Megson, David, Tiktak, Guuske P, Shideler, Steve, Dereviankin, Mike, Harbicht, Lacey and Sandau, Courtney D (2023) Source apportionment of polychlorinated biphenyls (PCBs) using different receptor models: A case study on sediment from the Portland Harbor Superfund Site (PHSS), Oregon, USA. Science of the Total Environment, 872. p. 162231. ISSN 0048-9697

DOI: https://doi.org/10.1016/j.scitotenv.2023.162231

Publisher: Elsevier

Version: Published Version

Data Access Statement: Raw data is available in the SI and is publicly available at http://ph-public-data.com

Science of the Total Environment 872 (2023) 162231

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Source apportionment of polychlorinated biphenyls (PCBs) using different receptor models: A case study on sediment from the Portland Harbor Superfund Site (PHSS), Oregon, USA

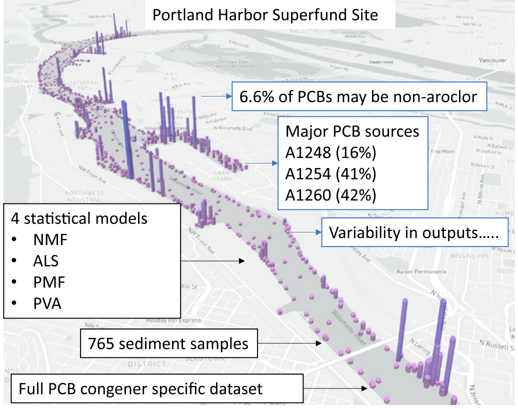
David Megson [a,b,*], Guuske P. Tiktak [a], Steve Shideler [b], Mike Dereviankin [b], Lacey Harbicht [b], Courtney D. Sandau [b]

[a] *Manchester Metropolitan University, Manchester, UK*
[b] *Chemistry Matters Inc., Alberta, Canada*

HIGHLIGHTS

- Different PCB signatures identified by 4 different statistical models.
- Major PCB sources in sediments attributed to A1248 (16 %), A1254 (41 %) and to A1260 (42 %).
- Novel residual method may be used to identify inadvertently produced PCBs.
- 6.6 % of PCBs in sediments may be attributed to non-aroclor sources.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Jay Gan

Keywords:
Polychlorinated biphenyl (PCB)
Source apportionment
Receptor modelling
Inadvertent PCBs
Non-aroclor PCBs
Portland Harbor Superfund Site (PHSS)

ABSTRACT

Multivariate modelling techniques are used by a wide variety of investigations in environmental chemistry. It is surprisingly rare for studies to show a detailed understanding of uncertainties created by modelling or how uncertainties in chemical analysis impact model outputs. It is common to use untrained multivariate models for receptor modelling. These models produce a slightly different output each time they are run. The fact that a single model can provide different results is rarely acknowledged. In this manuscript, we attempt to address this by investigating differences that can be generated using four different receptor models (NMF, ALS, PMF & PVA) to perform source apportionment of polychlorinated biphenyls (PCBs) in surface sediments from Portland Harbor. Results showed that models generally had a strong agreement and identified the same main signatures that represented commercial PCB mixtures, however, subtle differences were identified by; different models, same models but with a different number of end members (EM), and the same model with the same number of end members. As well as identifying different Aroclor-like signatures, the relative proportion of these sources also varied. Depending on which method is selected it may have a significant impact on conclusions of a scientific report or litigation case and ultimately, allocation on who is responsible for paying for remediation. Therefore, care must be taken to understand these uncertainties to select a method that produces consistent results with end members that can be chemically explained. We also investigated a

---

* Corresponding author at: Manchester Metropolitan University, Manchester, UK.
E-mail address: d.megson@mmu.ac.uk (D. Megson).

http://dx.doi.org/10.1016/j.scitotenv.2023.162231
Received 19 September 2022; Received in revised form 9 February 2023; Accepted 9 February 2023

novel approach to use our multivariate models to identify inadvertent sources of PCBs. By using a residual plot produced from one of our models (NMF) we were able to suggest the presence of approximately 30 different potentially inadvertently produced PCBs which account for 6.6 % of the total PCBs in Portland Harbor sediments.

1. Introduction

1.1. Chemical fingerprinting

Chemical fingerprinting is an important tool that allows scientists to identify similarities and differences among environmental samples to apportion pollution sources and age date contamination events (Johnson et al., 1964; Morrison and Murphy, 2010; Megson et al., 2014; Murphy and Morrison, 2014). Chemical fingerprinting has been widely used for numerous pollutants including volatile organic compounds (VOCs) (Srivastava, 2004; Srivastava et al., 2005; Cai et al., 2010), polycyclic aromatic hydrocarbons (PAHs) (Khalili et al., 1995; Boehm et al., 1997; Baldwin et al., 2020), polychlorinated dibenzo-p-dioxin (PCDD) and dibenzofuran (PCDF) (Kjeller et al., 1996; Buekens et al., 2000; Sundqvist et al., 2010) and heavy metals (Pekey et al., 2004; Muhammad et al., 2011; Men et al., 2018).

Chemical fingerprinting can be at its most effective when there are many chemicals present in a mixture and when there are a number of conserved source patterns of chemicals that have been produced through a batch process. This is the case for polychlorinated biphenyls (PCBs) which contain a group of 209 congeners that were largely produced in batches as commercial mixtures. In the U.S., PCBs were manufactured by Monsanto and given the trade name Aroclor. However, it is important to appreciate different manufacturers produced similar products in other countries (e.g. Clophen in Germany and Kanechlor in Japan). Our ability to perform better chemical fingerprinting has increased over time alongside advances in analytical chemistry techniques (Wait, 2000; Megson et al., 2013). Initial PCB fingerprinting methods were performed by visually inspecting chromatograms and using trained analysts to identify different patterns and relate these to Aroclors (e.g. Environmental Protection Agency (EPA) methods 8082A & 608) and other commercial mixtures (e.g. Clophen and Kanechlor). This historical technique worked well when releases of PCB mixtures was new and only consisted of a single type. Most releases being investigated now, including analysis of sediments in large river systems such as Portland Harbor, consists of old, highly weathered releases of PCBs from multiple commercial mixtures and incidental PCB sources. The simple Aroclor identification using EPA 8082 is no longer a simple and viable task. Today, modern fingerprinting techniques use multivariate models that rely on accurate and quantitative congener specific data which aim to separate as many polychlorinated biphenyls (PCBs) as possible. These can be obtained from multidimensional chromatography methods or high-resolution gas chromatography coupled with high resolution mass spectrometry (e.g. EPA method 1668C) (Litten et al., 2002; US EPA, 2010; Muscalu et al., 2011; Megson et al., 2013).

Most PCB studies use truncated datasets that only measured selected PCBs, such as the most abundant PCBs commonly analysed, or indicator PCBs. These limited datasets have been used estimate total PCB concentrations with some success (Longnecker et al., 2003; Fitzgerald et al., 2005; Wolff et al., 2005; Ishikawa et al., 2007; Jain and Wang, 2010) or to calculate health risks (van den Berg et al., 2006), however chemical fingerprinting techniques are most effective with larger congener datasets (Megson et al., 2019). The main reason for this is that PCB signatures change over time in the environment, and therefore separating and identifying specific PCB congeners is crucial when ascertaining important processes such as bioaccumulation, volatilisation and microbial dechlorination (Field and Sierra-Alvarez, 2008; Saba and Boehm, 2011; Rodenburg et al., 2015; Erickson, 2018, 2020; Capozzi et al., 2019). Using a reduced dataset, such as one that focuses on PCBs that are most common to specific commercial mixtures (such as Aroclors), also increases the likelihood of missing new sources of PCBs such as PCBs that are produced as by-products in other chemical processes (these are referred to as by-product PCBs here but are also commonly called inadvertently produced or incidental PCBs). There are currently over 100 published and certified methods for analysing PCBs, the majority of which have been summarized (Muir and Sverko, 2006; Clement et al., 2012; Reiner et al., 2013; Guo and Kannan, 2015; Megson et al., 2019). The absence of one universally used method presents a significant challenge when combining and comparing datasets for environmental fingerprinting. This is often approached by filtering the data by combining or removing certain PCBs to match co-elutions between the datasets. However, the more data are filtered, the more its diagnostic power is potentially reduced.

1.2. Use of multivariate models

As well as differences in analytical methods used to determine PCBs, there are also a wide range of mathematical methods that are used to analyse and interpret datasets and perform chemical fingerprinting. These range from manual visual comparison of PCB profiles to performing detailed multivariate analysis. Often the analyst will use a method that is tried and tested or a method that is most familiar to them. Within the scientific literature or during litigation proceedings, there is rarely a comparison of the effectiveness and sensitivity of different multivariate (statistical) models. Over the last two decades, advanced statistical and multivariate approaches have become more commonplace in PCB forensics. These are used to model congener specific data from environmental samples, allowing for more accurate source apportionment, and identifying other factors such as degradation and incidental PCBs (Johnson et al., 2015). Some studies may still rely on visual inspection of data and do not present multivariate models (Hagemann et al., 2018), whereas others are starting to use a combination of models (Mao et al., 2020). The following multivariate models have been used for PCB fingerprinting: Principal Component Analysis (PCA) (Megson et al., 2013, 2014, 2015; Ranjbar Jafarabadi et al., 2018; Foster et al., 2019; Guerra et al., 2019; Mao et al., 2020; Santos et al., 2020), US EPA Positive Matrix Factorization Model (PMF) (US EPA, 2014; Karakas et al., 2017; Capozzi et al., 2018, 2019; Cetin et al., 2018; Rodenburg and Delistraty, 2019; Yurdakul et al., 2019; Mao et al., 2020; Rodenburg et al., 2020) and Polytopic Vector Analysis (PVA) (Johnson et al., 2000, 2015; DeCaprio et al., 2005; Magar et al., 2005). As well as Partial Least Squares (PLS) (Zhang and Harrington, 2015), Alternating Least Squares (ALS) (Salau et al., 1997) and Non-negative Matrix Factorisation (NMF) (Kfoury et al., 2016; Trindade et al., 2017; Vesselinov et al., 2018). Receptor modelling relies on the concept that a dataset matrix (A) containing concentrations can be factored by the number of potential end members represented as (k), producing both a width (W) and a height matrix (H), where the rows of the original matrix are samples (m) and the columns are the congeners (n) (Baldwin et al., 2020). This is summarized in Eq. (1):

$$A(m*n) = W(m*k)*H(k*n) \tag{1}$$

Equation 1. *Generation of a data matrix for receptor modelling*

Many multivariate models are designed to find principal components attributing to the most variability of the data, so it is common for analysts to remove certain PCBs from their dataset to improve the variance they can explain. However, this may not always be the best approach and if any PCB is omitted it should be fully justified, and not just performed to improve the overall variance (Johnson et al., 2015). A detailed overview of a proposed approach for data screening and treatment was produced by Johnson et al. (2015). There are a variety of different appropriate approaches that have been successfully used, however there is no standard approach or set of rules. Therefore, the environmental forensic analyst must be systematic and non-bias in their approach and document each decision (Erickson, 2018, 2020). This leaves room for variability and can result in data analysts failing to perform adequate pre-treatment to those focusing on trying to optimise the predictive capability of their models rather than trying to understand what their data set is revealing. In most environmental samples, main sources of PCBs are from commercial PCB mixtures (such as Aroclors), therefore, PCBs prevalent in commercial mixtures often drive the formation of the model. By removing certain PCBs that don't fit this model (because they are not prevalent in commercial mixtures) there is the potential to bias data to focus on Aroclor sources (Erickson, 2018). This approach could miss important inputs from by-product PCBs (PCBs produced inadvertently as by-products in other chemical processes) or degraded/dechlorinated PCB's from weathered sources (Magar et al., 2005; Capozzi et al., 2019; Rodenburg and Delistraty, 2019). In application of this bias assessment of data in litigation proceedings, it can mean wrongly implicating certain polluters over others resulting in millions of dollars of clean-up costs being allocated to the wrong PCB emitter. There is also potential to create this bias when dealing with non-detected contaminants in datasets. Different approaches to this include removing PCBs with a high percentage of values below a certain limit of detection (LOD) as recommended by Helsel (2006), substituting LODs with a nominal value and causing a left censored dataset (e.g. 0.5 LOD), or imputing values between zero and the detection limit to provide a continuous distribution below the detection limit. By running iterative models with and without these changes, it is easier for an analyst to identify what effect this pre-data processing has on their overall outputs.

Most PCB fingerprinting studies have been focused on identifying which commercial PCB mixtures are responsible for the contamination that is observed. Since 2011, there has been a broadening of this approach when Hu and Hornbuckle (2010) were able to identify a new inadvertent source of PCBs that did not originate from the traditional commercial PCB mixtures. Since then, by-product PCBs, for example PCBs 11, 28, 52, 77 and 209, have gained more interest due to their ubiquitous nature and ability for long range transport (i.e. atmospheric transport). These PCBs have since been identified in other environmental datasets as well as consumer products (Hu and Hornbuckle, 2010; Hu et al., 2011; Garmash et al., 2013; Guo et al., 2014; Vorkamp, 2016; Liu and Mullin, 2019; Megson et al., 2019; Hermanson et al., 2020; Rodenburg et al., 2020; Mao et al., 2021a, 2021b). It is important to note that by-product PCBs are not likely to have one single specific signature, unless there is a ‘plume’ originating from a known source where they show up in higher concentrations in certain samples. It can be easier to identify by-product PCBs that are not present in most commercial mixtures (e.g. PCBs 11 and 209). Some inadvertently produced PCBs (e.g. PCBs 28 and 52) are also abundant in technical mixtures, but it has been possible to detect a non-Aroclor source of these congeners in arctic samples (Bartlett et al., 2019).

1.3. PCBs in Portland Harbor Superfund Site (PHSS)

The Portland Harbor Superfund Site (PHSS), located in Portland, Oregon, is a well characterised contaminated site with multiple chemicals of concern that occupies over 10 miles of the lower Willamette River. The PHSS has been contaminated with many environmental contaminants (including PCBs) following decades of industrial use along the river (Sethajintanin et al., 2004; Rodenburg et al., 2015; Fitzpatrick et al., 2018; Rodenburg and Delistraty, 2019; US EPA, 2019). It has been the international subject of numerous studies on pollutants in environmental samples (e.g. sediment, fish, birds and water) (Sethajintanin et al., 2004; Henny et al., 2009; Luxon et al., 2014). High concentrations of PCBs have been found at the PHSS, which helped contribute to the EPA placing it on the National Priority List (NPL) in 2000 (Sethajintanin et al., 2004; Rodenburg et al., 2015; US EPA, 2019). There are several existing studies in the scientific literature on source apportionment at Portland Harbor (http://ph-public-data.com, 2020) in a variety of sample types (Rodenburg et al., 2015; Rodenburg and Delistraty, 2019) which makes for an excellent comparison of different approaches. Data from the PHSS were collected using EPA method 1668C so contains a comprehensive congener specific dataset (data for all 209 PCBs as 162 variables, due to coelutions). Details on sample collection, analysis and all raw data used in this manuscript is publicly available (http://ph-public-data.com, 2020) which makes it an excellent testing ground for other analysts to compare outputs from their own chemical fingerprinting methods.

1.4. Aims and objectives

The main aim of this manuscript is to use data from the Portland Harbor Superfund Site (PHSS) to compare the effectiveness of different PCB fingerprinting techniques and receptor models used for source apportionment. Specifically, this study aims to 1) compare the variation produced by four different multivariate models (PMF, NMF, PVA, ALS) on the same PCB dataset and evaluate their effectiveness for source apportionment, 2) use the most appropriate fingerprinting method to apportion PCB sources in surface sediment from Portland Harbor, Oregon, USA 3) and establish if multivariate models can help to identify sources of incidental PCBs.

2. Materials and methods

2.1. Obtaining a sample database

The Portland Harbor Pre-Remedial Design Investigation and Baseline Sampling (PDI) database was obtained online from the Portland Harbor Environmental Data Portal (http://ph-public-data.com, 2020). All samples were collected from surface sediments (0–30 cm depth) during March 30th 2018 and May 1st 2019, and analysed by the same laboratory using the same analytical method. The data were acquired in Microsoft Access database format, extracted into Microsoft Excel, and imported into statistical software JMP. All samples with a sediment matrix were selected and exported as a subset. The subset was distilled further by selecting only the 765 samples with PCB results. Coeluting congeners were combined as one variable (e.g. PCBs 12 and13 coeluted so were reported as PCB 12/13). This resulted in a total of 162 PCB variables (reported as concentrations) that were used to begin the data analysis. This dataset is available in the Supplementary information (SI1).

In preparation for comparison of the source apportionment results to the Aroclor profiles the PCB congener lists were converted to relative proportions for comparison against the 2004 Rushneck et al. Aroclor profiles and 1996/1999 Frame et al. Aroclor profiles. To allow for a direct comparison of the source apportionment results to the Aroclor data, applicable PCBs were combined so that coelutions were identical in the PDI data and Aroclor profiles reported by Rushneck (Rushneck et al., 2004) and Frame (Frame et al., 1996). If one PCB was reported as a non-detect when combining PCBs, then only the detected value was used; the detection limit was not included in the calculation. If both values were non-detect, then the maximum detection limit was used. This process resulted in the creation of a dataset with 148 different variables, comprised of the same individual or coeluting PCBs. We accept that there will be errors in PCB measurements associated with sample collection, preparation and analysis and these have been previously investigated in the Portland Harbor Environmental Data Portal (http://ph-public-data.com, 2020). As data was collected for a period of over 1 year we performed exploratory data analysis (using UMAP) to assess the impact of temporal variation. No distinct sample clusters or end members could be linked to sample date and therefore any variation with time did not appear to have a significant impact on the results. Whilst there will be some small degree of variation and error within the data this is desirable as the aim of this manuscript was to use data from real environmental samples, rather than a fabricated dataset. The same final dataset was used by each model to perform source apportionment.

2.2. Imputations

The PDI dataset included values below detection limit as identified by a 'U' or 'UJ' laboratory qualifier. Approximately 28 % of all data points in the dataset were below the laboratory detection limit. This included 20 PCB congeners with over 90 % of the data points below detection limit. These specific PCB congeners were removed from analysis following investigation with an initial PCA analysis as their inclusion added noise to the dataset and did not provide any additional diagnostic value. This resulted in the production of a final dataset containing 128 different variables, comprised of the same individual or coeluting PCBs, with 17 % of the values below the limits of detection. The remaining data points reporting below the detection limit were imputed using the Kaplan-Meier (KM) method (Palarea-Albaladejo and Martín-Fernández, 2015). The KM method is a nonparametric method that does not assume any underlying probability distribution of the data.

2.3. Data analysis

The dataset was examined using four different receptor models, Polytopic Vector Analysis (PVA), Positive Matrix Factorisation (PMF), Non-negative Matrix Factorisation (NMF) and Alternating Least Squares (ALS).

Polytopic Vector Analysis (PVA) is an eigenvector decomposition model resolved in terms of oblique vectors (i.e., rotated "factors") as source compositions. The PVA algorithm is a collection of MATLAB scripts developed by Dr. Glen Johnson and summarized by Ehrlich et al. (1994). The PVA receptor modelling was performed according to procedures from Johnson et al. (2015). The EXTENDED QMODEL version of PVA was run with 150 iterations, DENEG value set to 0.25 and the maximum negative missing proportions set to 0.05.

Positive Matrix Factorization (PMF) Model is a multi-dimensional mathematical receptor model developed by Paatero and colleagues (Paatero and Tapper, 1994; Paatero, 1999; Paatero et al., 2005; Paatero and Hopke, 2008; Paatero et al., 2014) by the United States Environmental Protection Agency (US-EPA). Results are obtained using the constraint that no sample can have significantly negative source contributions and using user calculated uncertainties. The PMF model software was downloaded directly from the US-EPA (US EPA, 2018). The model was run three different times each with 150 iterations using a random seed number. An uncertainty matrix was calculated following guidance in Baldwin et al. (2020). The uncertainty of the individual congener concentration was calculated using Eq. (2), where the error term was calculated as the median relative percentage difference between duplicate samples. Thirty-nine duplicate sample pairs were identified in the PDI database and the median relative percent difference was calculated for use in the uncertainty calculation. The combined unit uncertainty $U_c(x)$ on the total sample concentration was computed as the root sum of the squares of the individual compound uncertainties (Eq. (3)):

$$\text{Congener uncertainty} = \sqrt{\text{DL}^2 + (\text{error} \times \text{concentration})^2} \tag{2}$$

Equation 2. *Congener uncertainty*

$$U_c(x) = \sqrt{\left(U_1(x)^2 + U_2(x)^2 + U_3(x)^2 + \ldots\right)} \tag{3}$$

Equation 3. *Uncertainty matrix calculation*

Non-negative Matrix Factorisation (NMF) is a multi-dimensional mathematical receptor model where a matrix is factorised into two matrices with the constraints that all three matrices have no negative elements. NMF was developed on the same basis as the original PMF algorithm by Paatero and colleagues. The algorithm was named, 'non-negative matrix factorization' after Lee and Sung (1999) investigated the properties of the algorithm and published some simple and useful extensions of the PMF foundation. NMF was run three separate times using [R package: NMFN 2.0] (Lee and Seung, 2001). The model (method = "nnmf_mm") was run untrained, using default settings with maximum iterations (argument maxiter) equal to 1000.

Alternating Least Squares (ALS) is a further extension of the NMF algorithm that aims to reduce the root mean squared error (RMSE) between the actual and predicted matrices. ALS was run three separate times using an application developed by Chemistry Matters. The model was run untrained, with defaulted settings and maxiter set to five.

The actual versus predicted concentrations reconstructed from the receptor models were compared through a cosine theta (cos-Θ) similarity metric (herein referred to as cosine similarity) [R package: LSA]. This similarity metric can be used to compare two sets of data (e.g., two histograms or samples) by treating each congener distribution as a multi-dimensional vector and to calculate the cosine of the angle (Θ) between the two vectors. If two samples are identical, the vectors will be parallel, the angle between the vectors will be 0 degrees, and the cosine of that angle is 1.0. Similarity, if two samples share no common vector similarity, the angle defined between them is 90 degrees and the cosine of that angle is zero. Thus, cos-Θ is bounded between zero and one where zero is indicative of 0 % vector similarity and 1 is indicative of 100 % vector similarity.

3. Results and discussion

3.1. Comparison of multivariate models

An important first step in source apportionment is deciding how many end members (EM) should be included when modelling the data. This was initially assessed graphically using scree plots and non-graphically using; optimal coordinates, the acceleration factor, the parallel analysis, and the Keyser-Guttman rule, as described in Raîche et al. (2006). The results indicated the data could be best represented by a four or five end member solution (SI2). Therefore, to identify the most appropriate solution the four models (NMF, ALS, PMF and PVA) were run for both four and five andend member solutions. Each model was run three times to assess variability between runs, except for PVA which was run once (because the polytype is calculated the same way so each run produces identical answers). Each end member produced by these models was compared visually against Aroclor profiles from Rushneck et al. (2004) and Frame et al. (1996) (SI3), in addition cosine theta values were calculated to aid the comparison. The profiles produced by Rushneck and Frame were very similar, but cosine similarity provided a slightly better relationship using the Rushneck data, therefore the Rushneck data are reported for comparison.

Under the four end member scenario the four multivariate models generally produced the same four end member patterns which were identified as; three Aroclor-like signatures A1248, A1254, and A1260, and a weathered A1260 pattern (A1260W) (Table 1). However, PMF and one of the ALS runs failed to identify the weathered A1260 signature and instead reported two similar A1260 signatures. The multivariate models are untrained which is why the numbering of the end members changed for each signature, but importantly the same four signatures were consistent (cosine similarity of 0.878–0.969). Although each of the four models produced similar end members, there were some differences in the outputs for the A1260 and A1260W type signatures (see Fig. 1). The weathering of A1260 appears to have occurred through microbial dechlorination that has targeted the PCBs with a chlorine in the flanked meta position (Wu et al., 1998; Bedard, 2001) (SI2). This resulted in relative decreases of PCBs 180, 174, 170, 153, 149 and 138, and relative increases in PCBs 100, 91, 54, 53, 51, 49, and 47. As there is no definitive "source" of A1260W (it is linked to A1260) the models varied in predicting the exact composition of PCBs in the A1260 and A1260W end members. For example, when run through PVA, the A1260W signature contained higher proportions of the hexachlorinated congeners than in other models (e.g.

NMF), which had the resulting effect of making the original A1260 signature appear more like A1262 (Fig. 1).

By using five end members, the overall variance explained by the models increased, as did the mean cosine similarity result. However, the profiles produced from the five end member models were less consistent (Table 1). Most models appeared to split one of the Aroclor-like end members into two similar but distinct signatures of A1254, A1260 or A1260W (that could not be explained chemically), or they identify an additional signature that is indicative of A1242.

The results highlight the different outputs that can be produced by using different multivariate models on the same dataset. Seven different end members (or potential signatures) were identified by the different models; A1242, A1248, A1254, A1260, A1260W, A1262 and a signature that didn't match any known commercial PCB mixtures, “no match”. The results also show how different outputs can be produced by using the same model (e.g. ALS with five end members gave three different results each time). The largest variation in source apportionment results was observed for the PMF outputs. Our assessment was performed using default convergence criteria (Paatero et al., 2014), yet the results indicate that the model did not reach convergence. As our final source apportionment was performed by NMF, we did not perform optimisation to achieve convergence of the PMF model. However, we would recommend that anyone using PMF for source apportionment should follow the guidance from Brown et al. (2015) to achieve this. Differences can also be observed when comparing these findings against valid approaches taken by other researchers. For example, Rodenburg et al. (2015) used PMF with four end members to model the sediment data, processed to contain 83 PCB variables (comprised of 116 congeners or congener groups). Rodenburg et al. (2015) identified the same three Aroclor-like signatures (A1248, A1254 and A1260) as identified by our current study, however they identified a different fourth end member that was classified as an incidental PCB signature with a high proportion of PCB 11 (3 %). A similar signature to this was observed in the PVA 5 endmember model (identified as S5 “no match” in Table 1 and SI3). The weathered Arolcor 1260 signature (not observed by Rodenburg et al., 2015) was only observed by PMF in our current study when by a five end member model which may explain why it was not recorded by the 4 end member PMF model used in Rodenburg et al. (2015).

The different outputs reported by these different approaches (both within this manuscript and the wider scientific literature) were all reached using robust and scientifically accepted methods. However, these differences highlight the importance of understanding uncertainty in multivariate modelling as different methods and models can produce different outputs. In some scientific studies and courtroom scenarios the data analyst will only use one model with a fixed number of end members (sometimes run one time without varying model parameters) and have complete confidence in the output. By performing a more detailed assessment it is possible to better understand the uncertainty in the data (which may be more difficult to explain in a courtroom) but will ultimately allow us to make more informed scientific decisions. To achieve this, we must take due care and consideration to understand the predictive uncertainty in models, establish the most appropriate number of end members to use, and make sure the signatures identified make sense from a chemical perspective.

3.2. Portland Harbor source apportionment

Seven different end members were identified by the four receptor models, these included Aroclor-like signatures; A1242, A1248, A1254, A1260, A1260W, A1262 and “no match”. The appropriateness of these end members was assessed statistically, through cosine similarity analysis, and chemically, by interpreting and comparing the signatures visually. The A1242 signature was recorded by two models with five end member solutions (NMF run 2 and PVA). The cosine similarity results of these end members to A1242 were strong (~0.93), but the fact that it was not consistently identified casts doubt on its presence as a major source of PCBs for the entirety of Portland Harbor sediments as compared to other end members. By reviewing the data spatially, it is considered more likely that it is a minor contributor to the overall PCB load in Portland Harbor but may have some significant inputs on a localised scale in specific areas within Portland Harbor. A1248 and A1254 were consistently detected across most models with strong cosine similarity matches of >0.97, therefore these are likely the major sources of PCBs. The A1260 and A1260W signatures are linked as discussed in Section 3.1 and Fig. 1. Both signatures were identified in nearly every model with the exceptions of A1260W not appearing in the PMF four end member models or some of the ALS runs. Therefore, the results indicate that A1260 is considered a likely major source of PCBs. A1262 and A1260 have a very similar composition, however, A1262 (<1 % U.S. sales) was used less than A1260 (>10 % U.S. sales) (Durfree et al., 1976). From interpreting all signatures together it appears more likely that the A1262 signature is an artefact created by the variation in outputs for the A1260 and A1260W signatures as discussed in Section 3.1 and Fig. 1. Due to its low usage and the fact it was only identified intermittently

Table 1

Aroclor-like source apportionment of PCBs derived from four and five end member models, and the mean cosine similarity for each model. S1–5 corresponds to each end member identified by the model (presented in SI3).

Model	A1221	A1232	A1016	A1242	A1248	A1254	A1260	A1260W	A1262	A1268	no match	Cosine Theta
Four end members												
NMF Run 1					S2	S1	S3	S4				0.969
NMF Run 2					S4	S3	S2	S1				0.969
NMF Run 3					S1	S3	S2	S4				0.969
ALS Run 1					S4	S2	S1	S3				0.888
ALS Run 2					S3	S4	S2	S1				0.885
ALS Run 3					S2	S3	S4		S1			0.878
PMF Run 1					S1	S4	S2 S3					0.969
PMF Run 2					S4	S2	S1 S3					0.969
PMF Run 3					S1	S4	S2 S3					0.969
PVA Run 1 - 3					S3	S1	S2*	S4				NA
Five end members												
NMF Run 1					S4	S3	S1 S2	S5				0.97
NMF Run 2				S3	S2	S5	S4	S1				0.973
NMF Run 3					S1	S2	S3 S4	S5				0.97
ALS Run 1					S2	S4	S1 S3 S5					0.885
ALS Run 2					S5	S2 S4	S1 S3					0.891
ALS Run 3					S3	S4	S1	S2 S5				0.892
PMF Run 1					S2	S1	S4 S5	S3				0.972
PMF Run 2					S1	S3	S4 S5	S2				0.972
PMF Run 3					S5	S4	S1 S2	S3				0.972
PVA Run 1 - 3				S3		S1	S2*	S4			S5	NA

*Although cosine similarity was a better match for A1262 we believe it is more likely to represent an Aroclor 1260 type source as explained in Fig. 1.

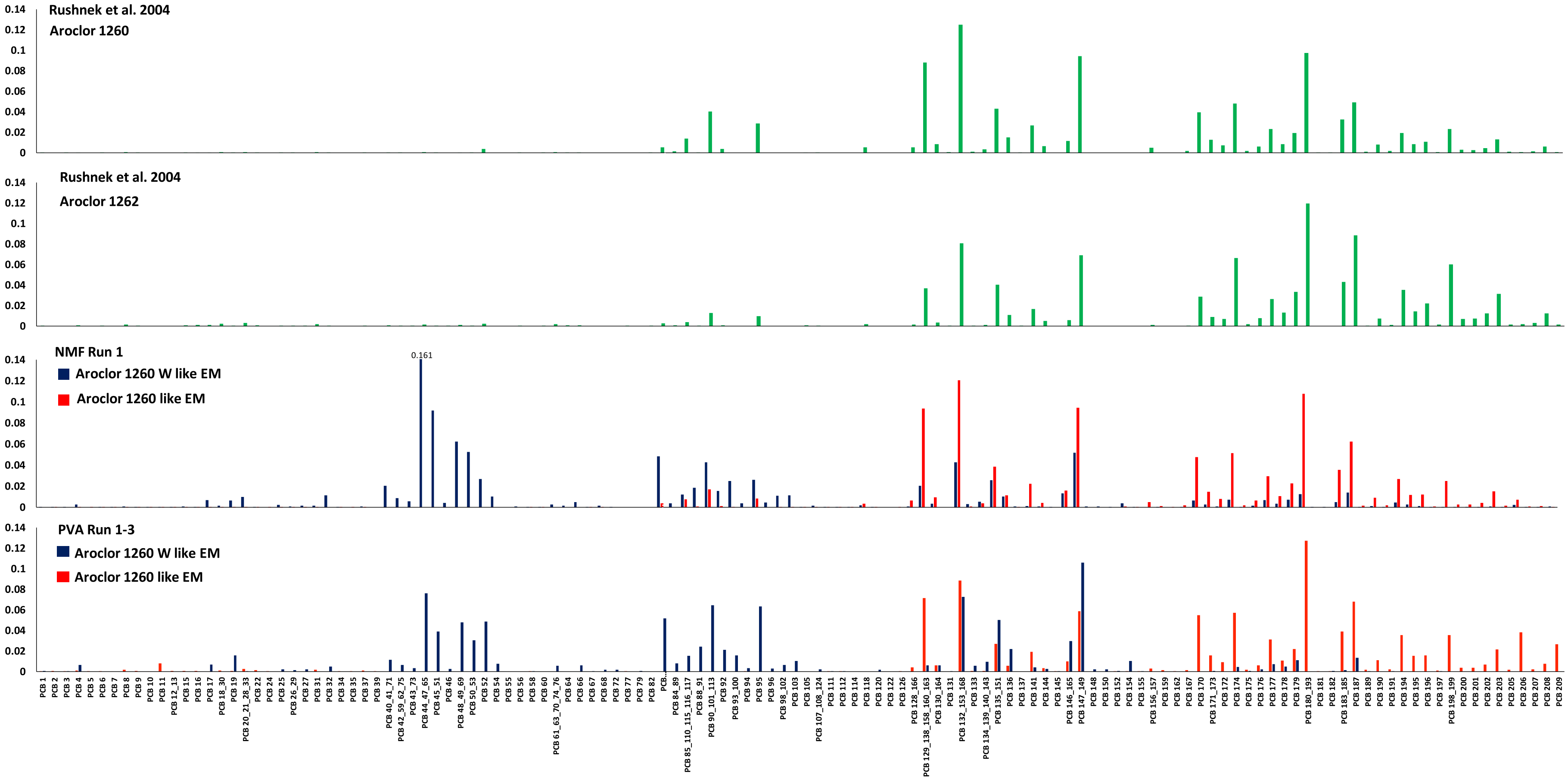


Fig. 1. Comparison of the 2004 Rushnek PCB Aroclors 1260 and 1262 signatures and observed patterns for the four EM NMF run 1 S3 and S4, and PVA run 1 S2 and S4.

casts doubt on A1262 being a significant potential source of PCBs and this contribution is more likely to have arisen from A1260. The final potential signature was only detected by PVA in the five end member iteration (and by Rodenburg et al., 2015 with a four end member PMF model). This signature did not match any known commercial mixture and contains high proportions of PCBs that have been associated with by-product PCB production (e.g. 4 % PCB 11, 19 % PCB 206 and 14 % PCB 209). This indicates that there is a source of by-product PCBs in the river system and using this end member could provide a useful indication of the input from by-product PCBs, however it may not be the most robust approach. There is no one common specific signature for a by-product PCB source, there are wide variety of different by-product PCB sources that each have their own unique signature that can contain one to tens of different PCBs all in different proportions. Therefore, to assess this all together as one signature may provide misleading results. Each time one of these congeners is detected in a sample, the model is restricted to adding the identified end member (and the other congeners in the end member) to the sample in order to account for the presence of the single congener. This will result in overprediction of some congeners from the by-product PCB end member.

Based on a review of all model outputs and potential signatures we believe the data are indicating that three different commercial PCB mixtures represent the main sources of PCBs in this system. This is best characterised by four end members that match aroclor like signatures for 1) A1248, 2) A1254, and 3) A1260 & A1260W, with localised inputs of A1242 and sources of by-product PCBs. In this assessment the four end member system produced by NMF best represented this dataset as it resulted in a consistent output where end member could be chemically explained. Using this model, the major sources of PCBs in this sample set for Portland Harbor river-wide could be attributed to 16 % A1248, 41 % A1254 and 42 % to A1260. A summary of outputs for all models is presented in Table 2. This highlights variability in modelling and shows that the relative proportion of each signature was variable between and within models. A1254 and A1260 were identified as the main sources of PCBs with estimates of A1254 inputs varying from 41 % to 63 % and A1260 from 27 % to 46 %.

The data from four end member NMF (Run 1) were assessed geospatially to see if PCB signatures were homogenously distributed within the river basin or if there were hotspots. By plotting total PCB concentrations (SI4) it was possible to identify that the highest PCB concentrations were located two areas along the north bank, with several additional hotspots located along the course of the river. These hotspots appear to correlate with high proportions of either A1254 or A1260 along with a hotspot of A1242 in the north west section of the river (Fig. 2). Results show A1260W was present in lower concentrations and was distributed relatively evenly along the river.

Table 2

Summary of source apportionment for all models (% values for each end member are rounded to the nearest whole integer, but all sum to exactly 100 %). Where more than one end member matched the same commercial mixture the two contributions for each end member were summed in bold and allocations for each end member presented in brackets "()".

Model	A1242	A1248	A1254	A1260	A1260W	A1262	no match
Four end members							
NMF Run 1		**16%**	**41%**	**35%**	**7%**		
NMF Run 2		**16%**	**41%**	**35%**	**7%**		
NMF Run 3		**16%**	**42%**	**35%**	**7%**		
ALS Run 1		**9%**	**63%**	**26%**	**2%**		
ALS Run 2		**7%**	**63%**	**26%**	**3%**		
ALS Run 3		**9%**	**59%**	**27%**		**6%**	
PMF Run 1		**5%**	**40%**	**55%** (35%) (20%)			
PMF Run 2		**19%**	**29%**	**52%** (37%) (15%)			
PMF Run 3		**5%**	**54%**	**41%** (21%) (20%)			
PVA Run 1 - 3		**16%**	**37%**	**34%**	**13%**		
Five end members							
NMF Run 1		**16%**	**37%**	**41%** (28%) (13%)	**7%**		
NMF Run 2	**5%**	**13%**	**41%**	**32%**	**9%**		
NMF Run 3		**15%**	**36%**	**43%** (12%) (31%)	**5%**		
ALS Run 1		**9%**	**44%**	**47%** (12%) (22%) (12%)			
ALS Run 2		**7%**	**63%** (49%) (14%)	**30%** (21%) (9%)			
ALS Run 3		**5%**	**63%**	**26%**	**6%** (2%) (4%)		
PMF Run 1		**21%**	**27%**	**36%** (22%) (14%)	**16%**		
PMF Run 2		**6%**	**15%**	**55%** (16%) (39%)	**24%**		
PMF Run 3		**26%**	**34%**	**24%** (7%) (17%)	**16%**		
PVA Run 1 - 3	**13%**		**35%**	**29%**	**19%**		**4%**

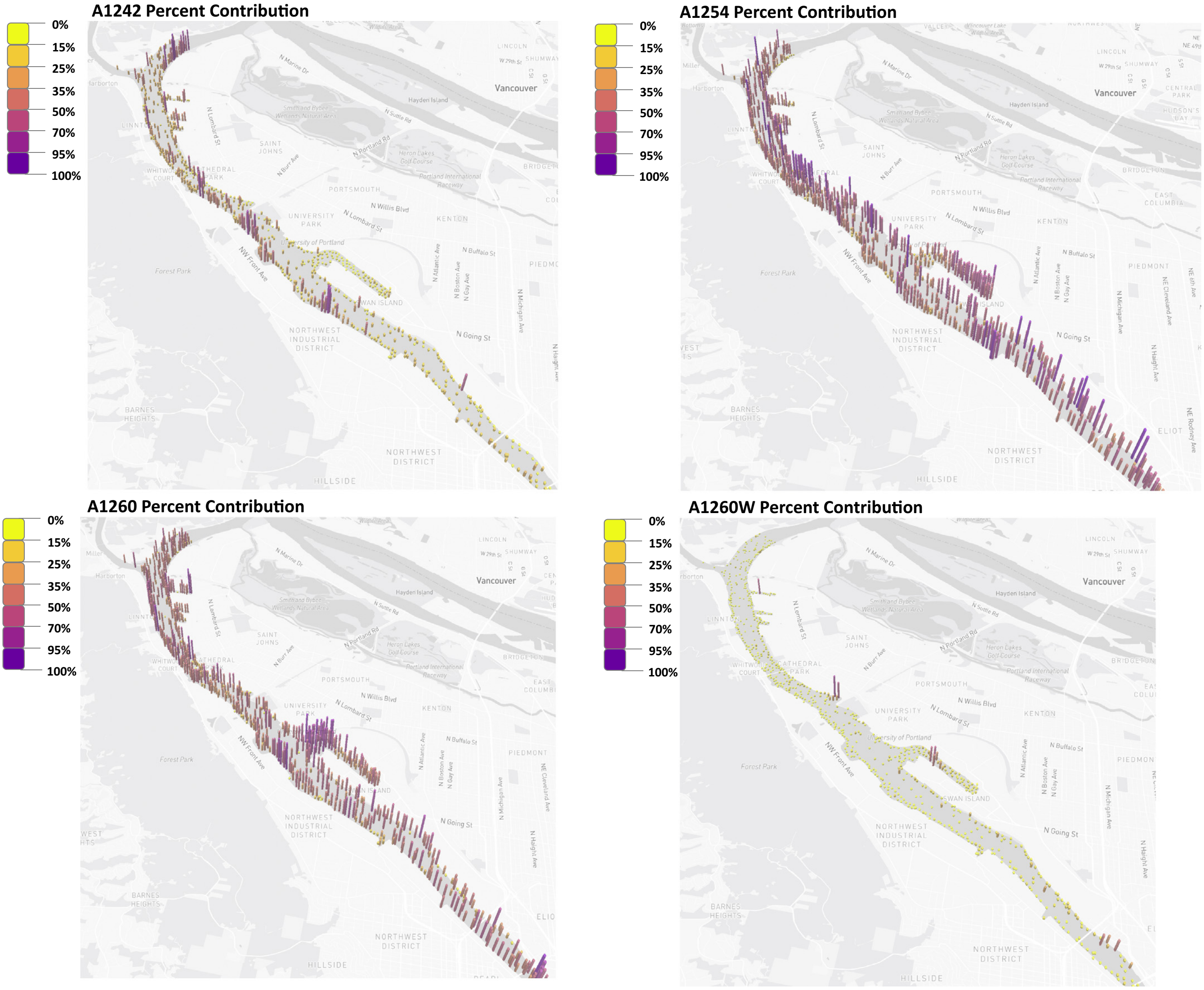


Fig. 2. Geospatial plots displaying the relative proportion of each end member (A1242, A1254, A1260, and A1260W, Percent Contribution) within the river section.

3.3. Identifying sources of incidental PCBs

By-product PCBs in most studies have been identified as anomalies in datasets by identifying something unusual in the PCB profiles of their samples, or as unusual end members in multivariate models (SI3 5EM PVA S5; Rodenburg et al., 2015). In many studies, they may go unnoticed or unreported with many investigations targeting and focusing on, indicator, WHO12 or common PCBs found in commercial mixtures (e.g. Aroclors). Existing studies have identified by-product PCBs in environmental samples including PCBs 5, 11, 52, 206, 207, 208 and 209 (Hu and Hornbuckle, 2010; Bartlett et al., 2019; Rodenburg et al., 2020; Megson et al., 2022). However, these PCBs are only a small subset of the wide range of potential by-product PCBs that may be present in the environment. Many different by-product PCBs may be present in low (but potentially significant) concentrations, unless near a specific point source, and so can be easily missed in chromatograms and spreadsheets. In litigation and source apportionment studies these PCBs need to be appropriately accounted for. We investigated whether a comparison between the observed and predicted PCB concentrations in our multivariate models could help identify inputs from by-product PCBs. We have observed that by-product PCBs often do not fit multivariate models well, presumably as the model is focused on explaining the most variance or the majority of the data. Therefore, models focus on common trends and patterns for high concentration PCBs, which are often linked to commercial mixture signatures that are present in the majority of samples. Our four end member NMF model was based on four end members that represent Aroclor-like signatures (A1248, A1254, A1260 and A1260W). We hypothesised that due to this, PCBs from commercial mixture signatures would have a normal distribution with no significant difference between observed and predicted concentration from the models (as they fit the model well). Whereas, PCBs that have a skewed distribution that consistently underpredicted observed concentrations by the models, may indicate the presence of a source of by-product PCBs. Using the four end member NMF (Run 1), residual percentages were calculated using Eq. (4) so that a negative residual corresponds with the model underpredicting the observed concentration:

$$\frac{(\text{Predicted Concentration} - \text{Actual Concentration})}{\text{Actual concentration}} \times 100 \qquad (4)$$

Equation 4. *Residual percentage calculation*

Only a handful of by-product PCBs have ever been identified in environmental samples in the scientific literature e.g. PCBs 5, 11, 52, 206, 207, 208 and 209 (Hu and Hornbuckle, 2010; Bartlett et al., 2019; Rodenburg et al., 2020; Megson et al., 2022). Hannah et al. (2022) identified a total of 149 different incidental PCBs in paints which indicates there are likely to be many that are missed. Megson et al. (2019) identified over 40 PCBs that are major components from a variety of sources to contain inadvertently produced PCBs (pigments and dyes, waste incineration and impurities in other chemical processes). By using this information, the residual plot (Fig. 3) was presented to identify potential; by-product PCBs not commonly found in commercial mixtures (green), common Aroclor PCBs (red), and mixed PCBs (blue) that may have originated from either by-product or Aroclor sources.

The PCBs that best fit the model (and hence have the lowest residual values) tended to be those that are present in relatively high concentrations in commercial mixtures and were the most resistant to weathering such as, PCBs 180 and 153 (Fig. 3). Two of the PCBs which had the largest underpredictions in observed concentrations were py-product PCBs 11 and 209 (Fig. 3). However, there were many more PCBs which were greatly underpredicted by the model and may be a result of a by-product PCB source. Nineteen of these PCBs were identified as inadvertently produced PCBs by authors referenced in Megson et al. (2019). These included PCBs 1–8, 9, 12/13, 15, 16, 26/29, 31, 34, 35, 126, 206 and 208. We believe this shows the strongest evidence to date of the presence of these PCBs from by-product sources in environmental samples. PCBs created from dechlorination of highly chlorinated analogues were generally well predicted by the model (e.g. PCBs 44/47/65, 45/51, 49/69, 50/53, 52 and 88/91), therefore it is reasonable to expect that the model is working well at predicting PCB concentrations produced by dechlorination. The A1260W end member also produces an excellent match (SI 1) with A1260 dechlorination experiments performed by Wu et al. (1998). This is important as two key congeners in this end member (PCBs 47 and 51) are associated with inadvertent PCBs from silicone (Jan and Peridh, 1991; Herkert et al., 2018). However, if a significant silicone source was present then there would also be elevated proportions of PCB 68 (Jan and Peridh, 1991; Herkert et al., 2018) (which there is not). Whilst we cannot discount the possibility that some of the mixed source PCBs are dechlorination products (especially many of the lighter PCBs), the underprediction of observed concentrations does suggest that there are many more inadvertently produced PCBs than just PCB 11 and 209 present within Portland Harbor sediments. The residual plot also identified several PCBs that were not identified as by-product PCBs by Megson et al. (2019), but were highly underpredicted by the model, (PCBs 111, 112, 120, 155, 181 and 182). These PCBs were generally detected in very low concentrations around the detection limit, which may explain these slight underpredictions.

By investigating the residuals produced by our NMF model, we identified the following 27 PCBs that did not fit the model well and had an observed concentration that was consistently underpredicted: PCBs 1–8, 9, 11 12/13, 15, 16, 26/29, 31, 34, 35, 111, 112, 120, 126, 155, 181, 182, 206, 208 and 209. When compared against the end member produced by the five end member PVA S5 profile with a high proportion of by-product PCBs (SI3) there is a poor correlation between PCBs identified by the residuals analysis and those present in the end member. The differences may be due to the S5 profile picking up one specific by-product PCB source (accounting for ~4 % of the total PCBs), whereas the residual analysis picks up multiple different inadvertent sources of these individual PCBs, which means there is no specific PCB signature that covers all by-product PCBs. By-product PCBs are produced by many unrelated industries and processes so they are often not all co-located which means multiple by-product PCB sources may be identified by multivariate modelling (Mao et al., 2021a,

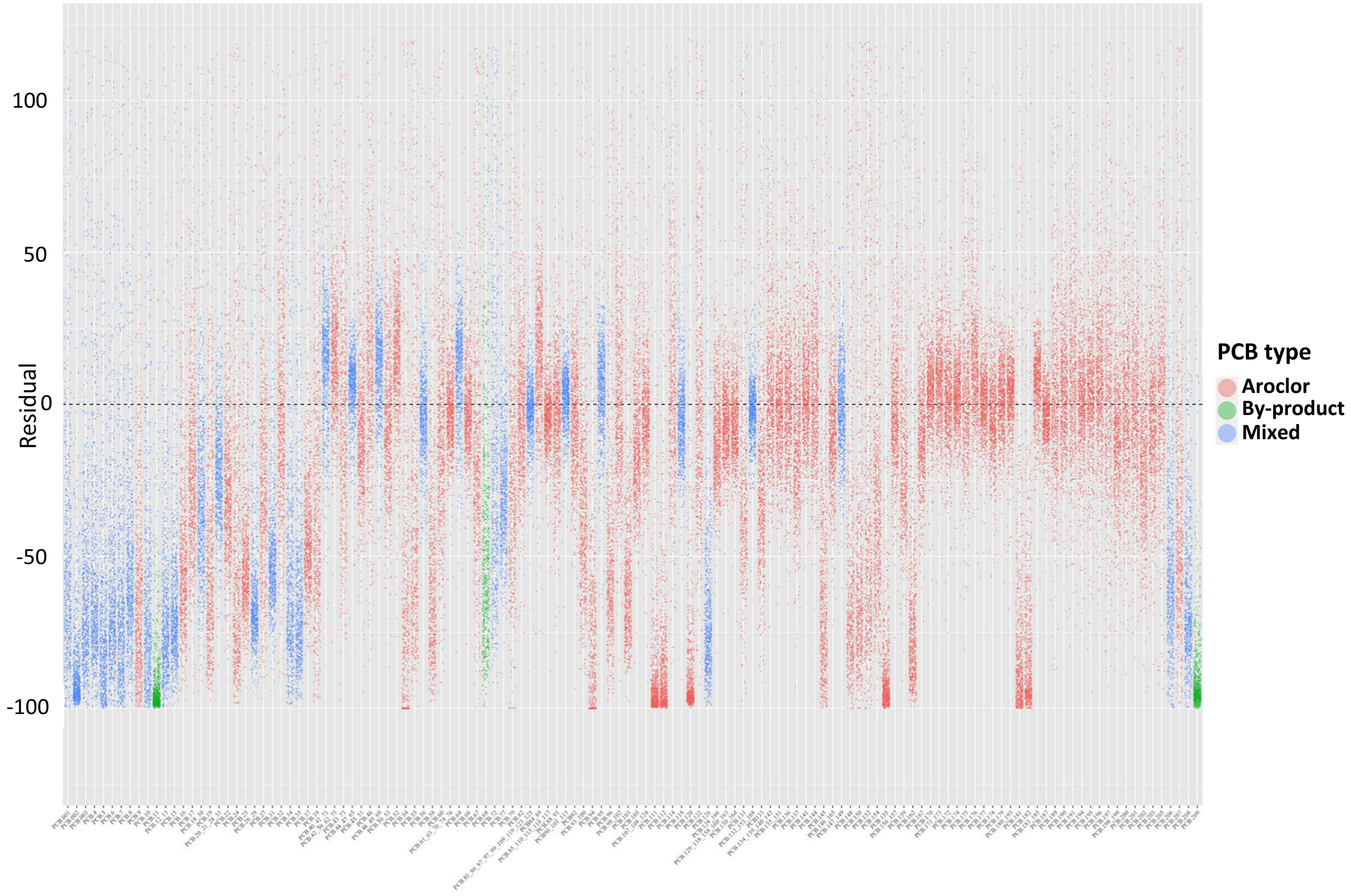


Fig. 3. Residual plot for all PCBs with By-product PCBs (green), Aroclor PCBs (red) and PCBs that may have originated from mixture of both sources (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2021b). End members identified in models that are a poor match for commercial mixtures and contain elevated proportions of PCBs 11 or 209 are a good indication that by-product PCBs are present. But considering the residuals may be more useful to identify potential by-product PCB sources especially for PCBs present in both by-product sources and commercial mixtures. Residuals accounted for 6.6 % of the total PCBs mass that could not be attributed to commercial mixtures by the NMF model, within this 52 % were attributed to known by-product PCBs resulting in a conservative estimate that at least 3.4 % of PCBs in Portland harbor may have arisen from by-product sources.

4. Conclusion

Seven different potential PCB signatures were identified by analysing the same dataset using four different receptor models (NMF, ALS, PMF & PVA) based on four and five end member solutions Although the five end member models had a higher average cosine similarity result, there was less consistency between the runs and resulted in the creation of very similar end members where differences could not be chemically explained. The four end member model was more consistent and generally apportioned the same four signatures (Aroclor 1248, 1254, 1260 and 1260W). This assessment shows that there can be a high degree of variability between (and within) multivariate models. We therefore recommend that data analysts investigate the use of more than one model and run that model on several iterations to better understand the potential uncertainty in this type of analysis.

A four end member NMF source apportionment model was used to allocate the PCBs in the surface sediments of Portland Harbor to 16 % A1248, 41 % A1254 and 42 % to A1260. However, this apportionment is based on the PCBs identified by the model. Residual analysis identified that 6.6 % of PCBs observed in the river system could not be identified by this model. This allowed us to revise our source apportionment as: 15 % A1248, 38 % A1254, 40 % A1260, 3.4–6.6 % non-Aroclor sources, and 0–3.2 % other commercial mixtures.

By using residual plots produced from NMF we have provided the first evidence for the potential presence of approximately 30 different by-product PCBs in environmental samples (PCBs 1–8, 9, 11 12/13, 15, 16, 26/29, 31, 34, 35, 111, 112, 120, 126, 155, 181, 182, 206, 208 and 209). Whilst we accept it is unrealistic to claim that the PCBs that did not fit this NMF model are all definitively due to by-product PCB sources, it will be interesting to see if subsequent research will follow this approach and identify similar trends, thus linking these PCBs to by-product PCB sources.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.162231.

CRediT authorship contribution statement

David Megson: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. **Guuske P. Tiktak:** Formal analysis, Writing – original draft, Writing – review & editing. **Steve Shideler:** Formal analysis, Writing – review & editing. **Mike Dereviankin:** Formal analysis, Writing – review & editing, Visualization. **Lacey Harbicht:** Data curation, Writing – review & editing. **Courtney D. Sandau:** Conceptualization, Writing – review & editing.

Data availability

Raw data is available in the SI and is publicly available at http://ph-public-data.com

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Co-authors Megson D, Shideler S, Dereviankin M, Harbicht L and Sandau C work for Chemistry Matters undertaking legal cases that involve source apportionment of technical PCB mixtures and inadvertent sources of PCBs.

References

Baldwin, A.K., Corsi, S.R., Oliver, S.K., Lenaker, P.L., Nott, M.A., Mills, M.A., Norris, G.A., Paatero, P., 2020. Primary sources of polycyclic aromatic hydrocarbons to streambed sediment in Great Lakes tributaries using multiple lines of evidence. Environ. Toxicol. Chem. 00 (00), 1–17.

Bartlett, P.W., Isaksson, E., Hermanson, M.H., 2019. “New” unintentionally produced PCBs in the Arctic. 5. Elsevier Ltd, pp. 9–14 Elsevier Ltd.

Bedard, D.L., 2001. Microbial dechlorination of PCBs in aquatic sediments. PCBs: Recent Advances in Environmental Toxicology and Health Effects, pp. 27–33.

Boehm, P.D., Douglas, G.S., Burns, W.A., Mankiewicz, P.J., Page, D.S., Bence, A.E., 1997. Application of petroleum hydrocarbon chemical fingerprinting and allocation techniques after the Exxon Valdez oil spill. Mar. Pollut. Bull. 34 (8), 599–613.

Brown, S.G., Eberly, S., Paatero, P., Norris, G.A., 2015. Methods for estimating uncertainty in PMF solutions: examples with ambient air and water quality data and guidance on reporting PMF results. Sci. Total Environ. 518, 626–635.

Buekens, A., Cornelis, E., Huang, H., Dewettinck, T., 2000. Fingerprints of dioxin from thermal industrial processes. Chemosphere 40, 1021–1024.

Cai, C., Geng, F., Tie, X., Yu, Q., An, J., 2010. Characteristics and source apportionment of VOCs measured in Shanghai, China. Atmos. Environ. 44, 5005–5014 Elsevier Ltd.

Capozzi, S.L., Rodenburg, L.A., Krumins, V., Fennell, D.E., Mack, E.E., 2018. Using positive matrix factorization to investigate microbial dehalogenation of chlorinated benzenes in groundwater at a historically contaminated site. Chemosphere 211, 515–523 Elsevier Ltd.

Capozzi, S.L., Jing, R., Rodenburg, L.A., Kjellerup, B.V., 2019. Positive matrix factorization analysis shows dechlorination of polychlorinated biphenyls during domestic wastewater collection and treatment. Chemosphere 216, 289–296 Elsevier Ltd.

Cetin, B., Yurdakul, S., Gungormus, E., Ozturk, F., Sofuoglu, S.C., 2018. Source apportionment and carcinogenic risk assessment of passive air sampler-derived PAHs and PCBs in a heavily industrialized region. Sci. Total Environ. 633, 30–41 Elsevier B.V.

Clement, R.E., Reiner, E.J., Bhavsar, S.P., 2012. Organohalogen contaminants of emerging concern in Great Lakes fish: a review. Anal. Bioanal. Chem. 404 (9), 2639–2658.

DeCaprio, A.P., Johnson, G.W., Tarbell, A.M., Carpenter, D.O., Chiarenzelli, J.R., Morse, G.S., Santiago-Rivera, A.L., Schymura, M.J., 2005. Polychlorinated biphenyl (PCB) exposure assessment by multivariate statistical analysis of serum congener profiles in an adult Native American population. Environ. Res. 98, 284–302.

Durfree, R.L., Contos, G., Whitmore, F.C., Barden, J.D., Hackman, E.E.I., Westin, R.A., 1976. PCBs in the United States - Industrial Use and Environmental Distribution.

Ehrlich, R., Wenning, R.J., Johnson, G.W., Su, S.H., Paustenbach, D.J., 1994. A mixing model for polychlorinated dibenzo-p-dioxins and dibenzofurans in surface sediments from Newark Bay, New Jersey using Polytopic Vector Analysis. Arch. Environ. Contam. Toxicol. 1994 (27), 486–500.

Erickson, M.D., 2018. Aroclor misidentification in environmental samples: how do we communicate more effectively between the laboratory and the data user? Environ. Sci. Pollut. Res. 25 (17), 16291–16299.

Erickson, M.D., 2020. Environmental PCB forensics: processes and issues. Environ. Sci. Pollut. Res. 27 (9), 8926–8937.

Field, J.A., Sierra-Alvarez, R., 2008. Microbial transformation and degradation of polychlorinated biphenyls. Environ. Pollut. 155 (1), 1–12.

Fitzgerald, E.F., Hwang, S.A., Lambert, G., Gomez, M., Tarbell, A., 2005. PCB exposure and in vivo CYP1A2 activity among native americans. Environ. Health Perspect. 113 (3), 272–277.

Fitzpatrick, A.G., Apitz, S.E., Harrison, D., Ruffle, B., Edwards, D.A., 2018. The Portland Harbor superfund site sustainability project: introduction. Integr. Environ. Assess. Manag. 14 (1), 17–21.

Foster, G.D., Walls, C., Mceachern, P.R., Huff, T.B., Mcbride, R., 2019. Sedimentary profiles of pollution marker chemicals along a large tributary of Chesapeake Bay (mid-Atlantic USA). J. Soils Sediments 1511–1526.

Frame, G.M., Cochran, J.W., Bøwadt, S.S., 1996. Complete PCB congener distributions for 17 aroclor mixtures determined by 3 HRGC systems optimized for comprehensive, quantitative, congener-specific analysis. J. High Resolut. Chromatogr. 19, 657–668.

Garmash, O., Hermanson, M.H., Isaksson, E., Schwikowski, M., Divine, D., Teixeira, C., Muir, D.C.G., 2013. Deposition history of polychlorinated biphenyls to the Lomonosovfonna Glacier, Svalbard: a 209 congener analysis. Environ. Sci. Technol. 47, 12064–12072.

Guerra, R., Pasteris, A., Righi, S., Ok, G., 2019. Historical record of polychlorinated biphenyls (PCBs) in the continental shelf of the Korea Strait. Chemosphere 237, 124438 Elsevier Ltd.

Guo, Y., Kannan, K., 2015. Analytical methods for the measurement of legacy and emerging persistent organic pollutants in complex sample matrices. Compr. Anal. Chem. 67, 1–56.

Guo, J., Capozzi, S.L., Kraeutler, T.M., Rodenburg, L.A., 2014. Global distribution and local impacts of inadvertently generated polychlorinated biphenyls in pigments. Environ. Sci. Technol. 48, 8573–8580.

Hagemann, L., Buchty-Lemke, M., Lehmkuhl, F., Alzer, J., Kümmerle, E.A., Schwarzbauer, J., 2018. Exhaustive screening of long-term pollutants in riverbank sediments of the Wurm River, Germany. Water Air Soil Pollut. 229 (6).

Hannah, T.J., Megson, D., Sandau, C.D., 2022. A review of the mechanisms of inadvertent PCB formation in pigments, dyes and paints. Sci. Total Environ. 852, 158529.

Helsel, D.R., 2006. Fabricating data: how substituting values for nondetects can ruin results, and what can be done about it. Chemosphere 65 (11), 2434–2439.

Henny, C.J., Kaiser, J.L., Grove, R.A., 2009. PCDDs, PCDFs, PCBs, OC pesticides and mercury in fish and osprey eggs from Willamette River, Oregon (1993, 2001 and 2006) with calculated biomagnification factors. Ecotoxicology 18, 151–173.

Herkert, N.J., Jahnke, J.C., Hornbuckle, K.C., 2018. Emissions of tetrachlorobiphenyls (PCBs 47, 51, and 68) from polymer resin on kitchen cabinets as a non-aroclor source to residential air. Environ. Sci. Technol. 52, 5154–5160.

Hermanson, M.H., Isaksson, E., Divine, D., Teixeira, C., Muir, D.C.G., 2020. Atmospheric deposition of polychlorinated biphenyls to seasonal surface snow at four glacier sites on Svalbard, 2013–2014. Chemosphere 243, 125324 Elsevier Ltd.

http://ph-public-data.com, 2020. Portland Harbor Environmental Data Portal. [Online] [Accessed on 28th May 2020]. http://ph-public-data.com.

Hu, D., Hornbuckle, K.C., 2010. Inadvertent polychlorinated biphenyls in commercial paint pigments. Environ. Sci. Technol. 44, 2822–2827.

Hu, D., Martinez, A., Hornbuckle, K.C., 2011. Sedimentary records of non-Aroclor and Aroclor PCB mixtures in the Great Lakes. J. Great Lakes Res. 27 (2), 359–364.

Ishikawa, Y., Noma, Y., Mori, Y., Sakai, S.ichi, 2007. Congener profiles of PCB and a proposed new set of indicator congeners. Chemosphere 67 (9), 1838–1851.

Jain, R.B., Wang, R.Y., 2010. Regression models to estimate total polychlorinated biphenyls in the general US population: 2001–2002 and 2003–2004. Chemosphere 79, 243–252 Elsevier Ltd.

Jan, J., Peridh, A., 1991. Formation of polychlorinated biphenyls and chlorinated benzenes by heating of bis(2,4-dichlorobenzoyl) peroxide. Toxicol. Environ. Chem. 31, 297–302.

Johnson, G.W., Quensen, J.F., Chiarenzelli, J.R., Coreen Hamilton, M., 1964. Polychlorinated biphenyls. Environmental Forensics: Contaminant Specific Guide.

Johnson, G.W., Jarman, W.M., Bacon, C.E., Davis, J.A., Ehrlich, R., Risebrough, R.W., 2000. Resolving polychlorinated biphenyl source fingerprints in suspended particulate matter of San Francisco Bay. Environ. Sci. Technol. 34 (4), 552–559.

Johnson, G.W., Ehrlich, R., Full, W., Ramos, S., 2015. Principal components analysis and receptor models in environmental forensics. Introduction to Forensic Chemistry. Elsevier Ltd., pp. 609–653.

Karakas, F., Imamoglu, I., Gedik, K., 2017. Positive matrix factorization dynamics in fingerprinting: a comparative study of PMF2 and EPA-PMF3 for source apportionment of sediment polychlorinated biphenyls. Environ. Pollut. 220, 20–28 Elsevier Ltd.

Kfoury, A., Ledoux, F., Roche, C., Delmaire, G., Roussel, G., Courcot, D., 2016. PM2.5 source apportionment in a French urban coastal site under steelworks emission influences using constrained non-negative matrix factorization receptor model. J. Environ. Sci. (China) 40.

Khalili, N.R., Scheff, P.A., Holsen, T.M., 1995. PAH source fingerprints for coke ovens, diesel and gasoline engines, highway tunnels, and wood combustion emissions. Atmos. Environ. 29 (4), 533–542.

Kjeller, L.O., Jones, K.C., Johnston, A.E., Rappe, C., 1996. Evidence for a decline in atmospheric emissions of PCDD/Fs in the U.K. Environ. Sci. Technol. 30 (4), 1398–1403.

Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401 (6755), 788–791.

Lee, D., Seung, S., 2001. Algorithms for Non-negative Matrix Factorization.

Litten, S., Fowler, B., Luszniak, D., 2002. Identification of a novel PCB source through analysis of 209 PCB congeners by US EPA modified method 1668. Chemosphere 46 (9–10), 1457–1459.

Liu, X., Mullin, M., 2019. Inadvertent polychlorinated biphenyls (PCBs) in consumer products. The 2019 ISEE & ISIAQ Joint Conference. ISES-ISIAQ 2019, Kaunas, Lithuania.

Longnecker, M.P., Wolff, M.S., Gladen, B.C., Brock, J.W., Grandjean, P., Jacobson, J.L., Korrick, S.A., Rogan, W.J., Weisglas-Kuperus, N., Hertz-Picciotto, I., Ayotte, P., Stewart, P., Winneke, G., Charles, M.J., Jacobson, S.W., Dewailly, E., Boersma, E.R., Altshul, L.M., Heinzow, B., Pagano, J.J., Jensen, A.A., 2003. Comparison of polychlorinated biphenyl levels across studies of human neurodevelopment. Environ. Health Perspect. 111 (1), 65–70.

Luxon, M., Toll, J., Hanson, C., 2014. Assessing effects of PCB exposure on American mink (Mustela vison) abundance in Portland Harbor. Integr. Environ. Assess. Manag. 10 (1), 60–68.

Magar, V.S., Johnson, G.W., Brenner, R.C., Quensen, J.F., Foote, E.A., Durell, G., Ickes, J.A., Peven-McCarthy, C., 2005. Long-term recovery of PCB-contaminated sediments at the Lake Hartwell superfund site: PCB dechlorination. 1. End-member characterization. Environ. Sci. Technol. 39 (10), 3538–3547.

Mao, S., Zhang, G., Li, J., Geng, X., Wang, J., Zhao, S., Cheng, Z., Xu, Y., Li, Q., Wang, Y., 2020. Occurrence and sources of PCBs, PCNs, and HCB in the atmosphere at a regional background site in east China: implications for combustion sources. <journal-title>Environ. Pollut.</journal-title>Environ. Pollut. 262, 114267 Elsevier Ltd.

Mao, S., Liu, S., Zhou, Y., An, Q., Zhou, X., Mao, Z., Wu, Y., Liu, W., 2021. The occurrence and sources of polychlorinated biphenyls (PCBs) in agricultural soils across China with an emphasis on unintentionally produced PCBs. Environ. Pollut. 271.

Mao, S., Liu, S., Zhou, Y., An, Q., Zhou, X., Mao, Z., Wu, Y., Liu, W., 2021. The occurrence and sources of polychlorinated biphenyls (PCBs) in agricultural soils across China with an emphasis on unintentionally produced PCBs. Environ. Pollut. 271, 116171.

Megson, D., Kalin, R., Worsfold, P.J., Gauchotte-lindsay, C., Patterson, D.G., Lohan, M.C., Comber, S., Brown, T.A., Sullivan, O., 2013. Fingerprinting polychlorinated biphenyls in environmental samples using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry. J. Chromatogr. A 1318, 276–283 Elsevier B.V.

Megson, D., Brown, T.A., Johnson, G.W., O'Sullivan, G., Bicknell, A.W.J., Votier, S.C., Lohan, M.C., Comber, S., Kalin, R., Worsfold, P.J., 2014. Identifying the provenance of Leach's storm petrels in the North Atlantic using polychlorinated biphenyl signatures derived from comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry. Chemosphere 114, 195–202 Elsevier Ltd.

Megson, D., Focant, J.F., Patterson, D.G., Robson, M., Lohan, M.C., Worsfold, P.J., Comber, S., Kalin, R., Reiner, E., O'Sullivan, G., 2015. Can polychlorinated biphenyl (PCB) signatures and enantiomer fractions be used for source identification and to age date occupational exposure? Environ. Int. 81, 56–63 Elsevier Ltd.

Megson, D., Benoit, N.B., Sandau, C.D., Chaudhuri, S.R., Long, T., Coulthard, E., Johnson, G.W., 2019. Evaluation of the effectiveness of different indicator PCBs to estimating total PCB concentrations in environmental investigations. Chemosphere 237, 124429 Elsevier Ltd.

Megson, D., Brown, T., Jones, G.R., Robson, M., Johnson, G.W., Tiktak, G.P., Sandau, C.D., Reiner, E.J., 2022. Polychlorinated biphenyl (PCB) concentrations and profiles in marine mammals from the North Atlantic Ocean. Chemosphere 288, 132639.

Men, C., Liu, R., Xu, F., Wang, Q., Guo, L., Shen, Z., 2018. Pollution characteristics, risk assessment, and source apportionment of heavy metals in road dust in Beijing, China. Sci. Total Environ. 612, 138–147 Elsevier B.V.

Morrison, R.D., Murphy, B.L., 2010. Environmental Forensics: Contaminant Specific Guide. Elsevier.

Muhammad, S., Shah, M.T., Khan, S., 2011. Health risk assessment of heavy metals and their source apportionment in drinking water of Kohistan region, northern Pakistan. Microchem. J. 98, 334–343 Elsevier B.V.

Muir, D., Sverko, E., 2006. Analytical methods for PCBs and organochlorine pesticides in environmental monitoring and surveillance: a critical appraisal. Anal. Bioanal. Chem. 386 (4), 769–789.

Murphy, B.L., Morrison, R.D., 2014. Introduction to Environmental Forensics. Third edition. Introduction to Environmental Forensics: Third Edition.

Muscalu, A.M., Reiner, E.J., Liss, S.N., Chen, T., Ladwig, G., Morse, D., 2011. A routine accredited method for the analysis of polychlorinated biphenyls, organochlorine pesticides, chlorobenzenes and screening of other halogenated organics in soil, sediment and sludge by GCxGC-μECD. Anal. Bioanal. Chem. 401, 2403–2413.

Paatero, P., 1999. The multilinear engine-a table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. J. Comput. Graph. Stat. 8, 854–888.

Paatero, P., Hopke, P.K., 2008. Rotational tools for factor analytic models. Chemometrics 23 (2), 91–100.

Paatero, P., Tapper, U., 1994. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. Envirometrics 5, 111–126.

Paatero, P., Hopke, P., Begum, B., Biswas, S., 2005. A graphical diagnostic method for assessing the rotation in factor analytical models of atmospheric pollution. Atmos. Environ. 39, 193–201.

Paatero, P., Eberly, S., Brown, S.G., Norris, G.A., 2014. Methods for estimating uncertainty in factor analytic solutions. Atmos.Meas.Tech. 7, 781–797.

Palarea-Albaladejo, J., Martín-Fernández, J.A., 2015. ZCompositions - R package for multivariate imputation of left-censored data under a compositional approach. Chemom. Intell. Lab. Syst. 143.

Pekey, H., Karakaş, D., Bakoğlu, M., 2004. Source apportionment of trace metals in surface waters of a polluted stream using multivariate statistical analyses. Mar. Pollut. Bull. 49, 809–818.

Raîche, G., Riopel, M., Blais, J.-G., 2006. Non graphical solutions for the Cattell's scree test. International Meeting of the Psychometric Society, pp. 1–12.

Ranjbar Jafarabadi, A., Riyahi Bakhtiari, A., Aliabadian, M., Laetitia, H., Shadmehri Toosi, A., Yap, C.K., 2018. First report of bioaccumulation and bioconcentration of aliphatic hydrocarbons (AHs) and persistent organic pollutants (PAHs, PCBs and PCNs) and their effects on alcyonacea and scleractinian corals and their endosymbiotic algae from the Persian Gulf, Iran. Sci. Total Environ. 627, 141–157 Elsevier B.V.

Reiner, E.J., Jobst, K.J., Megson, D., Dorman, F.L., Focant, J.F., 2013. Analytical Methodology of POPs. Environmental Forensics for Persistent Organic Pollutants.

Rodenburg, L.A., Delistraty, D.A., 2019. Alterations in fingerprints of polychlorinated biphenyls in benthic biota at the Portland Harbor Superfund Site (Oregon, USA) suggest metabolism. Chemosphere 223, 74–82 Elsevier Ltd.

Rodenburg, L.A., Krumins, V., Curran, J.C., 2015. Microbial dechlorination of polychlorinated biphenyls, dibenzo-p-dioxins, and -furans at the Portland Harbor superfund site, Oregon, USA. Environ. Sci. Technol. 49, 7227–7235.

Rodenburg, L.A., Hermanson, M.R., Sumner, A.L., 2020. Sources of polychlorinated biphenyl blank contamination and their impact on fingerprinting. Environ. Forensic 21 (1), 99–112 Taylor & Francis.

Rushneck, D.R., Beliveau, A., Fowler, B., Hamilton, C., Hoover, D., Kaye, K., Berg, M., Smith, T., Telliard, W.A., Roman, H., Ruder, E., Ryan, L., 2004. Concentrations of dioxin-like PCB congeners in unweathered Aroclors by HRGC/HRMS using EPA Method 1668A. Chemosphere 54, 79–87.

Saba, T., Boehm, P.D., 2011. Congener-based analysis of the weathering of PCB aroclor 1242 in paper mill sludge. Chemosphere 82, 1321–1328 Elsevier Ltd.

Salau, J.S.I., Tauler, R., Bayona, J.M., Tolosa, I., 1997. Input characterization of sedimentary organic contaminants and molecular markers in the northwestern Mediterranean Sea by exploratory data analysis. Environ. Sci. Technol. 31, 3482–3490.

Santos, F.R., Neves, P.A., Kim, B.S.M., Taniguchi, S., Lourenço, R.A., Timoszczuk, C.T., Sotão, B.M.T., Montone, R.C., Figueira, R.C.L., Mahiques, M.M., Bícego, M.C., 2020. Organic contaminants and trace metals in the western South Atlantic upper continental margin: anthropogenic influence on mud depocenters. Mar. Pollut. Bull. 154 (December 2019), 111087 Elsevier.

Sethajintanin, D., Johnson, E.R., Loper, B.R., Anderson, K.A., 2004. Bioaccumulation profiles of chemical contaminants in fish from the lower Willamette River, Portland Harbor, Oregon. Arch. Environ. Contam. Toxicol. 46, 114–123.

Srivastava, A., 2004. Source apportionment of ambient VOCS in Mumbai city. Atmos. Environ. 38, 6829–6843.

Srivastava, A., Sengupta, B., Dutta, S.A., 2005. Source apportionment of ambient VOCs in Delhi City. Sci. Total Environ. 343, 207–220.

Sundqvist, K.L., Tysklind, M., Geladi, P., Hopke, P.K., Wiberg, K., 2010. PCDD/F source apportionment in the Baltic Sea using positive matrix factorization. Environ. Sci. Technol. 44, 1690–1697.

Trindade, G.F., Abel, M.L., Watts, J.F., 2017. Non-negative matrix factorisation of large mass spectrometry datasets. Chemom. Intell. Lab. Syst. 163.

US EPA, 2010. Method 1668B - Chlorinated Biphenyl Congeners in Water, Soil, Sediment, Biosolids, and Tissue by HRGC/HRMS.

US EPA, 2014. EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guid. Environmental Protection Agency, Research Triangle Park, NC.

US EPA, 2018. Positive Matrix Factorization Model for environmental data analyses. [Online] [Accessed on 24th Februar https://www.epa.gov/air-research/positive-matrix-factorization-model-environmental-data-analyses.

US EPA, 2019. Portland Harbor Superfund Site. [Online] [Accessed on 10th June 2020] https://cumulis.epa.gov/supercpad/SiteProfiles/index.cfm?fuseaction=second.Cleanup&id=1002155#bkground.

van den Berg, M., Birnbaum, L.S., Denison, M., De Vito, M., Farland, W., Feeley, M., Fiedler, H., Hakansson, H., Hanberg, A., Haws, L., Rose, M., Safe, S., Schrenk, D., Tohyama, C., Tritscher, A., Tuomisto, J., Tysklind, M., Walker, N., Peterson, R.E., 2006. The 2005 World Health Organization reevaluation of human and mammalian toxic equivalency factors for dioxins and dioxin-like compounds. Toxicol. Sci. 93 (2), 223–241.

Vesselinov, V.V., Alexandrov, B.S., O'Malley, D., 2018. Contaminant source identification using semi-supervised machine learning. J. Contam. Hydrol. 212.

Vorkamp, K., 2016. An overlooked environmental issue? A review of the inadvertent formation of PCB-11 and other PCB congeners and their occurrence in consumer products and in the environment. Sci. Total Environ. 541, 1463–1476 Elsevier B.V.

Wait, A.D., 2000. Evolution of organic analytical methods in environmental forensic chemistry. Environ. Forensic 1, 37–46.

Wolff, M.S., Deych, E., Ojo, F., Berkowitz, G.S., 2005. Predictors of organochlorines in New York City pregnant women, 1998–2001. Environ. Res. 97, 170–177.

Wu, Q., Sowers, K.R., May, H.D., 1998. Microbial reductive dechlorination of Aroclor 1260 in anaerobic slurries of estuarine sediments. Appl. Environ. Microbiol. 64 (3), 1052–1058.

Yurdakul, S., Çelik, I., Çelen, M., Öztürk, F., Cetin, B., 2019. Levels, temporal/spatial variations and sources of PAHs and PCBs in soil of a highly industrialized area. Atmos. Pollut.Res. 10, 1227–1238.

Zhang, M., Harrington, P.D.B., 2015. Simultaneous quantification of aroclor mixtures in soil samples by gas chromatography / mass spectrometry with solid phase microextraction using partial least-squares regression. Chemosphere 187–193.