyword co-occurrence analysis, emergence graph, clustering and time line analysis and time zone map analysis for this part of literature, in order to show the current situation of academic research, research development context, knowledge flow in this academic field, and to find out the deficiencies of research and expand the space domain.

3. Results and discussion
3.1. Co-occurrence of key words
Key words are the extraction of the main content of the paper, which can reflect the author's main point of view and academic thought. Tracking keywords can grasp the development status of the object concerned in the literature. In the map, 95 nodes (keywords) and 244 lines (co-occurrence relationship) are obtained, and the threshold in node labels is set to 19 to get a clear visualization effect. The main keyword co-occurrence knowledge map is as follows. Click the network summary table to get the detailed parameter information of keywords, and sort out the keyword co-occurrence frequency. The top 10 keywords of keyword co-occurrence frequency are arranged in descending order, which information is shown as table 1.

Fig 1. Distribution of hot spots of the research on PAHs pollution
From the overall distribution of the figure, the main key words are around the center of the map, which shows that the foreign research on the subject of "polymeric aromatic hydrocarbon pollution" is relatively concentrated. It can be seen from the figure that the size of different nodes is different, which means that the frequency of keyword co-occurrence is different. By measuring the number of any shortest path through the node in the network, the connection effect of the node in the whole network can be calculated, which is quantified as centrality. Nodes with large centrality are relatively easy to become key nodes in the network. Among them, polycyclic aromatic hydrocarbon nodes are the largest, followed by PAH, segmentation, air pollution, source apportionment, pollution, particle matter, soil, heavy metal, water, polychromatic biphenyl, surface division, exposure, contact, China, risk assessment, etc. These high-frequency keywords shows that the research direction of "polycyclic aromatic hydrocarbon pollution" is "source apportionment".

3.2. Key words time zone chart

The cold and warm colors of the nodes represent the different time of the first appearance of the nodes. The center color of the node in the above figure is close to yellow, indicating that the keyword appears earlier; if the center color of the node is close to red, it means that the keyword appears later. The cold and warm colors of nodes represent the time sequence of the first appearance of nodes. Click layout in the control panel to get the context change of keyword distribution, that is, keyword time zone graph. The nodes in a two-dimensional time zone are located in different time axes according to their positions on the time axis. A knowledge evolution diagram from the left is shown as Fig. 2.
Fig 2. Key words time zone chart of the research on PAHs pollution.

In the figure, we can see that the main keywords appear in the order of polycyclic aromatic hydrocarbon, PAH, division, air pollution, source apposition, particle matter, emission, Organo chorine pest, pollution, risk assessment, sorting, soil, biodegradation, phenanthrene, urban, surface water, city, pollutant, extraction, bioremediation, organic matter, agricultural soil, sea, polybrominated diphenyl ether, spatial distribution, toxicity, seasonal variation, environment, PM2.5, health risk, impact, source identification. It can be seen that PM2.5, health risk, impact, source identification are the research hotspots in recent years.

3.3. Keyword emergence
Emergence words are the key words in this field, which can reflect the turning of hot spots in the research field and the hot spots in recent years.

Fig 3. Map of burst words in different periods.
Fig 3 shows the top 25 keywords with the strongest citation bursts. We can see that emergence can be roughly divided into three stages. (1) the first stage: 2005-2011, aerosol, matter, organic compound, Hong Kong, United States, air pollution, source, sorting, phenanthrene, ESTAR, Pearl River Delta, pollutant, bioremediation, this stage focuses on the pollution of Hong Kong, the United States, the Pearl River Delta and other places. (2) the second stage: from 2012 to 2016, agricultural soil, PM2.5, atmosphere, river, persistent organic pollutant, ambient air, identification, organochlorine pest, polymerized diphenyl. (3) The third stage: from 2016 to 2020, toxicity, spatial distribution, health risk, PM2.5, in which toxic, spatial distribution, health risk and PM2.5 will increase sharply from 2015, and health will increase rapidly Risk and PM2.5 increased sharply from 2019, and there was no sharp decrease. Therefore, “pm2.5, health risk, impact, source identification” are the hot spots of scholars in this field in recent years. These words are with a certain degree of research heat.

3.4. Keyword clustering
Each cluster is a group of closely coupled documents, representing different research directions and backgrounds in the field. From the topics in this group of documents, representative keywords were extracted as the group’s labels. Fig 4 shows the keyword clustering analysis obtained by CiteSpace.

From the keyword clustering graph above, we can see that there are eight red labels. These tags represent eight clusters, and the tags of each cluster are keywords in the co-occurrence network. The larger the number of cluster tag number is, the less keywords are contained in the cluster. On the contrary, the smaller the number of cluster tag number is, the more keywords are contained in the cluster. The cluster labels are as follows organochlorine pesticides, biodegradation, particle, organochlorine pesticides residues, pm2, positive matrix factorization, diagnostic ratios.

The more the lines between nodes, the higher the co-occurrence degree of keywords in the domain. organochlorine pesticides, biodegradation, particle, organochlorine pesticides residues, pm2, positive matrix factorization. There are more nodes in the cluster, which indicates that the co-occurrence degree of keywords in these research fields is high. However, the label titles #5, soot and #6#7 diagnostic ratios. There are few connections among the nodes in these clusters, which reflects the low co-occurrence degree of keywords in the research field.

3.5. Keyword timeline
On the basis of the cluster graph, in order to further study the evolution of keyword context within each cluster, click the layout - Timeline view in the control panel to get the time line graph of keyword distribution. The red text on the right is clustering tags, and each tag is a keyword in the co-occurrence network. These keywords are spread out in the cluster according to their own year of occurrence, showing the development of keywords in each cluster.
Fig 5. Keyword timeline analysis obtained by CiteSpace.

Most of the lines in the graph are concentrated in the corresponding clusters, but there are also some cross cluster lines used as a new breakthrough point for research. For example, organochlorine pesticides, biodegradation, particle positive matrix factorization. These four cross cluster lines are indicating that the co-occurrence degree between these research directions is high. But PM2 cross cluster connection is relatively few, it is obviously relatively independent, almost no cross clusters co-occurrence.

4. Conclusion
Based on the bibliometric data of polycyclic aromatic hydrocarbons (PAHs) in the field of environment, this paper analyzes the related literatures collected from 2005 to 2020. the key research trend of this field is "pollution sources of polycyclic aromatic hydrocarbons". In recent years, "PM2.5, health risk, impact, source identification" have become hot spots. The keywords with the strongest citation bursts are including toxicity, spatial distribution, health risk, pm2.5. Among them, toxicity and spatial distribution have increased sharply since 2015, health risk and PM2.5 have increased sharply since 2019, and there is no sharp decrease. These words are the research hotspots of scholars in this field in recent years. There are many cross clusters links among organochlorine peptides, biodegradation, particle, positive matrix factorization, which indicates that there is a high degree of cross cluster co-occurrence among these research directions. PM2 and spot has few cross clusters links, which is obviously relatively independent and almost no cross clusters co-occurrence. Therefore, the follow-up cross cluster research can be used as a new entry point for exploration and research.

Acknowledgments
Supported by the Academy–Locality Cooperation Project of Sanya (2019YD04).

References
[1] Aziz. F, Syed J. H., Malik R. N., Katsoyiannis A., Mahmood A. and Li J 2014 Occurrence of polycyclic aromatic hydrocarbons in the soan river, pakistan: insights into distribution, composition, sources and ecological risk assessment. *Ecotox. Environ. Safe.* 109 77-84.
[2] Ravindra K., Sokhi R. and René Van G. 2008 Atmospheric polycyclic aromatic hydrocarbons: source attribution, emission factors and regulation. *Atmos. Environ.* 42 2895-2921.
[3] Loska K., Wiechula D. and Korus I.M. 2004 Metal contamination of farming soils affected by industry. *Environ. Int.*, 30 159-165.
[4] Yao H., Zhang S.C., Liu M.L., Wang J., Lu Y.T. and Yu X.H. 2018 Sources and risk assessment of Polycyclic Aromatic Hydrocarbons from the urbanization process of topsoil. *Chin. J. Envir. Sci.* **39** 143-146.

[5] Duan F.K., He K.B. and Ma Y.L. 2009 Concentration and sources of atmospheric polycyclic aromatic hydrocarbons (PAHs) in PM2.5 in Beijing, *Acta. Scien. Circum.* **29** 1363-1371.

[6] Tao S., Luo Y.M., Zhu L.Z., Wang X.J., Cao J., Li B.G., Xu F.L. and Zhang Z.H. 2006 Regional environmental processes of typical trace organic pollutants. *Acta. Scien. Circum.* **26** 170-173.

[7] Qiu J.P., Duan Y.F., Chen J.Q., Song E.M. and Ji L. 2003 The retrospect and prospect on bibliometrics in China, *Stud. Sci. Sci.* **21**143-148.

[8] Sun Y.Q., Wu S.M. and Gong G.Y. 2019 Trends of research on polycyclic aromatic hydrocarbons in food: A 20-year perspective from 1997 to 2017. *Trends Food Sci. Tech.* **83** 86-98.

[9] Wu M. J., Wang H., Wu Y.N., Wang W.H. and Zou H.F. 2020 Research Status of Polycyclic Aromatic Hydrocarbon Degradation in Soil Environment Based on Bibliometrics. *Chinese Agr. Sci. Bull.* **36** 60-67.

[10] Gu X.Y., Wu J., Wang M., Wu J.Q., Xu Z.H., Tang H. and Su J.H. 2016 Research advance of polycyclic aromatic hydrocarbons in environment: a bibliometric analysis based on Web of Science. *Environ. Sci. Technol.* **39** 196-204.

[11] Zhao Q., Zhang W., Wang R., Wang Y. and Ouyang, D. 2017 Research advances in molecular modeling in cyclodextrins. *Curr. Pharm. Design.* **23** 522-531.

[12] Chen C.M. 2006 CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *J. A.M. Soc. Inf. Sci. Tec.* **57** 359-377.