

Molecular tracers of soot and sewage contamination in streams supplying New York City drinking water

Anthony K. Aufdenkampe¹, David B. Arscott², Charles L. Dow³, AND
Laurel J. Standley⁴

Stroud Water Research Center, 907 Spencer Road, Avondale, Pennsylvania 19311 USA

Abstract. A molecular tracer method was used to assess the extent and sources of pollution to 60 stream sites that were distributed across the watersheds that supply drinking water to the greater New York City area. Samples were collected from each site annually from 2000 to 2002 during summer baseflow conditions. Twelve polycyclic aromatic hydrocarbons (PAH), 2 fragrance materials (FM), caffeine (CAF), and 7 fecal steroids (FS) were measured using a modification of EPA method 8270, which quantified concentrations to laboratory reporting levels ranging from 0.00009 to 0.016 $\mu\text{g/L}$ or 3 to 5 orders of magnitude lower than method detection levels (MDL) given by EPA 8270. In 54 of 180 stream samples, concentrations of ≥ 1 PAHs exceeded suggested, nonregulatory EPA guidance values for water supplies (0.0038 $\mu\text{g/L}$ for the 5 most toxic PAHs), and PAH signatures (ratios) and spatial patterns suggested that soot from local urban/suburban combustion was the primary source. CAF, FM, and FS all showed their highest concentrations at the 3 sites with large, failing sewage treatment plants, but more complex relationships to landscape variables at remaining sites suggested a variety of anthropogenic point and nonpoint sources. Concentrations of all molecular tracers measured were strongly negatively correlated with % forest cover (= all forest variables used) in the watershed.

Key words: water-quality, fecal steroids, caffeine, polycyclic aromatic hydrocarbons, Catskills, land use, spatial distribution.

Degradation of water quality can occur from a variety of point and nonpoint sources originating from both anthropogenic and natural factors including industrial effluent, sewage from waste water treatment plants (WWTP) or septic leakage, road and agricultural runoff, atmospheric deposition, and wildlife. The range of contaminants includes excessive nutrient loading, metals, pesticides, other toxic organic compounds, and pathogens. To best maintain the quality of drinking-water resources, targeted efforts to reduce or eliminate primary contamination sources first require accurate identification and quantification of all contaminant sources that contribute to the degradation of water quality.

The use of molecular tracers to identify sources of

¹ E-mail address: aufdenkampe@stroudcenter.org

² Present address: National Institute of Water and Atmospheric Research, P.O. Box 8602 Christchurch, New Zealand. E-mail: d.arscott@niwa.co.nz

³ E-mail address: cdow@stroudcenter.org

⁴ Present address: Silent Spring Institute, 29 Crafts Street, Newton, Massachusetts 02458 USA. E-mail: standley@silentspring.org

contaminants is an emerging technique that qualitatively links chemical fingerprints unique to these sources with contaminants of concern (Leeming et al. 1996, Standley et al. 2000, Kolpin et al. 2002, Yunker et al. 2002, Buerge et al. 2003, Glassmeyer et al. 2005). These tracer compounds do not themselves have to be toxic or contribute directly to water-quality degradation. Rather they need only to enable discrimination between different sources and, therefore, act as proxies for contaminants originating from those same sources. For example, a recent and increasingly used proxy to detect potential sewage contamination is caffeine. Caffeine is not considered toxic to humans or aquatic life at any measured environmental concentration. However, in temperate latitudes the only source of caffeine to surface waters is from the urine of consumers of caffeinated beverages and pharmaceuticals (Buerge et al. 2003). Therefore, high aquatic concentrations of caffeine are a strong indicator of sewage or septic contamination (Standley et al. 2000, Buerge et al. 2003, Vogel et al. 2005).

The practice of using source-specific organic compounds, or biomarkers, to estimate source contribu-

tions has been well developed by the field of organic geochemistry. It was used originally to identify natural organic matter sources to petroleum formation and more recently to elucidate a wide range of processes in the cycling of natural organic C (Eglinton 1969, Hedges and Prahl 1993). Organic geochemists have long recognized that a compound must meet a certain set of criteria to be useful as a biomarker or as a tracer of sources (Hedges and Prahl 1993). The ability for quantitative interpretation of tracer data increases as each of the following criteria are met: 1) the tracer must be detectable at a concentration well below that of interest, 2) ambient concentrations of the tracer molecule must be accurately quantified, 3) all sources of the tracer are known and relatively unique, and 4) environmental diagenesis or degradation of the tracer compound is either minimal, well understood, or proportional to other tracer compounds to which it might be compared (e.g., ratios do not change with degradation).

Enhanced monitoring projects of ambient water quality could benefit substantially by adopting the biomarker approach of organic geochemistry. However, most organic contaminant monitoring efforts, even in watersheds with many unknown sources, focus on regulated compounds and pay little attention to proxies for source identification. Furthermore, the most widely used EPA methods (i.e., 625 or 8270; USEPA 1996 and 1998, respectively) have method detection limits (MDL; generally 1–100 $\mu\text{g/L}$) that are often orders of magnitude higher than EPA water-quality criteria for those same compounds (USEPA 2002b). Last, most recent studies using more sensitive methods do not quantify ambient concentrations by correcting for known analytical biases, such as extraction recovery (Kolpin et al. 2002, Glassmeyer et al. 2005). Extraction recovery and many other analytical biases vary by sample, so failure to quantify and apply sample-specific recovery corrections introduces unnecessary analytical imprecision and inaccuracy.

The goal of our paper is to introduce a contaminant-monitoring study designed using the biomarker approach of organic geochemistry. We present a modified method that quantifies 12 polycyclic aromatic hydrocarbons (PAH), 2 fragrance materials (FM), caffeine (CAF), and 7 fecal steroids (FS) to laboratory reporting levels (LRL) ranging from 0.00009 to 0.016 $\mu\text{g/L}$ (Table 1). For most of these compounds, this method was the most sensitive published method of which we were aware. In addition, ambient stream-water concentrations were more accurately quantified by correcting for analytical biases for each sample using a suite of internal surrogate recovery standards. Concentration distributions of these compounds mea-

sured at 60 stream sites for 3 consecutive summers (2000–2002) are described as part of a multidisciplinary, large-scale enhanced water-quality monitoring project (the Project) in New York City (NYC) drinking-water-supply watersheds (Blaine et al. 2006). Potential contamination sources to waters upstream of study sites are inferred from patterns in tracer compounds. Last, mean tracer concentrations measured at these sites are related to watershed landscape characteristics.

Methods

Field

NYC draws drinking water from reservoirs capturing stream water from 2 primary regions: east of Hudson River (EOH) and west of Hudson River (WOH) (Blaine et al. 2006). Thirty stream sites were visited in each region once annually from 2000 to 2002 during summer baseflow to collect water samples for quantification of various chemical constituents (e.g., nutrients, major ions, suspended solids, dissolved organic C) as described by Dow et al. (2006) and Kaplan et al. (2006). Characteristics of sites are described in detail by Arscott et al. (2006).

Water samples were collected for molecular tracer analyses in 4-L glass jugs concurrently with summer baseflow sampling (see appendix in Arscott et al. 2006) for basic chemistry (except for some WOH sites in 2000 where tracer and basic chemistry samples were not collected on the same day). Field blanks and duplicates were each collected at 3 sites during each year of summer baseflow sampling (5% of sites). All glass sampling equipment and sample jugs were washed with detergent, rinsed with nanopure water, and finally combusted in a kiln at 480°C for 4 h to remove all remaining organic compounds. Metal and teflon sampling equipment (including jug cap liners), which could not be combusted in a kiln, were cleaned with detergent and nanopure water, dried, and finally rinsed with hexane/acetone (1:1) followed by dichloromethane, as was any field equipment that was reused between sites.

Laboratory

Molecular tracers were extracted by liquid–solid extraction onto EmporeTM C-18 disks and analyzed by capillary gas chromatography–mass spectrometry (GC–MS) using modified protocols based on EPA Methods 3535 and 8270 (USEPA 1998). An earlier version of this method was presented by Standley et al. (2000). In brief, surrogate recovery standards (perdeuterated phenanthrene [PHE-D10], perdeuter-

TABLE 1. Compounds chosen as molecular tracers of contamination, abbreviations used in text and figures, their Laboratory Reporting Levels (LRL; equivalent to 99% confidence levels for no false negatives), and their 75% and 95% confidence levels (CL) for no false positives (derived from distributions of measured blanks). Ambient water-quality criteria are listed for reference. PAH = polycyclic aromatic hydrocarbon, EPA = US Environmental Protection Agency, H(WS) = source of drinking water (water supply), H(FC) = human consumption of fish, A(C) = fish propagation-aquatic life (chronic), A(A) = fish survival-aquatic life (acute). Blank indicates no criterion available.

Analyte	Abbreviation	LRL ($\mu\text{g/L}$)	75% CL ($\mu\text{g/L}$)	95% CL ($\mu\text{g/L}$)	New York ambient water-quality guidance values ($\mu\text{g/L}$) ^a					EPA human health guidelines for consumption of: ($\mu\text{g/L}$) ^b
					H(WS)	H(FC)	A(C)	A(A)	Water + organism	
PAH										
Fluorene	FLU	0.00046	0.0006	0.0023	50		1	5	1100	5300
Phenanthrene	PHE	0.00068	0.0029	0.0067	50		5	45		
Anthracene	ANT	0.00033	0.001	0.0033	50		3.8	35	8300	40000
2-methyl phenanthrene	2MP	0.0013	0.0011	0.0035						
1-methyl phenanthrene	1MP	0.00090	0.0005	0.0022						
Fluoranthene	FLR	0.00042	0.0028	0.0099	50				130	140
Pyrene	PYR	0.00045	0.00081	0.0078	50		5	42	830	4000
Benz(a)anthracene	BAA	0.00068	0.00064	0.0012	0.002		0.03	0.23	0.0038	0.018
Chrysene	CHR	0.00033	0.00073	0.0011	0.002				0.0038	0.018
Benzo(b)fluoranthene	BBF	0.0003	0.00072	0.014	0.002				0.0038	0.018
Benzo(k)fluoranthene	BKF	0.00084	0.00048	0.0015	0.002				0.0038	0.018
Benzo(a)pyrene	BAP	0.00014	0.00055	0.0025	0.002				0.0038	0.018
Fragrances and caffeine										
Tonalide	HHCB	0.0037	0.0047	0.012						
Galaxolide	AHTN	0.0050	0.011	0.016						
Caffeine	CAF	0.0023	0.0023	0.016						
Steroids										
Coprostanol (5 β -cholestan-3 β -ol)	bCOP	0.000087	0.0016	0.022	0.3 ^c		0.3 ^c		0.3 ^c	
Epi-coprostanol (5 β -cholestan-3 α -ol)	EPI	0.00037	0.0026	0.026						
Cholesterol (cholest-5-en-3 β -ol)	CHOL	0.010	0.024	0.034						
Cholestanol (5 α -cholestan-3 β -ol)	aCOP	0.0017	0.0027	0.032						
Cholestanone (5 α -cholestan-3-one)	aONE	0.00085	0.00074	0.011						
Coprostanone (5 β -cholestan-3-one)	bONE	0.0085	0.0031	0.038						
Ethyl-cholestanol (24-ethyl-5 α -cholestan-3 β -ol)	SNOL	0.016	0.008	0.015						

^a NYS DEC (1998)

^b USEPA (2002a)

^c Based on a limit of 200 fecal coliforms/100 mL and the upper 95% confidence limit for the relationship between fecal coliforms and coprostanol as given by Leeming and Nichols (1996)

ated chrysene [CHR-D12], perdeuterated perylene [PER-D12], and ^{13}C -labeled cholesterol [CHO- $^{13}\text{C}_2$], all codissolved in water-miscible acetonitrile) were added to each water sample and mixed. The exact volume of the 4-L water sample was assessed gravimetrically prior to filtering through a 90-mm-diameter glass-fiber filter (precombusted GF/F) stacked on top of a 90-mm-diameter EmporeTM C-18 disk. Particulate tracer compounds were extracted from the dried filter by sonic extraction with dichloromethane (DCM), and dissolved tracers were eluted from the C-18 disk with a solvent series of acetone, DCM, and hexane. Dissolved and particulate extracts were then combined and back-extracted in a separatory funnel with a pre-extracted aqueous salt solution to remove ionic and highly polar impurities, mixed with anhydrous sodium sulfate to remove moisture, rotoevaporated to 1 mL, and transferred to auto-injector vials. Samples were gently dried under a stream of N_2 until just barely dried, immediately sealed and purged with N_2 , redissolved in 50 μL of MSTFA (N-methyl-N-(trimethylsilyl)trifluoroacetamide) derivatizing reagent with 1% TMCS (trimethylchlorosilane), spiked with 2.5 μL of 2-methylnaphthalene (2MN) internal standard, and silylated by heating to 70°C for 30 min in an aluminum heating block. These derivitized sample extracts were analyzed for each of the molecular tracer compounds by GC-MS in selective ion monitoring (SIM) mode, using a DB1701 column (30 m, 0.25-mm inner diameter, 0.25- μm film thickness; J&W Scientific/Agilent, Palo Alto, California) on an Agilent 6890 series GC interfaced with a 5973n series mass selective detector (MSD) using electron impact ionization. All sample-handling equipment and reagents were pre-cleaned by either combusting at 480°C for 4 h or exhaustive solvent rinsing.

Laboratory blanks and duplicates, laboratory-spiked matrix samples, and standard reference materials (SRM) were extracted and analyzed in conjunction with all sites having field blanks and duplicates (5% of sites).

Quantification

Each batch of 12 to 15 samples was analyzed by GC-MS along with 7 to 8 analytical standard mixes: 5 calibration standards at 0.2, 1.0, 4.0, 20, and 50 ng/ μL nominal concentration (exact concentrations for each compound were slightly different, but known to 3 significant figures) and 2 to 3 check standards at 4.0 ng/ μL nominal concentration. To enable the greatest consistency, tracer concentrations were quantified using an automated data-quantification system (Dow

and Aufdenkampe 2006). In brief, after confirmation by the analyst, for each compound, the peak areas of 1 quantitation and 1 to 2 confirmation ions were exported for all standards and samples from the Agilent GC-MS ChemStation chromatography software directly into our central server. These raw data were then manipulated with SAS-based scripts (SAS/Base, version 9.1, SAS Institute, Cary, North Carolina) to produce final concentration data. Thus, decisions regarding how to fit the calibration curve, when to drop outlying standards, whether or not peak identity was adequately confirmed, etc., all were made uniformly for the entire 3-y data set using the same objective criteria. Additional benefits of this quantitation system included documentation of all calibration decisions, which could be easily reviewed or revised at any time, less potential for error, and better quality control. If a compound concentration was above the highest calibration standard, the sample extract was diluted and reanalyzed. If a compound concentration was below the lowest calibration standard, the compound was flagged as estimated but nevertheless quantified using a linear fit from the origin to the lowest standard. If any compound in a check standard did not give a concentration within 20% of the known value, all samples analyzed after that check standard were reanalyzed for that compound.

All data presented here were corrected for extraction recoveries and other analytical biases measured for each sample using internal surrogate standards, which were added to each sample prior to filtration and extraction. Surrogate standard recoveries were assigned to tracer compounds, based on recovery data from laboratory-spiked matrix samples for all compounds, as follows (see Table 1 for analyte abbreviations): perdeuterated phenanthrene (FLU, PHE, ANT), perdeuterated chrysene (FLR, PYR, BAA, CHR, HHCB, AHTN), perdeuterated perylene (BBF, BKF, BAP), and $^{13}\text{C}_2$ -cholesterol (FS). 2MP and 1MP were corrected using the average recovery of perdeuterated phenanthrene and perdeuterated chrysene, and CAF was corrected with the average recovery of perdeuterated phenanthrene (most similar molecular weight) and $^{13}\text{C}_2$ -cholesterol (most polar surrogate). These assignments were confirmed with laboratory-spiked test samples (i.e., known amounts of compounds added to clean water), by matching measured recoveries of each tracer with the surrogate having the most consistently similar recovery.

Laboratory reporting levels (LRL) were assigned to each analyte using the definitions and methods of the US Geological Survey (USGS 1999). In brief, the LRL is defined as the concentration above which there is 99% confidence that reporting a false negative will be

avoided. In other words, if the ambient concentration is above the LRL, the laboratory is 99% confident to detect a concentration. The LRL is equivalent to $2\times$ the MDL as defined by EPA in CFR Title 40, Part 136, Appendix B (USGS 1999). In brief, CFR 40–136-B states that the MDL is $3\times$ the standard deviation (99% confidence interval of the mean) of >7 replicates of an analytical standard spiked at the lowest reliably detectable and quantifiable concentration in a clean sample matrix. For our method, where contamination of blanks and MDL study samples did occur occasionally, outlier high-concentration MDL sample data were removed using Chauvenet's criterion (Taylor 1997) before calculating our LRL. This step insured that our calculated LRL remained true to the intended definition as the 99% confidence limit for no false negatives.

Data were not censored below estimated MDL or LRL values a priori for most of our statistical analyses, with the exception of values for ratios of ≥ 2 compounds. A number of studies and reports have examined the numerous negative effects of censoring data (Gilliom et al. 1984, Porter et al. 1988, Helsel 1990, USGS 1999). These rationales for minimizing data censorship are examined in detail in the Discussion. Ratios of 2 tracer compounds were censored, or eliminated from consideration, if one compound had a concentration below its censorship limit, which was either the LRL or the 75% percentile of measured blanks, whichever was greater. For the ratio of high- to low-molecular-weight PAHs (HMW/LMW_{PAH}), the value was eliminated if the sum of measured concentrations in the either the numerator or denominator was less than the sum of the censorship limits for the same compounds.

Statistical analyses

Prior to all statistical analyses, concentration data for all 22 tracer compounds were $\log_{10}(x)$ transformed (after adding $0.00003 \mu\text{g/L}$, a value just below the minimum detected concentration, to all values). Log-normal distributions were confirmed with a Shapiro–Wilk test, where normality was rejected at level of $p < 0.01$, which corresponded to a Shapiro–Wilk statistic (W) > 0.90 for $n = 30$. (Pearson correlation, regression, and analysis of variance do not require strict conformation to normality [SAS/STAT, version 9.1, SAS Institute, Cary, North Carolina]). For the EOH data set ($n = 30$), 16 of 22 compounds showed log-normal distributions (the exceptions being 2 FMs and 4 FS). Removal of 3 obvious outlier sites, all with large sewage treatment plants upstream (sites 43, 49, and 58), improved the number of log-normal distributions

to 20 of 22 (the exceptions being bONE and aONE). This reduced EOH data set ($n = 27$) was commonly used for a number of statistical analyses. For the WOH data set, 21 of 22 tracers showed log-normal distributions (the exception being HHCB). In general, tracer ratio data did not require log transformation, with the exception of HMW/LMW_{PAH} .

Principal Components Analysis (PCA).—PCA was used to assess interannual variability and consistencies in spatial typology in molecular tracer compounds. Specifically, within- and between-year PCAs (ADE-4 software; Thioulouse et al. 1997) were used to partition and quantify within-year variance (i.e., among sites within a year) and between-year variance (i.e., among years within sites) for molecular tracer concentrations measured at both WOH and EOH sites (separate analyses). One WOH and 3 EOH sites were eliminated from this analysis because of missing data in ≥ 1 of 3 y of data collection. All concentration data were $\log_{10}(x)$ transformed, as described above, prior to analysis. Foucart (1978) and Benzécri and Benzécri (1986) first introduced between-class and within-class PCAs, and several authors have shown their utility in distinguishing between seasonal/annual and spatial effects in aquatic sciences (Dolédec and Chessel 1989, Burgherr and Ward 2001). An intermediate step in this procedure used 1-way analysis of variance (ANOVA) to test for spatial and temporal effects on each molecular tracer compound.

Redundancy Analyses (RDA).—RDA is a direct gradient ordination technique (Jongman et al. 1995) that constrains sample scores, in this case generated from molecular tracer concentrations measured at each site, to linear combinations of explanatory variables (i.e., land-use/cover and other watershed-characteristic variables [hereafter just watershed-characteristic variables]). RDA is akin to the more common practice of correlating factors 1, 2, ... n from a PCA with independent variables. RDA was chosen over Canonical Correspondence Analysis (CCA) because RDA is a linear rather than a unimodal method and, thus, is more appropriate for chemical rather than community data.

RDAs were done at various scales to select watershed-characteristic variables that best explained among-site differences in molecular tracer concentrations. Twenty-seven molecular tracer variables were included in each RDA model-building process (all compounds in Table 1 plus the sum of PAHs = tPAH, sum of soot PAHs = sPAH, sum of volatile PAHs = vPAH, sum of fragrance materials = tFM, and sum of fecal steroids = tFS). All tracer concentration variables were summarized as 3-y geometric means for a given site, created from the arithmetic mean of log-trans-

formed data as described above. Five RDA models were built using different geographically defined subsets of the 60-site database (all EOH and WOH sites, EOH and WOH sites minus 3 outliers, EOH sites only, EOH sites minus 3 outliers, and WOH sites only). As many as 78 possible watershed-characteristic variables (Table 2) could have been included in the selection procedure (see forward selection procedure below), and these were described by 4 general categories: watershed scale (all watershed area upstream of the site), riparian scale (30-m riparian buffer on both sides of the stream network upstream of the site), reach scale (30-m riparian buffer on both sides of the stream network upstream of the site, truncated 1-km above the site), and scale independent. All watershed-characteristic variables were derived using a geographic information system (GIS). The first 3 categories contained landuse variables presented in Arscott et al. (2006) and Dow et al. (2006). The scale-independent category included watershed area, number of permitted point-source discharges, and the mean annual watershed-area-normalized effluent volumes (point-source discharge; Arscott et al. 2006). All landuse percentages were arcsine \sqrt{x} transformed to minimize bimodality, and density data were $\log_{10}(x)$ transformed to reduce magnitude effects. CANOCO version 4.02 (Microcomputer Power, Ithaca, New York; ter Braak and Šmilauer 1998) was used to build all models, and the manual forward selection procedure was used to select watershed-characteristic variables that significantly contributed to a model explaining molecular tracer concentrations. The manual forward selection procedure used a partial Monte Carlo permutation test (permuted 1000 times) to assess the usefulness of explanatory variables ($p < 0.10$) in the ordination.

Variance partitioning (Borcard et al. 1992) was done for each RDA model to quantify the variability that could be attributed to each general scale-based category of watershed characteristic (i.e., watershed-, riparian-, and reach-scale, and scale independent). Variance partitioning decomposes the total variability into parts that can be explained solely by a given category and by the interaction among categories. This partitioning was accomplished by defining the suite of variables from each category as covariables and rerunning each RDA to generate a partial RDA. The procedure was repeated holding each category's suite of selected model components as covariables until all 4 categories were partitioned.

Multiple linear regression (MLR).—MLR analysis was used to test relationships between 3-y mean molecular tracer data (geometric means for concentration data as for the PCA analysis) and molecular tracer ratios (no

data transformations) vs watershed-characteristic variables. Stepwise variable selection with a variable significance cutoff of $\alpha = 0.05$ was used in selecting independent variables in the MLR models (SAS/STAT). This statistical evaluation occurred for each variable at each step of the variable-selection process, whether the variable was retained from a previous step or not. Thus, a variable could be selected initially as a statistically significant part of the model but then be rejected as not statistically significant at a subsequent variable-selection step. This process can select 2 independent variables that are collinear or correlated with one another. Multicollinearity was assessed using the variance inflation factor (VIF), where VIF values >10 for any independent variable suggest that the associated slope between it and the dependent variable in the MLR model may be biased by multicollinearity effects (Myers 1990). These biases are expressed in unstable slope values that can change dramatically either in magnitude or direction when other collinear independent variables are added to or removed from the model. Multicollinearity does not affect the variance explained by each individual independent variable (i.e., partial r^2 value); these values remain representative of the variability in the dependent variable explained by each independent variable while accounting for the presence of all other independent variables in the model.

Separate MLRs were built for each molecular tracer compound and for 19 molecular tracer ratios with potential usefulness for source tracking. Three different geographically defined site groupings were used: all WOH sites, all EOH sites, and EOH sites minus sites 43, 49, and 58. Each tracer/region-specific MLR was built 3 times using landuse variables derived from 3 different scales (watershed, riparian, and reach). Seventeen independent variables were available for selection in each MLR iteration. These variables included 16 scale-defined landscape variables (4 urban/suburban, 4 agricultural, 5 shrub/forested, and 1 wetland categories as % total area, and population and road densities) plus 1 scale-independent variable, point-source discharge (SPDE) (see Table 2 for variables and abbreviations).

The model having the highest overall adjusted R^2 value without any significant multicollinearity issues, as assessed by the VIF, was selected as the best model across the 3 landscape scales. Furthermore, the selected best model was considered unique from the remaining scale models for a given tracer value if: 1) the adjusted R^2 of the best model was $\geq 10\%$ greater than either adjusted R^2 for the remaining scale models, and 2) no strong regression (e.g., $R^2 > 0.50$) could be found between the strongest predictor from the best model

TABLE 2. Landuse and other watershed-characteristic variables used to explain molecular tracer concentrations in east of Hudson River (EOH) and west of Hudson River (WOH) sites. Most variables (land use, density) were quantified at 3 scales: watershed (W), 30-m riparian buffer (each side of stream) along the entire upstream network (b), and 30-m buffer truncated 1-km upstream from the sampling point (1k) (see Arscott et al. 2006, fig. 8 in Dow et al. 2006). The rest were scale independent. Redundancy analysis was done using 5 different subsets of the 60 sites: 1) all 60 sites (EOH + WOH), 2) all sites minus 3 outliers (EOH and WOH [- 43, 49, 58]), 3) WOH sites, 4) EOH sites, and 5) EOH sites minus outliers (EOH [- 43, 49, 58]). o = variable included in manual forward selection procedures, X = variable selected ($p < 0.1$), blank = variable not included in manual forward selection procedures.

Abbreviation	Variable	EOH + WOH			EOH + WOH			WOH			EOH			EOH		
		EOH + WOH			(- 43, 49, 58)			WOH			EOH			(- 43, 49, 58)		
		W	b	1k	W	b	1k	W	b	1k	W	b	1k	W	b	1k
Scaled variables																
Landuse variables (%)																
CMIN	Commercial + industrial	o	o	o	o	o	o	o	o	o	X	o	X	X	o	o
COMM	Commercial	o	o	o	o	o	o	o	o	o	X	o	o	o	o	X
CONF	Coniferous forest	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
CROP	Cropland	o	o	o	o	o	o	o	o	o	o	X	X	o	X	o
DECD	Deciduous forest	o	o	o	o	o	o	o	o	o	X	o	o	o	o	o
FMCR	Farmstead + cropland	X	o	o	o	o	o	X	o	o	X	o	o	o	o	o
FMST	Farmstead	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
GRAS	Herbaceous (grassland)	o	o	o	o	o	o	o	o	o	X	o	X	o	o	X
INDU	Industry	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
MBRH	Mioed brushland (herbaceous + brush)	o	o	o	X	o	o	o	o	o	X	o	o	X	o	o
MCON	Mioed forest + deciduous forest	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
MFOR	Mioed forest	X	o	o	X	o	o	o	o	o	o	X	o	o	o	o
ORCH	Orchards	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
OURB	Other urban	o	X	o	o	o	o	o	X	o	o	o	o	o	o	o
RESD	Residential	o	o	o	o	X	o	o	o	o	X	o	o	o	o	X
RSDT	Residential + transportation											o	o	o	o	X
SHRB	Shrub + brush	o	o	o	o	o	o	X	o	o	o	o	o	o	X	X
TRAN	Transportation, communications, and utilities										o	o	X	X	X	o
WETL	Wetland	o	o	X	o	o	X	o	o	o	o	X	o	o	o	o
WTER	Water	o			o			o			o			o		
Density variables																
NOSE	Number of septic systems								o	o	o					
SEPT	Septic system density								o	o	o					
PDNS	Population density	o	X	X	o	o	X	o	o	o	o	X	X	o	X	X
RAIL	Railroad density (m/km ²)	o	X	o	X	o	o	o	X	o	o	o	o	o	o	o
RDNS	Road density	o	o	o	o	X	o	o	o	o	o	X	o	X	o	X
Scale-independent variables																
WTSD	Watershed area (km ²)	o	o	o	o	o	o	X	o	o	o	o	o	o	o	o
SPDTE	Total number of State Pollutant Discharge Elimination System (SPDE) dischargers	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o
SPDE	Mean annual watershed-area normalized State Pollution Discharge Elimination effluent volume	X	o	o	o	o	o	o	o	o	o	X	o	o	o	o

with either group of significant predictors from the remaining scale models.

Results

Analytical method

Relatively modest modifications to widely used EPA methods for semivolatile organic compounds (i.e., 625 or 8270) resulted in LRLs (and therefore MDLs) that were 3 to 5 orders of magnitude lower than the given MDL for those methods (Table 1, Fig. 1). These large increases in sensitivity were achieved by: 1) increasing

sample volumes from 1 to 4 L by the use of solid-phase extraction rather than liquid-liquid extraction (4× sensitivity), 2) decreasing final extract volumes from 1 to 0.0525 mL (20× sensitivity), 3) setting the Mass Selective Detector to Selective Ion Monitoring (SIM) mode rather than full scan (~10× sensitivity), 4) using calibration standards with a lower minimum concentration and using a separate, quadratic calibration curve when analyte concentrations were in the range of 0.2 to 4.0 ng/μL in the injected extract (5× sensitivity), and 5) correcting for differences in sample-to-sample analytical recovery with internal

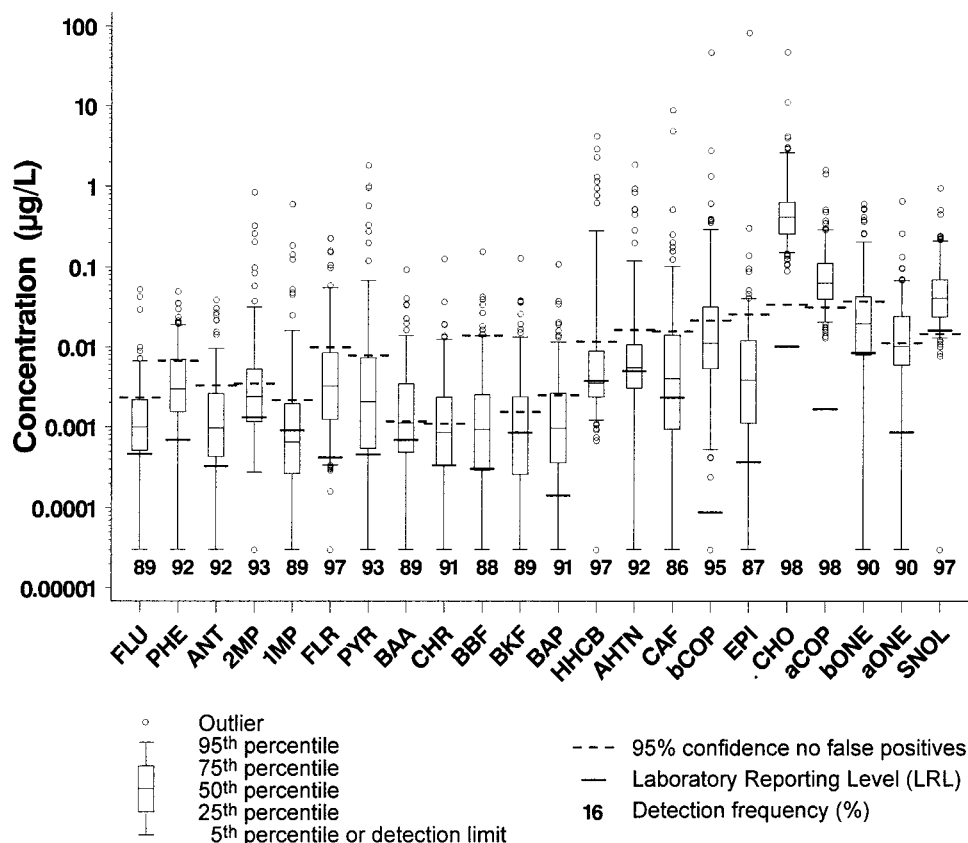


FIG. 1. Box-and-whisker plots for summer baseflow concentrations for each of 22 tracer compounds for 2000 to 2002 ($n = 180$), relative to measured Laboratory Reporting Levels (LRL; equivalent to the 99% confidence level for no false negatives) and the 95% confidence level for no false positives (see Methods). Detection frequencies, as % total samples, are given as bold numbers beneath each plot. 0.00003 was added to all values to allow log scale. See Table 1 for molecular tracer abbreviations.

surrogate standards ($\sim 2\times$ sensitivity, as observed by a halving of standard deviations in our MDL study). Cumulatively, these modifications resulted in a theoretical $8000\times$ ($= 4 \times 20 \times 10 \times 5 \times 2$) improvement in sensitivity of our method over EPA 625 or EPA 8270, which closely matched the observed improvements in our LRL over given MDL values for these methods. LRL values were generally a factor of 5 to $10\times$ greater than instrument detection concentrations.

A consequence of having a sensitive method is that analytical blanks are much more likely to contain detectable concentrations. Of the 22 blanks analyzed with our 180 study samples, the % of these blanks exceeding our LRL values ranged from 4% for SNOL to 82% for PHE and CHO. The laboratory- and field-blank data sets are pooled for subsequent discussion because no significant difference was observed between laboratory and field blanks ($p < 0.05$ for paired t -tests with pairing by extraction week). For 8 of our 22 tracer compounds, $<27\%$ of analyzed blanks exceeded LRLs. For 8 compounds, 30% to 48% of blanks exceed LRLs, and for 6 compounds 48% to 82% of blanks

exceeded LRLs. On the other hand, for 18 of 22 compounds, $>25\%$ of blanks were undetectable (CHO, HHCb, PHE, and SNOL were exceptions). Contamination was highly variable between blanks, exhibiting a 50- to 1000-fold range in magnitude. Contamination also was random; concentrations of tracers in blanks generally were not correlated with the minimum, mean, maximum, or coefficient of variance of concentrations of tracers in samples from the same extraction batch (exceptions were the minimums and means of 1MP and PHE, where p -values ranged from 0.02 to 0.0012). Therefore, blanks were not useful for determining whether a specific sample might be contaminated or for estimating the magnitude of that contamination. On the other hand, concentration distributions measured in blanks did allow estimation of probabilities of contamination to a given concentration level.

The 75th and 95th percentile concentrations from our blank data were calculated for each analyte, and these percentiles were used as the levels above which we have a 75% and 95% confidence, respectively, of not

reporting a false positive. Using a 99% confidence interval was equivalent to using the highest blank observed during the study (Table 1, Fig. 1). The 95% confidence levels ranged from 0.0011 to 0.038 $\mu\text{g/L}$, still well below the given MDL for EPA methods and below or comparable to detection limits from more recent methods (Kolpin et al. 2002, Glassmeyer et al. 2005). The 75% confidence levels for no false positives were below the median sample concentrations for 19 of 22 analytes. Data were not censored a priori below any of these levels prior to statistical analyses, following the recommendations of numerous studies (Gilliom et al. 1984, Porter et al. 1988, Helsel 1990, USGS 1999). The rationale for not censoring data is presented at length in the Discussion section.

Recoveries of internal surrogate standards added to each sample typically ranged from 73% to 107% for PHE-D10, 63% to 98% for CHR-D12, 56% to 86% for PER-D12, and 47% to 86% for CHO- $^{13}\text{C}_2$ (given values are for the 25th and 75th percentile of measured recoveries). In all but rare cases (<5% of occurrences), recoveries of all surrogate standards were within the range of 20% to 160%. Recoveries deviated below (–) or above (+) 100% possibly because of incomplete extraction from the sample water or particle matrix (–), volatilization during evaporation (–), adsorption to glassware (–), incomplete transfers from one vessel to another (–), incomplete dissolution by final solvents (–), inaccurate volumes of final solvents (– or +), unintended evaporation of final extract solvents (+), or contamination from other samples or standards (+). Recoveries typically were lower for higher molecular weight compounds. ANOVA confirmed that PHED10 recoveries were significantly higher than all other surrogates, and CHRd12 recoveries were significantly higher than recoveries of PERd12. These observations suggest that evaporative losses of tracers were less important than hydrophobic interactions with stream particles, glassware, etc.

Internal surrogate standard recoveries appear to track adequately variability resulting from real differences in sample-to-sample handling. Relative % differences (RPD) between sample duplicates almost always decreased by correcting for surrogate recoveries (with the exception of duplicates of samples with low concentrations near the LRL, which, on average, showed no improvement in precision after surrogate recovery corrections). Correcting for average recoveries was not adequate; despite utmost care in uniform treatment, recoveries varied as much between samples within an extraction batch as between batches (ANOVA showed significant differences for only 2 of 15 batches). These results together support the practice of using internal surrogate standard recoveries measured

for each sample to correct for sample-specific analytical biases to more accurately estimate ambient concentrations in the environment.

Tracer concentrations and ratios in streams

Measured concentrations of molecular tracers ranged over 6 orders of magnitude, from below instrument detection to values approaching 10 or 100 $\mu\text{g/L}$, depending on the compound (Fig. 1). Detection frequencies for each compound ranged from 86 to 98% of the 180 samples collected at 60 sites over the 3-y project (Fig. 1). However, generally only 50 to 80% of measurements for a given compound were above the respective LRLs (except for bCOP, aCOP, and CHO, which were 94–98% above LRLs).

Between-site variability in tracer concentrations generally was much greater than interannual variability within sites. For example, the summed concentrations of all 12 analyzed PAHs varied by 3 orders of magnitude between sites, whereas interannual variability within a site was generally less than a factor of 5 (data not shown). Individual PAH compounds, such as BAA, showed similar ranges (Fig. 2A). Coprostanol concentrations ranged over 5 orders of magnitude between sites, with only 1 to 2 order of magnitude variations in concentrations within a site (Fig. 3A). Similarly, caffeine concentrations spanned almost 4 orders of magnitude between sites, with intrasite variability ~ 1 order of magnitude (Fig. 4A). Because intersite variability exceeded intrasite variability by orders of magnitude, geometric means of tracer concentrations at each site were a valid way to make intersite comparisons or to synthesize site-specific results.

Concentrations of the 2 FMs (HHCB and AHTN) were generally much less spatially variable than the other tracers, with concentrations between 0.001 to 0.05 $\mu\text{g/L}$, except for 3 polluted sites with concentrations between 0.5 and 5 $\mu\text{g/L}$ (HHCB: Fig. 4B). This limited range of observed variability may have resulted from our high LRL for these compounds (0.0037 and 0.0050 $\mu\text{g/L}$, respectively) relative to ambient concentrations (Table 1).

Molecular tracers are not necessarily toxic compounds, but 10 of the 12 PAHs analyzed for the Project were listed by the EPA as Priority Toxic Pollutants and 5 of these are known human carcinogens (USEPA 2002a, b). These 5 most toxic PAHs (BAA, CHR, BAP, BBF, BKF) have been given exceptionally low National Recommended Water Quality Criteria for Human Health of 0.0038 $\mu\text{g/L}$ for the consumption of water or 0.018 $\mu\text{g/L}$ for the consumption of organisms living in the water (USEPA 2002b; Table 1). Similarly, New

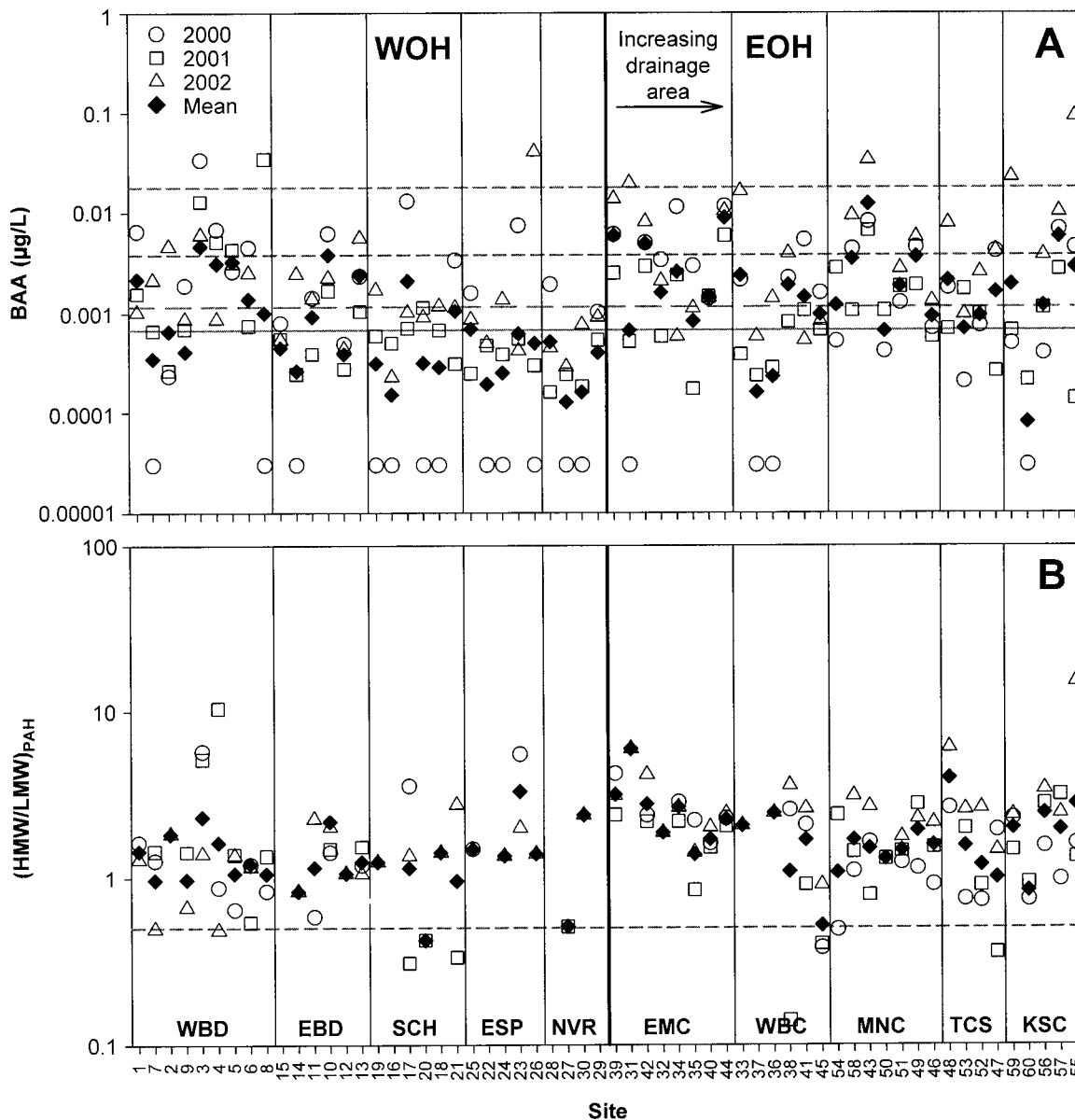


FIG. 2. Summer baseflow stream water concentrations of BAA (A) and ratios of high (HMW) vs low (LMW) molecular weight PAHs ($\text{HMW}/\text{LMW}_{\text{PAH}}$, see text for details) (B). See Table 1 for molecular tracer abbreviations. Solid and long-dashed lines in (A) show the Laboratory Reporting Level (LRL) and the 95% confidence level for no false positives, respectively. Short-dashed lines in (A) show US Environmental Protection Agency water-quality criteria levels for the consumption of water (lower line) and for the consumption of organisms living in the water (upper line) (USEPA 2002a). The short-dashed line in (B) shows a suggested delimitation between combustion or asphalt sources ($\text{HMW}/\text{LMW}_{\text{PAH}} > 0.5$) vs petroleum sources ($\text{HMW}/\text{LMW}_{\text{PAH}} < 0.5$) (Zakaria et al. 2002). Geometric means of all measured values at each site are given by solid diamonds. Nondetectable concentrations were assigned a value of $0.00003 \mu\text{g/L}$. Sites are arranged for west of Hudson (WOH) and east of Hudson (EOH) regions by watershed in order of increasing watershed area (see table 1 in Arscott et al. 2006 for site numbers and names). WBD = West Branch Delaware River, EBD = East Branch Delaware River, SCH = Schoharie Creek, ESP = Esopus Creek, NVR = Neversink River and Rondout Creek, EMC = East and Middle Branch Croton River, WBC = West Branch Croton River, MNC = Muscoot River and north of Croton Reservoir, TCS = Titicus, Cross, and Stone Hill rivers, and KSC = Kensico Reservoir and other sites south of Croton Reservoir.

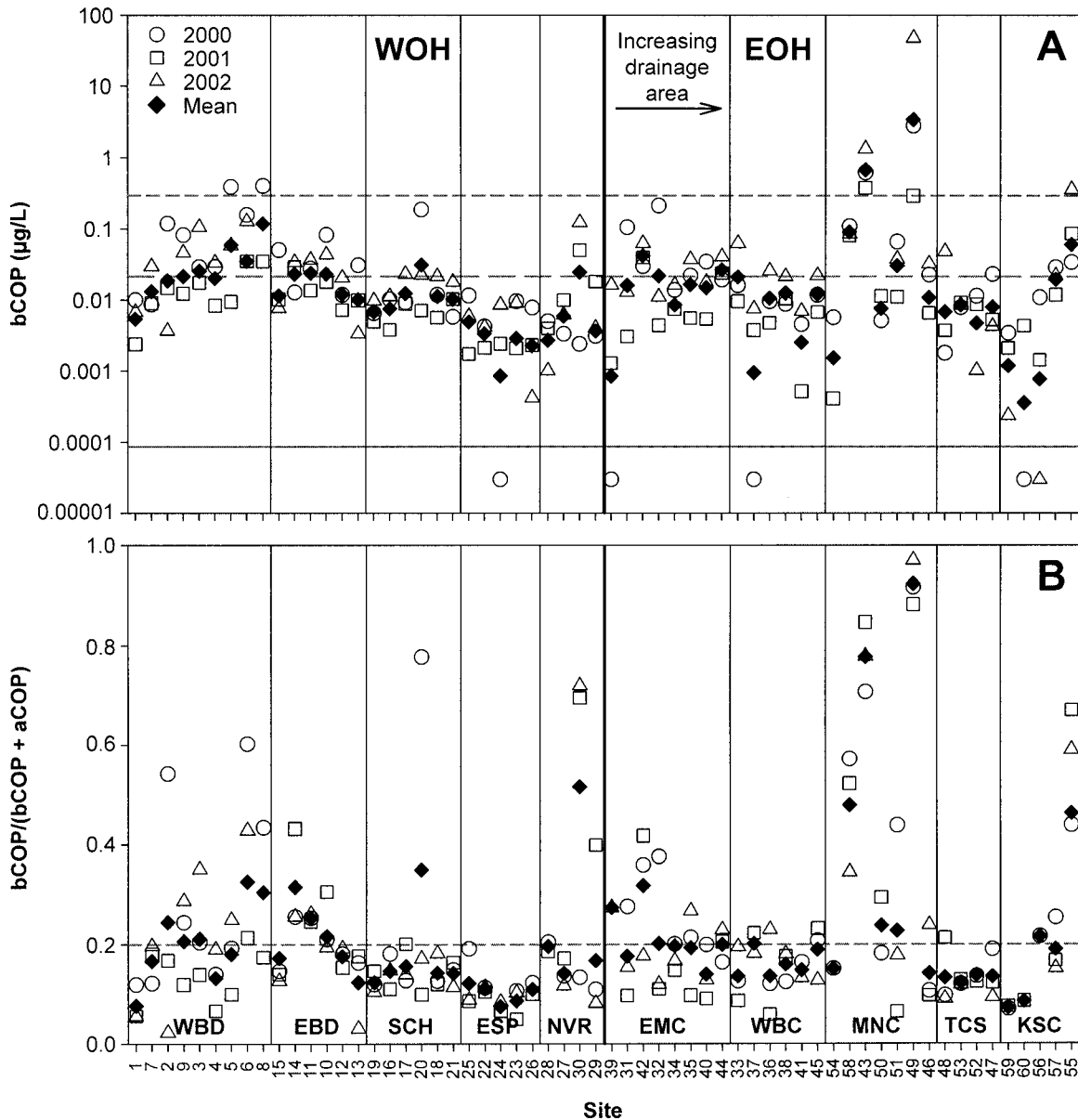


FIG. 3. Summer baseflow stream water concentrations of bCOP (A) and ratios of bCOP/(bCOP + aCOP) (B). See Table 1 for molecular tracer abbreviations. All other symbols and abbreviations as in Fig. 2. The short-dashed line in (A) shows concentration of bCOP corresponding to 200 fecal coliforms/100 mL, using the upper 95% confidence limit of the relationship given by Leeming and Nichols (1996). The short-dashed line in (B) shows a suggested delimitation between human fecal sources (bCOP/[bCOP + aCOP] > 0.2) and wildlife sources (bCOP/[bCOP + aCOP] < 0.2) in watersheds with minimal livestock (Grimalt et al. 1990, O'Leary et al. 1999).

York State Department of Environmental Conservation (NYS DEC) has set water-quality guidance values of 0.002 µg/L for BAA, CHR, BAP, BBF, and BKF for ambient waters directly feeding water supplies, 0.0012 µg/L for BAP in waters used for fish consumption, and 0.03 µg/L for BAA as a flag of chronic toxicity to aquatic life (NYS DEC 1998). These limits are non-regulatory guidance values that are not enforceable, and none of our sites were near water-supply intakes,

but the guidance values are useful for placing measured PAH concentrations in the context of potential human and ecosystem toxicity. When compared to the EPA guidance values, measured concentrations of ≥1 of these 5 compounds exceeded the lower EPA guidance level of 0.0038 µg/L in 54 of 180 summer baseflow samples and exceeded the upper guidance level of 0.018 µg/L in 10 of 180 samples (see Fig. 2A for example). The mean number of these

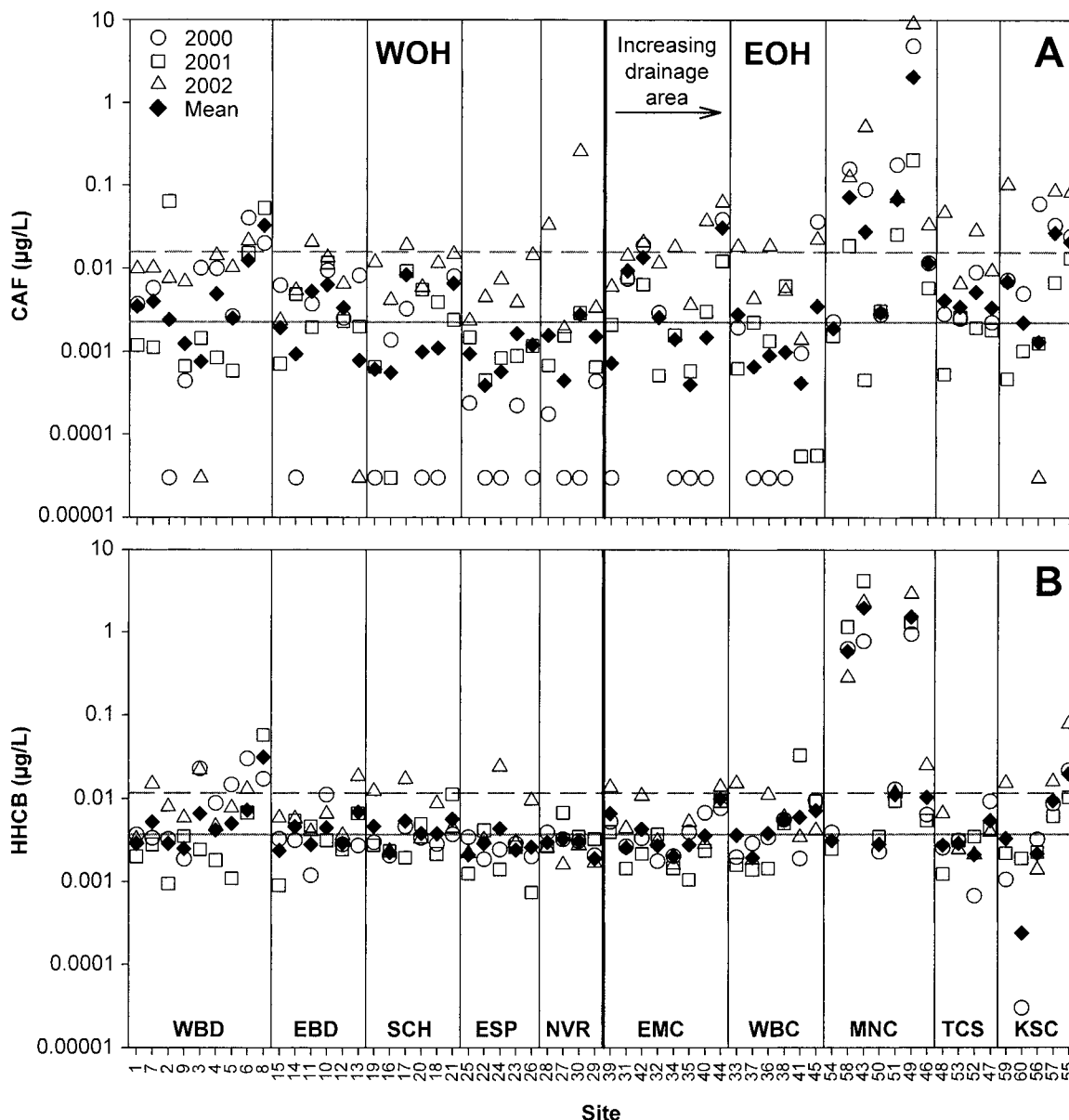


FIG. 4. Summer baseflow stream water concentrations of CAF (A) and HHCB (B). See Table 1 for molecular tracer abbreviations. All other symbols and abbreviations as in Fig. 2.

compounds to exceed a limit in a sample was 3.7 of 5 (i.e., if 1 compound exceeded a limit, so did most of the other 5). Geometric mean site concentrations exceeded the lower EPA guidance level ($0.0038 \mu\text{g/L}$) for an average 10 of 60 sites per compound.

PAH ratios, which are useful in distinguishing petroleum from combustion sources (Dickhut et al. 2000, Yunker et al. 2002, Zakaria et al. 2002), generally showed high values relative to suggested source delimitations (Figs 2B, 5A, B). Considering only reliable PAH ratios (see Methods regarding censoring of ratio data), 111 of 121 samples had ratios of high-

molecular-weight (HMW) PAHs (FLR, PYR, BAA, CHR, BBF, BKF, BAP) to low-molecular-weight (LMW) PAHs (FLU, PHE, ANT, 2MP, 1MP) ($\text{HMW}/\text{LMW}_{\text{PAH}} > 0.5$) (Fig. 2B), 58 of 61 samples had $\text{ANT}/(\text{ANT} + \text{PHE})$ ratios > 0.1 (Fig. 5A), and 80 of 95 samples had $\text{FLR}/(\text{FLR} + \text{PYR})$ ratios > 0.4 (Fig. 5B).

FSs are not themselves toxic compounds and, as such, they do not have EPA guidance values that can be used as a reference to evaluate measured concentrations. However, coprostanol, which is the dominant fecal steroid found for humans but not other animals (Leeming and Nichols 1996), has been correlated with human sewage

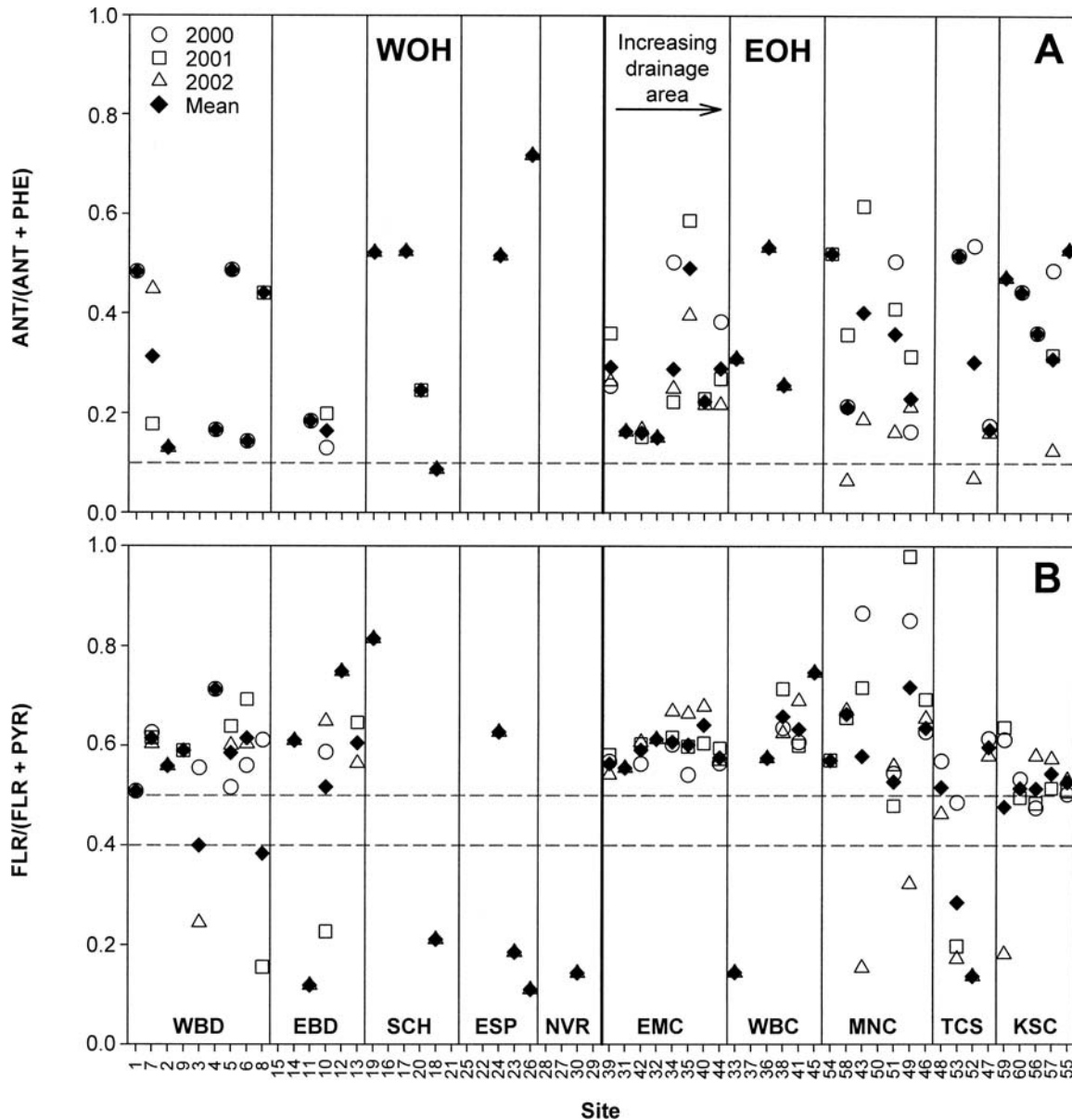


FIG. 5. Summer baseflow stream water ratios of ANT/(ANT + PHE) (A) and FLR/(FLR + PYR) (B). See Table 1 for molecular tracer abbreviations. All other symbols and abbreviations as in Fig. 2. The short-dashed line in (A) shows a suggested delimitation between combustion sources (ANT/[ANT + PHE] > 0.1) and petroleum sources (ANT/[ANT + PHE] < 0.1) (Yunker et al. 2002). Short-dashed lines in (B) show a suggested delimitation between combustion sources (FLR/[FLR + PYR] > 0.5) and petroleum sources (FLR/[FLR + PYR] < 0.4) (Yunker et al. 2002). Arithmetic means of all measured values at each site are given by solid diamonds.

inputs, thermotolerant (fecal) coliforms, and fecal streptococci (Leeming and Nichols 1996). Thus, given an EPA guidance value of 200 fecal coliforms/100 mL and using the upper 95% confidence limit of the relationship between the fecal coliforms and coprostanol given by Leeming and Nichols (1996), we calculated an approximate coprostanol guidance limit of 0.3 µg/L (Table 1). Only 9 of 180 samples exceeded this value (Fig. 3A). However, at the 2 EOH sites with known sewage

treatment plant problems (43 and 49) each of 3 samples were >0.3 µg/L coprostanol.

Similar to PAHs, ratios of FSs can help distinguish potential sources. For example, high values of the ratio bCOP/(bCOP + aCOP) has been used to identify sites where fecal material from humans dominates over other fecal sources (Grimalt et al. 1990, O'Leary et al. 1999). Fifty-six of 163 samples had bCOP/(bCOP + aCOP) ratios >0.2 (Fig. 3B).

PCA

PCAs were used to assess patterns of intersite variability vs interannual variability within a site. The initial WOH PCA (Fig. 6A, B) explained 52% of the total variance on the first 2 PCA factors ($F1 = 42.0\%$, $F2 = 10.5\%$). All molecular tracers loaded heavily on positive $F1$, whereas $F2$ distinguished between LMW PAHs (positive $F2$) and HMW PAHs (negative $F2$). Decomposition of the explained variance showed that 87.6% of the explained variance could be attributed to among-site variance within each year, whereas 12.4% could be attributed to among-year variance within sites. This result illustrated that, within any given year, among-site differences were stronger than interannual variability. However, interannual variability did cause significant differences between years in the ranking of sites by either of the first 2 PCA factors. Specifically, 1-way ANOVAs used to test spatial and temporal effects for each molecular tracer compound showed that only 4 of 22 tracer compounds had significantly different ($p < 0.05$) annual rankings among sites, whereas 12 of 22 tracer compounds had significantly different site rankings among years.

The initial EOH PCA (Fig. 6C, D) explained 57.6% of the total variance on the first 2 PCA factors ($F1 = 44.7\%$, $F2 = 12.9\%$). All molecular tracers loaded on the positive $F1$ dimension, whereas PAHs contributed to the positive $F2$ dimension, and FS contributed to the negative $F2$ dimension. Among-site variance within each year also accounted for most of the EOH explained variance (85.8%). Nine of 22 compounds had significantly different annual rankings among sites (multiple ANOVAs: $p < 0.05$), whereas 18 of 22 compounds had significantly different site rankings among years.

RDA

Five RDA iterations were done using various subsets of the 60-site data matrix. Common to all analyses was that most of the constrained variance loaded on $F1$, and molecular-tracer–landuse correlations for $F1$ and $F2$ generally had $r \geq 0.75$ (Table 3, Fig. 7A–I). PAHs and FS (along with CAF and FM) typically loaded on the opposite ends of the $F2$ axis (Fig. 7A, D, G). Several landuse variables were positively correlated with molecular tracers loading on the $F1$ axis, and all of these variables described land uses that replaced forest cover (either agricultural or rural communities in the WOH or urban/suburban land uses in the EOH) (Fig. 7B, E, H). More-forested sites had lower concentrations of all molecular tracer compounds than less-forested sites (Fig. 7B). The 60-site RDA model distinctly separated most EOH sites

from WOH sites (Fig. 7C), with most EOH sites loading towards higher PAH concentrations relative to WOH sites (Fig. 7C). Exceptions to this general pattern included 3 EOH sites with high concentrations of FS, CAF, and FM that were associated with known large point-source discharges. Removal of these 3 sites (43, 49, and 58) accentuated regional differences (Table 3, Fig. 7G, H, I) and resulted in a slightly different selection of explanatory variables. In particular, point-source discharge was not included in this 57-site model despite being dominant in the full 30-site EOH model (Tables 2 and 3).

Comparing explanatory power among models, particularly between models with differing numbers of explanatory variables, is not valid. However, comparing the independent variables that were selected in different models is instructive. Variance partitioning (Fig. 8) indicated the importance of scale-independent variables (particularly point-source discharge) in explaining molecular tracer concentrations at EOH sites 43, 49, and 58. Variance partitioning also suggested that landuse variables summarized at the reach scale were more important for the EOH model than the WOH model. In the WOH model, only 1 of 6 selected variables in the model was a reach-scale variable (Table 3) and that variable, % commercial land use, was positively associated with molecular tracer concentrations (Fig. 7E). On the other hand, in the EOH model (– 43, 49, and 58), 8 of 14 selected variables were reach-scale variables, and several of these variables were negatively associated with molecular tracer concentrations (Table 3, Fig. 7H).

Statistical analyses—MLR

MLR analyses of molecular tracer concentrations, sums, and ratios were used as an initial step toward tracking sources of individual molecular tracers (Table 4). MLR is particularly suited to the task of examining how several independently varying landscape characteristics are together related to one dependent variable such as the concentration of a tracer. MLR systematically determines which subset of independent variables combine linearly to best explain variability of the chosen dependent variable, taking into account covariance within the independent set (Neter et al. 1996). However, in data sets exhibiting substantial covariance between independent variables, it is typical that 2 MLR models using different sets of variables might yield very similar results. For example, if A and B are strongly correlated they most likely will not both be significant variables in a single MLR model for predicting Y, but an MLR model using variables A, C, and D might predict Y almost equally well to a MLR

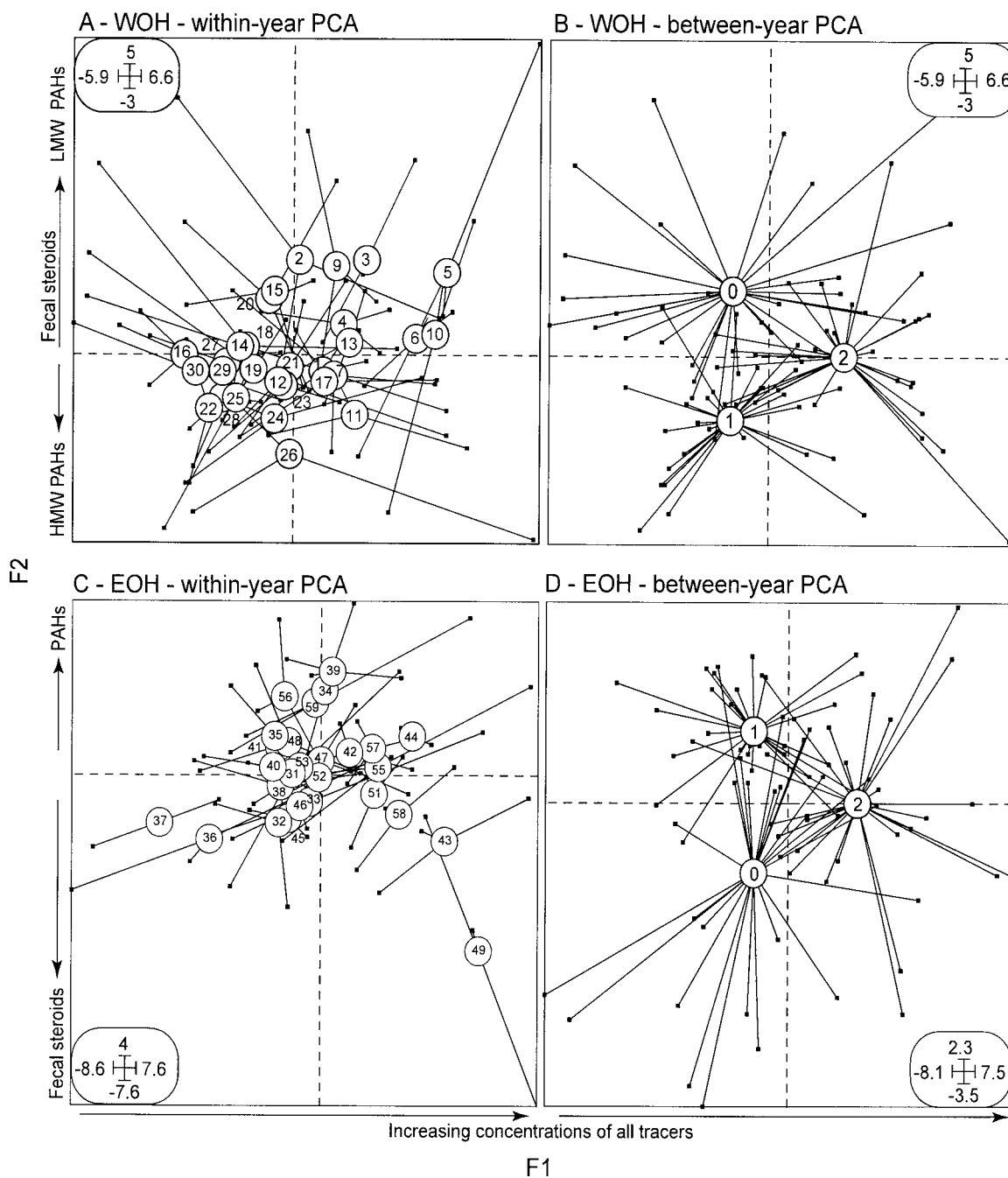


FIG. 6. Within- and between-class (years) Principal Components Analyses (PCA) done using $\log_{10}(x)$ transformed concentrations of 22 molecular tracers (see Methods) measured from 29 west of Hudson River (WOH) sites (A, B) and 27 east of Hudson River (EOH) sites (C, D). One WOH and 3 EOH sites were eliminated from this analysis because of missing data. (A) and (C) illustrate the central tendency (open circles) based on 3 summer baseflow measurements (small black squares) of the 22 tracers at each site after the year effect was removed. (B) and (D) present the same data but with open circles representing the central tendency based on tracer concentrations found in 2000 (0), 2001 (1), or 2002 (2) during summer baseflow sampling after the site effect was removed. Molecular tracer loadings are not shown, but arrows along axes F1 and F2 describe general patterns of increasing concentrations of molecular tracers. Insets show the axis scales. See table 1 in Arcscott et al. (2006) for site numbers and names. PAH = polycyclic aromatic hydrocarbon, HMW = high molecular weight, LMW = low molecular weight.

TABLE 3. Eigenvalues, correlations, number of variables at each scale included in redundancy analyses, and the 4 landuse variables most strongly correlated with Factor 1 (F1; in order of increasing concentration; numbers in parentheses are correlation coefficients) for analyses using subsets of the 60-site east of Hudson River (EOH) and west of Hudson River (WOH) database for 27 molecular tracer variables. The 1st and all canonical axes were significant at $p < 0.001$ (Monte Carlo permutations = 1000). See Table 2 for explanations of data subsets, variables included in all models, and their abbreviations.

	EOH + WOH	EOH + WOH (– 43, 49, 58)	WOH	EOH	EOH (– 43, 49, 58)	WOH Delaware	WOH Catskills
No. sites	60	57	30	30	27	15	15
Canonical eigenvalues							
All axes	0.500	0.482	0.589	0.926	0.883	0.736	0.923
F1	0.397	0.431	0.553	0.646	0.675	0.703	0.790
F2	0.055	0.340	0.022	0.112	0.108	0.021	0.480
Molecular tracer–landuse correlations							
F1	0.747	0.754	0.818	0.991	0.972	0.920	0.993
F2	0.846	0.736	0.725	0.971	0.945	0.726	0.929
No. variables	8	7	6	18	14	4	8
Watershed-scale	2	3	2	7	3	2	3
Riparian-scale	3	1	2	4	4	1	0
Reach-scale	2	3	1	6	8	1	4
Scale-independent	1	0	1	1	0	0	1
Landuse variables most strongly correlated with molecular tracer F1							
	SPDE (0.39)	RAIL (0.44)	b-RAIL (0.55)	1k-PDNS (0.50)	CMIN (0.46)	RDNS (0.80)	WTSD (0.53)
	b-PDNS (0.36)	b-RDNS (0.33)	b-OURB (0.54)	SPDE (0.45)	b-PDNS (0.37)	COMM (0.72)	1k-RAIL (0.53)
	FMCP (0.35)	MBRH (0.32)	FMST (0.49)	CMIN (0.42)	1k-CROP (–0.35)	1k-MCON (–0.36)	FMST (–0.33)
	b-RAIL (0.34)	1k-RESD (0.23)	1k-COMM (0.48)	RESD (0.41)	1k-RDNS (0.29)	b-FMST (0.31)	MBRH (0.20)

using B, C, and D, and the standardized slopes and significance of A and B in their respective models will probably be similar. However, if A and B are moderately correlated, they might both be included in the model, but with reduced significance and potentially very different slopes relative to the slopes of models including A or B alone. Therefore, the models in Table 4 were selected using the VIF factor criteria to minimize unstable slopes the MLR, but these models were not the only significant models showing interesting trends.

For WOH sites, generally >50% of the variance in PAH concentrations could be predicted from 3 or 4 landscape characteristics (Table 4). Variables related to human development, such as point-source discharge, % industrial, % commercial, % residential, and % other urban land uses, had positive partial correlations to PAHs. Landscape characteristics related to natural vegetation, such as % mixed forest, and % coniferous forest were inverse predictors of PAH concentrations. These landuse groupings (variables related to human development vs natural vegetation) showed high internal covariance (fig. 8f in Dow et al. 2006) and, thus, these variables were largely interchangeable within WOH MLR models. Agricultural vegetation,

such as % grassland and % orchard, showed positive partial correlations with PAHs in WOH MLR models, and simple bivariate correlations between these variables and PAHs also showed positive correlations (data not shown). Road density had negative partial correlations with most HMW PAHs, but only for the reach-scale models. Simple bivariate correlations confirmed that these trends were robust. Four of the 5 HMW PAHs had significant negative ($p < 0.05$) correlations with reach-scale road density, whereas 3 of the 5 HMW PAHs had significant positive correlations with riparian-scale road density (data not shown), and all 12 PAHs had significant positive correlations with watershed-scale road density (data not shown).

WOH PAH ratios generally showed weak relationships to watershed-characteristic variables ($R^2 = 0.1$ – 0.39 , Table 4), largely because of the reduction of the data set from censoring data before calculating ratios. ANT/(ANT + PHE), an indicator of combustion soot over petroleum sources (Dickhut et al. 2000, Yunker et al. 2002, Zakaria et al. 2002), was negatively correlated to watershed-scale % shrub cover and HMW/LMW_{PAH} was positively correlated with point-source discharge.

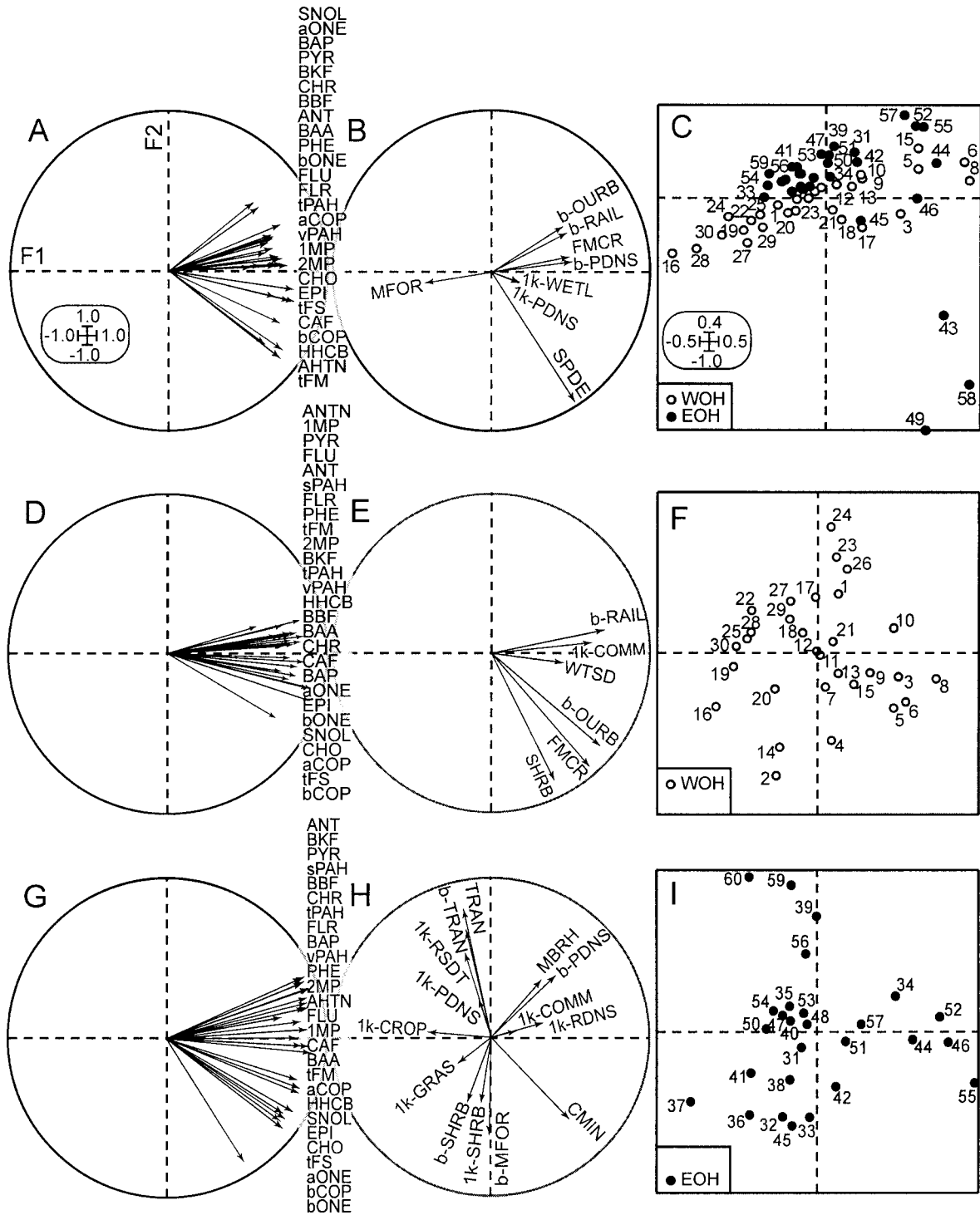


FIG. 7. Redundancy Analyses of molecular tracer compound concentrations vs watershed characteristics for 3 different sets of sites: all 60 sites (A, B, C), all 30 west of Hudson River (WOH) sites (D, E, F), and 27 east of Hudson River (EOH) sites (minus sites 43, 49, 58) (G, H, I). Ordinations for molecular tracer concentrations are shown in the left column (A, D, G), with compounds listed in order of their Factor (F) 2 loadings. See Table 1 for molecular tracer abbreviations. tFS = sum of fecal steroids, tPAH = sum of polyaromatic hydrocarbons, sPAH = sum of soot PAHs, vPAH = sum of volatile PAHs, tFM = sum of fragrance materials. Ordinations for landuse and watershed characteristics are shown in the center column (B, E, H). See Table 2 for landuse abbreviations. Site scores for each analysis are shown in the right column (C, F, I). See table 1 in Arscott et al. (2006) for site numbers, names, and locations. F1 and F2 axes are all scaled to lengths of ± 1.0 , except where insets show alternate lengths (C).

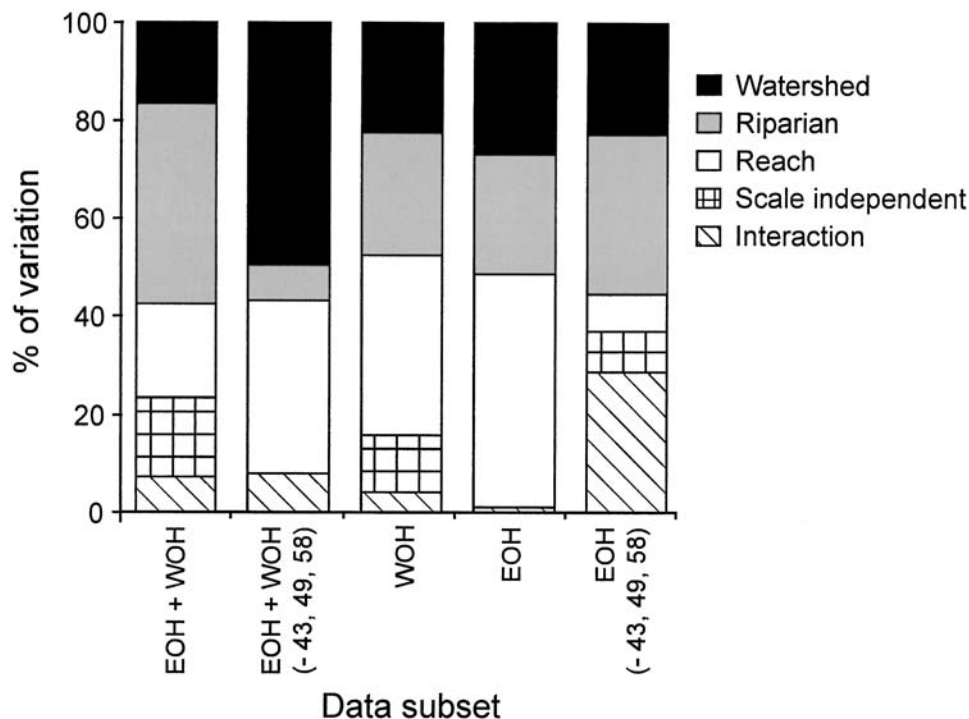


FIG. 8. Results of variance partitioning of the explained variance from the redundancy analysis of molecular tracer concentrations. Bars show the variance explained by watershed-, riparian-, and reach-scale and scale-independent landuse and watershed-characteristic variables and the interaction of these scales for each subset of sites (see Table 2 for definitions of subsets). Unexplained variance is not included in this figure.

At WOH sites, the fragrance HHCb was positively related to point-source discharge and negatively related to % mixed forest ($R^2 = 0.50$), whereas CAF showed a positive relationship with watershed-scale road density ($R^2 = 0.30$) (Table 4). Conversely, tFS, bCOP, and aCOP could be strongly explained ($R^2 = 0.60$ – 0.83) by watershed-scale agricultural land uses such as % farmstead and % grassland and by % other urban land uses. As with PAH ratios, reductions in the data set by censoring allowed only a few FS ratios to be tested adequately. However, bCOP/(bCOP + EPI) showed a very strong relationship ($R^2 = 0.80$) to % residential (+), % farmstead (+), and % other urban (–) land uses.

At EOH sites, 40 to 76% of the variability in PAH concentrations could be predicted from geographic variables (Table 4). Population density was the most important predictor for LMW PAHs, and road density and % industry were the most important predictive variables for HMW PAHs. Percent cropland or % orchard showed positive partial correlations with 4 of 12 PAHs (Table 4) and positive simple correlations with 8 of 12 PAHs (data not shown). Vegetation such as % deciduous forest, % mixed forest, % shrub, and %

grassland all showed negative partial correlations to PAHs.

At EOH sites, CAF, FM, and FS were strongly related to point-source discharges (mostly sewage treatment plants). Three sites (43, 49, and 58) with exceptionally high sewage discharge had substantial potential to skew regressions, and MLR analyses were rerun for EOH excluding these sites (Table 4). Despite removing these sites, 32 to 65% of variability in CAF, HHCb, bCOP, and FS was explained by land use. Point-source discharge was an important predictor for CAF and FS, % commercial land use was most predictive for HHCb, and % industry was most predictive for bCOP. The FS ratio bCOP/(bCOP + aCOP) was explained best by % industrial land use ($R^2 = 0.27$) and bCOP/(bCOP + EPI) was best explained by % mixed brush ($R^2 = 0.18$).

Discussion

Need for more sensitive analytical methods

A comparison of recommended ambient water criteria for priority organic pollutants (NYS DEC 1998, USEPA 2002a, b) vs MDL values for the most common analytical methods (USEPA 1998) indicated

TABLE 4. Adjusted R^2 values and significant predictive variables (partial R^2 , direction of slope) for the best (highest adjusted R^2 , no multicollinearity) model among scales for each analyte in each region/watershed. Models were derived from stepwise multiple linear regressions (MLR) of mean molecular tracer concentration as a function of watershed landuse characteristics quantified at the watershed (W), riparian (b), or reach (1k) scales. The best MLR is indicated in bold and marked with an asterisk (*). See text for definition of a unique model and for definitions of tracer ratio abbreviations. See Table 1 for tracer abbreviations. See Table 2 for landuse abbreviations. EOH = east of Hudson River, WOH = west of Hudson River, PAH = polycyclic aromatic hydrocarbon, HMW = high molecular weight, LMW = low molecular weight, CAF = caffeine, tFS = sum of fecal steroids, Y = yes, N = no, - = no significant model found.

Tracer compound	Model adjusted R^2					Significant predictors
	n	W	b	1k	Unique	
WOH						
FLU	30	0.33	0.44*	0.12	Y	RESD (0.38, +) ORCH (0.10, +)
PHE	30	0.37	0.59*	0.41	Y	MFOR (0.31, -) INDU (0.16, +) ORCH (0.09, +) SHRB (0.08, -)
FLR	30	0.54	0.55*	0.54	N	SPDE (0.27, +) CONF (0.25, -) ORCH (0.08, +)
PYR	30	0.80*	0.52	0.48	Y	INDU (0.41, +) MFOR (0.13, -) ORCH (0.12, +) SHRB (0.11, -) SPDE (0.05, +)
BAA	30	0.36	0.44	0.59*	Y	GRAS (0.29, +) OURB(0.27, +) RDNS (0.07, -)
CHR	30	0.32	0.35	0.71*	Y	OURB (0.21, +) RDNS (0.19, -) FMST (0.18, -) CROP (0.10, +) MBRH (0.08, +)
BAP	30	0.87*	0.77	0.70	N	CROP (0.57, +) INDU (0.12, +) ORCH (0.08, +) SHRB (0.04, -)
ANT/(ANT + PHE)	15	0.39*	0.32	-	N	SHRB (0.43, -)
(HMW/LMW) _{PAH}	30	0.10*	0.10	0.10	N	SPDE (0.13, +)
CAF	30	0.30*	0.25	0.17	N	RDNS (0.32, +)
HHCB	30	0.47	0.50*	0.43	N	SPDE (0.34, +) MFOR (0.20, -)
tFS	30	0.78*	0.77	0.71	N	GRAS (0.59, +) SPDE (0.14, +) ORCH (0.07, +)
bCOP	30	0.60*	0.59	0.11	N	FMST (0.55, +) SHRB (0.08, +)
aCOP	30	0.83*	0.83*	0.73	N	OURB(0.61,+) SPDE(0.20,+) FMST(0.05,+)
bCOP/(bCOP+aCOP)	-	-	-	-	-	No significant model found
bCOP/(bCOP+EPI)	27	0.80*	0.61	0.15	Y	RESD (0.65, +) OURB (0.09, -) FMST (0.09, +)
EOH						
FLU	30	0.58	0.76*	0.71	Y	PDNS (0.36, +) CROP (0.22, +) MBRH (0.12,-) CONF (0.09, +)
PHE	30	0.28	0.42*	0.38	Y	PDNS (0.23, +) CROP (0.15, +) RESD (0.11, -)
FLR	30	0.40*	0.35	0.18	N	PDNS (0.34, +) MBRH (0.10, +)
PYR	30	0.41*	0.30	0.28	Y	RDNS (0.37, +) ORCH (0.09, +)
BAA	30	0.35	0.40	0.48*	Y	RDNS (0.21, +) SPDE (0.17, +) CROP (0.15, -)
CHR	30	0.46*	0.34	0.45	Y	RDNS (0.34, +) INDU (0.10, +) MBRH (0.08, +)
BAP	30	0.10	0.15	0.50*	Y	RDNS (0.20, +) SHRB (0.17,-) INDU (0.11, +) GRAS (0.10, -)
FLR/(FLR + PYR)	28	0.17	0.17	0.30*	Y	SPDE (0.20, +) OURB (0.14, -)
(HMW/LMW) _{PAH}	30	-	0.14	0.22*	Y	MFOR (0.15, +) SHRB (0.12, -)
CAF	30	0.60	0.46	0.67*	N	SPDE (0.41, +) COMM (0.13, +) ORCH (0.12, -) RESD (0.05, +)
HHCB	30	0.65	0.63	0.65*	N	SPDE (0.59, +) FMST (0.09, +)
tFS	30	0.46	0.53*	0.46	N	SPDE (0.48, +) CONF (0.08, +)
bCOP	30	0.60	0.59	0.69*	N	SPDE (0.52, +) PDNS (0.08,-) RESD (0.07, +) COMM (0.06, +)
aCOP	30	0.35	0.44*	0.35	N	SPDE (0.38, +) CONF (0.11, +)
bCOP/(bCOP + aCOP)	30	0.44	0.44	0.52*	N	SPDE (0.46, +) FMST (0.09, +)
bCOP/(bCOP + EPI)	27	0.20	0.21	0.51*	Y	SPDE (0.23, +) MFOR (0.13, +) OURB (0.13, +) COMM (0.10, +)
EOH (- 43, 49, 58)						
CAF	27	0.46	0.42	0.50*	Y	RESD (0.17, +) COMM (0.14, +) SPDE (0.14, +) ORCH (0.12, -)
HHCB	27	0.65*	0.59	0.21	N	COMM (0.46, +) RESD (0.15, +) PDNS (0.08, -)
tFS	27	0.44*	0.42	0.29	Y	SPDE (0.31, +) MFOR (0.17, +)
bCOP	27	0.43*	0.38	0.34	Y	INDU (0.45,+)
aCOP	27	0.27	0.25	0.32*	Y	WETL (0.19, +) CONF (0.18, +)
bCOP/(bCOP + aCOP)	27	0.24	0.27*	-	N	INDU (0.29, +)
bCOP/(bCOP + EPI)	24	-	0.18*	-	Y	MBRH (0.21, +)

that typical laboratories cannot detect these compounds at concentrations that are orders of magnitude higher than levels of concern. Furthermore, ambient water criteria have been continually revised down-

ward, despite the fact that EPA's *Guidelines for Deriving Numerical Criteria for the Protection of Aquatic Life* (USEPA 1985) have not been officially revised since 1985 (although a draft revision is in review; available

at: http://www.epa.gov/waterscience/criteria/alcg_sab_draft.pdf). Clearly, monitoring programs must use much more sensitive analytical methods to obtain useful data on organic pollutants to: 1) assess whether current criteria are being met, 2) provide data relevant to potential future decreases in water-quality criteria, 3) assess whether concentrations below criteria levels are increasing over time, and 4) interpret patterns of pollutant concentrations to identify potential sources before water-quality criteria are exceeded (see below).

USGS laboratories and monitoring programs have taken strong steps to increase analytical sensitivity for organic contaminants (Kolpin et al. 2002, Phillips and Bode 2002, Glassmeyer et al. 2005), but most federal and state monitoring programs still use methods with grossly inadequate sensitivity for monitoring contamination of ambient and drinking water. Evidence for this comes from Proficiency Testing (PT) programs in use by the 13 states participating in the National Environmental Laboratory Accreditation Conference (NELAC). All laboratories providing monitoring data in these states must regularly analyze a blind PT sample and provide the correct concentrations of all compounds for which they are certified. The concentration ranges in the blind test samples for nonpotable water accreditation (available at: http://www.epa.gov/nelac/pttables/npw_fopt_final111004.pdf) have minimum values of 0.5 to 20 µg/L for pesticides, 7 to 25 µg/L for volalite aromatics and halocarbons, 10 to 200 µg/L for semivolatile base/neutrals (which include PAHs), and 30 to 200 µg/L for acid organics (phenols). These values all closely match the MDLs published in EPA method 8270 (USEPA 1998). Thus, commercial laboratories have little incentive to enhance the sensitivity of their measurements. At the same time, most monitoring projects maintain a purely regulatory mindset, and they analyze only regulated pollutants and require that laboratories adhere to published EPA methods. This mindset, at its worst, may even discourage more sensitive analyses because they would increase the detection frequency of pollutants and alarm the public. Fueling this mindset is the practice of focusing on detection frequency as the primary result of a monitoring study (e.g., Kolpin et al. 2002), rather than distributions of quantified concentrations relative to water-quality criteria or other reference levels (Fig. 1).

Results from our method (Table 1, Fig. 1) demonstrate that relatively modest modifications to EPA method 8270D (using equipment that all analytical chemistry laboratories possess) can improve sensitivity by 4 to 5 orders of magnitude. In our study, we quantified only 22 compounds (including only 10

compounds listed as priority pollutants), but our modified method should provide similar increases in sensitivity to all 245 analytes listed in EPA method 8270D. These improvements were generally sufficient to quantify compounds below any current recommended water-quality criteria (USEPA 2004) and also below any probable future criteria. We believe that all monitoring programs should demand similar sensitivity levels from all contracted laboratories. Because the required modifications to EPA method 8270D were minor (and often listed as options in the method itself) and because NELAC certification is largely based on performance with blind test samples, laboratories have substantial freedom to make the necessary modifications and still conform to EPA analytical guidelines. Our laboratory was a case in point; we have maintained NELAC certification for PAHs through the New York State Environmental Laboratory Approval Program (ELAP, <http://www.wadsworth.org/labcert/elap/elap.html>).

The Delaware River Basin Commission's (DRBC) work with polychlorinated biphenyls (PCBs) provides an example of a local monitoring agency working with commercial laboratories to reduce detection levels to those appropriate to monitor total maximum daily load (TMDL) concentrations. Existing human-health water-quality criteria for different zones of the Delaware estuary range from 7.9 to 64 pg/L total PCBs, and the proposed Stage II TMDL for the entire estuary is a uniform 16 pg/L total PCBs (0.000016 µg/L) for all zones (DRBC resolution No. 2005-19). The most common EPA methods for PCBs (methods 608 and 8082) listed MDLs in water ranging from 0.054 to 0.90 µg/L for individual PCBs. Even the more recent EPA method 1668, which uses high resolution GC-MS instrumentation that is unavailable to many commercial laboratories, has listed MDLs of 4 to 455 pg/L for individual PCBs. Given these analytical inadequacies, the DRBC worked with several local laboratories to modify EPA 1668 slightly to yield quantitation levels ~5 pg/L for each of 209 PCBs (http://www.state.nj.us/drbc/PCB_info.htm). These minor modifications were to increase to 2-L sample volumes, decrease to 20-µL extract volumes, and add lower calibration standards to give a 5-level curve down to 0.5 ng/mL. As with our modified method 8270, analytical blanks were a big issue with this modified method 1668, but quantifying blanks by a more sensitive method was the first step to reducing blanks. Monitoring agencies could substantially maximize the value of data obtained with limited monitoring funds by following the DRBC example of demanding improvements in analytical sensitivity and thereby avoid empty data sets of nondetects.

Minimizing data censorship

It is relatively common for laboratories to censor concentration data if values fall below reporting levels set relative to their measured MDLs or blanks, but a number of reasons exist for minimizing or even eliminating this practice in water-quality studies. These reasons have been substantiated at length in the literature. Gilliom et al. (1984) demonstrated that censoring data at any level tends to eliminate valuable information, and that even when low-level data were substantially degraded by random noise, trends were more effectively detected in the uncensored data. Porter et al. (1988) and Helsel (1990) independently echoed these findings in their explorations of statistical treatments of nondetects in water-quality data and describe these “less than” values as a serious interpretation problem for data analysts.

As a result of these and many other studies, the USGS National Water Quality Laboratory has adopted the convention of not censoring any data from information-rich methods such as GC-MS, but rather simply flagging data as estimated if they are below the quantified 99% confidence limits for no false positives or the 99% confidence limits for no false negatives (USGS 1999). This reporting convention allows the data analyst to choose the approach best suited for the given analysis. For instance, for statistical analyses, concentration data should be left completely uncensored because analysis of the full data distribution, even if it contains large errors at low concentrations, is always preferable over the alternative of assuming a data distribution for the censored data (Helsel 1990). On the other hand, evaluation of data at a specific site relative to water-quality criteria for the purposes of regulatory action clearly requires consideration of reporting levels based on confidence limits. In this case, censorship is appropriate. Censorship of derived variables, such as concentration ratios, may be appropriate prior to statistical analyses because the reliability of the ratio is independent of its magnitude, and the purpose of a ratio is primarily to distinguish among sources where concentrations are high. For these examples, the convention of reporting all data requires that every reported concentration is linked to an associated confidence level for no false positives and no false negatives. The commonly used MDL approach quantifies neither of these limits (USGS 1999).

The situation with our data set, where false positives (contamination) can occur to levels substantially greater than reliable quantitation levels (the LRL), is not described elsewhere to our knowledge. However, for all trace-level water-quality analyses, the frequen-

cies of analyte detections in blanks will increase as analytical sensitivities are improved. Regardless, the general approach to reporting 99% (or 95%) confidence limits for no false positives or no false negatives provides a means for honestly evaluating data quality while maximizing data information.

Temporal variability

Baseflow streamwater concentrations of tracers showed relatively substantial temporal variability, generally ranging ~1 order of magnitude for individual compounds at each site over the 3-y study. Spatial variability was higher than temporal variability (Fig. 6), but temporal variability should be considered by monitoring programs that infrequently revisit sites, not only because site rankings are subject to temporal variability, but also because a once-every-5-y sampling program (for example) will be unable to discern all but the grossest temporal trends.

The observed temporal variability in our data set does not appear to be entirely random. Within and between PCA analyses showed that sample year explained 12.4% and 14.2% of our data matrix variance for WOH (29 sites \times 3 y) and EOH (27 sites \times 3 y) data sets, respectively. In general, tracer concentrations were greater at most sites in 2002 than in 2000 or 2001 (Fig. 6), corresponding to generally lower total annual flow in 2002 that was below historic means (1964–1999; Arscott et al. 2006). However, to frame the temporal variability in our data set as interannual variability ignores short-term variability in hydrologic and atmospheric conditions. For example, 2000 appeared to be the wettest of the 3 sample years in general, and total annual flows in 2002 were below historic means (fig. 4 in Arscott et al. 2006). Nevertheless, site-specific discharge (normalized by watershed area) on the day that tracer samples were collected at all East and West Branch Delaware River sites were much lower in 2000 than in 2002 (Arscott et al. 2006).

A closer look at patterns of temporal variability reveals that tracer concentrations might respond differently to hydrological variations in different watersheds. In the EOH region, tracer concentrations were highest during the lower baseflow discharge in 2002. At EOH sites, point-source discharges, which generally have relatively stable discharge rates, were more important and may have been less diluted during low streamflow conditions than during high streamflow conditions. On the other hand, in the WOH region, elevated tracer concentrations in 2002 were related to increased precipitation and stream baseflow discharge, which may have increased runoff from agricultural lands and subsurface flushing of septic

systems, which in turn increased FS, FM, and CAF loading to streams. Temporal variability probably also depends on a variety of other natural and anthropogenic factors, such as the spatial configuration of different land covers and land uses, with different effects on each tracer compound. These hypotheses remain untested, but our continued effort at many of these sites should allow more rigorous examination of interannual variability in the future.

Sources of contamination to streams

To improve water quality or to mitigate future degradation, watershed managers should value knowledge related to the contributions of each potential contamination source to their specific watershed. Sources can be categorized in 2 ways: by land use or by input process. Describing contaminant concentrations or loads as a function of land use can aid in planning by enabling the watershed manager to consider landuse-specific impacts on water quality. However, a land use is not itself a source, but rather a proxy for processes that mobilize contaminants into the aquatic system. For instance, the concentrations of a particular PAH may be positively related to residential land use in a watershed, yet it is the sum of characteristics of residential land use, such as automobile exhaust, runoff from paved areas, leaking fuel tanks, or improper disposal of petroleum products, that ultimately is the source of the PAH measured in the stream. Molecular tracers offer the potential to fingerprint these input processes, through the use of ratios, factor analysis, or other statistical techniques. Understanding the relative importance of these input processes is necessary for managing water quality in existing land uses. Thus, geographic and mechanistic perspectives complement one another, and our data set can evaluate sources from both perspectives.

The most obvious geographic pattern was that concentrations of all tracer compounds generally varied concurrently (F1 axes in Figs 6 and 7); where one tracer or group of tracers was high, the others also tended to be high. This result was not necessarily expected because of the very different input mechanisms of PAHs vs CAF, FM, and FS. However, only minor separation was found between PAHs and the tracers of wastewater contamination (F2 axes in Figs 6 and 7, Table 3). Also, no single master landscape variable consistently explained intersite differences in either RDA or MLR analyses, except that sites with a higher % of their watershed in forest generally had lower concentrations of all tracers.

Patterns of tracer concentrations as a function of land use were distinctly different between EOH and

WOH sites (Fig. 7). EOH sites tended to have higher proportion of PAHs relative to FS, CAF, and FS, and the most pristine sites were all located in the WOH region. Within the WOH region, sites with elevated concentrations of molecular tracers were associated with nearby rural community infrastructure (i.e., commercial or other urban categories), riparian railroad infrastructure, and farmstead and row-crop components of agricultural land uses. These sites also were typically on larger streams (lower West Branch Delaware) where larger communities were located. In the EOH region, sites with high concentrations of molecular tracers were either associated with point-source discharges or urban infrastructure, whereas sites with lower concentrations of molecular tracers were either more forested or had more agriculture in their watersheds. However, in general, agriculture was a very small component of any EOH watershed.

PAHs generally were related to riparian-scale urban land uses (e.g., % commercial, % industrial, % other urban, population density, point-source discharge) and riparian-scale rail density (Figs 6 and 7, Table 4). Three groups of mechanistic source pathways can be distinguished based on PAH ratios: 1) petroleum products, such as kerosene, diesel oil, lubricating oil, and crude oil, which are characterized by lower ratios of less-stable to more-stable isomers (i.e., ANT/[ANT+PHE] or FLR/[FLR+PYR]) and by generally lower HMW/LMW_{PAH} ratios (Yunker et al. 2002, Zakaria et al. 2002); 2) combustion byproducts, such as automotive exhaust, smelter emissions, coal-burning emissions and wood smoke, which are characterized by higher ratios of less-stable to more-stable isomers and by generally higher HMW/LMW_{PAH} ratios (Dickhut et al. 2000, Yunker et al. 2002); and 3) asphalt, which is characterized by low ratios of less-stable to more-stable isomers (similar to petroleum products) and by higher HMW/LMW_{PAH} ratios (similar to combustion byproducts) (Yunker et al. 2002). In our study, high ratios of ANT/(ANT+PHE), FLR/(FLR+PYR) and HMW/LMW_{PAH} all suggested that combustion emissions generally appeared to dominate over petroleum spills as the primary source of PAHs to most stream sites (Figs 2 and 5). High HMW/LMW_{PAH} ratios alone would be inconclusive evidence for pyrogenic sources because HMW PAHs are relatively abundant in asphalt (Yunker et al. 2002), because diesel soot has relatively low HMW/LMW_{PAH} ratios (Miguel et al. 1998), and because HMW and LMW PAHs have different fates and transport characteristics in the environment caused by different volatility and particle affinity. However, the other 2 ratios do not have these problems (molecular weights of ANT and PHE and of FLR and PYR are identical,

and volatility and particle affinity are very similar). Furthermore, because all 3 ratios suggest pyrogenic sources, and because mean site PAH concentrations were related to urban land uses, we have reasonable confidence in our conclusion.

Combustion soot has the potential to travel long distances (i.e., from mid-western coal burning), but the relatively strong relationships with riparian urban land use at our sites suggests that local sources may be more important than long-distance sources. Relatively high methyphenanthrene (1MP + 2MP) to phenanthrene ratios (generally >0.5 , data not shown) support this conclusion because low ratios (<0.2) have been associated with aging of soot aerosols (Nielsen 1996, Simó et al. 1997, Stein et al. 2006). The observation that EOH sites tended to exhibit slightly higher PAH ratios than WOH sites (Figs 2B and 5A, B) despite further transport distance from midwestern coal burning, was further evidence for the importance of local sources of combustion-derived PAHs.

Intrasite variability in FS concentrations showed the strongest relationships to point-source discharge at 3 EOH sites (43, 49 and 58; Table 4, Fig. 7). The remaining EOH sites showed relationships between most FS and industrial and other urban land uses and also to point-source discharges (Table 4, not all data shown). However aCOP and EPI, 2 FS that are relatively abundant in birds and wildlife (Leeming et al. 1996) were most related to wetland and coniferous forest land covers, respectively (Table 4, not all data shown). WOH FS concentrations were positively associated with residential and agricultural land uses and negatively related to forest cover (Table 4). These data suggest a wide mixture of point and nonpoint sources of fecal contamination to streams.

Previous studies have used FS ratios to distinguish whether fecal sources were primarily human. Ratios of coprostanol, a steroid that predominates in human feces but is relatively less abundant in feces of other animals (Leeming et al. 1996), to other steroids (e.g., $bCOP/[bCOP+aCOP]$ or $bCOP/[bCOP+EPI]$) responded strongly to known inputs of human sewage (i.e., sites 43, 49 and 58; Fig. 3B, Table 4). However, the suggested delimitation of $bCOP/(bCOP+aCOP) = 0.2$ to distinguish between human fecal sources (>0.2) and other animals (<0.2) may only apply to watersheds with limited livestock (Grimalt et al. 1990, O'Leary et al. 1999). In the northeastern USA, raw livestock feces appear to have relative bCOP concentrations nearly as high as WWTP effluent and septic leachate (Standley et al. 2000, AKA, unpublished data). A more-robust FS index of human contamination for North America will require multivariate analyses of the full suite of FS compounds (AKA, unpublished data), similar to that

done by Leeming et al. (1996, 1998) for Australia. With simple ratios alone, it is impossible to determine whether human or livestock inputs were most responsible for the high FS levels at the high-agriculture WOH sites in the East and West Branch Delaware watersheds (Fig. 3).

FM (HHCB and AHTN) and CAF were more related to land use and, specifically, to point-source discharge and urban/commercial land uses, than any of the other tracers. They are introduced to streams and rivers from relatively unambiguous sources. FM are anthropogenic compounds introduced to the environment primarily in domestic greywater sewage (Simonich et al. 2000). These polycyclic compounds have low biodegradability, so they are transported relatively conservatively through sewage treatment (Simonich et al. 2002, Artola-Garicano et al. 2003, Phillips et al. 2005). The primary source of CAF to streams and rivers in temperate climates is the urine of humans (and sometimes domestic animals) (Buerge et al. 2003). CAF is removed more effectively than HHCB and AHTN by wastewater treatment processes (Phillips et al. 2005), but it still displays relatively low rates of biodegradation in the environment, is nonvolatile, and has low particle affinity, all resulting in relatively efficient transport through waterways (Buerge et al. 2003, Glassmeyer et al. 2005). As a result of these properties, concentrations of FM and CAF appear to be robust indicators of sewage, septic, and greywater sources, with the distinction that CAF primarily indicates poorly treated sources including septic systems. These distinctions are supported by our MLR analyses of geographic sources.

Analysis of relationships between upstream watershed characteristics quantified at different scales and tracer concentrations using variance partitioning showed that the best predictor scale depended on the specific landscape variable (Table 2). Kratzer et al. (2006) showed that watershed- and riparian-scale variables explained most of the variance in macroinvertebrates collected from our sites, but results from our analyses suggest that prediction of molecular tracer concentrations benefits from consideration of reach-scale factors in addition to watershed- and riparian-scale factors. Arscott et al. (2006) and Dow et al. (2006) demonstrated that watershed- and riparian-scale variables were closely related and probably described similar phenomena, albeit with different absolute values, depending on the scale at which a variable was summarized. On the other hand, reach-scale variables were generally uncorrelated with values quantified for larger scales.

The ultimate goal of using organic compounds as molecular tracers of sources of contamination to our

water supplies should be site-by-site quantitative mixing model analysis of potential sources, using the biomarker approach of organic geochemistry. The data set presented here begins to meet the first 2 criteria of such an analysis in that: 1) tracers are detected at concentrations well below levels of interest, and 2) ambient concentrations are adequately quantified. However, further improvements in methods toward both of these goals would be beneficial. Approximately 50% of stream samples had tracer concentrations low enough to be potentially affected by random laboratory blank contamination. Any resulting errors increase as they propagate to ratios, and we believe this problem may explain the relatively high scatter in PAH ratios (Figs 2B and 5) and also the low ability of MLR to relate ratios to landuse variables. Meeting the other 2 criteria for quantitative mixing model source analysis requires further study. Source signatures are relatively well known for PAHs (Dickhut et al. 2000, Yunker et al. 2002, Zakaria et al. 2002), but data are limited for the full suite of FS for North American human and animal sources (Standley et al. 2000, Glassmeyer et al. 2005). Moreover, additional work on the fate and transport of CAF, FM, and FS in streams is needed (Glassmeyer et al. 2005), especially work with a focus on how ratios might change during different processes.

Acknowledgements

We thank Stroud Water Research Center staff members who participated in sample collection and processing. In particular, Karen Jansson, Thy Truong, Amanda Christian, and Arlene Casas were instrumental in both field and laboratory sample processing. Lara Hall (Martin), Dave Van Horn, and a number of interns also assisted during the field sampling effort of this project. The manuscript benefited substantially from the comments of Pamela Silver and 2 anonymous referees. The NYC Department of Environmental Protection provided many GIS data layers and reports. This project was funded by a grant under the Safe Drinking Water Act from the NYS DEC and the USEPA.

Literature Cited

- ARSCOTT, D. B., C. L. DOW, AND B. W. SWEENEY. 2006. Landscape template of New York City's drinking-water-supply watersheds. *Journal of the North American Benthological Society* 25:867–886.
- ARTOLA-GARICANO, E., I. BORKENT, J. L. M. HERMENS, AND W. H. J. VAES. 2003. Removal of two polycyclic musks in sewage treatment plants: freely dissolved and total concentrations. *Environmental Science and Technology* 37:3111–3116.
- BENZÉCRI, J. P., AND F. BENZÉCRI. 1986. *Pratique de l'analyse de données*. Bordas, Paris, France.
- BLAINE, J. G., B. W. SWEENEY, AND D. B. ARSCOTT. 2006. Enhanced source-water monitoring for New York City: historical framework, political context, and project design. *Journal of the North American Benthological Society* 25:851–866.
- BORCARD, D., P. LEGENDRE, AND P. DRAPEAU. 1992. Partialling out the spatial component of ecological variation. *Ecology* 73:1045–1055.
- BUERGE, I. J., T. POIGER, M. D. MÜLLER, AND H.-R. BUSER. 2003. Caffeine, an anthropogenic marker for wastewater contamination of surface waters. *Environmental Science and Technology* 37:691–700.
- BURGHERR, P., AND J. V. WARD. 2001. Longitudinal and seasonal distribution patterns of the benthic fauna of an alpine glacial stream (Val Roseg, Swiss Alps). *Freshwater Biology* 46:1705–1721.
- DICKHUT, R. M., E. A. CANUEL, K. E. GUSTAFSON, K. LIU, K. M. ARZAYUS, S. E. WALKER, G. EDGEcombe, M. O. GAYLOR, AND E. H. MACDONALD. 2000. Automotive sources of carcinogenic polycyclic aromatic hydrocarbons associated with particulate matter in the Chesapeake Bay region. *Environmental Science and Technology* 34:4635–4640.
- DOLÉDEC, S., AND D. CHESSEL. 1989. Rythmes saisonniers et composantes stationnelles en milieu aquatique II. Prise en compte et élimination d'effets dans un tableau faunistique. *Acta Oecologica* 10:207–232.
- DOW, C. L., D. B. ARSCOTT, AND J. D. NEWBOLD. 2006. Relating major ions and nutrients to watershed conditions across a mixed-use, water-supply watershed. *Journal of the North American Benthological Society* 25:887–911.
- DOW, C. L., AND A. K. AUFDENKAMPE. 2006. Using SAS to improve the quantification of environmental chemistry samples. Paper DA02 in *Proceedings of the Northeast SAS Users Group Conference, Philadelphia, Pennsylvania, September 20–22, 2006*. (Available from: <http://www.nesug.org/html/Proceedings/nesug06.pdf>)
- EGLINGTON, G. 1969. Organic chemistry: the organic chemist's approach. Pages 20–71 in G. Eglinton and M. T. J. Murphy (editors). *Organic geochemistry*. Springer-Verlag, New York.
- FOUCART, T. 1978. Sur les suites de tableaux de contingence indexés par le temps. *Statistique et Analyse des données* 2:67–84.
- GILLIOM, R. J., R. M. HIRSCH, AND E. J. GILROY. 1984. Effect of censoring trace-level water-quality data on trend-detection capability. *Environmental Science and Technology* 18:530–535.
- GLASSMEYER, S. T., E. T. FURLONG, D. W. KOLPIN, J. D. CAHILL, S. D. ZAUGG, S. L. WERNER, AND M. T. MEYER. 2005. Transport of chemical and microbial compounds from known wastewater discharges: potential for use as indicators of human fecal contamination. *Environmental Science and Technology* 39:5157–5169.
- GRIMALT, J. O., P. FERNANDEZ, J. M. BAYONA, AND J. ALBAIGES. 1990. Assessment of fecal sterols and ketones as

- indicators of urban sewage inputs to coastal waters. *Environmental Science and Technology* 24:357–363.
- HEDGES, J. I., AND F. G. PRAHL. 1993. Early diagenesis: consequences for applications of molecular biomarkers. Pages 237–253 in M. H. Engel and S. A. Macko (editors). *Organic geochemistry*. Plenum Press, New York.
- HELSEL, D. R. 1990. Less than obvious. *Environmental Science and Technology* 24:1767–1774.
- JONGMAN, R. H., C. J. F. TER BRAAK, AND O. F. R. VAN TONGEREN, (EDITORS). 1995. *Data analysis in community and landscape ecology*. Cambridge University Press, Cambridge, UK.
- KAPLAN, L. A., J. D. NEWBOLD, D. J. VAN HORN, C. L. DOW, A. K. AUFDENKAMPE, AND J. K. JACKSON. 2006. Organic matter transport in New York City drinking-water-supply watersheds. *Journal of the North American Benthological Society* 25:912–927.
- KOLPIN, D. W., E. T. FURLONG, M. T. MEYER, E. M. THURMAN, S. D. ZAUGG, L. B. BARBER, AND H. T. BUXTON. 2002. Pharmaceuticals, hormones, and other organic wastewater contaminants in U.S. Streams, 1999–2000: a national reconnaissance. *Environmental Science and Technology* 36:1202–1211.
- KRATZER, E. B., J. K. JACKSON, D. B. ARSCOTT, A. K. AUFDENKAMPE, C. L. DOW, L. A. KAPLAN, J. D. NEWBOLD, AND B. W. SWEENEY. 2006. Macroinvertebrate distribution in relation to land use and water chemistry in New York City drinking-water-supply watersheds. *Journal of the North American Benthological Society* 25:954–976.
- LEEMING, R., A. BALL, N. ASHBOLT, AND P. NICHOLS. 1996. Using faecal sterols from humans and animals to distinguish faecal pollution in receiving waters. *Water Research* 30:2893–2900.
- LEEMING, R., AND P. NICHOLS. 1996. Concentrations of coprostanol that correspond to existing bacterial indicator guideline limits. *Water Research* 30:2997–3006.
- LEEMING, R., P. D. NICHOLS, AND N. J. ASHBOLT. 1998. Distinguishing sources of faecal pollution in Australian inland and coastal waters using sterol biomarkers and microbial faecal indicators. Research Report No. 204. Water Services Association of Australia, Melbourne, Australia. (Available from: <http://www.fwr.org/wsaa/wsaa204.htm>)
- MIGUEL, A. H., T. W. KIRCHSTETTER, R. A. HARLEY, AND S. V. HERING. 1998. On-road emissions of particulate polycyclic aromatic hydrocarbons and black carbon from gasoline and diesel vehicles. *Environmental Science and Technology* 32:450–455.
- MYERS, R. H. 1990. *Classical and modern regression with applications*. 2nd edition. PWS-Kent Publishing Company, Boston, Massachusetts.
- NETER, J., M. H. KUTNER, C. J. NACHTSHEIM, AND W. WASSERMAN. 1996. *Applied linear regression models*. 3rd edition. IRWIN/McGraw-Hill Professional Publishing, Chicago, Illinois.
- NIELSEN, T. 1996. Traffic contribution of polycyclic aromatic hydrocarbons in the center of a large city. *Atmospheric Environment* 30:3481–3490.
- NYS DEC (NEW YORK STATE DEPARTMENT OF ENVIRONMENTAL CONSERVATION). 1998. Ambient water quality standards and guidance values and groundwater effluent limitations. Division of Water Technical and Operational Guidance Series (1.1.1). New York State Department of Environmental Conservation, Albany, New York. (Available from: http://www.dec.state.ny.us/website/dow/togs/tog_cont.htm)
- O'LEARY, T., R. LEEMING, P. D. NICHOLS, AND J. K. VOLKMAN. 1999. Assessment of the sources, transport and fate of sewage-derived organic matter in Port Phillip Bay, Australia, using the signature lipid coprostanol. *Marine and Freshwater Research* 50:547–556.
- PHILLIPS, P. J., AND R. W. BODE. 2002. Concentrations of pesticides and pesticide degradates in the Croton River watershed in southeastern New York, July–September 2000. WRIR 02–4063. US Geological Survey, Denver, Colorado.
- PHILLIPS, P. J., B. STINSON, S. D. ZAUGG, E. T. FURLONG, D. W. KOLPIN, K. M. ESPOSITO, B. BODNIEWICZ, R. PAPE, AND J. ANDERSON. 2005. A multi-disciplinary approach to the removal of emerging contaminants in municipal wastewater treatment plants in New York State, 2003–2004. Pages 5095–5124 in *Proceedings of the 78th Annual Water Environment Federation Annual Technical Exhibition and Conference (WEFTEC) Washington, D.C., Oct 29–Nov 2*. Water Environment Federation, Alexandria, Virginia. (Available from: <http://www.wef.org/ScienceTechnologyResources/ConferenceProceedings/TablesOfContents/2005.htm>)
- PORTER, P. S., R. C. WARD, AND H. F. BELL. 1988. The detection limit. *Environmental Science and Technology* 22:856–861.
- SIMÓ, R., J. O. GRIMALT, AND J. ALBAIGÉS. 1997. Loss of unburned-fuel hydrocarbons from combustion aerosols during atmospheric transport. *Environmental Science and Technology* 31:2697–2700.
- SIMONICH, S. L., W. M. BEGLEY, G. DEBAERE, AND W. S. ECKHOFF. 2000. Trace analysis of fragrance materials in wastewater and treated wastewater. *Environmental Science and Technology* 34:959–965.
- SIMONICH, S. L., T. W. FEDERLE, W. S. ECKHOFF, A. ROTTIERS, S. WEBB, D. SABALIUNAS, AND W. DE WOLF. 2002. Removal of fragrance materials during U.S. and European wastewater treatment. *Environmental Science and Technology* 36:2839–2847.
- STANDLEY, L. J., L. A. KAPLAN, AND D. SMITH. 2000. Molecular tracers of organic matter sources to surface water resources. *Environmental Science and Technology* 34:3124–3130.
- STEIN, E. D., L. L. TIEFENTHALER, AND K. SCHIFF. 2006. Watershed-based sources of polycyclic aromatic hydrocarbons in urban storm water. *Environmental Toxicology and Chemistry* 24:373–385.
- TAYLOR, J. R. 1997. *An introduction to error analysis: the study of uncertainties in physical measurements*. 2nd edition. University Science Books, Sausalito, California.
- TER BRAAK, C. J. F., AND P. ŠMILAUER. 1998. CANOCO reference manual and user's guide to Canoco for Windows:

- software for canonical community ordination (version 4). Microcomputer Power, Ithaca, New York.
- THIOULOUSE, J., D. CHESSEL, S. DOLÉDEC, AND J.-M. OLIVIER. 1997. ADE-4: a multivariate analysis and graphical display software. *Statistics and Computing* 7:75–83.
- USEPA (US ENVIRONMENTAL PROTECTION AGENCY). 1985. Guidelines for deriving numerical criteria for the protection of aquatic life. EPA PB85–227049. Office of Research and Development, US Environmental Protection Agency, Duluth, Minnesota.
- USEPA (US ENVIRONMENTAL PROTECTION AGENCY). 1996. Method 625: base neutrals and acids. Pages 862–883 in *Methods for organic chemical analysis of municipal and industrial wastewater*. EPA-821-B-96–005. Office of Water, US Environmental Protection Agency, Washington, DC. (Available from: <http://www.epa.gov/waterscience/methods/guide/methods.html>, <http://nepis.epa.gov/pubtitleOW.htm>)
- USEPA (US ENVIRONMENTAL PROTECTION AGENCY). 1998. Method 8270d: semivolatile organic compounds by gas chromatography/mass spectrometry (GC/MS). In *Test methods for evaluating solid waste, physical/chemical methods (SW-846)*. US Environmental Protection Agency, Washington, DC.
- USEPA (US ENVIRONMENTAL PROTECTION AGENCY). 2002a. 2002 edition of the drinking water standards and health advisories. EPA 822-R-02–038. Office of Water, Office of Science and Technology, US Environmental Protection Agency, Washington, DC.
- USEPA (US ENVIRONMENTAL PROTECTION AGENCY). 2002b. National recommended water quality criteria: 2002. EPA-822-R-02–047. Office of Water, Office of Science and Technology, US Environmental Protection Agency, Washington, DC.
- USEPA (US ENVIRONMENTAL PROTECTION AGENCY). 2004. Current national recommended water quality criteria. EPA 4304T. Office of Water, Office of Science and Technology, US Environmental Protection Agency, Washington, DC.
- USGS (US GEOLOGICAL SURVEY) 1999. New reporting procedures based on long-term method detection levels and some considerations for interpretations of water-quality data provided by the U.S. Geological Survey National Water Quality Laboratory. Open-File Report 99–193. US Geological Survey, Reston, Virginia.
- VOGEL, J. R., L. B. BARBER, E. T. FURLONG, T. B. COPLEN, I. M. VERSTRAETEN, AND M. T. MEYER. 2005. Occurrence of selected pharmaceutical and non-pharmaceutical compounds, and stable hydrogen and oxygen isotope ratios in a riverbank filtration study, Platte River, Nebraska, 2002 to 2005. Volume 2. Data Series 141. US Geological Survey, Reston, Virginia.
- YUNKER, M. B., R. W. MACDONALD, R. VINGARZAN, R. H. MITCHELL, D. GOYETTE, AND S. SYLVESTRE. 2002. PAHs in the Fraser River basin: a critical appraisal of PAH ratios as indicators of PAH source and composition. *Organic Geochemistry* 33:489–515.
- ZAKARIA, M. P., H. TAKADA, S. TSUTSUMI, K. OHNO, J. YAMADA, E. KOUNO, AND H. KUMATA. 2002. Distribution of polycyclic aromatic hydrocarbons (PAHs) in rivers and estuaries in Malaysia: a widespread input of petrogenic PAHs. *Environmental Science and Technology* 36:1907–1918.

Received: 7 February 2006

Accepted: 25 July 2006