Unraveling Sediment PCB Signatures in Disparate Data Sets for Source Allocation

By **Katherine Heal, Ph.D.,** *Scientist, Integral*
Neil Swanson, *Consultant, Integral*

Introduction

Polychlorinated biphenyls (PCBs) are contaminants of concern in many industrial waterways because of their prevalent historical use and persistent nature in sediment and soil. In urban waterways with multiple industrial sources, costly remediation is driven by both the overall concentration of contaminants (including PCBs) and the allocation of liability to parties responsible for the releases. Understanding that liability relies, in part, on differentiating source signatures. Because there are 209 individual PCB compounds, the chemical nature of PCBs lends itself to forensic analyses to tease apart source signatures in sediment and soil—*if data handling and analysis are performed carefully.*

product line, several distinct mixtures of congeners were used and each mixture (i.e., Aroclor) consisted of a characteristic congener "fingerprint." Because Aroclors were manufactured for specific industrial uses, the observation of certain Aroclors in sediment or soil can point to likely industrial contributors to the PCB mixture measured in sediment or soil.
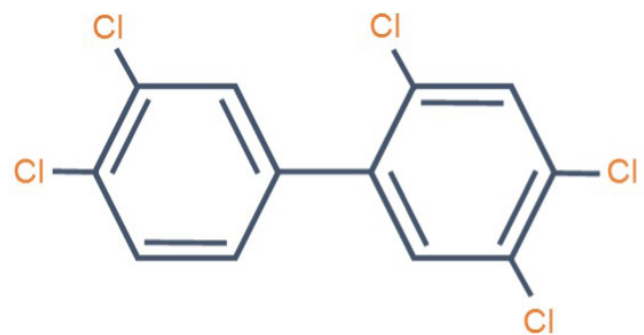
What Are PCBs?

PCBs are a mixture of biphenyls with 1 to 10 chlorine atoms (Figure 1). Each variation of this structure is called a congener, and there are a total of 209 congeners. Within any environmental sample impacted by PCBs, there is a "fingerprint" of individual congener concentrations that can provide information about the sources of the PCBs. PCBs were historically sold and used in the United States between 1930 and 1977 under the trade name Aroclor[1]. Within the Aroclor

Figure 1. Example PCB compound structure (PCB No. 118).

One Challenge of Working with PCBs: Congener Coelutions

The type of laboratory analysis performed informs the forensic methods that can be used to examine sources. Congener analyses provide the most information, relative to Aroclor or other analyses[2]. However, the chromatographic peaks for two congeners cannot always be fully separated from one another—this phenomenon is called coelution. Instead of reporting concentrations of individual congeners, coelutions are reported as the sum of the concentrations of the co-eluting congeners. The co-eluting congeners are often laboratory- and study-specific, with differences arising from slight differences in the methodologies. **The variability in coelutions poses a challenge to forensic analyses when PCB congener data from multiple studies need to be compared.**

A Case Study in PCB Forensics: Making the Most of Coelutions

For an industrial waterway with historical PCB usage that Integral has examined, seven different PCB congener data sets were available. Over the years of sample collection, methodologies have changed slightly and multiple laboratories have analyzed the samples, resulting in data sets with different individual and co-eluted congeners reported. In order to perform statistical evaluations and machine learning methods with the complete data set, a common list of congeners (or coelutions) is necessary. One approach is to compare only exact matching congeners and coelutions among all studies, which sacrifices several variables in the analyses and limits the understanding of the characteristics of each mixture. In the example industrial waterway system, this would limit the analysis to fewer than half of the congeners and ignore some congeners most associated with specific industrial sources. In lieu of sacrificing data that may clarify source signatures, Integral uses R software to combine congeners and coelutions, producing a list of consistent analytes across sample sets. Figure 2 demonstrates this concept in greater detail; in the example industrial waterway system, this results in a list of 101 congener groups encompassing all 209 congeners.

Study 1 Congeners	Study 2 Congeners	Matching Congeners	Combined Congeners
A	A	A	A
B	B+C		B+C
C			
D	D	D	D
E+F	E		E+F
	F		
G+H+I	G+H+I+J		G+H+I+J+K
J+K	K		
L+M	L+M	L+M	L+M
N	N+O		N+O+P
O+P	P		
Q	Q	Q	Q
11 congeners	*11 congeners*	*4 congeners*	*8 congeners*

Figure 2. Example of congener treatment to accommodate different congener coelutions in two hypothetical studies. Each PCB congener is represented by a letter, coelutions are summed letters.
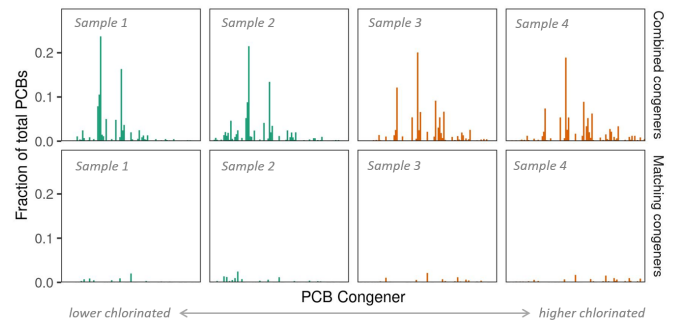

Figure 3. Fingerprint comparisons of four samples after combining congeners (top row) and with using only perfectly matched congeners (bottom row)

With a consistent set of congener groups, analysts at Integral are able to compare samples from all available data sets and employ advanced statistical and machine learning methods like fingerprinting, principal components analysis, unmixing (non-negative matrix factorization), and clustering. All of these forensic techniques can be used together to help elucidate the sources of PCBs within contaminated sediment or soil.

The importance of careful congener data handling to reveal patterns is illustrated in Figure 3. Each subplot is the PCB fingerprint composed of relative congener abundance in an individual sample. The top row represents the congeners and coelution groups Integral used for comparing across the sample sets (after combining as demonstrated in Figure 2), and the bottom row represents *the same samples* but with only the congeners and coelutions that were constant across all sample sets. It is clear in the top row that the two samples represented by green bars (Samples 1 and 2) are similar in composition and richer in lower chlorinated congeners compared to the samples represented by orange bars (Samples 3 and 4). However, when looking at only the perfectly matching congeners in the bottom row, we are missing key information needed to differentiate samples—both by eye and statistically.

Finding meaning in disparate data sets is challenging and can be time intensive. Integral has streamlined the workflow by storing data in a central database with a consistent structure and handling, enabling analysts to pull data from multiple sources and use in-house scripts to translate congeners into a common coelution space for forensic analysis. When a new data set with different congener coelutions is added to the database, the process is easily adapted to update the forensic analyses and regenerate reportable figures. In systems with active data generation, this enables Integral to efficiently provide up-to-date insights to help clients make decisions using robust analyses and the most data available.

By combining a deep understanding of environmental chemistry and advanced data science techniques, Integral is at the forefront in using environmental forensics methods in complex systems.

References

[1]Erickson, M.D., and R.G. Kaley. 2011. Applications of polychlorinated biphenyls. *Environmental Science and Pollution Research* 18(2):135–151.

[2]Shields, W.J., T. Saba, P.D. Boehm, and J. Pietari. 2015. Congeners: A forensics analysis. In: *Introduction to Environmental Forensics.* B.L. Murphy and R.D. Morrison (eds). Third Edition. pp. 347–393. Academic Press.

For more information, please contact:

Katherine Heal, Ph.D.
Scientist
(206) 957-0307
kheal@integral-corp.com
View Bio

Neil Swanson
Consultant
(206) 957-0334
nswanson@integral-corp.com
View Bio