

# Multivariate Analyses to Improve Understanding of NAPL Pollutant Sources

by Susan E. Powers, James F. Villaume, and John A. Ripp

## Abstract

Current site assessment techniques do not always generate adequate information regarding the presence, type, or distribution of nonaqueous phase liquids (NAPLs) at sites with ground water contamination. Without this information, however, the design of remediation methods is uncertain, often resulting in costly and/or failed attempts to minimize risks associated with the contamination. In this work, it is proposed that a thorough multivariate analysis of data from ground water sampling efforts could improve our overall understanding of these sites.

Multivariate analyses can provide considerable insight into the contaminant source characteristics by elucidating correlations in ground water concentrations that identify recurring chemical patterns or "signatures." These correlations are related to the type of NAPL and proximity to the contaminant source.

Multivariate plots and a principal components analysis (PCA) are used to interpret ground water data from a manufactured gas plant site in Iowa contaminated with both gasoline and coal tar. Conclusions from these analyses regarding the distribution of NAPL contaminants were generally consistent with those derived from direct physical evidence of the NAPL sources. The multivariate analyses, however, provide an additional level of interpretation regarding the distribution of coal tar in the subsurface that was not possible with the standard evaluation techniques used during the remedial investigation (RI). Comparison of chemical signatures of ground water samples among wells identified two distinct regions of coal tar contamination, and suggests which wells are impacted by each source providing greater confidence in the location of the DNAPL sources.

## Introduction

NAPLs present in the subsurface are recognized as long-term sources of ground water pollution (Mackay and Cherry 1989). The effectiveness of any attempt to remediate these contaminated sites is limited by our ability to first identify the distribution of NAPLs in the subsurface, a problem that is particularly complex at sites with multiple contaminant sources, or when the NAPL has a density higher than that of water (DNAPL). When the geologic strata include even small changes in permeability, DNAPLs undergo a significant amount of lateral spreading as they migrate (Kueper et al. 1993), making it extremely difficult to predict their distribution in the subsurface.

Extensive ground water data are collected at most hazardous waste sites during remedial investigations. These data can provide considerable insight into the nature and extent of site contamination if thoroughly analyzed and used in conjunction with adequate knowledge of the site history. The U.S. EPA (1993) and Cohen et al. (1993) provide guidelines to infer the presence of a NAPL source based on ground water data. It is now a common practice to use ground water concentrations that are greater than 1 percent of the effective solubility of a species to suggest that a NAPL is present. These guidelines offer a rudimentary means of interpreting ground water data to provide additional information regarding the source characteristics. There are

numerous advantages of this "inverse" approach as an alternative to direct observation of subsurface NAPL contamination. By focusing on the ground water plume, there is less risk of increasing the extent of DNAPL contamination during drilling activities. In addition, the collection and analysis of aqueous phase samples is often easier and cheaper than direct characterization of the source material.

One common attribute among many contaminated sites is the presence of NAPLs composed of a complex mixture of many organic species. The resulting concentrations of multiple chemical species in ground water contaminated by these NAPL mixtures depend on the chemical composition of the contaminant source, proximity of the sampling location to the source, and transformation processes that occur in the subsurface. Patterns, or chemical "signatures," emerge among the contaminant concentrations in the ground water that are directly related to the source of contamination and subsequent transformation processes.

Ground water in the vicinity of a NAPL becomes contaminated following the leaching of components from the organic phase into the water. In many cases, there are classes of NAPLs—such as gasolines—present in the subsurface that have similar compositions. Although there are variations in the chemical composition among brands and grades of gasoline, there are some commonalities among all gasolines, including the presence and relative proportions of BTEX (benzene, toluene, ethyl benzene, xylenes) species.

The aqueous phase concentration of species equilibrated with an ideal NAPL is estimated by Raoult's law (Feenstra 1992):

$$C_i^e = X_i^n C_i^{*w} \quad (1)$$

where  $C_i^e$  is the equilibrium concentration of species  $i$  in the aqueous phase;  $X_i^n$  is the mole fraction of species  $i$  in the organic phase; and  $C_i^{*w}$  is the aqueous phase solubility of  $i$  in its pure liquid state. The absolute concentration of species found in the subsurface may be significantly less than predicted by this equilibrium equation due to dilution or rate-limited mass transfer processes (Mackay and Cherry 1989). However, the relative proportions of an individual chemical in ground water in the vicinity of a NAPL source should be directly related to its solubility and mole fraction in the organic phase. This implies that the relative ground water concentrations of species leached from NAPLs of similar composition should also be similar. In addition, since the mole fractions of common contaminants vary among different types of petroleum products, ground water samples exposed to different types of NAPLs will have their own unique chemical signatures. With increasing distance from a NAPL source, however, chemical signatures will change as differences in biodegradation rates and sorption capacities result in the reduction of certain chemical species.

The identification of patterns in chemical concentrations among ground water samples requires the use of

multivariate analyses to determine correlations among variables within a data set. The multivariate data from each monitoring well at suspected NAPL sites typically include concentrations of a wide variety of organic and inorganic contaminants, properties such as pH and redox potential, time of sampling, and well location. The patterns or signatures identified by the multivariate analysis depend on the chemical composition of the contaminant source, proximity of the sampling location to the source, and sorption or biodegradation processes. Interpretation of these patterns to infer the specific factors contributing to the observed spatial and temporal distribution of the ground water contaminants is critical to the successful application of the multivariate techniques.

Multivariate analysis techniques have been widely used in the field of geochemistry to determine the sources of inorganic ground water constituents, air pollutants, and contaminated marine sediments (e.g., Novak and Eckstein 1988; Hopke 1991; Shine et al. 1995). Recently, a few multivariate techniques have been employed to evaluate gasoline contamination (Lesage and Lapcevic 1990; Anderson et al. 1992). Neither researchers nor practitioners, however, have developed these techniques sufficiently for their application to ground water contamination at a wider range of NAPL sites.

The objectives of this work were to develop and illustrate the application of multivariate data analysis tools for identifying patterns or chemical signatures in ground water contaminant data at suspected NAPL sites. It is hypothesized that with identification and interpretation of specific chemical patterns that emerge in samples collected from complex NAPL spill sites, our understanding of the nature and extent of these contaminant sources can be improved.

## Approach

Two specific multivariate techniques were applied to ground water data from a former manufactured gas plant (MGP) site in Iowa to evaluate the potential contribution of these methods to the overall site characterization process. The Iowa site is contaminated with distinct gasoline and coal tar sources, providing a good example to develop and illustrate the use of multivariate analyses to differentiate between these NAPL sources.

The first technique considered here involved the use of multivariate plots representing complex data sets to enable comparison of chemical signatures between actual ground water samples and those of known contaminants. Initially, a literature review was conducted to compile a database of chemical signatures for comparative purposes. Representative data collected during this review are included in Table 1. Data from both field studies and laboratory-scale experiments quantifying the concentration of solutes in aqueous phase samples exposed to NAPL sources were included. Due to the nature of contamination at the MGP site considered here and the

**Table 1**  
**Composition of Water Equilibrated with NAPL Sources**

NAPL Source	Concentrations (mg/L)						Reference
	Benzene	Toluene	EB+Xyl*	Naphthalene	Acenaphthylene	Phenanthrene	
<b>Gasolines:</b>							
** summer	30.7	34.8	10.8	0	NA	NA	Shiu et al. 1990
** unleaded	24.2	36.3	23.9	0.2	NA	NA	Shiu et al. 1990
unleaded	58.7	33.4	26.6	0	0	0	API 1985
super unleaded	67.0	107.4	24.6	NA	NA	NA	API 1989
weathered	5.8	63.0	36.9	NA	NA	NA	Hinchee and Reisinger 1987
Amoco	32.4	34.3	6.6	NA	NA	NA	Bruce et al. 1991
** 89 oct.	41.2	40.8	17.2	NA	NA	NA	Luhrs and Pyott 1992
** 89 oct., weathered	9.8	39.1	24.7	NA	NA	NA	Luhrs and Pyott 1992
<b>Heavier Hydrocarbons:</b>							
Fuel oil #2	0.14	0.44	0.7	0.11	NA	NA	Shiu et al. 1990
Fuel oil #4	0.22	1.01	0.73	0	NA	NA	Shiu et al. 1990
Fuel oil #2	0.646	1.68	8.89	NA	NA	0.296	Heath et al. 1993
** diesel oil	0.2	0.5	0.86	0.09	NA	NA	Shiu et al. 1990
lube oil - used	0.1	0.2	0.2	0.01	NA	NA	Shiu et al. 1990
<b>Coal-derived NAPLs:</b>							
** wood preserving chemicals	0.17	0.23	0.48	10.4	10.4	2	Goerlitz 1992
wood preserving chemicals	0.033	0.048	0.133	3.312	0.059	1.825	Rosenfeld and Plumb 1991
** average coal tar	0.01	0.005	0.016	0.028	0.02	0.007	GRI 1987
** Stroudsburg coal tar	2.4	1.1	2.99	3.9	0.7	0.16	Peters and Luthy 1993
NA— Not available							
*Ethyl benzene + xylenes							
**Samples used for representative star plots (Figure 3)							

extent of data available from the literature, this review was limited to organic phases originating from petroleum or other fossil fuel sources. Thus, the primary constituents measured in these studies were aromatic hydrocarbons (BTEX) and polycyclic aromatic hydrocarbons (PAHs).

The second technique considered relied on mathematical interpretation of correlations among chemical species present in ground water samples from this site. Application of this method does not require a priori knowledge of representative chemical signatures. Instead, correlation patterns within the data set were identified mathematically and interpreted based on knowledge of the site history and physical, chemical, and biological processes affecting contaminant concentrations.

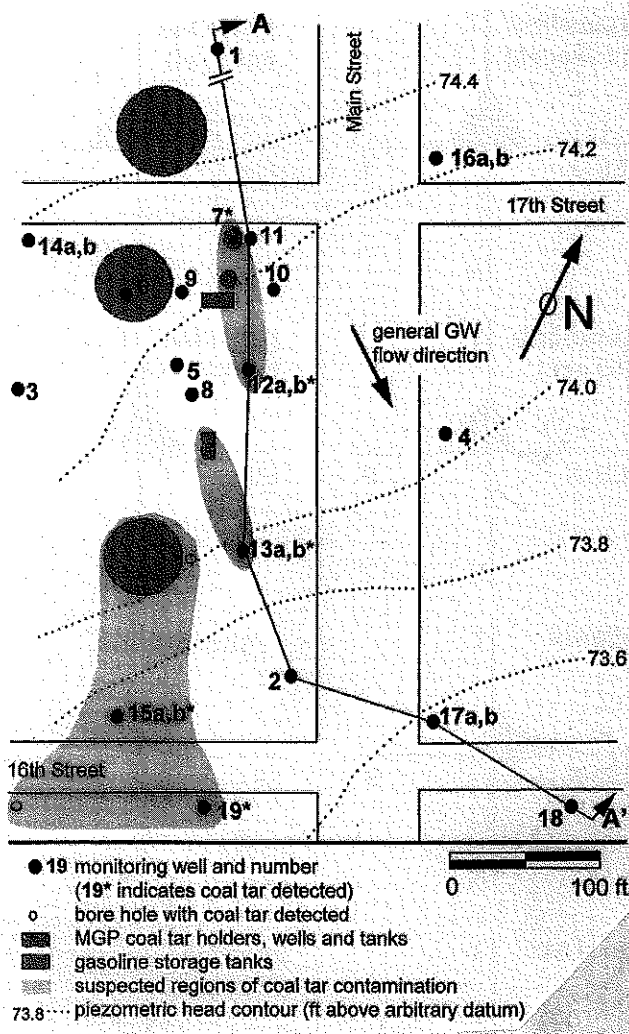
### Site Description

Investigation of the subsurface contamination at the site in Iowa was initiated following the 1987 discovery that gasoline had leaked from an underground storage tank. Subsequent field activities and review of historical documents showed that an MGP facility also used this property from the mid-1850s to the late 1940s (Figure 1). A four-phase remedial investigation conducted at this site from 1989 to 1991 included an evaluation of the ground water hydrology and an extensive

collection of soil, water, and air samples from borings, test pits, monitoring wells, and an array of soil gas sampling locations (Atlantic 1991).

The cross-sectional plot presented in Figure 2 identifies the primary aquifer as an outwash sand and gravel unit overlain by a clayey silt layer. Piezometric head contours shown in Figure 1 indicate that ground water flow is generally in a southeasterly direction across this site with a seepage velocity of approximately 66 m/year (Atlantic 1991). Thirty-six ground water samples from both the upper and lower portion of the sandy aquifer were collected from a total of 24 wells during the RI. In many cases, the wells were drilled in pairs with one screened in the shallow portion of the sand and gravel unit and the other screened near the base of this aquifer (Figure 2). During different phases of the RI, samples were analyzed for volatile aromatics, PAHs, cyanide, phenol, and RCRA metals. Ground water samples from downgradient wells (all except Wells 1, 3, and 14) contained elevated concentrations of organic contaminants with naphthalene, acenaphthylene, benzene, and xylenes detected in more than 95 percent of the downgradient ground water wells.

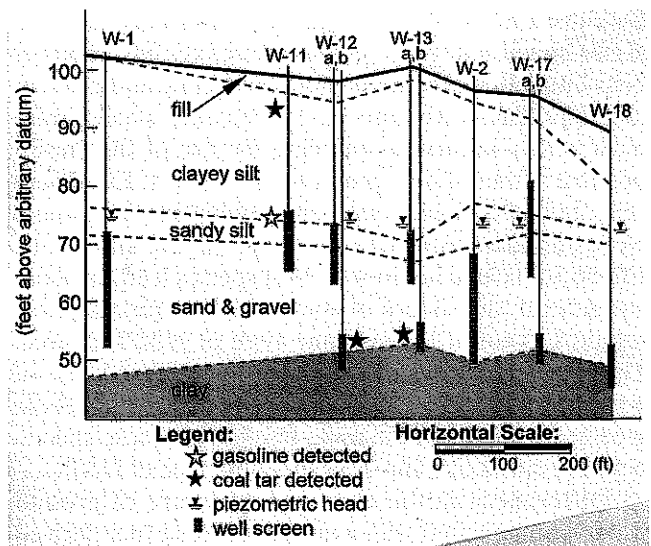
Both gasoline and coal tar were detected in the subsurface during the RI. The gasoline was detected at the capillary fringe in the immediate vicinity of the former underground storage tank (Wells 9, 10, 11). Coal tar was



**Figure 1:** Base map for the former MGP site summarizing results of the RI (Atlantic 1991). Shaded areas indicate regions of suspected NAPL contamination based on test pit and soil boring samples, and an understanding of the site stratigraphy and ground water hydrology.

found in samples collected from surficial soils in the region of many of the former MGP structures, in several boreholes drilled directly into gas holders and coal tar wells, and at the bottom of the sandy aquifer in Wells 12b, 13b, 15, and 19 (Figure 1). Four coal tar samples were analyzed by infrared spectral analysis. Based on their composition, tar samples from deep Well 19, the shallow Well 7 drilled directly into a tar well, and a borehole in the southernmost gas holder were identified as carburetted water gas tars. The sample collected from deep Well 13b was identified as a heavier coal tar residue based on its lack of BTEX species (Atlantic 1991).

Questions regarding the distribution and original sources of coal tar in the subsurface remained unanswered following completion of the RI. Based on limited data defining the elevation of the bottom of the aquifer and analysis of site stratigraphy and ground water hydrology, the distribution of DNAPL presented in Figure 1 was inferred. The high concentrations of total PAH (average ~ 4000 µg/L) measured in ground



**Figure 2:** Subsurface stratigraphy along cross section A-A' (Figure 1).

water samples from Well 2 were not well understood. Coupled with the dilution expected in this fully penetrating well and the relative immobility of most PAHs in ground water, these results suggested the presence of coal tar in the immediate vicinity of this well. The elevation of the bottom of the aquifer at this well is lower than that at either Well 15b or 13b, indicating that there could be DNAPL migration toward Well 2 from either of these areas.

## Multivariate Plots

### Chemical Signatures for Known NAPL Contaminants

The application of multivariate plots was based on comparison of ground water concentration data at the Iowa site to those found associated with common NAPLs (Table 1). A type of multivariate plot called a "star" or "radar" plot was used to illustrate the chemical signatures for these aqueous phase data. These plots use polar coordinates with the number of axes equal to the number of variables (Figure 3). Either the absolute or relative concentration of each species is plotted as a datum on its axis and the points are connected to form a polygon. The shape of the polygon is the "signature" for a particular pollutant source (Lesage and Lapcevic 1990; Anderson et al. 1992).

Star plots for different sets of chemical species included in Table 1 were initially generated to determine the best representation of these data. It was desired to use a maximum amount of available data to produce star plots that have consistent shapes for aqueous phase samples exposed to the same type of NAPL, yet are unique between NAPLs. In order to generate plots of similar size, all concentrations were normalized by dividing the absolute concentration of each individual species by the sum of all constituents used in the plots. The best results for identifying unique signatures

for different NAPL sources were four-axis star diagrams generated with aqueous phase concentrations of benzene, toluene, combined ethyl benzene and xylenes, and naphthalene (Figure 3). These contaminants were most widely occurring, enabling the use of the largest possible subset of data presented in Table 1. Ethyl benzene and xylenes have the same molecular weight and were summed to create one axis on the star plots. Efforts to include additional PAH species lead to inconsistencies in the shapes of star plots for aqueous phase samples exposed to the same type of NAPL source. These inconsistencies could stem from higher levels of uncertainty in analysis of low concentrations of PAHs or greater variability in the mole fractions of these species between different samples of the same type of NAPL.

Differences in the chemical signature of ground water contaminated with a variety of fossil-fuel-derived NAPLs are clearly apparent based on the different shapes of the polygons presented in Figure 3. Four gasoline and two coal tar sources are included to provide an indication of the consistency in the polygon shapes for each of these NAPLs. The inclusion of naphthalene concentrations provides an indicator chemical to differentiate between coal-derived NAPLs and heavy petroleum products. Star plots for the two coal tars illustrate the consistency in shape between samples collected from field (GRI 1987) and laboratory (Peters and Luthy 1993) studies. The chemical signatures of these two samples are the same, although absolute concentrations vary by nearly two orders of magnitude (Table 1).

All of the data presented in Table 1 and Figure 3 are for aqueous phase samples directly impacted by an organic phase. An increase in the distance between the ground water sample and the NAPL source, or changes in the NAPL composition over time, will affect the chemical signatures. This can be illustrated by comparing star plots of "weathered" and fresh gasoline samples (Figure 3). Benzene, the most soluble aromatic hydrocarbon in gasoline, is preferentially leached from the source, resulting in a reduction of both the mole fraction of this component in the gasoline and concentration in the nearby aqueous phase. The shape of the star plot resulting from the weathering process is apparent from the shorter length of the polygon along the benzene axis for the weathered gasoline relative to the fresh gasolines.

### Analysis of Chemical Signatures

Ground water contamination data from the former MGP site in Iowa were initially analyzed by comparing star plots for sample data to those associated with known NAPLs. For comparative purposes, star plots with the same format as shown in Figure 3 were generated for ground water samples with detectable concentrations of BTEX and naphthalene and are included on a base map in Figure 4. In general, and as expected, the deep wells show elevated relative concentrations of naphthalene indicating contamination from a coal tar source, while the shallower wells generally have much

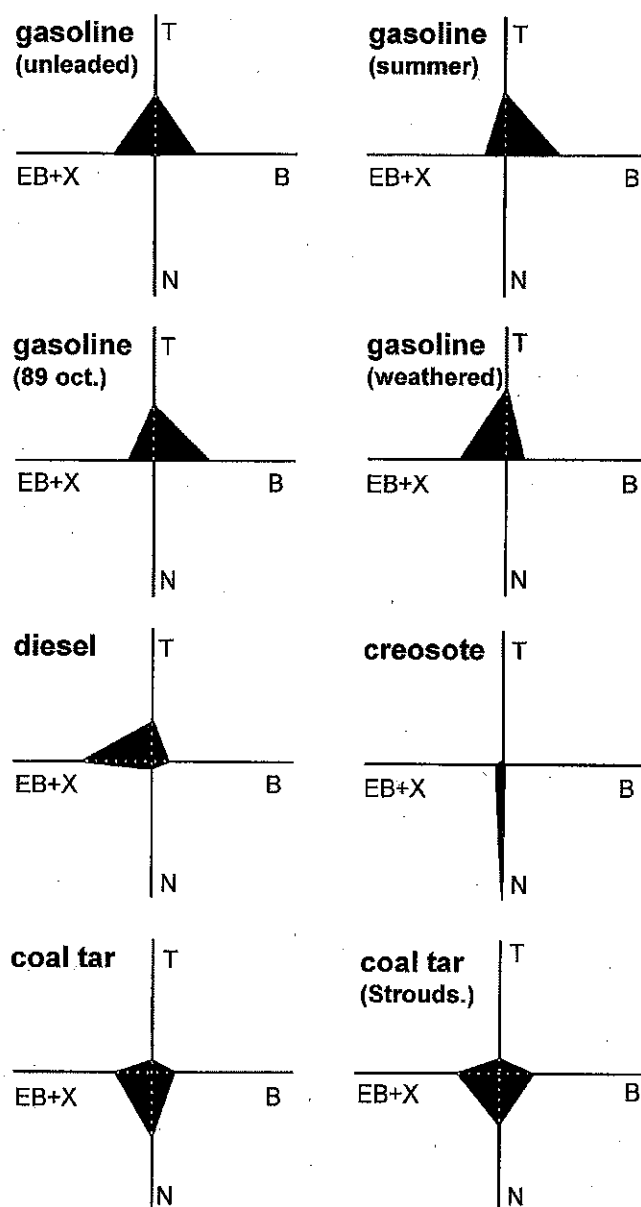
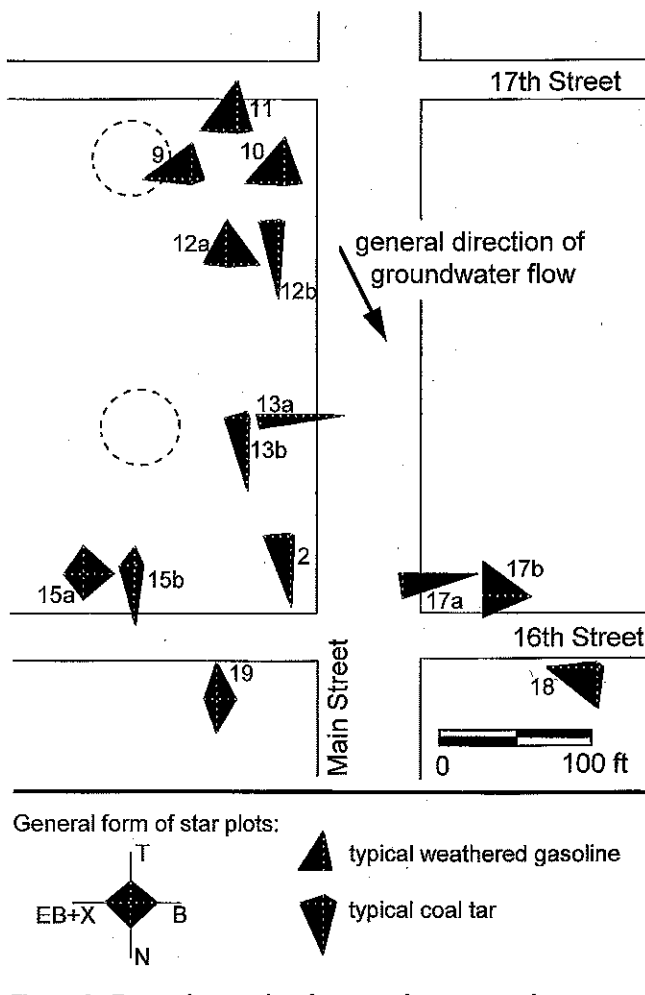


Figure 3: Four-axis star plots of chemical constituents in water near NAPL pollutants. Axes include normalized concentrations of benzene (B), toluene (T), the sum of ethylbenzene plus xylenes (EB+X), and naphthalene (N). Differences in the characteristic shapes of these plots among different types of NAPL provide clearly unique chemical signatures.

lower relative concentrations of this constituent. Wells 9 through 12a have a triangular shape, similar to that of weathered gasoline shown in Figure 3, indicating that each of these wells is in the close proximity to pooled or entrapped gasoline in the capillary fringe.

A comparison of the shapes of star plots in shallow wells downgradient from the gasoline source illustrates the change in chemical signatures due to subsurface sorption and transformation processes. Wells 13a and 17a both have much higher relative concentrations of benzene than in the wells immediately adjacent to the gasoline storage tank location. When considering the potential for sorption, benzene has the highest solubility and lowest organic carbon partitioning coefficient. These characteristics indicate that this species would be



**Figure 4:** Four-axis star plots for ground water samples are plotted on the base map in the vicinity of their sampling locations. Axes are the same as depicted in Figure 3 but are removed for the purposes of this figure. Comparison of shapes with those from "typical" NAPLs illustrates that samples from the shallower regions of the aquifer are contaminated primarily with gasolines while the deeper wells are affected by coal tars.

the most mobile in a subsurface system and, thus, would be observed in the highest relative concentrations as the distance from the NAPL source is increased. Ethyl benzene, however, would be transported at a rate that is much slower than benzene due to its order-of-magnitude higher organic carbon partitioning coefficient and the moderate organic carbon fraction of the soil ( $f_{oc} \sim 0.003$ ). The decreased length of the axis for ethyl benzene plus xylenes on the star plots from the generally downgradient Wells 13a and 17a is consistent with the anticipated effect of sorption on the overall chemical signature.

Analysis of the shape of star plots for downgradient Wells 13a and 17a also shows a significant reduction in the relative concentration of toluene in these wells. Toluene, which has a mobility intermediate between benzene and ethyl benzene, was found in these wells at concentrations of only a few  $\mu\text{g/L}$ . There is limited evidence to suggest that toluene is degraded at a faster rate than other BTEX species under both aerobic and anaerobic conditions at petroleum spill sites (Wilson et al. 1991; 1994). Thus, the reduction in the relative concentration of this species could be due to a combination

of greater retardation and biodegradation in comparison with benzene.

Coal tar was directly observed in several wells (12b, 13b, 15b, and 19) penetrating the deeper portion of the sandy aquifer (Figure 1). Star plots for the aqueous phase samples from these wells (Figure 4) generally have the same shape as for the known coal tar sources presented in Figure 3. Variability in the polygon shapes among these wells could indicate different ages or types of MGP wastes. The infrared spectral analysis of the sample collected from Well 19 was identified as a carburetted water gas tar or fraction thereof, while the analysis of a sample collected from Well 13b indicated a much heavier coal tar residual with low concentrations of BTEX (Atlantic 1991). The shapes of the star plots describing the ground water concentrations in the vicinity of these two wells are consistent with the measured DNAPL compositions. That is, the plot for Well 19 shows much higher relative concentrations of benzene and toluene than the plot for Well 13b.

Incorporating the information provided by analysis of the star plots with analysis by more standard approaches (absolute concentrations, elevations of aquifer bottom, etc.) allows interpretation of the original sources of coal tar and the spatial distribution of coal tar throughout the subsurface with a higher level of confidence. Specific inferences for this site that differ from the original assessment presented in Figure 1 include the following:

- There is a coal tar source close to Well 2. This is supported by the lack of change in chemical signature between Well 13b and Well 2 that would have been expected due to differences in the relative mobility of naphthalene vs. benzene.
- There is probably not significant DNAPL contamination outside of the suspected leaking tar well (Well 7). Greater influence of naphthalene in the star plot from Well 11 would have been expected if this were not the case.
- Based on similarities in the shape of star plots between Wells 12b, 13b, and 2, the same coal tar source is in close proximity to each of these wells.
- Low relative concentrations of the mobile benzene and toluene species at Well 2 relative to those at Well 15b indicate that Well 2 has not directly impacted the carburetted water gas tar found at Wells 15b and 19.

While the first two of these inferences could also have been made based on absolute concentrations of contaminants in ground water samples from Wells 2 and 11, confidence in these conclusions is strengthened by the added analysis of multivariate plots. The second two conclusions could not have been made without analysis of the chemical signatures.

## Mathematical Techniques

Although the graphical techniques provide an easy means of increasing the level of insight obtained from ground water data, there are some instances where the



complexities of a site may limit the utility of these methods. For example, ground water from a site with multiple spills and overlapping contaminant plumes might not have a characteristic fingerprint of any one of the individual sources. Sorption or biological transformations of the NAPL may also confound interpretation of graphical representations of the data. In many of these complex situations, techniques that rely on more quantitative methods can be employed to analyze correlations among chemical concentration data.

Factor analysis (FA) is a multivariate tool that can be used to identify recurring patterns in data. With this method, correlated species are grouped into a few "factors" that describe most of the variability among the samples, effectively reducing the number of variables that need to be considered (Davis 1986). Each factor could include several related chemical species that would ideally represent a unique pollutant source. Unlike the graphical methods that identify the source by comparing environmental concentrations to known chemical signatures, factor analysis designates a number of significant factors and the variables contributing to each of those factors based on mathematical manipulation of the data. The scientist/engineer must then interpret these factors to identify the type of pollutant source or transformation processes associated with each. This assessment is based on an understanding of the physical and chemical processes in the environment and the nature of the possible pollutant sources. Therefore, knowledge of chemical characteristics and the history of spills at a site is required.

Application of factor analysis and other receptor modeling techniques in environmental site assessments has increased in the recent past. Early use of these techniques included the determination of the source of chemical constituents in ground water from basic geochemical parameters (e.g., Dalton and Upchurch 1978; Lawrence and Upchurch 1982; Usunoff and Guzmán-Guzmán 1989). Others have used various forms of factor analysis to differentiate between natural and anthropogenic sources with similar chemical compositions (Barker et al. 1988; Olmez et al. 1994). The application of these multivariate techniques to sites with soil or sediments polluted by a complex array of industrial sources has only recently been considered (Baum 1993; Wenning et al. 1993; Shine et al. 1995).

Although multivariate mathematical methods have rarely been applied to NAPL spill sites, there is ample evidence from related studies to suggest that these techniques are potentially useful. For example, Baum (1993) identified potential NAPL source locations at a Superfund site contaminated with several chlorinated solvents by using nonlinear multivariate methods to minimize the errors associated with physical, chemical, and biological processes.

#### **Application of Principal Components Analysis**

Data from the former MGP site considered in this work were analyzed with a type of factor analysis termed principal components analysis (PCA). This type

of analysis can provide additional insight, especially in situations where patterns in the chemical fingerprints are not known a priori or when graphical techniques produce ambiguous results. PCA is readily accessible through statistical software packages written for personal computers (e.g., SYSTAT, Statgraphics). As with other forms of factor analysis, principal components (loosely termed "factors") are identified that group chemical species which are correlated. These principal components are the eigenvectors of the variance-covariance or correlation matrix (Davis 1986).

The 26 ground water samples from 20 wells with detectable concentrations of at least a few ground water solutes were considered in the PCA. Since BTEX and PAHs were the only species common in all four rounds of sampling, the PCA was initially limited to these compounds. Ten chemical species, including BTEX and six PAH compounds, were detected in more than half of the samples. A log-transformation of these data was used to generate normally distributed data as suggested for PCA implementation (Davis 1986).

As with the mathematical or statistical analysis of any ground water data, some method of dealing with concentrations below detection limits was required. Two approaches were used in the analysis presented here to provide a numerical value for samples with nondetectable concentrations: (1) a concentration of one-half of the reported detection limit was assigned; or (2) for each nondetectable chemical species in each sample, a random number from a uniform distribution of values between 0.0 and 1.0 was generated and multiplied by the detection limit (Hopke 1991). Although the factor loadings were somewhat affected by the value used for nondetectable concentrations, the conclusions drawn from the PCA were not sensitive to this choice. For all results presented in this paper, nondetectable concentrations were assigned a value of one-half of their detection limit.

SYSTAT for Windows (version 5.0) was used to calculate the correlation matrix and its eigenvectors for the ground water data. The Varimax rotation option within this statistical package was used to maximize interpretation of the data (Wilkinson et al. 1992). By rotating the orthogonal factors so that they are aligned with the primary source of variability among the samples, it is often easier to interpret the physical or chemical source of variability contributing to each of the principal components (Roscoe et al. 1982). This rotation is analogous to the rotation of a coordinate system in a ground water flow problem so that the axes are aligned with the principal directions of anisotropy (e.g., Bear 1979).

#### **Interpretation of Results from PCA**

Table 2 presents the eigenvalues and component loadings for four of the 10 factors identified by the PCA. Evaluation of the magnitude of each eigenvalue is one method to determine which of these factors are significant. The first two factors presented in Table 2 are clearly significant. They both have eigenvalues greater than 1.0 and jointly explain nearly 70 percent of

**Table 2**  
**Results from Initial Application of PCA to the**  
**Former MGP Site Data**

	Factor 1	Factor 2	Factor 3	Factor 4
Eigenvalues	5.2	3.1	0.90	0.31
Component Loadings*:				
Benzene	0.13	<b>0.91</b>	-0.13	0.11
Toluene	0.09	<b>0.98</b>	0.02	-0.03
Ethyl Benzene	0.02	<b>0.88</b>	0.02	-0.48
Xylenes	0.18	<b>0.93</b>	0.20	0.11
Naphthalene	<b>0.86</b>	0.33	0.23	0.03
Acenaphthene	<b>0.91</b>	-0.06	0.33	0.01
Acenaphthalene	<b>0.95</b>	0.15	0.22	0.00
Fluorene	<b>0.71</b>	0.16	<b>0.51</b>	-0.08
Phenanthrene	0.48	0.02	<b>0.85</b>	-0.00
Anthracene	0.26	0.02	<b>0.96</b>	0.00
% Total Variance Explained	33.3	35.8	21.7	2.5
Probable Source	coal tar	gasoline	sampling or analytical errors	none

\*Chemical species considered significant (loading > 0.5) in each factor are **highlighted**.

the total variance among sample concentrations. The third factor also appears to be significant. Its eigenvalue is 0.90 and the relatively high fraction of the total variance explained by this factor (21.7 percent) suggests that it should be considered further.

The component loadings included in Table 2 are essentially correlation coefficients between the chemical species of concern and the factor (Olmez et al. 1994). Thus, component loadings of 1.0 indicate a perfect correlation between the species and factor. In other studies, component loading values greater than 0.70 (Roscoe et al. 1982), 0.50 (Kerndorff et al. 1992), and 0.25 (Olmez et al. 1994) have been used to identify significant species within a factor. In this study, chemical species with loadings greater than 0.5 were considered significant to a particular factor. Hence, the presence and concentrations of many of the PAH compounds are closely correlated as evidenced by their high factor loadings associated with Factors 1 and 3 (Table 2). The monoaromatic species (BTEX) are also correlated and combined in Factor 2. These types of factors were expected based on knowledge of the site history and chemical composition of the NAPLs present at this site.

The last step in the PCA involves the interpretation of factor scores calculated by the SYSTAT software. Factor scores for each of the samples indicate the importance of a given factor to that sample and can be used to cluster samples that are influenced by similar factors or combinations of factors (Davis 1986). The log-transformed data used in this analysis generated normally distributed factor scores with a mean of 0.0 and a range of about  $\pm 3$  standard deviations. Thus, samples with positive factor scores are correlated to the particular factor, and factor scores greater than about +1 reflect an intense response to the processes or source associated with that factor (Dalton and Upchurch 1978).

Review of factor loadings and factor scores was used to further evaluate the significance of the third factor.

Phenanthrene, anthracene, and fluorene have high loadings in the third factor (Table 2), and samples from Wells 12b, 13b, 2, and 6 had scores for Factor 3 greater than 1.0. Coal tar was found in Wells 12b and 13b, and the analysis of star plots also indicated a high probability of coal tar near Well 2. From this information and the high scores for Factor 3 in these wells, it could be concluded that this coal tar was of different composition than in wells farther to the south. Further review of the factor scores, however, show that the elevated Factor 3 scores for Samples 12b, 13b, and 2 were only calculated for the third sampling round. Careful analysis of the raw data revealed that, in fact, almost all samples collected during this phase of the investigation had between three and 15 times higher concentrations of anthracene and phenanthrene than samples collected from the same wells at other times. Given that the DNAPL entered the subsurface decades ago and that the ground water processes are typically slow, it is unlikely that these major changes in concentration can be attributed to physical, chemical, or biological processes in the subsurface. Therefore, it was concluded that the third factor was probably the result of variability in sample analysis and that this factor should not be considered further.

With the conclusion that the third factor was not related to NAPL composition or distribution, the PCA was repeated with the exclusion of anthracene and phenanthrene, and was limited to two factors. This analysis indicated that there are two distinct factors associated with BTEX species (Factor 1) and the remaining PAH species (Factor 2). These two factors account for 91 percent of the variance among the log-transformed concentrations. The inclusion of eight species in the PCA vs. four variables with the star plots increases the possibility of differentiating sources based on PAH compositions.

By plotting scores for Factor 1 against Factor 2, samples with similar chemical compositions can be grouped by their location on the plot indicating they have similar influences from each of the factors. Factor scores are shown in Figure 5 with three groupings identified. Samples outside these grouping generally show low scores for both factors indicating a lower degree of correlation between the chemical signatures of these wells and those in the immediate vicinity of the three predominant NAPL sources.

The first ("a" group) has high values for Factor 1 (high in BTEX species) and includes most of the shallow wells contaminated primarily by gasoline. The factor scores for these samples cover a considerable range of Factor 2 scores, with almost no influence of this factor in the two most northerly wells (10 and 11). This low influence of PAHs on the ground water sample collected from Well 11 adds confidence to the conclusion that there is minimal coal tar contamination of the aquifer in the vicinity of the nearby tar well (Well 7).

The second distinct grouping ("b" group) has elevated scores for both factors and includes deep Wells 15b and 19. As described previously, a DNAPL was



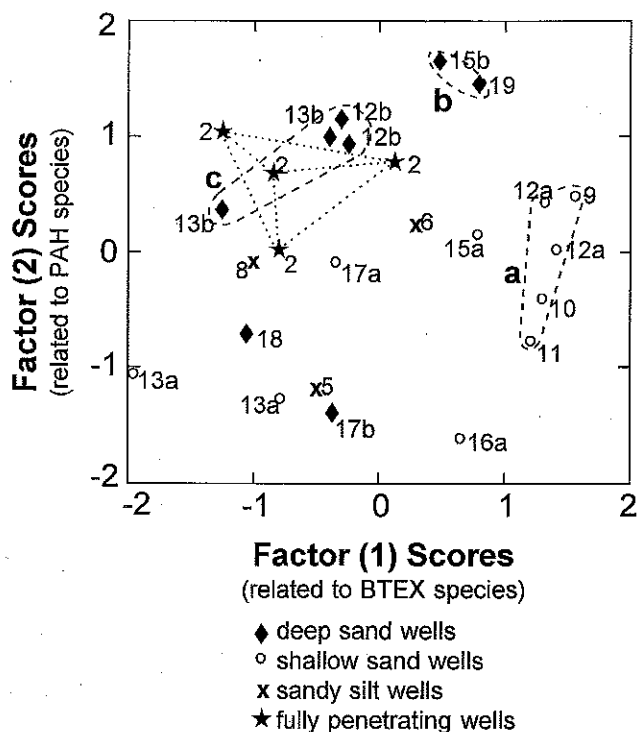


Figure 5: Factor scores are plotted for each ground water sample. Clusters of wells associated with the different pollutant sources are identified: (a) gasoline; (b) carburetted water gas tar (or fraction thereof); and (c) heavy coal tar. Note that in the final PCA, the factors associated with BTEX vs. PAHs are switched relative to those included in Table 2.

detected in both of these wells. Infrared analysis of the coal tar from Well 19 showed a similar composition to that collected from the southernmost gas holder. Results of the PCA, coupled with similarities in the star plots for these two wells and knowledge of the elevations of the bottom of the aquifer, can be used to conclude that the coal tar at both of these wells is a carburetted water gas tar from the nearby gas holder.

A third grouping ("c" group), including Wells 12b and 13b, and one of the Well 2 samples, is also identified in Figure 5. The spread in the area of this grouping illustrates the impact of variability in concentrations between sampling rounds. This group was drawn to include both Well 13b samples since both of these are clearly influenced by the same coal tar source. Concentrations for all analytes used in the PCA in the Well 13b sample during the third round of sampling were significantly less than during the fourth sampling round. With coal tar DNAPL at the base of this well, however, there are no chemical or physical explanations for a substantial drop in ground water concentrations. This trend in variability between the third and fourth sampling rounds was also observed between samples from a few other wells, but not for all wells sampled at both times. These inconsistencies suggest possible variability in sample collection techniques.

The range of factor scores calculated for Well 2 are also identified in Figure 5. There is significant variation in the factor scores for the four samples from this well.

Yet, based on the location of these samples on the factor score plot, it can be concluded that the chemical signatures of ground water samples from this well are much closer to that of Wells 12b and 13b than Wells 15b and 19. These results increase our confidence in the conclusion that Well 2 is not influenced by the carburetted water gas tar and that the actual areal extent of the heavier coal tar contamination extends from Well 12b to close to Well 2.

## Summary and Conclusions

In this work, it was proposed that a thorough analysis of the chemical patterns imbedded in data from ground water sampling efforts can improve our overall understanding of the nature and extent of subsurface NAPL sources, thereby increasing the effectiveness of any attempts to minimize risk associated with the site. Multivariate plots and principal components analyses were used to identify and interpret chemical signatures or patterns in the composition of ground water data from an MGP site in Iowa. Both techniques substantiated field evidence that there are three distinct NAPL sources at this site. Through analysis of the chemical signatures associated with ground water samples, however, greater confidence in the interpretation of the extent of each NAPL source was gained. Figure 6 identifies the wells with common chemical signatures and the suspected original NAPL sources. In comparison with the interpretation of NAPL contamination following more conventional approaches (Figure 1), the

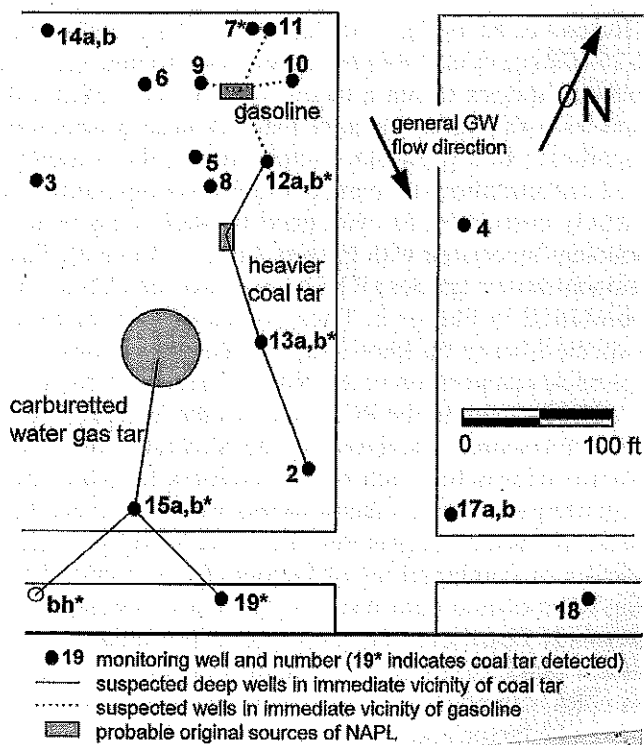


Figure 6: Revised interpretation of probable coal tar source areas resulting from graphical and mathematical multivariate analyses shows a greater number of wells affected by the heavier coal tar DNAPL relative to results of the more traditional interpretation (Figure 1).

added benefit of multivariate analyses of ground water samples includes the interpretations that:

- Coal tar contained in the tar well at Well 7 has not migrated to the aquifer
- The heavier coal tar has migrated to within a close proximity of Well 2
- The heavier coal tar at Wells 13b, 12b, and close to Well 2 probably leaked from the tar tanks south of Well 12b, not the tar well to the northwest nor the gas holder west of Well 2.

The work presented here illustrates that multivariate analyses can be used as an additional tool in an overall site assessment to increase our confidence and understanding regarding the type and location of DNAPL sources. Although the multivariate techniques do not provide a quantitative assessment of the predominant mechanisms affecting the spatial and temporal distribution of ground water contaminants, chemical patterns emerge which clearly show differences between the source and downgradient areas. The conclusions reached from both techniques used at this site were the same. The incorporation of additional PAH species in the PCA, however, provided even greater confidence that critical differences among the chemical fingerprints were identified.

Application of PCA to this data set also helped to reveal inconsistencies in measured concentrations between sampling rounds. The fact that there was significant variability in these data was evident from viewing the raw data. Through the PCA, however, specific analytes with the most significant variations were identified. As with any statistical or mathematical interpretation of environmental data, the quality of the inferences drawn from application of these methods are directly correlated to the quality of data available.

The example used here was meant to illustrate some of the capabilities of multivariate data analysis techniques. Application of these techniques could also be used to ensure optimal placement of additional wells or borings during the iterative process of designing sampling efforts, or could reduce the need for disturbing, and possibly increasing, the extent of DNAPL contamination during drilling. At sites without direct physical evidence of NAPL contamination, the use of multivariate plots could identify ground water samples with chemical signatures indicative of a nearby NAPL source.

## Acknowledgments

The authors wish to express their sincere gratitude to David Pluhar, Union Electric Co., for providing the MGP site data used in this paper. The useful suggestions of two reviewers of this manuscript are also gratefully acknowledged.

## References

- Atlantic. 1991. Remedial investigation report. Project # 1444-01-02. Colchester, Connecticut: Atlantic Environmental Services Inc.
- Anderson, K.S., A.A. Riley, and R.I. Thun. 1992. Analysis of multivariate plots as one tool in the delineation of hydrocarbon plumes. In *Proceedings: 1992 Petroleum Hydrocarbons and Organic Chemicals in Ground Water: Prevention, Detection, and Restoration*, 573-582, by National Ground Water Association. Dublin, Ohio: NGWA.
- API. 1985. Laboratory study of solubilities of petroleum hydrocarbons in ground water. API Publication No. 4395. Washington, D.C.: American Petroleum Institute.
- API. 1989. A guide to the assessment and remediation of underground petroleum releases. API Publication No. 1628. Washington, D.C.: American Petroleum Institute.
- Barker, J.F., J.E. Barbash, and L. Labonte. 1988. Ground water contamination at a landfill sited on fractured carbonate and shale. *J. Cont. Hydrol.* 3, no. 1: 1-25.
- Baum, E.J. 1993. Receptor models to study ground water contamination. *Anal. Chim. Acta* 277, no. 4: 421-429.
- Bear, J. 1979. *Hydraulics of groundwater*. New York: McGraw-Hill.
- Bruce, L., T. Miller, and B. Hockman. 1991. Solubility versus equilibrium saturation of gasoline compounds: A method to estimate fuel/water partition coefficient using solubility or *k<sub>oc</sub>*. In *Proceedings: 1991 Conference on Petroleum Hydrocarbons and Organic Chemicals in Ground Water: Prevention, Detection, and Restoration*, 571-582, by National Ground Water Association. Dublin, Ohio: NGWA.
- Cohen, R.M., J.W. Mercer, and J. Matthews. 1993. *DNAPL site evaluation*. C.K. Boca Raton, Florida: C.K. Smoley Publishing.
- Dalton, M.G., and S.B. Upchurch. 1978. Interpretation of hydrochemical facies by factor analysis. *Ground Water* 16, no. 4: 228-233.
- Davis, J.C. 1986. *Statistics and data analysis in geology*, 2nd ed. New York: John Wiley & Sons.
- Feenstra, S. 1992. Evaluation of multi-component DNAPL sources by monitoring of dissolved-phase concentrations. In *Subsurface contamination by immiscible fluids*, ed. K.U. Weyer, 65-72. Rotterdam: A.A. Balkema.
- GRI. 1987. *Management of manufactured gas plant sites*, Volume III. Report #GRI-87/0260.3. Chicago, Illinois: Gas Research Institute.
- Goerlitz, D.F. 1992. A review of studies of contaminated ground water conducted by the USGS organics project. In *Ground water contamination and analysis at hazardous waste sites*, ed. S. Lesage and R.E. Jackson, 295-355. New York: Marcel Dekker.
- Heath, J.S., K. Koblis, and S.L. Sager. 1993. Review of chemical, physical, and toxicological properties of components of total petroleum hydrocarbons. *J. Soil Contam.* 2, no. 1: 1-25.
- Hinchee, R.E., and H.J. Reisinger. 1987. A practical application of multiphase transport theory to ground water contamination problems. *Ground Water Monitoring Review* 7, no. 1: 84-92.
- Hopke, P.K. (ed.). 1991. *Receptor modeling for air quality management*. Amsterdam: Elsevier.

- Kerndorff, H., R. Schleyer, G. Milde, and R.H. Plumb, Jr. 1992. Geochemistry of ground water pollutants at German waste disposal sites. In *Ground water contamination and analysis at hazardous waste sites*, ed. S. Lesage and R.E. Jackson, 245-271. New York: Marcel Dekker.
- Kueper, B.H., D. Redman, R.C. Starr, S. Reitsma, and M. Mah. 1993. A field experiment to study the behavior of tetrachloroethylene below the water table: Spatial distribution of residual and pooled DNAPL. *Ground Water* 31, no. 5: 756-766.
- Lawrence, F.W., and S.B. Upchurch. 1982. Identification of recharge areas using geochemical factor analysis. *Ground Water* 20, no. 6: 680-687.
- Lesage, S., and P.A. Lapcevic. 1990. Differentiation of the origins of BTX in ground water using multivariate plots. *Ground Water Monitoring Review* 10, no. 2: 102-105.
- Luhrs, R.C., and C.J. Pyott. 1992. Trilinear plots: A powerful new application for mapping gasoline contamination. In *Proceedings of the 1992 Conference on Petroleum Hydrocarbons and Organic Chemicals in Ground Water: Prevention, Detection, Restoration*, 85-99, by National Ground Water Association. Dublin, Ohio: NGWA.
- Mackay, D.M., and J.A. Cherry. 1989. Groundwater contamination: Pump-and-treat remediation. *Environ. Sci. Technol.* 23, no. 6: 630-636.
- Novak, S.A., and Y. Eckstein. 1988. Hydrochemical characterization of brines and identification of brine contaminated aquifers. *Ground Water* 26, no. 3: 317-324.
- Olmez, I., J.W. Beal, and J.F. Villaume. 1994. A new approach to understanding multiple-source ground water contamination: Factor analysis and chemical mass balances. *Water Research* 28, no. 5: 1095-1101.
- Peters, C.A., and R.G. Luthy. 1993. Coal tar dissolution in water-miscible solvents: Experimental evaluations. *Environ. Sci. Technol.* 27, no. 13: 2831-2843.
- Roscoe, B.A., P.K. Hopke, S.L. Dattner, J.M. Jenks. 1982. The use of principal components factor analysis to interpret particle compositional data sets. *APCA Journal* 32, no. 6: 637-642.
- Rosenfeld, J.K., and R.H. Plumb Jr. 1991. Ground water contamination at wood treatment facilities. *Ground Water Monitoring Review* 11, no. 4: 133-140.
- Shine, J.P., R.V. Ika, and T.E. Ford. 1995. Multivariate statistical examination of spatial and temporal patterns of heavy metal contamination in New Bedford Harbor marine sediments. *Environ. Sci. Technol.* 29, no. 7: 1781-1788.
- Shiu, W.Y., M. Bobra, A.M. Bobra, A. Maijanen, L. Suntio, and D. Mackay. 1990. The water solubility of crude oils and petroleum products. *Oil & Chem. Poll.* 7, no. 1: 57-84.
- U.S. EPA. 1993. Evaluation of the likelihood of DNAPL presence at NPL sites. EPA/540/R-93/073. Washington, D.C.: Office of Solid Waste and Emergency Response.
- Usunoff, E.J., and A. Guzmán-Guzmán. 1989. Multivariate analysis in hydrochemistry: An example of the use of factor and corresponding analyses. *Ground Water* 27, no. 1: 27-34.
- Wenning, R., D. Paustenbach, G. Johnson, R. Ehrlich, M. Harris, and H. Bedbury. 1993. Chemometric analysis of potential sources of polychlorinated dibenzo-P-dioxins and dibenzofurans in surficial sediments from Newark Bay, New Jersey. *Chemosphere* 27, nos. 1-3: 55-64.
- Wilkinson, L., M.A. Hill, J.P. Welna, G. K. Birkenbeuel. 1992. *SYSTAT for Windows: Statistics*, Version 5 edition. Evanston, Illinois: SPSS Inc.
- Wilson, B.H., J.T. Wilson, D.H. Kambell, B.E. Bledsoe, and J.A. Armstrong. 1991. Biotransformation of monoaromatic and chlorinated hydrocarbons at an aviation gasoline spill site. *Geomicrobiology Journal* 8, no. 2: 225-240.
- Wilson, B.H., J.T. Wilson, D.H. Kambell, B.E. Bledsoe, and J.A. Armstrong. 1994. Traverse City: Geochemistry and intrinsic bioremediation of BTX compounds. In *Proceedings: Symposium on Intrinsic Bioremediation of Ground Water*, 94-102. Washington, D.C.: U.S. EPA Office of Research and Development.

## Biographical Sketches

**Susan E. Powers** has been an assistant professor in the Department of Civil and Environmental Engineering at Clarkson University (Potsdam, NY 13699-5710; (315) 268-6542; (315) 268-7636 (fax); sep@draco.clarkson.edu) since receiving a doctorate in environmental engineering from The University of Michigan, Ann Arbor, in 1992. Her research focuses on many aspects of multiphase flow and the fate and transport of contaminants in porous media. She is a registered professional engineer and is a member of NGWA, as well as numerous other professional societies.

**James F. Villaume** is senior scientist within the Environmental Management Division of Pennsylvania Power and Light Co. (Two N. Ninth St., Allentown, PA 18101-1179). He holds a B.A. in geology from Franklin and Marsh College and an M.S. in geochemistry from The Pennsylvania State University. He is a registered professional geologist and a member of the American Institute of Professional Geologists and the American Institute of Engineering Geologists.

**John A. Ripp** is a vice president of Atlantic Environmental, a division of GEI Consultants Inc. (GEI/Atlantic Environmental Services Inc., P.O. Box 297, Colchester, CT 06415). His 15 years of service to the electric and gas utility industry have included applied research, technology assessment, litigation support, and management of MGP site restoration projects. He started his career with a B.A. in physics from Holy Cross and an M.S. in ocean engineering from Northeastern University.