

Modified Polytopic Vector Analysis To Identify and Quantify a Dioxin Dechlorination Signature in Sediments. 1. Theory

NOÉMI BARABÁS,*
PETER ADRIAENS, AND
PIERRE GOOVAERTS†

Department of Civil and Environmental Engineering,
University of Michigan, Ann Arbor, Michigan 48109-2125

Risk-based sediment management decisions require the characterization of contamination sources and fate processes in the field. Polytopic vector analysis (PVA) is a multivariate technique based on a linear mixing model, used to resolve chemical fingerprints and suited for forensic investigations of environmental contamination. The traditional algorithm is constrained to positive fingerprint (end-member) components and cannot resolve fingerprints with both positive and negative values required for a reactive end-member. We developed a modified algorithm (M-PVA) to resolve a dioxin dechlorination fingerprint, indicative of biotic/abiotic transformations in field samples of sediments. The new procedure isolates from the dioxin pattern net compositional changes due to dechlorination in a separate end-member. Using two artificial data sets for which the composition and sample contribution of all end-members are known, the dechlorination fingerprint was reproduced with a root mean square error of 28–41%. The dechlorination end-member contribution to the total variability (set at 4.0 and 10.0%, respectively) was overestimated 1–5-fold. The ability of M-PVA to reproduce the dechlorination pattern and its variability contribution depends on the actual contribution of dechlorination to variability. At an actual contribution of 4.0%, the model outcome deviates more strongly from the original than is the case for a contribution of 10.0%. As such, application of M-PVA to environmental data should include an uncertainty analysis to distinguish variability due to dechlorination from variability due to error. The development of the modified PVA procedure is an important step toward the field characterization of fate processes in dioxin-impacted sediments.

Introduction

Contaminated sediment management presents one of the most challenging environmental problems due to the magnitude of contaminated mass and the uncertainty in source contributions (1). Remedial action plans that can be implemented depend on a number of decision factors. On one hand, one must identify contributing sources, apportion the contamination burden, and quantify its associated risk. On the other hand, an assessment of natural attenuation (fate)

processes and the potential for their enhancement is crucial for an optimal investment of remediation resources in light of the long-term fate of residual and/or stored contaminated sediments. Many severely contaminated sediment sites have extensive, spatially (and sometimes temporally) referenced databases of contaminant concentrations. These offer an opportunity to apply multivariate data analysis techniques capable of extracting the information relevant to these decision factors and of integrating laboratory data for a proper assessment of attenuation in the field.

Identifying sources and transformation reactions in the field can be difficult and expensive and is most often a forensic exercise in which multiple lines of indirect and direct evidence are necessary for the ultimate testing of a hypothesis. Multivariate, variance-based techniques can offer broad (e.g., high spatial or temporal coverage) insight into mixing systems at a very high level of detail (sample-level resolution) using only data about the response variables of a system. These “black-box” type models—with numerous assumptions and sources of uncertainty—complement other approaches of forensic analysis (such as less broad, but direct, sources of evidence) with which it is possible to narrow down the range of possible answers to the relevant questions about contaminant sources and fate.

Polytopic vector analysis (PVA) is perhaps the most sophisticated statistical procedure that has been applied in the forensic investigation of sediment contamination (2, 3). This technique is a self-training receptor/mixing model that unmixes complex mixtures into several contributing patterns (i.e., the distribution of the contaminants in different emission source categories) and their contribution to each sample (2). Applications of PVA have been limited to the identification of positively contributing, contaminant sources as is appropriate with the traditional PVA (T-PVA) algorithm, which is mathematically constrained to positive contributions only.

The contribution of transformation mechanisms to contaminant distribution patterns can be important for understanding long-term fate and for formulating efficient remediation strategies. While polychlorinated dibenzo-*p*-dioxins (PCDD) are persistent in the environment, they are not completely inert. The reactivity of dioxins under reducing (sulfidogenic and methanogenic) conditions, resulting in their (cometabolic) dechlorination, has been repeatedly demonstrated in various incubation regimes (reviewed in ref 4). Generally, dechlorination proceeds to mono- and dichlorinated isomers via sequential removal of chlorines from the environmentally predominant octa- and hepta-CDDs (OCDD and HpCDD). 2,3,7,8-tetraCDD (TCDD), the most toxic, tetra-substituted congener, has been shown to transiently accumulate prior to further dechlorination (5, 6). Whereas these reactions have been demonstrated under artificial laboratory conditions, using either freshly spiked or historical PCDD contamination, the importance of this mechanism on the field scale has not been validated.

The effect of dechlorination on PVA performance was investigated by Johnson and Quensen (7) in a dechlorination experiment on PCBs involving a single Aroclor (1248). Samples were taken weekly during the 20-week experiment. PVA, performed on the resulting PCB congener data, yielded three fingerprints, two of which resembled intermediate compositions of the dechlorinating Aroclor as measured at two different times during the 20-week experiment, showing a progression from higher to lower chlorinated congeners. The third end-member resembled Aroclor 1248 itself. This method indicates that it may be possible to see transitional end-members resulting from dechlorination in field data for

* Corresponding author phone: (734)615-5905; fax: (734)763-2275; e-mail: barabas@engin.umich.edu.

† Current address: Biomedware Inc., 710 Ridgmont Ln., Ann Arbor, MI 48103-1535.

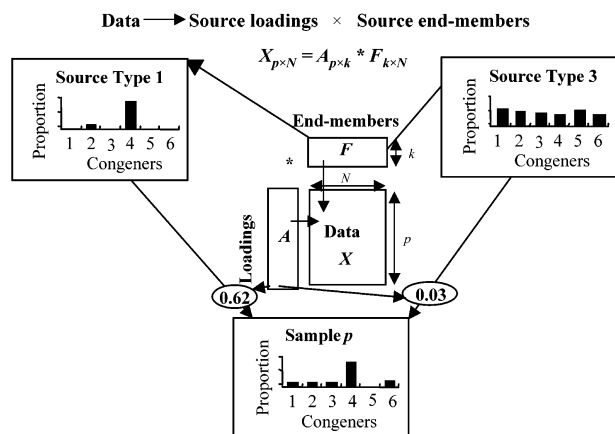


FIGURE 1. Conceptual representation of the PVA model. Source types 1 and 3 are end-members derived by PVA and contained in rows 1 and 3 of matrix F . The numbers in circles are the mixing proportions of each source to the p th sample contained in columns 1 and 3 of the loading matrix A .

PCBs without confounding source end-member compositions. Using field data, inferences by other authors about the presence of dechlorination reactions for PCBs using PVA have relied on the difference between estimated end-members EMs and known Arochlor source fingerprints (8). This method can be valid when the original source fingerprint patterns are well-defined with little variability (e.g., PCB distribution of Arochlors) and the nature and contribution of dechlorination is not of primary interest. In the case of dioxins, however, source fingerprints are numerous (as opposed to just three or four for PCBs) and are highly variable, and the patterns of original source fingerprints are uncertain. Therefore, neither transitional dechlorination patterns nor differences in expected and actual end-members can be expected to yield reliable evidence of dechlorination for dioxins, and a direct, numerical representation of the signature of dechlorination is necessary.

The goal of this paper is to describe and validate a modification of PVA according to mathematical criteria based on the properties of dioxin dechlorination, allowing for the recognition and estimation of a signature pattern representative of this attenuation process.

Theory

Mixing of various source emissions and biochemical alteration processes leads to the final mixture of contaminants detected in a sample (Figure S1, Supporting Information). In phenomenological terms, PVA unmixes complex mixtures into the contributing source fingerprints (end-members) and their corresponding source contributions (sample loadings) in each sample (Figure 1). Mathematically, PVA is based on traditional principal component analysis (PCA) and subsequent iterative rotation and shifting of principal component axes (end-members) until all end-members and their sample loadings are positive.

Although the link between the mathematical representation (EMs and their loadings) and real world phenomena (fingerprints and their contribution to each sample) is not rigorous, it has been shown repeatedly that the PVA procedure can decompose constructed data sets such that the mathematical representation is in agreement with the phenomenological representation (2, 9, 10). To allow the mathematical representation and real world phenomena to be equivalent, appropriate mathematical constraints must be applied to capture the nature of the phenomena modeled. Finally, the PVA analysis is completed only after comparing the EMs to known fingerprints of individual sources or source categories.

Principles of Polytopic Vector Analysis. The fundamental, mathematical principles of polytopic vector analysis were developed for geological applications by Miesch in 1976 (10) and implemented by Full (1981, 1982). In more recent applications, it has been used for data reduction in the field of oceanic current monitoring (3) and adopted in environmental forensics (2). PVA has three objectives: (1) to determine the number of contributing sources or EMs in the system; (2) to resolve each EM fingerprint (with a sum of 1 or 100%); and (3) to resolve the loading or mixing proportion of each EM on each sample (Figure 1).

PVA starts with PCA and is a special case of factor analysis in which measurements in N -dimensional variable space, represented in a data matrix $X_{p \times N}$ of p samples and N variables, are translated by decomposition of a similarity matrix (variance-covariance, correlation, etc.) to a factor (or vector) space of k factors, represented in a loading matrix $A_{p \times k}$ and a factor matrix $F_{k \times N}$. To capture 100% of the variability in real data and due to a multitude of sources and error, $k = N$ for complete mapping. Usually, k is reduced to $k < N$ factors (leading to reduced factor space), sufficient to capture most variability contained in the data, by eliminating factors that contribute very little to variability and, hence, are unresolvable within reasonable uncertainty and/or are assumed to represent error. The data matrix X is then modeled by incorporating an error term $E_{p \times N}$:

$$X_{p \times N} = A_{p \times k} F_{k \times N} + E_{p \times N} \quad (1)$$

Each row r in the matrix F is a factor, and its relative contribution to each sample is given in the corresponding column $c = r$ of the loading matrix A .

Generally, a problem of multivariate decomposition has an infinite set of solutions. Thus, the principal component axes (F) can be rotated to any new position and the loadings (A) recalculated for a new solution to the same data matrix. In traditional PVA, the axis rotation is performed such that all values in both matrices become positive. This constraint is necessary for an additive system of linearly mixing components, as is the case when contaminant sources contribute positive (or zero) concentrations in positive (or zero) proportions to each sample in the data matrix (orthogonality is not necessarily preserved) (Figure S2, Supporting Information). In the traditional algorithm, it is the constraint of positive values that results in the (assumed) equivalence of the mathematical and phenomenological representations. There are cases, however, when this constraint must be amended by others based on knowledge of the chemical behavior of the system, such as charge balance in a closed system consisting of ionic species (9).

Traditional Polytopic Vector Analysis (T-PVA) Algorithm. The algorithm can be broken into 8 steps: (1) transformation procedures on the data matrix; (2) PCA; (3) iterative determination of the number of EMs (k) by re-estimation of the data matrix using an increasing number of PCs, from 1 to N , and evaluating certain re-estimation criteria; (4) identification of outliers (erroneous samples), repeat steps 1–3 if these are found and corrected or removed; (5) Varimax (orthogonal) rotation of the retained axes; (6) initial polytope rotation (oblique) toward the most extreme samples; (7) iterative polytope rotation and expansion until positivity constraint and any additional constraints are satisfied (Note S1, Supporting Information); (8) mathematical procedure followed by identification of EM patterns using information about fingerprints of known sources. (Refer to refs 2, 3, and 9 for complete algorithm.)

Development of Dechlorination Constraints. The current work uses the concept of end-members to describe both "static" source fingerprints and reactive, dechlorination fingerprints. Two changes are introduced to the mathematical

treatment of EMs: (i) allow negative values in a subset of EMs (relaxed EMs) and (ii) constrain one relaxed EM to represent the chemical relationships among variables in the context of dechlorination for one of the relaxed EMs. Appropriate constraints are essential to narrow the infinite solution set to those consistent with the physical system modeled. Kim and Henry (11) demonstrated how insufficient constraints to model the chemical properties and interactions of factor components can easily lead to erroneous results and misguided conclusions. Separate constraints are possible on the matrices **A** and **F**. Constraints on the loading matrix **A** specify the relationship among sources, while constraints on the EM matrix **F** specify the relationship among variables. Traditional PVA uses the positivity constraint on both matrices. The loading matrix **A** is not usually constrained beyond the requirement of positive values since emissions from one source are not usually a function of another. In all previous work with additional constraints, these were applied equally to each EM in matrix **F**. They include charge balance (9), stoichiometry (11), total mass (12), and similar relationships among the variables.

The current modification for dechlorination is novel from two perspectives: (i) Through chemical transformations, some variables are reduced, not added, so the system is no longer strictly mixing. (ii) The EM matrix includes transformation processes in addition to source inputs, and these two types of system components have inherently different properties. As a consequence, a subset of EMs of the matrix **F** is constrained differently from the others.

Dynamics of Dioxin Dechlorination and the Dechlorination End-Member. Dioxin dechlorination proceeds, by definition, from higher to lesser chlorinated congeners, and dioxins and furans are not interconverted. Experimental evidence for dechlorination reactions is limited to studies on 1,2,3,4-TCDD or hepta- and octachlorodioxins (4). Several pathways of dechlorination were shown to exist, and depending on the prevailing geochemical conditions (electron acceptors and donors, microbial consortia, salinity), dechlorination proceeds to mono- and dichlorinated congeners or encounters a transient accumulation of TCDD (i.e., the concentration of TCDD increases because its formation rate exceeds its dechlorination rate) in case of low availability of reducing equivalents (electrons from electron donors) (5). The presence of dissolved organic carbon (DOC) was shown to both enhance the dechlorination reaction and shift the dechlorination pathways. HpCDD is formed rapidly, and hexaCDD (HxCDD) and pentaCDD (PeCDD) are transient intermediate products. The rates of dechlorination are generally higher in the presence of excess organic or inorganic electron donor. The observed lack of mass balance during dechlorination fate assessments indicates that reactions other than dechlorination (e.g., polymerization, transchlorination) may be relevant fate processes, in particular for OCDD (6).

Properties of the Dechlorination End-Member. Given the reactivity dynamics described above, the following properties identify a likely dechlorination end-member:

- (a) Dechlorination always implies that lesser-chlorinated congeners increase at the expense of higher-chlorinated ones.
- (b) HpCDD can be a product of OCDD dechlorination, but its further dechlorination could exceed this positive contribution. Thus this component may have a negative value in the EM.
- (c) TCDD has been observed to be a bottleneck species under most geochemical conditions. It should have a positive value.
- (d) As relative dechlorination rates and pathways are still unknown, it is not possible to predict how HxCDDs and PeCDD behave: either positive or negative values are acceptable in a dechlorination EM.

(e) Since PCDD dechlorination has only been demonstrated in the parts-per-trillion (ppt) to sub-parts-per-million (ppm) concentration range, the net effect of dechlorination reactions on OCDD are masked by the order of magnitude dominance in concentration of this congener relative to all other congeners in most sediment patterns. Hence, OCDD is not included.

(f) Furans and dioxins do not interact: changes in dioxins are independent of changes in furans.

A likely dechlorination EM is identified using the first five properties, and property f is implemented mathematically. The knowledge gaps with regard to dechlorination pathways may thus affect the interpretation of M-PVA results. The mathematical model itself is less limited because a minimum number of constraints should be applied to prevent solutions that violate basic requirements of the system (such as the property under point f). The solution space needs to be open enough to allow discovery of processes that are not anticipated.

Implementation of Modifications (M-PVA). The implementation consists of changes within the code and in its execution. There are two changes within the code: relaxation of the positivity constraint for one or more EMs and constraining one of the relaxed EMs (identified as a candidate dechlorination EM) to the first requirement of dechlorination (point a above). Because dioxins and furans must not interact, the dechlorination EM is rescaled such that the changes in furans sum to zero (0) and the changes in dioxins sum to one (1, or 100%) (PVA requires that the values of the congeners in each EM sum to 1). The dioxin sum is calculated by separately normalizing the negative and positive contributions to -1 and 2, respectively, thereby ensuring a total sum of 1 while limiting each individual value to range between -1 and 2. In the dechlorination row, f_d^+ and f_f^- represent the dioxin and furan components of $F_{k,dechl}$, and f^+ and f^- denote positive and negative components, respectively, such that

$$f_d^+ \rightarrow 2 \frac{f_d^+}{\sum_{d=1}^{N(d^+)} f_d^+} \quad \text{and} \quad f_d^- \rightarrow \frac{-f_d^-}{\sum_{d=1}^{N(d^-)} f_d^-}$$

$$f_f^+ \rightarrow \frac{f_f^+}{\sum_{f=1}^{N(f^+)} f_f^+} \quad \text{and} \quad f_f^- \rightarrow \frac{-f_f^-}{\sum_{f=1}^{N(f^-)} f_f^-} \quad (2)$$

The remaining relaxed but unconstrained EMs leave open the possibility to interpret them for other processes.

The execution of the new PVA process (see flowchart in Figure 2) consists of running T-PVA followed by M-PVA: the outcome of T-PVA serves as the starting point for M-PVA. T-PVA yields the number of positive EMs ($N(\text{EM})^+$), the EM profiles (matrix **F**), and the sample contributions (matrix **A**). After fixing $N(\text{EM})^+$, M-PVA yields the number of relaxed EMs ($N(\text{EM})^-$), profiles for the relaxed EMs and the dechlorination EM (EM_{dechl}), as well as updated profiles for the sources with

$$k = N(\text{EM})^+ + N(\text{EM})^- \quad \text{and} \quad N(\text{EM})_{\text{max}}^- = N - N(\text{EM})^+ \quad (3)$$

M-PVA is first run with relaxed EMs but without dechlorination constraints because it is not possible to know a priori which relaxed EM will represent dechlorination. The resulting relaxed EM profiles are visually evaluated for their compatibility with dechlorination processes. In subsequent runs (models), the constraint is applied to the EM that appears compatible with dechlorination (EMs for a given data set

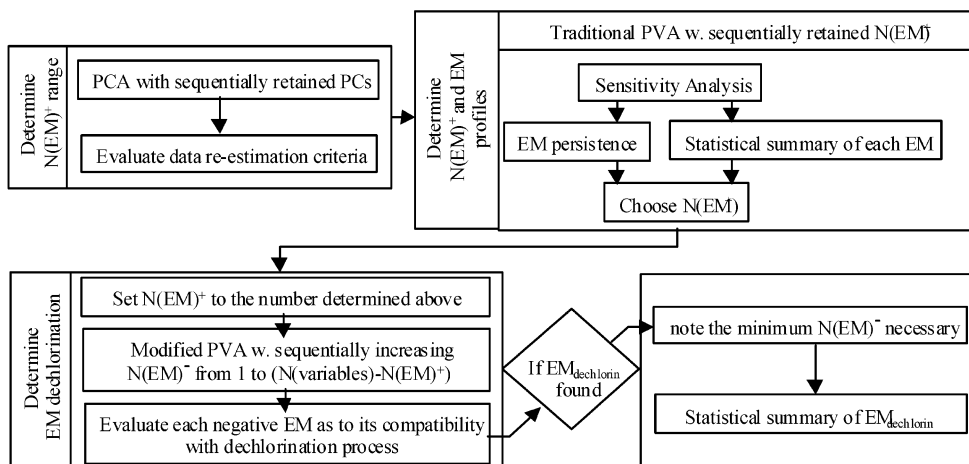


FIGURE 2. Flowchart of modified PVA algorithm. Explanations in text.

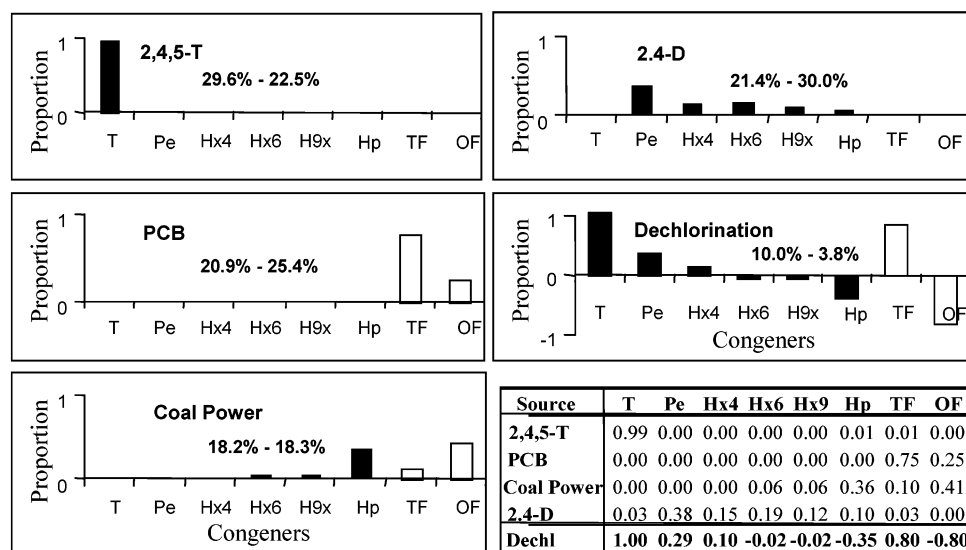


FIGURE 3. Fingerprints of sources and dechlorination. Percentages indicate variability contribution to data sets X_1 and X_2 , respectively. Dioxins are shown in black; furans are shown in white. The dechlorination end-member is double-boxed in. See text for sources. Congener codes: T = 2,3,7,8-tetraCDD; Pe = 1,2,3,7,8-pentaCDD; Hx4 = 1,2,3,4,7,8-hexaCDD; Hx6 = 1,2,3,6,7,8-hexaCDD; Hx9 = 1,2,3,7,8,9-hexaCDD; Hp = 1,2,3,4,6,7,8-heptaCDD; TF = 2,3,7,8-tetraCDF; and OF = 1,2,3,4,6,7,8,9-octaCDF.

tend to appear in the same relative position in all runs, although several runs may be necessary). In each run, $N(EM)^-$ is reduced sequentially from $N(EM)^-_{max}$ to a single EM_{dechl} . For each model, EM_{dechl} should consistently appear with minor changes. If for any of the models the procedure does not converge or does not yield EM_{dechl} , the constraint can be reapplied to the other relaxed EMs.

Other Processes That Can Lead to Shifts in Sediment Dioxin Patterns. Several other processes may also shift dioxin distribution patterns (e.g., alternate dechlorination pathways, preferential sorption of higher chlorinated species, and preferential desorption of lesser chlorinated ones in the course of sediment transport). If the changes are different from dechlorination and important, then they could appear as separate EMs.

Materials and Methods

A PVA code was written using Matlab and first tested for code errors using previously published data sets and results in refs 2, 9, and 10. Then, to validate the modified PVA code for its ability to extract known end-members of sources and dechlorination, the following procedure was adopted. Two error-free, artificial data matrices X_1 and X_2 were generated,

as PVA performance may be affected by the nature of the data set. These data matrices consisted of 8 variables (6 dioxins and 2 furans) and 131 (X_1) and 147 (X_2) samples generated using four source EMs and one dechlorination EM. Thus, both the loadings and the EMs are perfectly known. X is obtained by multiplying a randomly generated loading matrix A and the EM matrix F according to eq 1. The PVA procedure performed on X should then recreate these two matrices A and F within "reasonable" error.

The fingerprints of the four sources were selected from actual source profiles found in the literature (Figure 3): 2,4,5-trichlorophenoxyacetic acid (2,4,5-T) production (13), PCB production (14), coal-fired power plant emissions (15), and 2,4-dichlorophenoxyacetic acid (2,4-D) production (16). The congener contributions of the dechlorination fingerprint were selected to reflect dechlorination of HpCDD to TCDD with a highly negative value for HpCDD and a highly positive value for TCDD. The exact values are arbitrary, but the total sum of the dioxins and furans is 1 and 0, respectively. As the dechlorination processes for furans are less well understood, only two congeners were included, with a negative contribution for OCDF and an equal but opposite contribution for TCDF. These contributions were arbitrarily set to an absolute

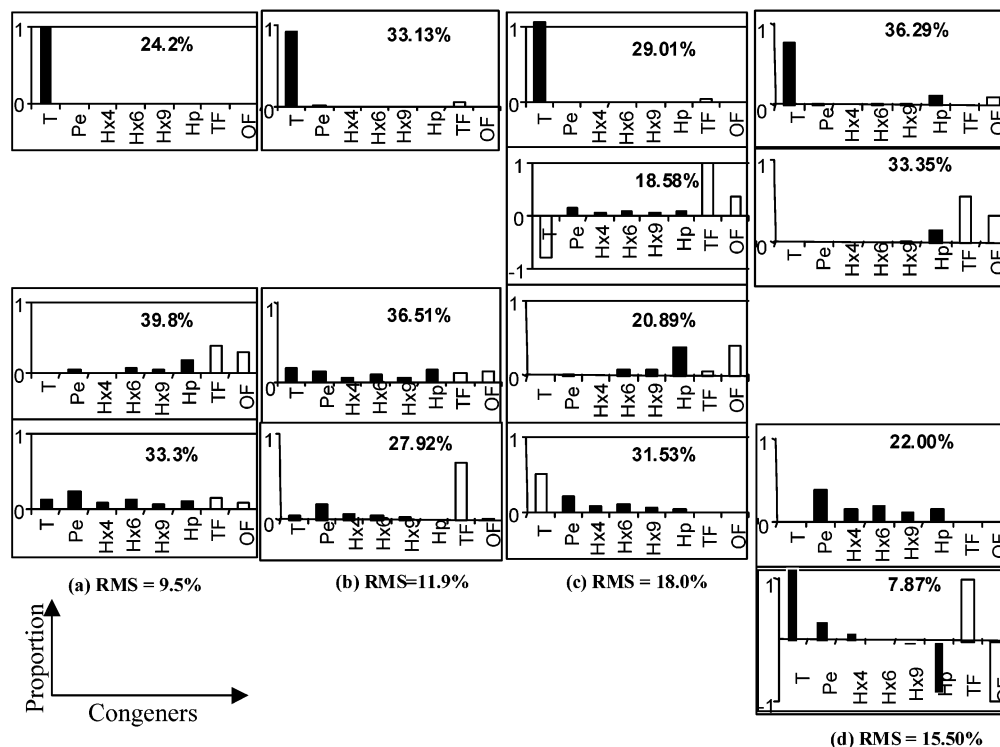


FIGURE 4. Best results for control case (X_1') (a); T-PVA with dechlorination (X_1) (b); M-PVA without dechlorination (X_1') (gray component in end-members 2 and 4 highlight the effect of relaxation) (c); and M-PVA with dechlorination (X_1) (d). Percentages in subfigures as in Figure 3.

value of 0.8 after repeated iterations for positive concentrations in \mathbf{X} .

The loadings were generated randomly by sampling Gaussian distributions with arbitrary variance–covariance matrices and means (Table S1, Supporting Information) and by eliminating samples with negative random loadings or loadings that resulted in the calculation of negative concentrations. The loadings' histograms can be seen in Figure S3, Supporting Information.

The two data sets are distinguished by the correlation pattern among the loadings (Table S2, Supporting Information) and the contribution of dechlorination to overall variability. Dechlorination contributions were selected at 10.0% to \mathbf{X}_1 and 4.0% to \mathbf{X}_2 . In particular the latter value is very low from the point of view of resolvable variance by eigenanalysis (PCA), but it is important to test the ability of M-PVA to detect such low contributions since expected values in dioxin-contaminated environments are similarly low. The data sets for \mathbf{X}_1 and \mathbf{X}_2 and their histograms are shown in Figures S4 and S5 in Supporting Information.

A nondechlorination counterpart (\mathbf{X}') of each data matrix \mathbf{X} was calculated without the dechlorination EM and its loadings. The remaining matrix \mathbf{A} was rescaled to sum to 1 for each sample, such that

$$\mathbf{X}_{p \times N} = \mathbf{A}_{p \times (k-1)} \mathbf{F}_{(k-1) \times N} \quad (4)$$

Since it is not necessarily known whether dechlorination takes place at a field site, a prior decision is necessary to use T-PVA or M-PVA. To evaluate the potential impact of this decision, four combinations of data set and algorithm were compared. The data set without dechlorination, analyzed with T-PVA, was the basis of comparison (control). T-PVA on dechlorination data shows how the false assumption of no dechlorination affects the performance of T-PVA, and M-PVA on data without dechlorination shows how results are skewed with the erroneous assumption of dechlorination. Then the

modified algorithm was applied to the full data sets to assess the ability of the modifications to extract the dechlorination EM from the data.

The number of retained source EMs was varied in order to establish the sensitivity of the results to this parameter. In every case, only a single EM was relaxed because there are no sources of error nor any additional pattern-shifting processes in the system. The impact of multiple relaxed EMs and the EM elimination procedure are demonstrated in a companion paper (17). Performance was assessed using the following criteria: RMS error (RMSE), reproduction error of the variability explained for each EM (d_{var}), and identifiability of sources (visual). RMSE of matrix \mathbf{F} and the ratio d_{var} were calculated according to

$$\text{RMSE} = \sqrt{\frac{\sum_{n=0}^N (f_n^* - f_n)^2}{N}} \quad \text{RMSE} \in [0, 1] \quad (5)$$

$$d_{\text{var}} = \frac{\text{var}^*}{\text{var}} \quad (6)$$

where f_n and f_n^* are the true and estimated concentration proportions; var^*/var are true and estimated variability contributions, respectively, calculated on columns of \mathbf{A} ; N_{res} is the total number of residuals ($f_n^* - f_n$). RMSE was calculated for the entire matrix, for each EM across congeners, and for each variable across EMs. The values of f_n , and thus RMSE itself, range from 0 to 1.

Results and Discussion

Performance of T-PVA. The results for the control data matrix \mathbf{X}_1' , shown in Figure 4a (complete results and results for \mathbf{X}_2' in Figures S5 and S6, respectively, in Supporting Information), indicate that both EMs and variabilities are reproduced well. Variability error ranges from a factor of 0.9 to a factor of 2.3

for $2 \leq N(\text{EM})^+ \leq 4$. Total RMSE ranges from 10 to 16% for \mathbf{X}_1' and from 5 to 17% for \mathbf{X}_2' . For both data sets the sources in the three- and four-EM models are identifiable. In terms of RMSE, the best model is not necessarily the one with all EMs retained: the three-EM model is best for \mathbf{X}_1' , and the four-EM model is best for \mathbf{X}_2' . Once k is reduced to half of N , more uncertainty is introduced as some sources appear combined in a single EM. In the case of data set \mathbf{X}_2' , a two-EM model would no longer be conclusive, while for \mathbf{X}_1' , the two EMs can still be identified with two of the sources.

These EMs represent the best-case expectation of the other combinations of PVA and data set. The reason for the variation between data sets \mathbf{X}_1' and \mathbf{X}_2' could be partly due to the differences in the statistical structure of the data set. In previous work with artificial data sets (2, 9) as well as in the testing of the current code with the same data sets, EMs were reproduced accurately (visual observation, error was not quantified). This indicates that certain features of the data set can affect the performance of PVA. Similarity of EMs, correlation among loadings, and variables are some possible factors which PVA may be sensitive to.

The false assumption of no dechlorination (T-PVA for data *with* dechlorination) still leads to good reproduction of source EMs. Results for \mathbf{X}_1 are shown in Figure 4b (complete results and results for \mathbf{X}_2 are in Figures S7 and S8, respectively, in Supporting Information). The RMSE ranges from 12 to 30% for \mathbf{X}_1 and from 6 to 28% for \mathbf{X}_2 , for $3 \leq N(\text{EM})^+ \leq 5$. Thus, relative to the control case, the increase in RMSE is minimal in the best case and 2-fold in the worst case. The best models are the three- and four-EM models for \mathbf{X}_1 and \mathbf{X}_2 , respectively. Variability is reproduced within a factor of 0.3–2 for both data sets. The positive components of EM_{dechl} tend to appear as additional congeners or increased contribution in the source EMs, but these contributions are distributed randomly among the source EMs: one positive congener may appear in one source, and another in another source. This is true even when dechlorination only contributes 4.0% to the variability. In previous applications (8, 18) of PVA, some source patterns displayed slight alterations, which were attributed to possible chemical reactions and aging in the sediment. These results support this conclusion but add the caveat that differences may not be consistent within a single EM. Some source EMs can come out unaltered, others may be altered in only one or a few of the congeners that are involved in the reactions. The sources are better identifiable with the lower dechlorination contribution of 4.0% (\mathbf{X}_2). Results from these particular data sets tentatively indicate that, when one is not interested in quantifying chemical reactions, T-PVA can still yield EMs that are identifiable as certain sources provided that the reactions have a low contribution, as is the case with persistent compounds such as dioxins and PCBs where PVA has been applied.

Performance of M-PVA. The false assumption of dechlorination (i.e., M-PVA for data *without* dechlorination) also leads to acceptable results (see Figure 4c for \mathbf{X}_1' ; results for \mathbf{X}_2' are in Figure S9, Supporting Information.) The RMSE is 18% for \mathbf{X}_1' for $N(\text{EM})^+ = 4$ and ranges from 10.0 to 15.3% for \mathbf{X}_2' for $2 \leq N(\text{EM})^+ \leq 4$. This is comparable to the range of RMSE in the control case. Variability is reproduced within a factor of 0.8–1.3 (\mathbf{X}_1') and 0.8–1.9 (\mathbf{X}_2'). For \mathbf{X}_2' , the four-EM model is best, while for \mathbf{X}_1' only a four-EM model was calculated. The sources are identifiable, but the effect of a relaxed EM leads to significant changes in the profiles apparent as negative contributions in an otherwise unchanged source EM. These negative values may be compensated for by increased positive values in other EMs. Thus, this solution may represent a mathematical solution for the data but not one that also has a counterpart in the physical system. This effect could lead to erroneous identification of

sources/processes. As such, relaxed EMs should only be modeled when there is evidence to support equivalent processes.

In the case where the modified algorithm is correctly applied to data with dechlorination, reproduction of sources most closely approaches the reproduction in the control. RMSE for source EMs were 8–19% (\mathbf{X}_1) and 8–17% (\mathbf{X}_2) for $3 \leq N(\text{EM})^+ \leq 5$ (see Figure 4d for \mathbf{X}_1 ; complete results for \mathbf{X}_1 and \mathbf{X}_2 are in Figures S10 and S11, respectively, in Supporting Information.) Variability contributions were reproduced within a factor of 0.6–1.9 (\mathbf{X}_1) and 0.6–1.3 (\mathbf{X}_2). The dechlorination EM was reproduced very well in the four-EM model for both data sets with an RMSE of 28% (\mathbf{X}_1 and \mathbf{X}_2), and variability contribution was estimated within 1–1.9% for \mathbf{X}_1 and 2.2–5.4% for \mathbf{X}_2 . The estimation was better for the data set with a higher dechlorination contribution to variability, indicating that dechlorination must make a sufficiently large contribution to the overall variability (no less than the order of magnitude of error) in order for M-PVA to reproduce EM_{dechl} adequately, even in the absence of error. Nevertheless, the dechlorination EM was identifiable as dechlorination in all models.

Sensitivity and Uncertainty. (The numerical results of all four cases above are compared in Table S3, Supporting Information.) For data with dechlorination, source EMs are better reproduced with M-PVA. However, even when choosing the correct model, some uncertainty is always associated with PVA results. The reproduction of the actual patterns and their contribution to global variability depend on the choice of model parameter values, the number of EMs in the model, selection of the EMs to be constrained, the initial polytope of extreme values (e.g., whether varimax rotation is performed or not), the variability contribution of dechlorination, and may depend on the similarity of EMs, statistical properties of the data and loadings, etc. An additional source of uncertainty arises when the presence of dechlorination itself and thus the appropriate model is in question. For this reason, methods are necessary to quantify model uncertainty and to facilitate appropriate model choice.

How does the modeler decide whether to use M-PVA or T-PVA with real data? Although direct field evidence of dechlorination is preferable, in its absence, T-PVA is an appropriate starting point in determining source EMs regardless of the presence of chemical reactions provided that these do not dominate sources. Most sources are well reproduced and are identifiable. At this stage, it is important to determine the minimum number of EMs carefully as the outcomes become highly sensitive to this parameter once $N(\text{EM})^+$ approaches half or less the actual number of sources (17). Then, when dechlorination is included through the M-PVA model, the changes that occur in the source EMs indicate whether M-PVA is appropriate and the presence of significant dechlorination in the system. If dechlorination is absent from the system, a relaxed EM in M-PVA appears as a source EM with additional negative components leaving the positive components relatively unchanged. This relaxed EM will not correspond to the criteria of dechlorination. In addition, the negative components may be “canceled” out by additional positive components in other source EMs (Figure 4c). If dechlorination is truly present in the system, using M-PVA will lead to some rearrangement of the source EMs from T-PVA, and the relaxed EM will correspond to the criteria of dechlorination (rather than simple addition of negative values to an otherwise identifiable source EM). Note also that the end-members are not dechlorinated; it is rather the mixture as a whole. How the dechlorination is distributed among the end-members is part of the uncertainty of a model that does not allow separation of dechlorination from the end-members.

The comparison of the four cases above points to a strategy involving the *persistence* of EMs and the *variability* of their compositions across different models to measure sensitivity to k . "True" EMs should appear in all or most models, and their variability should be low for congeners that are truly part of the EM. Even in these artificial data sets with few EMs, some EMs persist into low k models with limited variability in composition. Such EMs can be assumed to be correctly identified. Consequently, a sensitivity analysis involving a range of models from high to low k is possible and desirable with real data, as demonstrated for the Passaic River in a companion paper (17). Johnson et al. (19) have performed a bootstrapping-based uncertainty analysis by randomly subsampling the data set with replacement for repeated PVA runs. This approach quantifies uncertainty with respect to the representativeness of the samples, and it is useful for assessing the impact of uncertainties on the results. In the case of Johnson et al. (19), the resolved EMs remained stable during the bootstrapping.

Limitations of M-PVA. M-PVA is developed and demonstrated here for systems with a low degree of dechlorination, as is appropriate for dioxin/furan mixtures. By performing T-PVA first, positive end-members of unaltered sources are assumed to account for the first extracted principal components with the highest variability contribution. The contribution of EM_{dechl} to a sample is also limited by the available source-supply of congeners with negative values in EM_{dechl} . Mathematically, then, the value of any dechlorination loading in matrix **A** is a function of loadings and factors that together define congener availability in a sample.

The variability contribution is the most difficult parameter to estimate in the model. Prior to PVA, all samples are normalized to contribute equally to the variability regardless of whether their concentrations were very high or very low. This is a statistical necessity as well as a strength of the procedure allowing increased sensitivity to contributing phenomena. The importance of very strong signals is diminished, and less important contributors can be detected, which would otherwise be masked by the strong signal of high absolute variances. The implications of normalization depend on analysis objectives. If the objective is to detect the presence or absence of patterns, this approach is an advantage. If the objective is to know which source was a more important contributor in the mixing system, then normalization is a limitation for both T-PVA and M-PVA. Contribution to variability is not contribution to absolute mass in the system. This has particular implications for M-PVA. The dechlorination end-member has negative contributions that must be counterbalanced by positive contributions. Normalization can artificially increase or decrease the availability of a dechlorinated congener. This can affect the dechlorination fingerprint itself and not just the variability contribution. Further experimental and artificial data sets should be evaluated to test the impact of normalization on the ability of M-PVA to reproduce dechlorination fingerprints.

Incomplete information on dioxin dechlorination pathways and conditions limits our ability to interpret the results of PVA. Relative to PCBs, much less is known about dechlorination pathways and fingerprints, field patterns, and source fingerprints (4, 20). This is an important reason that PVA be combined with other lines of evidence (such as follow-up field validation) toward making a final conclusion about the presence of both sources and transformation reactions in the system.

Finally, these artificial data sets mirror only some aspects of a real data set. A key difference is that in a real system many more sources of variability exist and any one identified EM is representative of a source category. Certain sensitivities of the procedure, such as the impact of the parameter k ,

probably become less pronounced with more actual sources of variability. Another difference is that in addition to biological/chemical reactions, random and systematic errors are a possible source of negative values, likely at a comparable level of contribution as dechlorination itself. This could result in dechlorination patterns being masked by error. Implications of the presence of error need to be further investigated with artificial data sets that include random noise, similar to data sets used in ref 2. Consequently, when applying M-PVA to real data, the identification and elimination procedure must involve more than one relaxed EM to allow end-members related to error to arise separately from potential dechlorination end-members, and results must be further tested for the impact of error using bootstrapping uncertainty analysis. A companion paper (17) demonstrates such an approach.

Acknowledgments

We thank the National Science Foundation (Grant DMS-9905454) and the Office of Naval Research for financial support. The authors express their gratitude for constructive comments provided by Sharon Jaffes and Jon Josephs of the U.S. EPA and the anonymous reviewers.

Supporting Information Available

Conceptual mixing model, graphical illustration of rotation procedure, parameter sensitivities of PVA, variance-covariance matrices and means used to generate loadings matrices, histograms of the loadings for artificial data sets, comparison of correlation patterns of loadings for X_1 and X_2 , statistics and histograms for artificial data, complete T-PVA/M-PVA results for X_1 , X_1' and X_2 , X_2' . This material is available free of charge via the Internet at <http://pubs.acs.org>.

Literature Cited

- (1) U.S. National Research Council (NRC). *Contaminated Sediments in Ports and Waterways: Cleanup Strategies and Technologies*; National Academy Press: Washington, DC, 1997; ISBN 03090-54931 (alk. paper).
- (2) Johnson, G.; Ehrlich, R.; Full, W. In *Introduction to Environmental Forensics*; Murphy, B., Morrison, R., Eds.; Academic Press: New York, 2002; pp 461–515; ISBN 0-12-511355-2.
- (3) Evans, J. C.; Ehrlich, R.; Krantz, D.; Full, W. *J. Geophys. Res.* **1992**, *97*, 2365–2378.
- (4) Gruden, C.; Fu, Q. S.; Barkovskii, A. L.; Albrecht, I. D.; Lynam, M. M.; Adriaens, P. In *Dehalogenation: Microbial Processes and Environmental Applications*; Häggblom, M. M., Bossert, I. D., Eds.; Wiley & Sons: New York, 2003.
- (5) Albrecht, I. D.; Barkovskii, A. L.; Adriaens, P. *Environ. Sci. Technol.* **1999**, *33*, 737–744.
- (6) Barkovskii, A. L.; Adriaens, P. *Appl. Environ. Microbiol.* **1996**, *62*, 4556–4562.
- (7) Johnson, G. W.; Quensen, J. F. *Organohalogen Compd.* **2002**, *45*, 280–283.
- (8) Johnson, G. W.; Jarman, W. M.; Bacon, C. E.; Davis, J. A.; Ehrlich, R.; Risebrough, R. W. *Environ. Sci. Technol.* **2000**, *34*, 552–559.
- (9) Johnson, G. Ph.D. Dissertation, University of South Carolina, 1997.
- (10) Miesch, A. T. *Q-mode Factor Analysis of Geochemical and Petrologic Data Matrices with Constant Row Sums*; Geological Survey Professional Paper 574-G; U.S. Geological Survey: Denver, CO, 1976.
- (11) Kim, B. M.; Henry, R. C. *Chemom. Intell. Lab. Syst.* **2000**, *52*, 145–154.
- (12) Kim, B. M.; Henry, R. C. *Chemom. Intell. Lab. Syst.* **1999**, *49*, 67–77.
- (13) Umbreit, T. H.; Hesse, E. J.; Gallo, M. A. *Science* **1986**, *232*, 497–499.
- (14) Malisch, R. *Organohalogen Compd.* **1994**, *20*, 209–214.
- (15) Riggs, K. B.; Brown, T. D.; Schrock, M. E. *Organohalogen Compd.* **1995**, *24*, 51–54.
- (16) Cleverly, D.; Schaum, J.; Schweer, G.; Becker, J.; Winters, D. *Organohalogen Compd.* **1997**, *32*, 430–435.

- (17) Barabás, N.; Adriaens, P.; Goovaerts, P. Modified polytopic vector analysis to identify and quantify a dioxin dechlorination signature in sediments. 2. Application to the Passaic River. *Environ. Sci. Technol.* **2004**, *38*, 1821–1827.
- (18) Jarman, W. M.; Johnson, G. W.; Bacon, C. E.; Davis, J. A.; Risebrough, R. W.; Ramer, R. *Fresenius J. Anal. Chem.* **1997**, *359*, 254–260.
- (19) Johnson, G. W.; Jarman, W. M.; Bacon, C. E.; Davis, J. A.; Ehrlich, R.; Risebrough, R. W. *Environ. Sci. Technol.* **2000**, *34*, 552–559.
- (20) Bedard, D. L.; Quensen, J. F. In *Microbial Transformation and Degradation of Toxic Organic Chemicals*; Young, L. Y., Cerniglia, C. E., Eds.; John Wiley & Sons: New York, 1995; pp 77–127.

Received for review October 9, 2002. Revised manuscript received December 12, 2003. Accepted December 18, 2003.

ES026228Z