Supplementary Information for:

Application of Non-Target High Resolution Mass Spectrometry Data to Quantitative Source Apportionment

Katherine T. Peter[1,2*], Christopher Wu[1,2], Zhenyu Tian[1,2], Edward P. Kolodziej[1,2,3]

[1] Interdisciplinary Arts and Science, University of Washington Tacoma, Tacoma WA, 98421 USA
[2] Center for Urban Waters, Tacoma WA, 98421 USA
[3] Department of Civil and Environmental Engineering, University of Washington, Seattle WA, 98195 USA

Revised October 10, 2019 for
Environmental Science & Technology

10 pages, 7 tables, and 2 figures

*Corresponding author contact information
Email: ktpeter@uw.edu
Mail: Center for Urban Waters
326 E. D Street, Tacoma, WA 98421

Table S1. Composition and relative concentration of roadway runoff sources and creek water in each of the nine mixtures representing the model watershed.
Table S2. Isotopically labeled internal standards.
Table S3. Profinder data extraction settings.
Table S4. Source concentration estimate results for various method iterations.
Table S5. Results of SR520 source concentration estimates in single-source mixtures for various iterations of the source dilution curve range. These data compare estimates using only compounds within the peak area range of the modified dilution curve to estimates made using all qualified compounds, regardless of whether the mixture peak area fell within the peak area range of the dilution curve. The numbers of non-target compounds falling within, above, and below the dilution curve peak area range are also provided.
Table S6. Median peak area response observed for 12 isotopically labeled internal standards in the SR520 sample dilution curve.
Table S7. Quantitative estimates for SR520 roadway runoff concentration (as % v/v) using source-specific chemical surrogates identified within the non-target data set. The identified surrogates met the SR520 sample dilution curve screening criteria, were detected in all single-source watershed mixtures, and were absent in the SR518 runoff.

Figure S1. Hierarchical cluster diagram of SR518 and SR520 road runoff.
Figure S2. SR520 sample dilution curves for identified compounds.

Chemicals. Hexa(methoxymethyl)melamine (HMMM; 95%, CAS 3089-11-0), 1,3-diphenylguanidine (DPG; 97%, CAS 102-06-7), N-methyl-dicyclohexylamine (97%, CAS 7560-83-0), 1-cyclohexyl-3-phenylurea (97%, CAS 886-59-9), N-cyclohexylbenzothiazolamine (95%, CAS 28291-75-0), cotinine (≥98%, CAS 486-56-6), carbamazepine (≥99%, CAS 298-46-4), prometryn (99%, CAS 7287-16-6), theobromine-d3 (98%, CAS 117490-40-1), and sulfadimethoxine-d6 (≥98%, CAS 73068-02-7) were purchased from Sigma Aldrich (St. Louis, MO, USA). Nicotine-d3 (>98%, CAS 69980-24-1), sulfamethoxazole-d4 (>98%, CAS 1020719-86-1), atrazine-d5 (>98%, CAS 163165-75-1), DEET-d7 (>98%, CAS 1219799-37-7), prometon-d3 (>98%, CAS 1219803-43-6), metolachlor-d6 (>98%, CAS 1219803-97-0), and bis(2-ethylhexyl)phthalate-d4 (>98%, CAS 93951-87-2) were purchased from CDN Isotopes (Pointe-Claire, Quebec, Canada). Cotinine-d3 (>98%, CAS 110952-70-0) and carbamazepine-d10 (>99%, CAS 132183-78-9) were purchased from Cerilliant (Round Rock, TX, USA). 5-methyl-1H-benzotriazole-d6 (>98%, CAS 1246820-65-4) was purchased from Toronto Research Chemicals (North York, Ontario, Canada). Acetic acid (>99.7%) and ammonium acetate (HPLC grade, 97.8%) were purchased from VWR Scientific (Radnor, PA, USA). Methanol (MeOH: OPTIMA® grade) and Triton X-100 (octylphenol ethoxylate, scintillation grade; Research Products International) were obtained from Fisher Scientific (Fair Lawn, NJ, USA). A water purification system (Thermo Barnstead Nanopure Diamond UV, Dubuque, IA, USA) provided 18 MΩ water.

Identifications. Identifications efforts relied on formula assignment in Agilent Qualitative Analysis (B.08) and comparison of MS/MS spectra to online databases (METLIN, mzCloud) and/or reference standards. If reference standards were available, retention times were also compared. Identification confidence was assigned according to criteria proposed by Schymanski *et al,* where level S1 indicates a reference standard match and level S2 indicates a match to MS/MS database (S2a) and/or expert assessment of MS/MS fragmentation patterns (S2b). Level S4 indicates confident assignment of the formula.[1]

Table S1. Composition and relative concentration of roadway runoff sources and creek water in each of the nine mixtures representing the model watershed.

Mix	% SR520	% Coulter Creek	% Kautz Creek	% Crescent Valley Creek #1 (Dec)	% Swan Creek #1 (Dec)	% SR518	% Crescent Valley Creek #2 (May)	% Swan Creek #2 (May)
1	30	70	--	--	--	--	--	--
2	18	42	40	--	--	--	--	--
3	10	23	22	45	--	--	--	--
4A	4	9	9	18	60	--	--	--
5A	1	2.3	2.2	4.5	15	--	75	--
6A	0.16	0.37	0.35	0.72	2.4	--	12	84
4B	4	9	9	18	50	10	--	--
5B	1	2.3	2.2	4.5	12.5	2.5	75	--
6B	0.16	0.37	0.35	0.72	2	0.4	12	84

Table S2. Isotopically labeled internal standards used for QA/QC of non-target HRMS data, intended to represent a range of physico-chemical properties (e.g., polarity, functional groups) and a corresponding range of retention times, ionization efficiencies, and analytical responses (to the extent possible).

Compound	Formula	Retention Time [min]	Concentration [ng/mL]
Nicotine-d3	$C_{10}H_{11}D_3N_2$	1.89	500
Theobromine-d6	$C_7H_2D_6N_4O_2$	3	200
Cotinine-d3	$C_{10}H_9D_3N_2O$	3.44	100
Carbamazepine-d10	$C_{15}H_2D_{10}N_2O$	6.4	25
Sulfamethoxazole-d4	$C_{10}H_7D_4N_3O_3S$	4.28	100
Sulfadimethoxine-d6	$C_{12}H_8D_6N_4O_4S$	5	100
5-methyl-1H-benozotriazole-d6	$C_7HD_6N_3$	5	100
Atrazine-d5	$C_8H_9D_5ClN_5$	5	100
DEET-d7	$C_{12}H_{10}D_7NO$	7.19	100
Prometon-d3	$C_{10}H_{16}D_3N_5O$	7.97	100
Metolachlor-d6	$C_{15}H_{16}D_6ClNO_2$	9.9	100
Bis(2-ethylhexyl)phthalate-d4	$C_{24}H_{34}D_4O_4$	16.8	100

Table S3. Profinder data extraction settings.

Profinder Batch Recursive Feature Extraction Wizard Step	Category	Setting
Molecular Feature Extraction (MFE) – Extraction Parameters	Noise peak height	>= 300 counts
	Ion Species	+H, +Na, +NH4 (salt-dominated positive ions)
	Isotope model	Common organic molecules
	Limit assigned charge states to	1-2
MFE - Compound Filters	Compound ion count threshold	Two or more ions
Compound Binning and Alignment	RT window	0.6 min
	Mass window	30 ppm
MFE – Post-Processing Filters	Absolute height	>= 5000 counts
	Score (MFE)	>= 70
	Satisfy MFE conditions in	>= 2 files
Find by Ion – Matching Tolerances and Scoring	EIC tolerance, possible m/z	Symmetric, +/- 50 ppm
	EIC tolerance, expected RT range	+/- 1.5 min
	Expected data variation	MS mass 2.0 mDa + 10 ppm or 15%
	Scoring	Mass score (100), isotope abundance score (60), isotope spacing score (50), retention time score (50)
	Do not match if overall score	< 30
Find by Ion – EIC Peak Integration and Filtering	Integration	Agile 2
	Peak height	>= 3000 counts
Find by Ion – Post-Processing Filters	Absolute height	>= 1000 counts
	Score (Tgt)	>= 50
	Satisfy Find by Ion conditions in	>= 2 files
Export to MPP	Exported value override	Export median RT/mass instead of measured RT/mass
		Export missing values as missing
	File Type	Profinder Archive

Table S4. Source concentration estimate results for various method iterations, with results shown as the ratio of estimated *vs.* actual percent source concentration. Estimates were made with (a) SR520 sample dilution curve, (b) SR520 extract dilution curve, (c) SR518 extract dilution curve, or (d) targeted contaminants. The number of non-target HRMS compounds used to derive each estimate are shown in parentheses and italicized. Mixtures in which targeted contaminants were not detected are indicated by an asterisk.

Dilution Curve Used:			**(a) SR520 Sample Dilution**			**(b) SR520 Extract Dilution**		**(c) SR518 Extract Dilution**		**(d) Targeted Contaminants, from SR520 Sample Dilution Curve[4]**						
Compound Group Used:			All Non-target Compounds[1]	All Non-target Compounds, Exclude outliers[2]	SR520 Unique Compounds, Exclude outliers[3]	All Non-target Compounds, Exclude outliers[2]	SR520 Unique Compounds, Exclude outliers[3]	All Non-target Compounds, Exclude outliers[2]	SR518 Unique Compounds, Exclude outliers[3]	DPG	HMMM	Di-F HMMM	NCBA	HA	DCMA	CPU
Mix	Actual % SR520	Actual % SR518	Ratio of Estimated to Actual % SR520 *(# compounds used)*					Ratio of Estimated to Actual % SR-518 *(# compounds*		Ratio of Estimated to Actual % SR520						
1	30		1.04 *(429)*	1.04 *(401)*	--	0.927 *(447)*	--	--	--	1.2	1.1	1.0	1.1	1.1	1.2	0.98
2	18		1.11 *(369)*	1.11 *(354)*	--	0.989 *(386)*	--	--	--	1.3	1.1	1.2	1.2	1.2	1.1	1.1
3	10		1.00 *(241)*	0.99 *(225)*	--	0.87 *(278)*	--	--	--	1.2	1.1	1.1	0.95	1.2	1.1	1.1
4A	4		1.13 *(177)*	1.13 *(166)*	--	1.0 *(185)*	--	--	--	1.5	1.2	1.4	0.9	1.6	1.6	1.2
5A	1		0.91 *(28)*	0.82 *(23)*	--	1.2 *(32)*	--	--	--	*	0.57	0.78	*	1.2	1.8	0.66
6A	0.16		2.13 *(15)*	1.38 *(12)*	--	1.4 *(14)*	--	--	--	*	0.38	0.69	*	*	*	*
4B	4	10	1.65 *(146)*	1.43 *(132)*	1.05 *(50)*	1.5 *(150)*	1.1 *(57)*	3 *(31)*	1.39 *(5)*	1.8	0.68	0.83	0.68	2.0	3.1	1.8
5B	1	2.5	2.30 *(54)*	1.80 *(44)*	1.40 *(13)*	3.1 *(70)*	2.9 *(25)*	2.28 *(11)*	0.96 *(1)*	*	0.50	0.62	*	2.1	3.4	0.8
6B	0.16	0.4	1.25 *(13)*	1.13 *(12)*	0.938 *(5)*	1.1 *(10)*	0.813 *(3)*	72 *(8)*	0.925 *(1)*	*	*	0.56	*	*	*	*

[1]Estimate made with all non-target compounds that met screening criteria in source dilution curve and watershed mixture.
[2]See note 1, with outlier estimates excluded (outliers were >1.5 times the interquartile range above or below the third or first quartile, respectively).
[3]See notes 1 and 2, with compounds also detected in the secondary source excluded.
[4]Compounds are 1,3-diphenylguanidine (DPG), hexa(methoxymethyl)melamine (HMMM), Di-formylated (Di-F) HMMM, N-cyclohexylbenzothiazolamine (NCBA), hexylamine (HA), N,N-dicyclohexylmethylamine (DCMA), and 1-cyclohexyl-3-phenylurea (CPU). Peak area data for source estimation were from the SR520 sample dilution curve.

Table S5. Results of SR520 source concentration estimates in single-source mixtures for various iterations of the source dilution curve range. These data compare estimates using only compounds within the peak area range of the modified dilution curve to estimates made using all qualified compounds, regardless of whether the mixture peak area fell within the peak area range of the dilution curve. The number of non-target compounds within, above, and below the dilution curve peak area range are also provided.

SR520 Sample Dilution Curve (concentration points)	Mixture	Actual % SR520	Number of Non-target Compounds Within | Above | Below Dilution Curve Peak Area (PA) Range			Estimated % SR520, using compounds that are:	
						Within PA range	Within, above, & below PA range
Full Range (0.1, 1, 2.5, 6, 16, 40, 100)	**1**	30	429	0	6	31.3	31.3
	2	18	369	1	54	20.0	19.4
	3	10	241	0	82	10	9.5
	4A	4	177	1	55	4.5	4.5
	5A	1	28	0	40	0.91	0.75
	6A	0.16	15	0	29	0.3	0.3
Remove top 1 (0.1, 1, 2.5, 6, 16, 40)	**1**	30	286	25	0	27.7	28.2
	2	18	310	2	2	18.6	18.6
	3	10	274	0	10	9.5	9.4
	4A	4	202	0	37	4.5	4.5
	5A	1	34	1	41	0.85	0.78
	6A	0.16	14	0	27	0.2	0.2
Remove top 2 (0.1, 1, 2.5, 6, 16)	**1**	30	16	227	2	9.5	27.9
	2	18	96	154	2	15.3	18.1
	3	10	230	4	3	9.1	9.1
	4A	4	188	0	15	4.4	4.3
	5A	1	34	1	43	0.9	0.76
	6A	0.16	15	0	27	0.2	0.2
Remove bottom 2 (1, 2.5, 6, 16, 40, 100)	**1**	30	430	0	6	31.3	31.2
	2	18	368	1	55	20.0	19.4
	3	10	240	0	83	10	9.5
	4A	4	177	1	55	4.5	4.5
	5A	1	20	0	49	1.3	0.77
	6A	0.16	6	0	38	3.2	0.4
Remove bottom 1 (2.5, 6, 16, 40, 100)	**1**	30	430	0	6	31.3	31.2
	2	18	368	1	55	19.9	19.3
	3	10	240	0	83	9.8	9.4
	4A	4	171	1	61	4.4	4.3
	5A	1	7	0	62	25.5	0.66
	6A	0.16	3	0	41	7.7	0.4
Remove 1 mid-point (0.1, 1, 2.5, 6, 40, 100)	**1**	30	425	0	11	30.3	30.2
	2	18	351	1	72	19.2	18.4
	3	10	196	0	127	9	8.7
	4A	4	159	1	73	4.1	4.1
	5A	1	29	1	40	0.99	0.73
	6A	0.16	15	0	29	0.31	0.3

Table S6. Median peak area response observed for 12 isotopically labeled internal standards in the SR520 sample dilution curve.

% SR520 (Sample Dilution) / ISTD	100%	40%	16%	6%	2.50%	1%	0.10%
Nicotine-d3	83%	85%	94%	88%	96%	97%	98%
Theobromine-d6	34%	30%	25%	18%	26%	24%	24%
Cotinine-d3	75%	83%	94%	89%	96%	97%	98%
Carbamazepine-d10	82%	136%	144%	104%	108%	112%	115%
Sulfamethoxazole-d4	53%	70%	85%	86%	94%	97%	98%
Sulfadimethoxine-d6	21%	40%	58%	72%	79%	90%	97%
5-methyl-1H-benozotriazole-d6	73%	83%	94%	95%	100%	102%	103%
Atrazine-d5	42%	75%	84%	86%	95%	94%	94%
DEET-d7	35%	51%	69%	83%	89%	92%	92%
Prometon-d3	36%	52%	74%	80%	91%	96%	99%
Metolachlor-d6	48%	79%	87%	85%	91%	93%	94%
Bis(2-ethylhexyl)phthalate-d4	64%	81%	95%	80%	100%	104%	95%
Median ISTD Peak Area Response	51%	77%	86%	85%	94%	97%	97%

Table S7. Quantitative estimates for SR520 roadway runoff concentration (as % v/v) using source-specific chemical surrogates identified within the non-target data set. The identified surrogates met the SR520 sample dilution curve screening criteria, were detected in all single-source watershed mixtures, and were absent in the SR518 runoff.

		Estimated % SR520 (*via* sample dilution curve)					
Mass @ RT	**Identity**	Mix 1	Mix 2	Mix 3	Mix 4A	Mix 5A	Mix 6A
282.1453@6.77 min	HMMM in-source fragment	32	25	12	5.4	1.14	0.18
282.1732@8.55 min	$C_{18}H_{22}N_2O$	25	15	11	7.4	1.02	0.09
420.2349@7.21 min	Di-formylated HMMM	30	21	11	5.6	0.78	0.11
772.5488@13.58 min	Tridecapropylene glycol	33	17	7	3.0	0.49	0.35
888.6326@14.54 min	Pentadecapropylene glycol	41	26	10	5.7	0.68	0.78
	Actual Source Concentration:	30	18	10	4	1	0.16

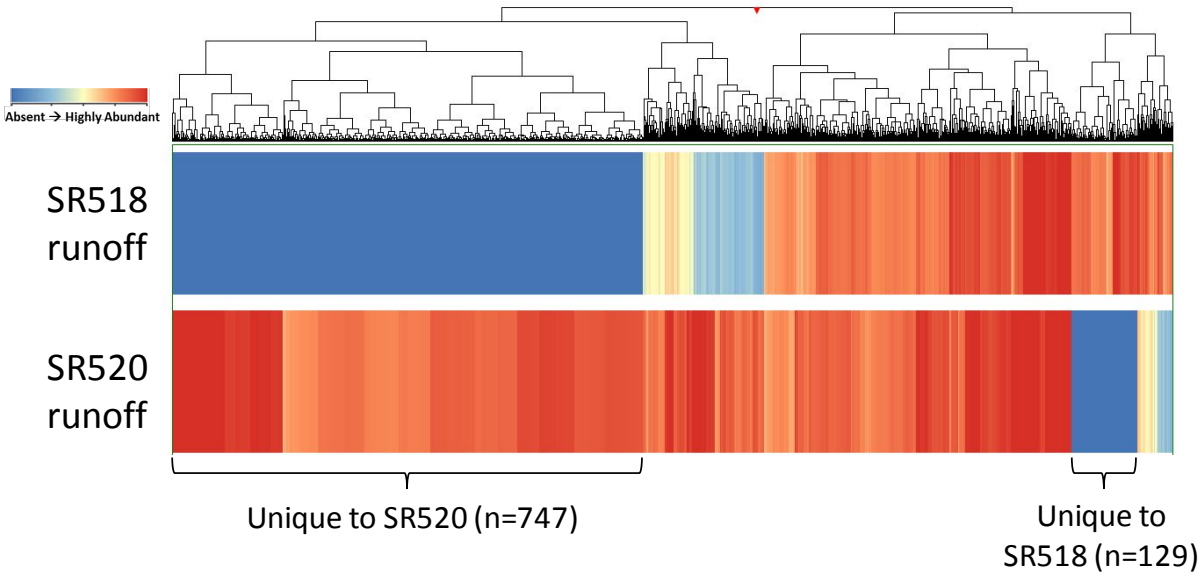

Figure S1. Hierarchical cluster diagram of SR518 and SR520 road runoff, where each vertical line represents an individual non-target compound and the color scheme represents peak area (abundance), increasing from absent = dark blue to light blue, yellow, peach, and red = maximum detected peak area. The features unique to each source are noted. Although only one sample of each roadway runoff was used in this study (representing 1 storm for SR518 or a composite of 2 storms for SR520), we note that based on our experience with sampling both SR520 and SR518 repeatedly for separate studies, the roadway runoff composition is surprisingly consistent across different seasons and storm event sizes.

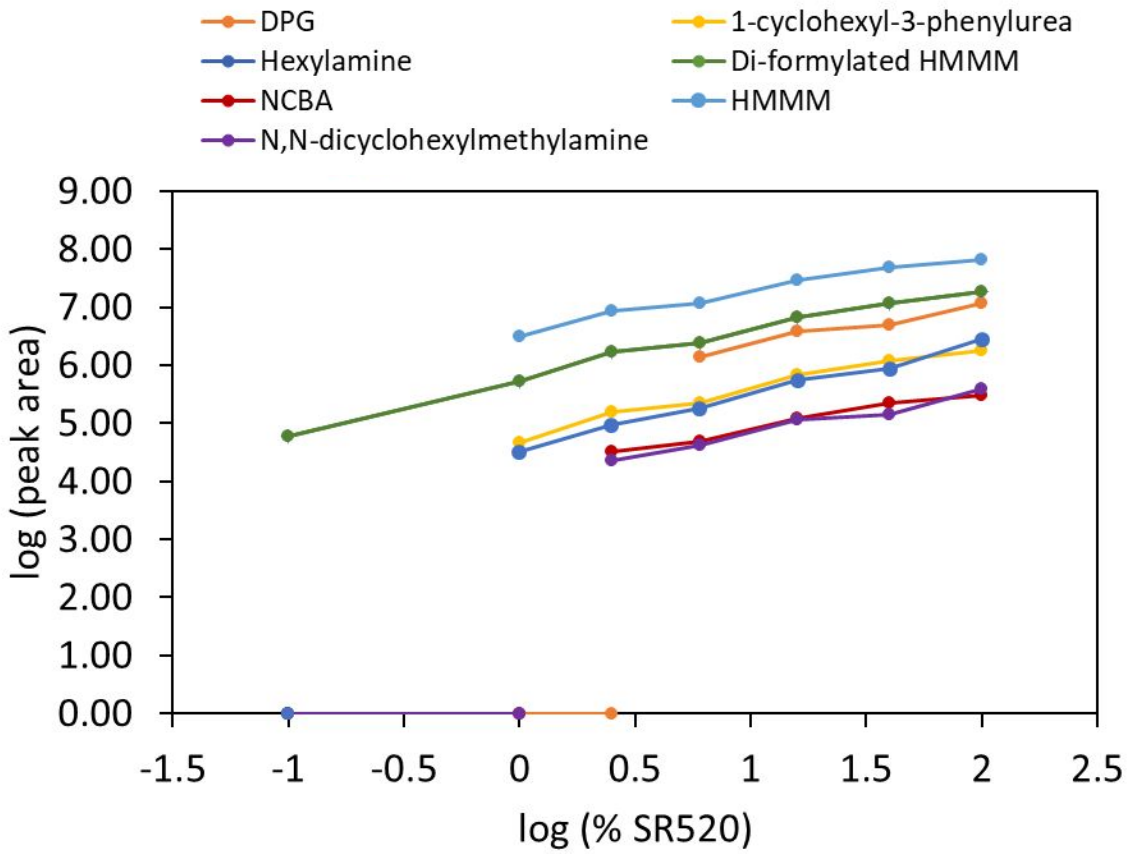

Figure S2. Log-scale SR520 sample dilution curves (peak area vs. SR520 concentration) for identified road runoff compounds, including diphenylguanidine (DPG), dicyclohexylmethylamine, N-cyclohexyl-benzothiazolamine (NCBA), hexa(methoxymethyl)melamine (HMMM), di-formylated HMMM, 1-cyclohexyl-3-phenylurea, and hexylamine.