An Overview of research and development themes in the measurement and occurrences of Polycyclic Aromatic Hydrocarbons in dusts and particulates

Mark R Cavea*, Joanna Wragga, Darren J Beriroa, Christopher Vanea, Russell Thomasb, Matthew Ridingb and Christopher Taylorc

a British Geological Survey, Keyworth, Nottingham, NG12 5GG, UK
b WSP Kings Orchard, 1 Queen Street, Bristol, BS2 0HQ, UK
c National Grid Property Holdings Ltd, National Grid House, Warwick Technology Park, Gallows Hill, Warwick, CV34 6DA, UK

*corresponding author mrca@bgs.ac.uk

Abstract

Polycyclic aromatic hydrocarbons (PAHs) are a group of organic compounds consisting of two or more fused aromatic rings and are probably one of the most studied groups of organic chemicals in environmental research. PAHs originate mainly from anthropogenic processes, particularly from incomplete combustion of organic fuels. PAHs are distributed widely in particulate matter. Due to widespread sources and persistent characteristics, PAHs disperse through atmospheric transport and exist almost everywhere. Human beings are exposed to PAH mixtures in gaseous or particulate phases in ambient air. Long-term exposure to high concentrations of PAHs is associated with adverse health problems. This review identifies the main research and development themes in the measurement and occurrences of PAHs in dusts and particulates using a new approach to carrying out a literature review where many peer-review publications have been produced. The review extracts the most important research themes from a literature search using a combination of text mining and a more detailed review of selected papers from within the identified themes.

Key words: review; PAH; particulates; dust; text mining; research themes

1 Introduction

The specific interest of our research group is the measurement of human exposure to polyaromatic hydrocarbons (PAHs) derived from airborne particulates and dust derived from coal-tar contaminated soil or related sources which can subsequently be used as part of a human health risk assessment. Whilst this is a relatively well defined research area, the aim of this review was not to narrow the search for published studies to this topic but to get as wide an overview of all the literature covering all aspects of the sources, sampling methods and measurement of PAHs in dusts and particulate matter to ensure that we could make use of any of the wider research developments in our studies.

PAHs are probably one of the most studied groups of organic chemicals in environmental research. A general search of papers on the Thompson Reuters WEB OF SCIENCE™ search engine containing PAH in the title or the abstract returns over 42000 papers published since 1970. Plotting the number of papers published per day (Figure 1) shows that the number of publication takes a sharp increase in 1991 to about 1 paper per day with an approximate linear increase 8 papers per day in 2016.
The Thompson Reuters research categories (Figure 2) show that over 30% of the papers come from environmental science and ecology research and 20% on chemistry research topics reflecting the fact that their effects on the environment are the main concern and that development of analytical chemistry protocols and sampling methods play a crucial role in quantifying PAHs in environmental matrices.
This very large data set which contains a huge store of information but accessing specific topics within this data has two main drawbacks:

i) It is very difficult to get an overview of the trends in a data set of thousands of papers as it is impossible to read and digest the information in that many abstracts let alone read all the papers; and

ii) Even if a team of reviewers were able to read the papers and produce an overview review within a period of a few weeks the review would be out of date as new papers are published daily (Figure 1).

One obvious way of making the task more manageable is to refine the search to be more specific for a given research area. In this study we are interested in PAH in airborne particulate matter therefore, in addition to PAH we searched specifically for dusts or particulates or PM10 or PM2.5 or geogenic dusts but excluded cosmic or interstellar or galactic or diesel or sediments or leaf or leaves or leafy (for the full search specification see the supplementary material). This reduced the number of papers down to 3041. Whilst inhalation of PAH contaminated soil particulates entrained in air at contaminated land sites is the particular focus of this study further refining the literature would exclude a large amount of material which may not be exactly aligned to our study but may contain important information that can have direct application to our research. It is for this reason that we decided to try and review this larger literature body.

Figure 3 shows the number of papers published per year for the refined PAH search which broadly follows the same pattern as the unrefined search with a linear increase in papers published between 1991 and 2016 but the rate of publication is between ca.0.7 and 5 papers per week. Despite a reduction in papers by a factor of 15, more than 3000 papers is still a large number of papers to read and be reviewed individually.
Text mining is the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning. The biological sciences, in particular, are making use of this method of extracting data from large amounts of scientific literature[1].

![Figure 3 Number of papers per year and their rate of production for the refined PAH search](image)

The aim of this study is to use a combination of data and text mining to review the relevant key scientific trends in the large volume of PAH particulate matter literature in a rapid and efficient manner and to drill down to give a more detailed review of the key papers covering PAH content of particulates.

2 Methods

The study will use the information supplied with the Thomson Reuters saved search as outlined in Table S1 and saved as a tab delimited text file.

The tab delimited text file will be split into the Thompson Reuters defined research areas so that differences between the research viewpoints and the overall data can be examined.

The review process will then proceed as follows:

Review the timescales, journal names and number of citations of the papers in each group;

Carry out text mining of the paper titles to find commonly occurring word pairs using the “tidytext” package in the R programming language[2] and visualising these as network graphs us the R “igraph” package [3] to show the underlying trends within the different research areas;

From each of the research areas select 5 of the highest cited papers most strongly associated with measuring PAH in particulate matter and carry out a detailed review by reading the original text of each paper; and

Use the R programming language [4] text string search packages “stringr” [5] and “stringi”[6] to find and summarise occurrences of key phrases and associated numerical entries identified as being
important by the text mining and the specific detailed reviews in the abstracts of all of the papers obtained from the Thompson Reuters review search.

3 Results

3.1 All research areas

Using the addresses of the corresponding authors to identify the countries associated with each paper, Figure 4 shows the top twenty countries with USA, China and Italy being the main contributors but there is good coverage from most of the major continents apart from Australasia. Figure 5 shows the number of publications categorised by journal title. Since we are focussing on fine particulate matter, it is not surprising that an atmospheric journal (Atmospheric Environment) is the most common journal. Of the 17 journals shown, 4 are related to air quality (Atmospheric Environment, Atmospheric Research, Aerosol Air Quality Research, Journal of the Air & Waste Management Association) 8 are environmental research journals (Environmental Science and Technology, Science of the total Environment, Chemosphere, Environmental Pollution, Environmental Monitoring and Assessment, Environment International and Environmental Research), 2 are related to human health (Journal of Hazardous Materials, Environmental Health Perspectives), 1 to compound specific research (Polycyclic Aromatic Compounds), 1 to toxicology (Mutation Research - Genetic Toxicology and Environmental) and 1 to analytical chemistry (Journal of Chromatography A). This again reflects PAH particulate research is predominantly related to monitoring PAHs in environmental applications, developing better analytical chemistry strategies for their analysis and assessing their effect on human health.

![Figure 4 Number of paper published on PAH particulates broken down by country](image-url)
Figure 5 Journals associated with the PAH particulate papers

Figure 6 summarises the quality of the papers as measured by their number of citation. Figure 6 A shows the number of citations per paper over time and indicates that the most cited studies were carried out in the mid-1990s through to the early 2000s. Figure 6 B shows that at least 50 of the papers have 10 citations each but also there are 49 individual papers that have between ca. 90 and 800 citations each.

Having gained an insight into the temporal trends, the countries responsible for publishing in this area, the main journals where the papers appear, and the quality of the papers, we now need to look into the scientific content of the papers. At a first stage this has been carried out by applying text mining techniques to the titles of the papers. The method used in this review closely follows the text mining procedure outlined in a recent text [7] and consists of the following steps:

i) “tokenise” the journal title into word pairs (bi-grams) to produce a data set of one word pair per row;
ii) The word pairs are split into two and pairs containing “stop words” (in text mining stop-words are words which are filtered out before or after processing of natural language text. Stop words usually refer to the most common words in a language e.g. “to” and “the”) are removed;
iii) The frequency of the recombined word pairs is then counted;
iv) The frequency word pair data set provides three pieces of information: the start word of the word pair (the node an edge is coming from); the end word of the word pair (the node an edge is going towards); and the frequency count which can be considered as a weighting factor associated with each edge;
v) The “igraph” package in the R programming language converts the frequency word pair data set into a network format that can be plotted.
Figure 6 Number of citations for each paper by year (A) and the number of citations for each paper (B)

Figure 7 Bigram network plot of paired word associations for all the papers in the PAH particulate literature search
Studies associated with the Guwahati, Analytical Chemistry Methods; Application areas; Health and Toxicology; Types of Particulates; and Geographical location. Whilst this gives a good overview it is possible to look into the themes of journal titles shown by bigram plots of the individual Thompson Reuters Web of Science research areas shown in Figure 8. The dominant research areas are Environmental Sciences and Ecology, Engineering, Meteorology and Atmospheric Sciences, Chemistry, Toxicology and Public Environmental Occupational Health. Bigrams were produced for each of these research areas (Figures S1-S11 in supplementary material) with a view to pulling out research themes not evident at the higher level bigram plot (Figure 7). Each bigram was examined for research themes, going from all papers in the search through the research categories shown in Figure 8 from the highest number of papers to the lowest. After the bigram for all the papers, only new research themes shown in the subsequent bigrams were noted. The results of this process are shown in Table S2 (supplementary material). Whilst Figure 4 shows that authors of papers on PAH in particulates come from a wide range of countries the titles the main geographies of the papers of Table S2 are in China, Mexico, Guwahati, Czech Republic and Rio de Janeiro with a number of papers on New York and the World Trade centre arising and the dusts produced from the “9/11” terrorist attack. The appearance of papers looking at seasonal variation show that monitoring over periods of months is common. Studies have been carried out on both indoor and outdoor dusts. The bigram (Figure 7) suggests that PAHs are often measured in conjunction with other contaminants, including heavy metals and polychlorinated biphenyls (PCBs). The source of PAH containing dust in many instances is from burning of fuels (wood and coal) and incineration of waste. In addition to this, there is a theme of human health research related to occupational exposure to PAHs many of these from coal-tar related products (e.g. road and pavement resurfacing) in which exposure is measured by the concentration of PAH metabolites in blood or urine (typically 1 hydroxy pyrene). Risk assessment associated with a cancer endpoints are also a common subject. Papers on the toxicology of PAHs
concentrate on different aspects of measuring damage to DNA. In many instances there are multiple sources of PAH which are investigated through source apportionment techniques. Finally, a key part of literature under study here is the methods used for extracting and analysing the PAHs in the particulate material. The main sample extraction protocols include: solid phase micro-extraction (SPME); microwave assisted extraction; accelerated solvent extraction (ASE); supercritical fluid extraction; and thermal desorption. The main method of analysis, post extraction, is gas chromatography interfaced to varying modes of mass spectrometry, although high pressure liquid chromatography allied to UV or fluorimetric detectors is also used.

3.2 Specific Research Areas

Having identified the main research themes from the bigram text mining of the journal titles, more specific aspects of the papers can be gained from querying the contents of the abstracts which should contain a succinct summary of each paper, although the quality of abstracts can vary significantly from journal to journal and within journals.

Since this review is specifically about the PAH family of compounds it would be good to know which PAHs were being analysed and what range of concentrations are being found. Figure 9 shows the number of papers which report the names of the EPA 16 PAHs in their abstracts. Clearly benzo[a]pyrene (BaP) has the highest frequency of occurrence appearing in over 300 of the paper abstracts. This would be expected since BaP is usually quoted as one of the highest toxicity PAH compounds [8] and often results are quoted as “BaP equivalents”.

**Figure 9 Number of papers in which the EPA 16 PAHs in appear in the abstracts**
The concentrations of PAHs found in the different studies can only be compared if they are all reported in the same units. Searching for the six most common units (ng m\(^{-3}\), μg m\(^{-3}\), mg m\(^{-3}\), ng kg\(^{-1}\), μg kg\(^{-1}\) and mg kg\(^{-1}\)) in all the abstracts shows that 489 contain one or more of these unit categories. Figure 10 shows a breakdown of the number of abstracts with each unit of concentration present. Interestingly, ng m\(^{-3}\) is the most common showing that in those papers the particulate samples must have been collected using volumetric samplers which would allow the amount of PAH in a particulate collected in a given volume of air.

![Occurrence of units of concentration of PAHs in paper abstracts](image)

**Figure 10** Occurrences of units of concentration in all papers

![Range of data reported in abstracts using units of ng m\(^{-3}\)](image)

**Figure 11** Range of data reported in abstracts using units of ng m\(^{-3}\)
Figure 11 shows the range of data reported in papers where the most frequently used unit of ng m\(^{-3}\) has been used. Although we cannot be sure that each of these data points refer to PAH concentrations and, if they do, which PAH compound it refers to, but the histogram gives a broad brush view the range of volumetric concentrations. Whilst most of these data are between 0 and 10 ng m\(^{-3}\) there are a few data points between 500-2000 ng m\(^{-3}\).

3.3 Literature associated with specific research areas

In this particular literature review, there is a specific interest in understanding the risk to human health from PAHs associated with coal tar related sources in soil dust. Whilst this is a relatively specific research area it has been important to consider the wider literature of PAHs in air entrained particulates as this will also have relevance to the more specific area of interest. Despite this, an overview of the literature related to the specific topics of interest is important to consider. Sub-samples of the original literature search were made by searching the literature abstracts for the words: “soil” (265 papers found), “coal-tar” (95 papers found), “Bioaccessibility” or “Bioavailability” (86 papers found), “risk assessment” (161 papers found). Bigram plots were produced each of these subsets (Figures S8-S12 in supplementary material). Table 1 summarises the additional themes, over and above those identified in Table S2. Whilst this does not identify many additional themes we do see that under the coal-tar related papers there are a number of studies on coal-tar related products and liver cancer shows up as an additional health hazard.

Table 1: Research themes from specific topic areas

<table>
<thead>
<tr>
<th></th>
<th>Chemical Compounds</th>
<th>Application areas</th>
<th>Health and Toxicology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil</td>
<td>low molecular weight PAH</td>
<td>principal component analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>parquet glue</td>
<td>liver cancer risk</td>
</tr>
<tr>
<td></td>
<td></td>
<td>coal tar pitch</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>asphalt based sealant</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>sealed/unsealed parking lots</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>bitumen fumes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>pavement sealants</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>manufactured gas</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>graphite electrodes</td>
<td></td>
</tr>
<tr>
<td>Coal-tar</td>
<td></td>
<td>geological origin</td>
<td>dermal penetration</td>
</tr>
<tr>
<td></td>
<td></td>
<td>contaminated soils</td>
<td>lung bioaccessibility</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>oral bioavailability</td>
</tr>
<tr>
<td>Bioavailability/Bioaccessibility</td>
<td></td>
<td>aerodynamic diameter</td>
<td>dose response</td>
</tr>
<tr>
<td></td>
<td></td>
<td>diagnostic ratios</td>
<td>inhalation exposure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cloud water</td>
<td></td>
</tr>
</tbody>
</table>

Additional source apportionment techniques appear in the soil and risk assessment categories (principal component analysis and PAH diagnostic ratios). Under bioavailability/bioaccessibility and risk assessment all three exposure pathways are mentioned although lung bioaccessibility seems to be related to cell line studies. In the risk assessment category there is a first appearance of aerodynamic diameters which are used to calculate the particle sizes entrained into air.
This illustrates that the text mining approach can give a very rapid overview of the main trends in a large volume of literature data but these overall trends may not supply enough information on a specific topic of interest. In this study, we have shown how we have drilled down further to get specific information on applications related to human health risk assessment. It may be, however, that researchers are more interested in specific aspects of sample collection rather than the overall applications, for example possible artefacts associated with volatilisation and degradation of PAHs during the sampling process [9, 10]. A search for words beginning with either “volatiliz” or “volatilis” in paper abstracts shows that only 46 papers out of the 3041 abstracts and a search for “degradation” 89 papers out of 3041 abstracts from the original search and therefore why these would not show up as a major theme of the broad search criteria used. This is not to say that this area is not important topic but would probably be best addressed by using more targeted initial search criteria."

4 Reviews of selected papers

Having used text mining to identify primary themes over all the papers resulting from the original literature search a selection of these papers have been reviewed in more detail. As already stated it is not feasible to review every paper so the papers chosen were based on the scientific quality of the paper measured, in this instance, by the number of times the paper had been cited in the scientific literature[11]. A total of 45 research papers from the seven overarching highlight topics (Environmental, Occupational, Chemical, Engineering, Atmospheric, ‘Other’ and Reviews) and with greater than 50 citations each, were identified to provide a snapshot of work related to human exposure to PAH via inhalation. Whilst the choice of the number citations is somewhat arbitrary, papers with 50 or more citations are usually considered to be well respected scientific studies and it provided a pragmatic method of selecting a manageable number of papers to review. Five papers per topic were chosen with a larger number (n=14) to represent the wide body of review papers available for different aspects of inhalation exposure. This choice was based on expert opinion with particular emphasis on papers with relevance to the specific interest of our research group i.e. the measurement of human exposure to PAHs derived from airborne particulates and dust derived from coal-tar contaminated soil or related sources. Within each highlight topic, individual papers provided key common areas for discussion: the PAHs of interest; PAH sources; sampling and analysis methodologies; along with the aims and outcomes. The publication period was between 1991 and 2013 and resulted in a total of 6312 citations. In the main, the range of PAH discussed here are listed as USEPA priority contaminants. A large proportion of the papers considered were not restricted to PAH alone as they often include co-contaminants such as PCBs, dioxins and metals, although only PAH have been included as part of this review. The studies (indoor and outdoor) considered here were undertaken on a city, country or continental basis, covered exposure exposure to <10μm particles or particles with an aerodynamic diameter <10μm (where aerodynamic diameter of an irregular particle is defined as the diameter of the spherical particle with a density of 1000 kg m−3 and the same settling velocity as the irregular particle) as well as associated vapour, and from sources such as traffic, asphalt, building rubble, fuel burning etc. This scope of the current review does not include PAH toxicity/mutagenicity/health outcomes.

4.1 Review papers

4.1.1 PAH and particulate sources and reactions
The body of work contained in the review articles was consistent in a number of themes/areas:

- **Sources of PAH in particulate matter (PM) and in the gas phase.** Sources include domestic burning of coal and wood for heating and cooking, biomass fuel burning power plants, industrial processes, road transport and urban environments [12-16], dust generation from office equipment [17] and low temperature combustion of e-waste for disposal purposes[18, 19] which also includes contaminating the surrounding soil. The concentration in air reflects the location of the source emitters [20]. After PAH compounds are released into the atmosphere they undergo a variety of processes and reactions for removal, transportation and degradation, which may result in the production of compounds that are more (or less) mutagenic and carcinogenic than the parent PAH [20]. Han et al.[13] report that PAHs are mainly bounded onto fine respiratory particles (pPAHs), especially ultra-fine particles, though PAH accumulation may occur on larger particle sizes later in the atmosphere. Wong et al.[19] summarised that the majority of PAH (70%) were 5 and 6 ring and concentrated in PM 2.5 μm size fraction for soils/dusts around a Chinese e-waste processing site with a range of PAH concentrations in the soil (from 44.8 to 3206 μg kg⁻¹ for the total amount of PAH) highlighting the heterogeneity of contamination and the data shown in Figure 11. The combination of outdoor and indoor source activities, with transport of outdoor particulates makes a significant contribution to settled house dust (SHD) [16] which is subsequently a significant source for indoor exposures. At ca. 26°C the fraction of the PAH considered in this review that would be expected to be adsorbed to indoor particulate matter is >80% [16].

Other review topics were:

- **Co-existing contaminants and source dependency.** PAHs in air/dust particulates from the sources described above co-exist with other contaminants such as PCBs [14, 18] are source dependant mixtures of different PAH with large intra-source variability [21] and react with other atmospheric gases (e.g. Ozone [20]) and particulate matter, for example organic matter. Robinson [18](2009) highlights PAH concentrations of up to 20,000 ng g⁻² in Chinese agricultural soils surrounding e-waste processing sites.

- **Partitioning of PAH between gas and particulate phases.** PAH are associated predominantly with particulate matter, but are also present in the gaseous phase [20]. Under ambient conditions two ring PAH exist, almost entirely in the vapour phase, whereas for 3 and 4 ringed PAH split between the vapour and particle phase and compound with a higher number of rings (and higher carcinogenicity [12] partitioning almost entirely in the particle phase [14]. Cousins et al.[22] reports that the distribution of SVOC between gas and particle phases depends on the compound’s vapour pressure and the concentration of particles suspended in the air. Adsorption is the dominant process for PAH association with soil/dust, with the degree of adsorption controlling the potential of PAH to subsequently evaporate from particulate matter, however partitioning/source apportionment of any given PAH is also related to changes in environmental conditions [12, 22]. In soils, PAH partition between soil solids, interstitial solutions and air-filled porosities; the distribution of mass between each phase can be characterised by employing a number of partition coefficients, and the exchange of mass, by diffusion coefficients [22].
Exposure scenarios and Human physiology. Exposure to indoor pollution is a high priority area according to organisations such as the USEPA [16]. In developed countries, most people spend the majority of their time indoors, inhaling around 22 m$^3$ of air per day [23]. The extent of human exposure to PAH is related to a number of parameters, including ambient air concentrations, prevailing atmospheric conditions, distribution between the gaseous and particulate phases and the size of the particles with which the particulate fraction are associated [20]. Maertons [16] reported that inhaled dust particles (>10μm) tend to be trapped in the nose, throat or upper respiratory tract. Smaller particles, <2.5μm, often contain higher concentration of the PAH of interest and are less likely to be eliminated from the respiratory system because they penetrate deeper.

4.1.2 Sampling strategy and analytical measurements

Barro et al. [23] have provided a substantial summary of research (from between 1999 and 2009) related to sampling of indoor air and dusts, and the subsequent extraction and analysis of up to 55 PAH. In summary, indoor air sampling has been generally undertaken by passive and/or active techniques over different timescales (22-48 h) and, where applicable, at flow rates between 1 and 224 l min$^{-1}$. Barro et al. [23] point out that the passive sampling approach is becoming increasingly attractive. This type of sampling is often used for collecting vapour phase PAH in occupational environments [24]. Collection methods for PAH in particulates use a variety of sorbent material (e.g. quartz filter fibre), with desorption by methods including soxhlet extraction and ultrasonication with various organic solvents [12, 17, 19, 23, 24]. Although filters and adsorbents have been used for sampling atmospheric PAH in a large proportion of the work to date, they may not be suitable if other volatile species are of interest as they may not be completely extracted once adsorbed [12]. According to Barro et al.[23] the preferred method for indoor dust sampling is using a vacuum cleaner. For both indoor air and dust samples the identification and quantification of PAH is, in the main, carried out using Gas Chromatography with Mass Spectrometry [17, 23, 24] which is also supported in the searches of the wider literature research data Figure 7 and Table S2.

The methodological approaches and quantification/speciation methods for the continuous measurement of particle mass and its chemical composition carried out as part of individual studies have been summarise [24, 25]. The outcomes of the report by Chow et al. [25] suggest that for organic speciation of PM samples, in-injection port thermal desorption-GC-MS offers a fast, simple, cost effective and sensitive approach which does not require sample pre-treatment and avoids the use of organic solvents thereby reducing sampling contamination. Peltonen and Kuljukka [24] discuss analysis using HPLC equipped with a fluorescence detector or gas chromatography mass spectrometry with electron and negative chemical ionization methods. They also consider that for qualitative analysis mass spectrometry lacks the ability to resolve the isomeric structures although this may not be such a problem with modern instrumentation. The difficulties in comparing studies using different analytical methods (with differing limits of detection) are highlighted by Chow et al.[25].

4.1.3 Conclusions on Review papers

This summary of review papers provides a snapshot of the range and complexity of sources of PAH, partitioning between solid and aqueous phases and, methods of sampling and analysis. They make the important comment that because of the growing emphasis on environmental monitoring, with respect to human health and the inhalation exposure route, there is a need to develop fit for purpose, rapid and less expensive methods to determine particulate bound PAH. This approach is considered
to be particularly applicable to any future development of practicable in vitro simulations of the respiratory environment.

4.2 Analytical Chemistry Methods

The body of work considered here highlights the range of dust/soil/vapour contaminants that have been studied. In some cases the work has concentrated on individual PAH and in others multiple PAH, with and without PCBs and metals. The snapshot of work carried out on soil/dust PAH shows the use of multiple sampling techniques to collect breathing zone vapour or particulate material of differing size, storage techniques, and extraction methodologies and solvents, which have been coupled to both HPLC and GC-MS detection. The dominant size particle fraction observed in this review was <2.5 µm, with 10 and 0.1 µm particulates featuring highly. In the main particulate material was collected using glass or quartz fibre filters with vapour collection on polyurethane foam plugs (PUF). Both static and high volume air flow samplers feature heavily, along with personal breathing zone samplers. The information on the analytical chemistry methods is best summarised in a tabular form (Table S3) [26-54]. These papers note the differences in extraction rates of the spiked and native PAHs were most dramatic for the lower molecular weight PAHs, indicating that relatively volatile species such as naphthalene must be tightly bound in order to remain associated with a real-world sample. Where they have been included the NIST SRMs 1648 (urban particulate matter) and 1649a (urban dust) appear to be the reference material of choice. The results show a large variability in recovered concentrations of individual PAHs and are a reflection of the extraction and detection methodologies applied.

4.3 Gas-Particle Partitioning

Whilst some reviews on this topic have already been discussed in Section 4.1.1 the bigram plot (Figure 7) shows that gas-particle partitioning (GPP) is an important theme. There are 72 papers of the original 3041 that actually use the “gas particle partitioning”. A bigram plot of these papers (Figure S12) shows a few of additional themes associated specifically with these topics which occur over and above those identified in Table S2. These are “pankow junge”, “dachs eisenreich”, “soot carbon”, “elemental carbon”, “subcooled liquid vapour pressure”, “octanol air partition coefficient” and “adsorption/absorption model”. These bigrams groupings are all related to attempts to understand the partitioning of the PAH between the gaseous state and the solid phase particulate.

Li et.al. [55] give a succinct overview of the processes involved which explain how the bigram groupings fit into GPP modelling. Once PAHs enter the atmosphere, they are redistributed between particulate and gaseous phases, according to vapour pressure, aerosol chemical and physical characteristics, and meteorological conditions. The gas-particle distribution of PAHs is therefore a very important parameter in describing their atmospheric transport and fate. Several models have been suggested in order to predict atmospheric concentrations of PAHs accurately. These include:

Junge-Pankow adsorption model [56] based on the super-cooled liquid vapour pressure ($P_c$) and related research [39, 57-69] with example applications [70-78];

Harnet-Bidleman organic matter (OM) absorption model [79] based on the octanol-air partition coefficient (Koa) and related research [80-82]; and
Dachs-Eisenreich, a dual blacksoot carbon (BC) adsorption and OM absorption model [83] and example related research [70, 71, 84, 85].

Many of the papers investigating GPP have applied one or more of these modelling strategies in their investigations [55, 62, 67, 70-78, 84-101].

During the course of their transport, PAHs are removed from the atmosphere by dry and wet deposition. Precipitation scavenging represents a significant fraction of the total atmospheric deposition of PAHs. The process is important for both the removal of particulate-bound and gaseous PAHs from the atmosphere [95]. Particulate bound PAHs are being washed out from the atmosphere by physical processes controlled by size, number density, and solubility of aerosol along with meteorological conditions and microphysics [95]. Gaseous phase PAHs dissolved into water are determined by Henry's law, by molecular diffusion and in-drop mixing [102].

4.4 Biomarkers of PAH exposure

Whilst the total concentration of PAH in the particulate matter is an important measure of the magnitude of contamination, the key measure of human exposure by inhalation is how much of the compound enters the systemic circulation of the body. A more realistic parameter for the individual exposure, therefore, may be obtained by measuring the excretion or blood content of the incorporated PAH and their metabolites formed by a number of different enzymes, e.g. in the liver, lung and bladder[103].

4.4.1 Biological samples to assess PAH exposure

Although not a feature of all of the selected papers included in this review, where biological markers have been included a brief overview follows. Samples were collected from two studies with male volunteers with different occupations by [46, 48]. Both male and female students were included another study on general environmental exposure [49]. In other studies [30, 50, 51, 54] the sex of workers and children was not reported.

4.4.2 Biological sample collection and storage

In some of these studies [30, 46, 48, 50, 51, 54] urine samples were collected for analysis. No details of collection methods were recorded, apart from one study that collected urine on a plastic bonnet placed under the toilet seat [30]. The urine was collected over each 24 h period [30, 50, 51], and, pre and post shift [46, 48]. The scheduling of sample collection was different (in most cases) in relation to both exposure time and trial length. The urine samples collected as part of the studies included in this review were frozen at different temperatures, for storage prior to analysis, in the range -20°C to -70°C (or in some cases not reported).

In some studies blood samples were collected instead of [49], or in conjunction with urine [46, 54] in both test and control subjects. As with the sampling of urine, limited details are available for collection of blood samples. The samples were taken on varying timescales but pre-shift in all cases.

4.4.3 Biological markers

The most frequently used marker of total PAH exposure in urine was 1-hydroxy pyrene (1-OHP) [30, 46, 48, 50, 54], which in some cases formed part of a suite of biomarkers that also included
dihydrodihydro-phenanthrenes and pyrenes as methoxy-PAHs, phenanthrene metabolites, pentachlorophenol (PCP), 2,4-dichlorophenoxyacetic acid (2,4-D). Four studies [30, 46, 48, 54] reported the 1-OHP in relation to creatinine. In each study the sample pre-treatments are given in with detection of 1-OHP by HPLC [46, 48, 54], reversed phase HPLC [51] and GC/GC-MS [30, 50]. DNA adducts extracted from blood samples were used [46, 49, 54] as these are considered to be the most widely used biomarkers for blood samples. As with the use of 1-OHP the full details of analysis are discussed by the individual authors.

4.4.4 Biological marker outcomes

The PAH metabolites 1- and 2-naphthol, 3-hydroxyfluoranthene, 1-hydroxypyrene, and 6-hydroxychrysene, as well as PCP, 2,4-D, and TCP in all urine samples, whereas the benzo[a]pyrene metabolites, 1- and 3-hydroxybenzo[a]pyrene, were only found in two of the nine composite urine samples[30]. The authors concluded, when considered in relation to the other measures included in this study that, in general, the children in this study had low, likely chronic, exposures to several of the target compounds.

The studies that considered occupational exposure to PAH [46, 48, 50, 51] reported that biological marker data was variable between workers and that in some cases not representative of the variation in inhaled particulate/vapour concentrations of PAH. However, other studies [46, 54] reported a significant correlation between 1-OHP and PAH exposure. In each case the highest concentration of 1-OHP were seen in evening collected biological samples but not necessarily immediately post shift.

The study by Georgiadis et al., [49] also showed bulky DNA damage was more prominent in male rather than female participants. The within study variation in biological monitoring samples highlights the non-standardised collection time of such samples, which are determined by shift pattern, length of shift and activity, along with the distance from exposure source.4.5 General overview

This section gives an overview of the research areas covered by the papers selected for more detailed review. The overview takes the form of a table (Table S4 supplementary material). The papers have been categorised under 7 headings (title, author, year, journal, aim, sources (of PAH) and outcomes).

5 Recent trends

The review has concentrated on the most cited articles which represent key milestones in the study of PAHs in particulate matter but, by their very nature, are not the most up to date publications (Figure 6). In order to get a view of the developments in the last five years in this field the original search criteria was applied for the years 2013 to 2018 gave rise to 1268 papers. To further refine the search to papers specifically targeting PAH in particulates the search was carried out on the paper titles only (excluding the abstracts) resulting in the identification of 164 papers. Using similar procedures to those used to examine the original literature search (Section 2) Figure 12 shows that the majority of recent papers come from China.
The number of citations per publication shows that some of the more modern papers from 2013 - 2015 have more than 20 citations with the highest of 56 [104] which studies the ecological and human health risks from PAH and heavy metals in road dust from Iran.
The bi-gram plot (word pairs with frequencies of greater than 7) for the abstracts of the recent papers is shown in Figure 14 and displays very similar themes to the overall bigram plot for all references Figure 7. Examining the different bigram groupings gives rise to specific recent themes.

![Figure 14 Bigram plot from the abstracts of the recent papers](image)

5.1 PAH analysed

The bigram plot (Figure 14) shows that 4,5,6,7 and 10 ring PAHs are commonly studied and that there is not only interest in the parent PAH compounds but also in the derivatives coming from secondary formation processes [105-121] with hydroxylated [111, 113, 122-124] and nitro/oxy PAH derivatives [105-133] being of most interest. These forms are sometimes considered more toxic than their parent compound. In one study [117] a Salmonella/microsome assay was used for the evaluation of air particulate matter (PM) mutagenicity and found that the nitro and oxy PAH were the major cause.

5.2 Sampling and Analytical Methods

Whilst developments in sampling methodology for PAHs in particulates has reached a plateau, a number of recent studies [9, 10, 134, 135] have highlighted potential inaccuracies in PAH quantification due volatilisation or degradation of the PAH once they have been immobilised on the collection medium. Melymuk et. al.,[135] reported that significant breakthrough in active air samplers occurred for compounds with vapor pressure >0.5 Pa at volumes <700 m³. Sample volumes between 700 and 10,000 m³ may lead to breakthrough for compounds with vapour pressures between 0.005 and 0.5 Pa. Breakthrough is largely driven by sample volume and compound volatility (therefore indirectly by temperature) and is independent of sampler type. The presence of significant breakthrough at “typical” sampling conditions is relevant for air monitoring networks, and may lead to under-reporting of more volatile PAHs. In a follow-up study looking at both passive and active samplers[134], Melymuk et. al. found that an ozone denuder in active
samplers significantly reduced the degradation of PAHs such as BaP and that passive samplers deployed for longer time periods showed a reduction of BaP concentration by a factor of four when comparing weekly and monthly sampling, again from degradation effects. In two related studies Balducci et al., [9, 10] assessed the effects of oxidative degradation and volatilization on PM10 bound polycyclic aromatic hydrocarbons (PAHs), collected at low volume condition according to the EU sampling reference method EN12341:2014 (flow rate 2.3m³ h⁻¹), on 47mm quartz filters. They also found that removal of ozone reduced the degradation particularly on benzo[a]pyrene, perylene and benz[a]anthracene. They found oxidative degradation led to a general decrease of PAH concentrations in both summer and winter. In the summer, the effect of volatilization exceeded that of oxidative degradation for light PAHs up to benzo[e]pyrene. They suggested that these losses could be mitigated by judicious choice of the sampling period during the day.

There do not appear to be any major developments in chemical analysis over the last five years with gas chromatography mass spectrometry being the main method of choice e.g. [105-110, 112, 114-121]. Zhang and co-workers [121] have, however, reported the use of a new atmospheric pressure gas chromatography-tandem mass spectrometry (APGC-MS/MS) to simultaneously separate, identify and quantify 18 nitro-polyaromatic hydrocarbons (NPAHs) in air fine particulate matter (PM2.5). Compared with traditional negative chemical ionization (NCI) or electron impact ionization (EI-MS/MS) methods, APGC-MS/MS equipped with an atmospheric pressure chemical ionization (APCI) source provided better sensitivity and selectivity for NPAHs analysis in PM2.5. 18 NPAHs were completely separated, with low instrumental detection limits.

5.3 PAH sources

Two main sources of PAHs in particulate matter are identified in Figure 14:


ii) Vehicular emissions [104, 108, 111-113, 116, 119, 125, 126, 139, 140, 143, 144, 147, 148, 152, 154, 155, 157, 159-162, 165, 171, 176-197]. Cao and co-workers [126] analysed 34 road dust samples collected from Xinxiang, China. Total PAHs concentrations indicated a significant positive correlation existed between PAH contamination and traffic density.

Although not so widely reported, there are also studies which show that firework displays produce significant amounts of PAH particulate matter

In addition to this, temporal variations are also important in controlling PAH concentrations [112, 137, 139, 142, 144, 148, 156, 159, 165, 166, 170, 171, 207-211]. In a study of three Chinese cities, Beijing, Shijiazhuang and Tianjin [166], a total of 671 air samples were collected in including a reference site, and four national air quality background sites. PAH concentrations varied significantly in different seasons, with similar trends in the three cities. The PAH concentrations measured during the central heating period of winter and spring greatly exceeded the Chinese national standards for 24-h average and annual average concentrations of BaP.

5.3.1 PAH source apportionment
The bigram indicates that considerable work goes into multivariate statistical methods for identifying the origin of different sources of PAHs found in particulate matter, the main methods being:

Molecular diagnostic ratios (MDRs) do not appear specifically in the recent bigram (Figure 14) but do appear alongside other methods for source apportionment in the recent papers. MDR are[212] unitless concentration ratios of specific individual pair-PAHs (usually of the same molecular weight), and the MDR source identification theory/methodology is based on the assumption that each specific PAHs emitting source will always release individual PAHs at the same ratios. Although their use has come under a number of criticisms e.g.[212] they have been used in recent studies either by themselves or in conjunction with other methods for PAH source apportionment in air particulates[105, 118, 137, 141-143, 145, 148, 149, 151, 158, 159, 161, 168, 176, 186-188, 213-224]

Principal component analysis (PCA) and multiple linear regression (MLR) have been used quite widely [104, 108, 137, 138, 140, 141, 143, 145, 148-150, 153, 154, 157, 160, 161, 167, 169, 173, 175, 176, 186, 187, 189, 196, 217, 222-227]. PCA is found to be more reliable than MDR although the results are usually semi-quantitative and sometimes difficult to interpret. An example application in a Beijing study [108] organic molecular composition and carbon and nitrogen isotope ratios of PM2.5 samples collected in November 2013 and analysed using gas chromatography/mass spectrometry and isotope ratio mass spectrometry. PCA/MLR analyses indicated that the main sources of PAHs were vehicle and coal combustion emissions, while n-alkanes had high contributions from petroleum emissions. These sources were supported by isotopic analyses.

Positive matrix factorisation (PMF) which uses an alternating least squares algorithm has become an increasingly popular choice for source apportionment [113, 138, 156, 159, 163, 165, 172, 186]. It has the advantage of producing fully quantified estimates of the source compositions along with uncertainties. In this example paper [165], PM2.5 samples were collected at nine sites in a city in north eastern China from September 2013 to October 2014. Sixteen priority PAHs in PM2.5 were analysed to determine their spatial and temporal distribution characteristics. The source apportionment of PAHs was conducted with the PMF model. Six factors were found: coke oven emissions; biomass burning; petrogenic sources; coal combustion; gasoline engine emissions; and diesel engine emissions. The results indicated that the PAHs in PM2.5 in the city are primarily caused by combustion processes and vehicle exhaust.

5.4 Risk assessment

Figure 14 shows that a common theme is human health risk assessment; out of the 164 recent papers 96 contain a reference to either human health or risk in their title or abstract. In particular the PAH in particulate data is used to calculate incremental lifetime cancer risk for the exposed population [104, 107, 119, 125, 126, 140, 141, 149, 150, 157, 160, 161, 176, 190-192, 214, 221, 222, 228-231]. Three pathways for exposure from PAH in particulates are discussed:

i) Ingestion [104, 141, 149, 150, 176, 190, 214, 220, 222, 223, 231-235];
ii) Dermal contact [104, 141, 149, 150, 176, 214, 220, 222, 223, 231, 233, 235]; and
An example of a combined pathway risk assessment [223] analysed 15 PAHs in two fractions (<63 μm and 63-850 μm) of street dust samples collected in different urban areas of Rio de Janeiro and Niterói cities, Brazil. The cancer risk levels for children and adults, for both particulate size fractions, were comparable for dermal contact and by ingestion, and ranged from $10^{-8}$ to $10^{-6}$, whereas in certain cases the cancer risk level for adults by dermal contact was 2 to 3 times larger than for children. The cancer risk levels via inhalation always ranged from $10^{-13}$ to $10^{-10}$, so this exposure pathway was almost $10^4$ less severe than through ingestion and dermal contact, and thus negligible.

Although not appearing specifically in the recent paper bigram (Figure 14), bioaccessibility and bioavailability of PAHs in particulates are important factors when carrying out risk assessments. Three recent papers contain these two keywords in their title or abstract [234, 240, 241]. Guney and co-workers [234] have produced a critical review of bioaccessibility and other key parameters in assessing oral exposure to PAH-contaminated soils and dust; Wang and co-workers [241] measured PAH in scalp hair as a measure of PAH body burden; and Al-Hunaiti and co-workers [240] looked at the relationship between floor dust bacteria and fungi and the concomitant PAH content of the dusts.

6 Conclusions

This review has outlined the difficulties of carrying out detailed literature searches on broad research areas where many peer-review publications have been produced and are continuing to be published at a rate of greater than 5 papers per week. This is the case for the research area of PAHs in particulate matter. This study has suggested an approach to extract the most important research themes from a literature search using a combination of text mining and a more detailed review of selected papers from within the identified themes. Whilst this approach will not identify every aspect of research on a particular topic it does allow a pragmatic approach to understanding the developments in a particular research topic in a timely manner. Text mining is a rapidly developing research area [242] and is likely to have an important part to play in reviewing scientific literature in the future.

Acknowledgements

This work was funded by National Grid Property Holdings Ltd and managed by WSP.

References


[119] C. Wei, B.A.M. Bandowe, Y.M. Han, J.J. Cao, C.L. Zhan, W. Wilcke, Polycyclic aromatic hydrocarbons (PAHs) and their derivatives (alkyl-PAHs, oxygenated-PAHs, nitrated-PAHs and azarenes) in urban road dusts from Xi’an, Central China, Chemosphere, 134 (2015) 512-520.


[239] Y.Y. Xie, B. Zhao, Y.J. Zhao, Q.Z. Luo, S.X. Wang, B. Zhao, S.H. Bai, Reduction in population exposure to PM2.5 and cancer risk due to PM2.5-bound PAHs exposure in Beijing, China during the APEC meeting, Environ. Pollut., 225 (2017) 338-345.