Supplement of
Source Apportionment of VOCs, IVOCs, and SVOCs by Positive Matrix Factorization in Suburban Livermore, California

R. A. Wernis et al.

Correspondence to: Rebecca A. Wernis (rwernis@berkeley.edu)

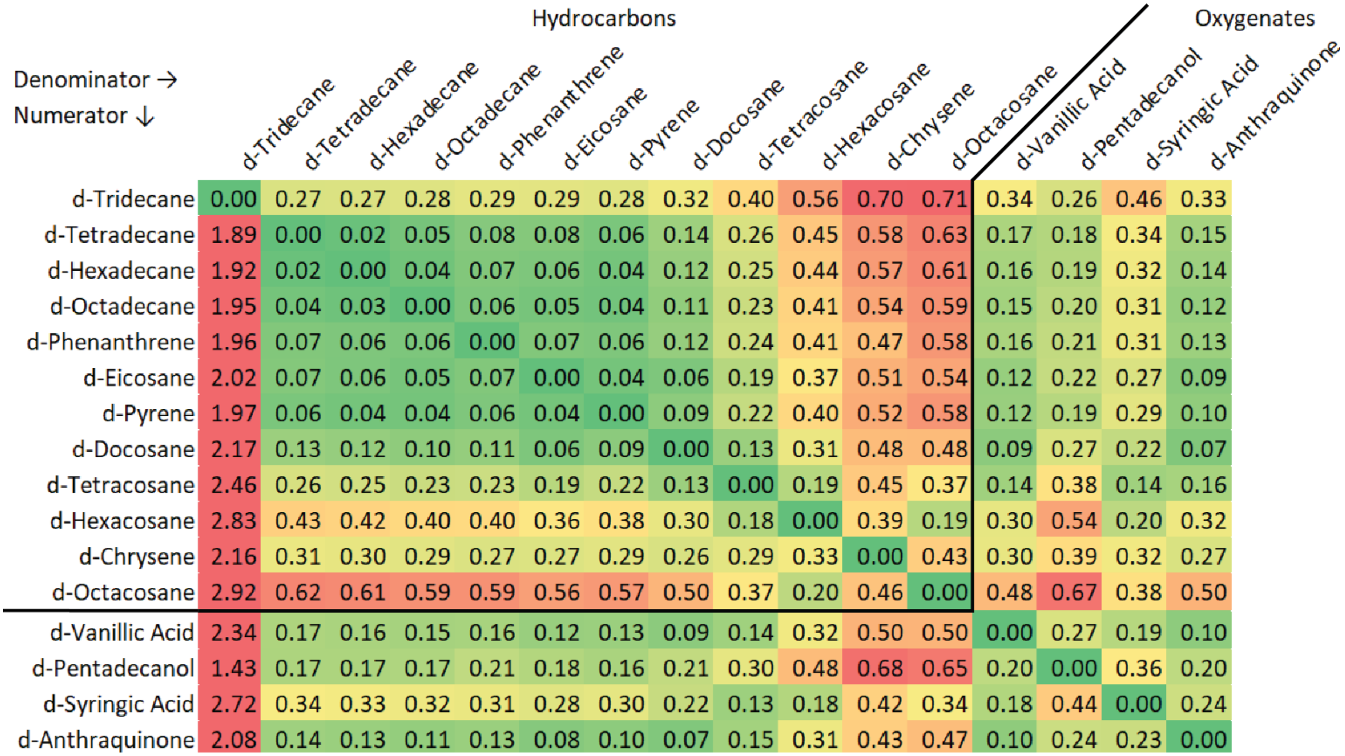
S1 Precision uncertainty histograms

Precision uncertainty is discussed in Sect. 2.2.3.2.

Figure S1 shows the relative standard deviations of the internal standard ratios for all internal standards used for normalization in this analysis. Hydrocarbons between *n*-tetradecane (C_{14}) and pyrene (elutes just before *n*-docosane (C_{22})) all correlate closely. Additionally, the distribution of internal standard ratios across samples is approximately normal (e.g., Fig. S2(a)). We thus infer that any ambient hydrocarbon observed in this volatility range could be normalized by any of the internal standards in this range with 10% or less precision uncertainty introduced. Though transfer efficiency suffers for hydrocarbons with lower volatility than *n*-docosane, leading to greater sensitivity of internal standard choice, 10% is still a satisfactory precision uncertainty for ambient compounds in this volatility range provided the analyte and internal standard retention indices differ by less than about 200. (In later deployments of cTAG, transfer efficiency remains high up to *n*-triacontane.) The best hydrocarbon pairings have a relative standard deviation of less than 5%. From this observation we conservatively estimate that compounds normalized by an isotopologue incur 5% precision uncertainty.

On the high volatility side, deuterated tridecane was lost in the refocusing step during occasional periods of high ambient temperature, causing its measured signal to drop near zero. Ambient compounds in this volatility range are therefore normalized by tetradecane with an assigned precision uncertainty of 25% based on the RSD of the tridecane-tetradecane ratio. The distribution of ratios is somewhat skewed (Fig. S2(b)); a few data points should likely be assigned a higher uncertainty and many could be assigned a lower uncertainty, but 25% represents an appropriate overall estimate for the most volatile compounds.

Oxygenated compounds exhibit greater RSD values regardless of whether they are paired with hydrocarbons or other oxygenates. Deuterated syringic acid, which elutes just before eicosane (C_{20}), produces an RSD of about 30% when paired with most other compounds, oxygenated or not, and an approximately normally distributed set of ratios (Fig. S2(c)). Ambient oxygenates are normalized by the nearest deuterated oxygenate in volatility if their retention indices are within 200 and the nearest hydrocarbon otherwise; in both cases the assigned precision uncertainty is 30%. Table S1 summarizes the categories of precision uncertainty assigned to ambient compounds for this analysis.



Denominator → Numerator ↓	Hydrocarbons												Oxygenates			
	d-Tridecane	d-Tetradecane	d-Hexadecane	d-Octadecane	d-Phenanthrene	d-Eicosane	d-Pyrene	d-Docosane	d-Tetracosane	d-Hexacosane	d-Chrysene	d-Octacosane	d-Vanillic Acid	d-Pentadecanol	d-Syringic Acid	d-Anthraquinone
d-Tridecane	0.00	0.27	0.27	0.28	0.29	0.29	0.28	0.32	0.40	0.56	0.70	0.71	0.34	0.26	0.46	0.33
d-Tetradecane	1.89	0.00	0.02	0.05	0.08	0.08	0.06	0.14	0.26	0.45	0.58	0.63	0.17	0.18	0.34	0.15
d-Hexadecane	1.92	0.02	0.00	0.04	0.07	0.06	0.04	0.12	0.25	0.44	0.57	0.61	0.16	0.19	0.32	0.14
d-Octadecane	1.95	0.04	0.03	0.00	0.06	0.05	0.04	0.11	0.23	0.41	0.54	0.59	0.15	0.20	0.31	0.12
d-Phenanthrene	1.96	0.07	0.06	0.06	0.00	0.07	0.06	0.12	0.24	0.41	0.47	0.58	0.16	0.21	0.31	0.13
d-Eicosane	2.02	0.07	0.06	0.05	0.07	0.00	0.04	0.06	0.19	0.37	0.51	0.54	0.12	0.22	0.27	0.09
d-Pyrene	1.97	0.06	0.04	0.04	0.06	0.04	0.00	0.09	0.22	0.40	0.52	0.58	0.12	0.19	0.29	0.10
d-Docosane	2.17	0.13	0.12	0.10	0.11	0.06	0.09	0.00	0.13	0.31	0.48	0.48	0.09	0.27	0.22	0.07
d-Tetracosane	2.46	0.26	0.25	0.23	0.23	0.19	0.22	0.13	0.00	0.19	0.45	0.37	0.14	0.38	0.14	0.16
d-Hexacosane	2.83	0.43	0.42	0.40	0.40	0.36	0.38	0.30	0.18	0.00	0.39	0.19	0.30	0.54	0.20	0.32
d-Chrysene	2.16	0.31	0.30	0.29	0.27	0.27	0.29	0.26	0.29	0.33	0.00	0.43	0.30	0.39	0.32	0.27
d-Octacosane	2.92	0.62	0.61	0.59	0.59	0.56	0.57	0.50	0.37	0.20	0.46	0.00	0.48	0.67	0.38	0.50
d-Vanillic Acid	2.34	0.17	0.16	0.15	0.16	0.12	0.13	0.09	0.14	0.32	0.50	0.50	0.00	0.27	0.19	0.10
d-Pentadecanol	1.43	0.17	0.17	0.17	0.21	0.18	0.16	0.21	0.30	0.48	0.68	0.65	0.20	0.00	0.36	0.20
d-Syringic Acid	2.72	0.34	0.33	0.32	0.31	0.28	0.30	0.22	0.13	0.18	0.42	0.34	0.18	0.44	0.00	0.24
d-Anthraquinone	2.08	0.14	0.13	0.11	0.13	0.08	0.10	0.07	0.15	0.31	0.43	0.47	0.10	0.24	0.23	0.00

Figure S1. Relative standard deviations of all internal standard ratios used in this study. Green = 0; red = 0.7 or greater.

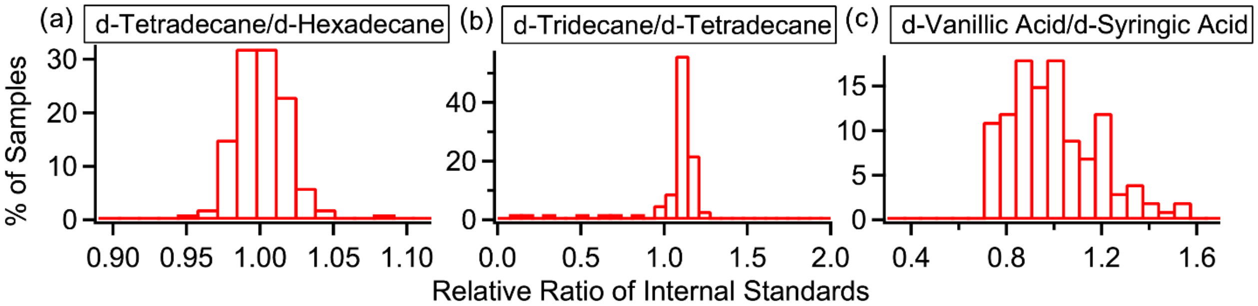


Figure S2. Example distribution of relative ratios of internal standards for (a) two hydrocarbons with a retention index difference of 200, (b) tridecane and tetradecane and (c) two oxygenated compounds.

Table S1. Summary of internal standard assignments for normalization of ambient compounds in this analysis and the precision uncertainty incurred from each assignment. Because the precision uncertainty is the only source of uncertainty that varies from sample to sample, it is the only uncertainty assigned to individual compounds as input to the PMF model.

Ambient Compound Category	Precision Uncertainty Assigned
Compounds with isotopologue internal standards	5%
Hydrocarbons with retention index at or above 1400 (no isotopologue)	10%
Compounds with retention index below 1400 (no isotopologue)	25%
Compounds that do not fall into any of the above categories, including most oxygenates	30%

S2 PMF evaluation of different factor solutions

S2.1 Q/Q_{exp}

A common consideration for deciding which solution best explains the data is to compare Q/Q_{exp} for solutions with different numbers of factors (Ulbrich et al., 2009). The expected value of Q, or Q_{exp}, is calculated for m compounds, n sample times and p factors as:

$$Q_{exp} = mn - p(m + n) \tag{S1}$$

Q/Q_{exp} is expected to decrease as the number of factors increases, as more of the data is able to be fit. A relatively large decrease in Q/Q_{exp} between successive solutions suggests the additional factor explains more of the data than would be expected and thus should be included. This effectively sets a minimum number of factors for the final chosen solution.

Q/Q_{exp} and the percent reduction in Q/Q_{exp} between successive solutions is shown in Fig. S3. Unusually large percent reductions are present with the addition of the 8th factor and, especially, with the addition of the 13th factor. This is one reason the 13 factor solution was chosen.

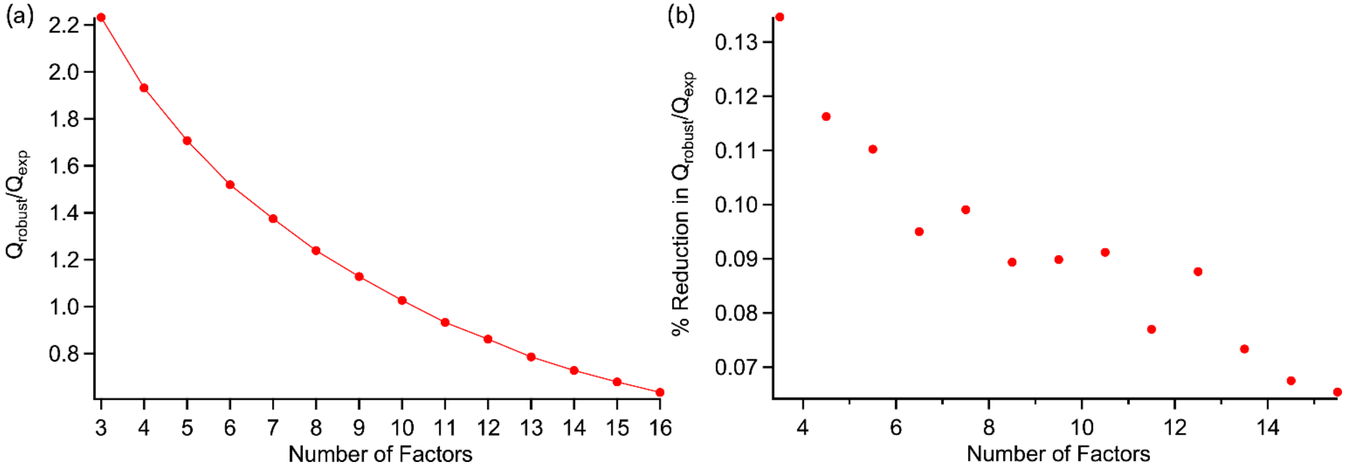


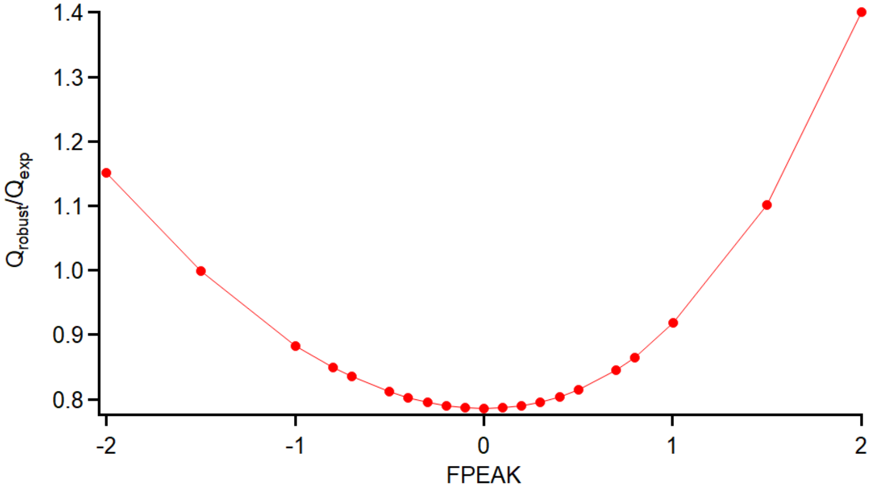
Figure S3. (a) Q/Q_{exp} for different numbers of factors. (b) % reduction in Q/Q_{exp} as the number of factors is increased.

S2.2 Beyond 13 factors: bootstrapping results

The 14 factor solution and beyond were ultimately rejected because the new factors contained low fractions of compounds from many factors, taking mass away from the other factors. Additionally, the factor timelines for 14th, 15th, and 16th factors were not very distinct from existing factors, as reflected in the decreasing average bootstrapping self-mapping fraction (89.77% for 13 factors, 88.21% for 14 factors, 86.07% for 15 factors and 87.19% for 16 factors). Finally, the additional factors had no clear physical interpretations that were distinct from factors present in the 13 factor solution.

S2.3 FPEAK rotation

FPEAK exploration is discussed in Sect. 3.1.



70 **Figure S4. Q/Q_{exp} for different values of the FPEAK parameter. The minimum value is at FPEAK = 0.**

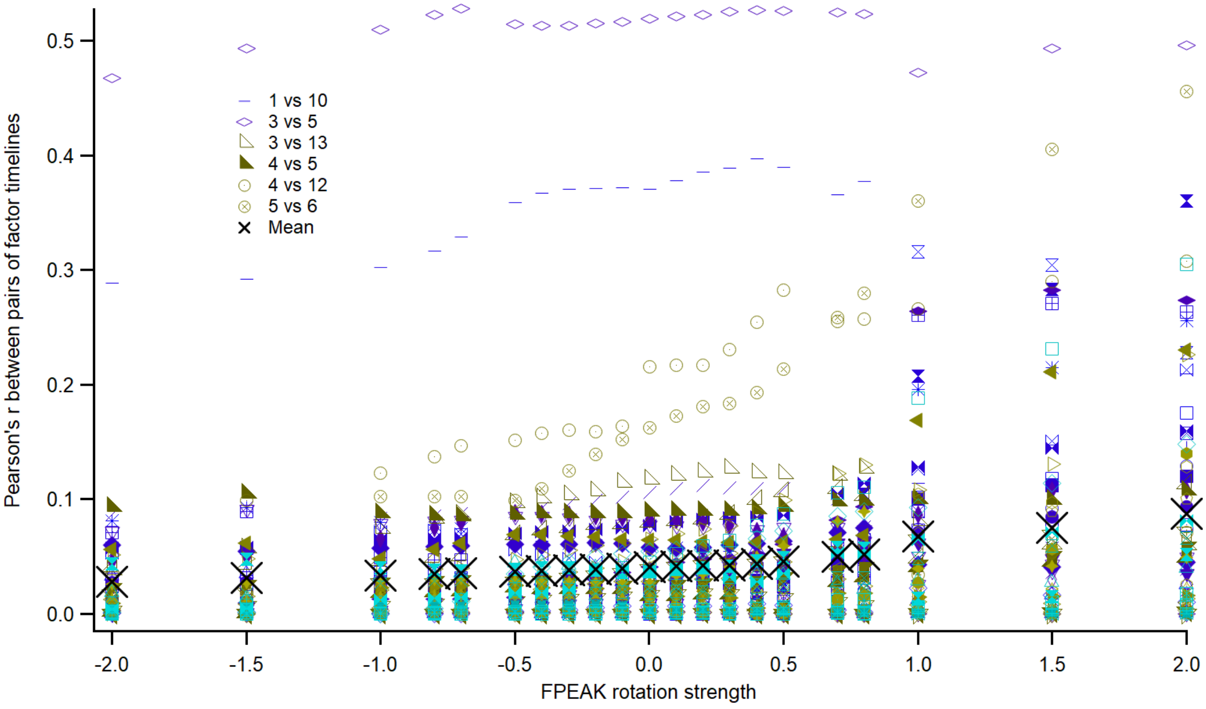


Figure S5. Pearson's r for the cross correlation between every pair of factors in the 13 factor solution, along with the mean r for each FPEAK value. Since there are a total of 78 pairings, only outliers are included in the legend. Negative values of FPEAK provide modest improvements (i.e. decreases) in r values, including the mean r, but not enough to outweigh the optimization of
75 **Q/Q_{exp} at FPEAK = 0 (Fig. S4).**

S3 Factor apportionment of VOCs, IVOCs and SVOCs

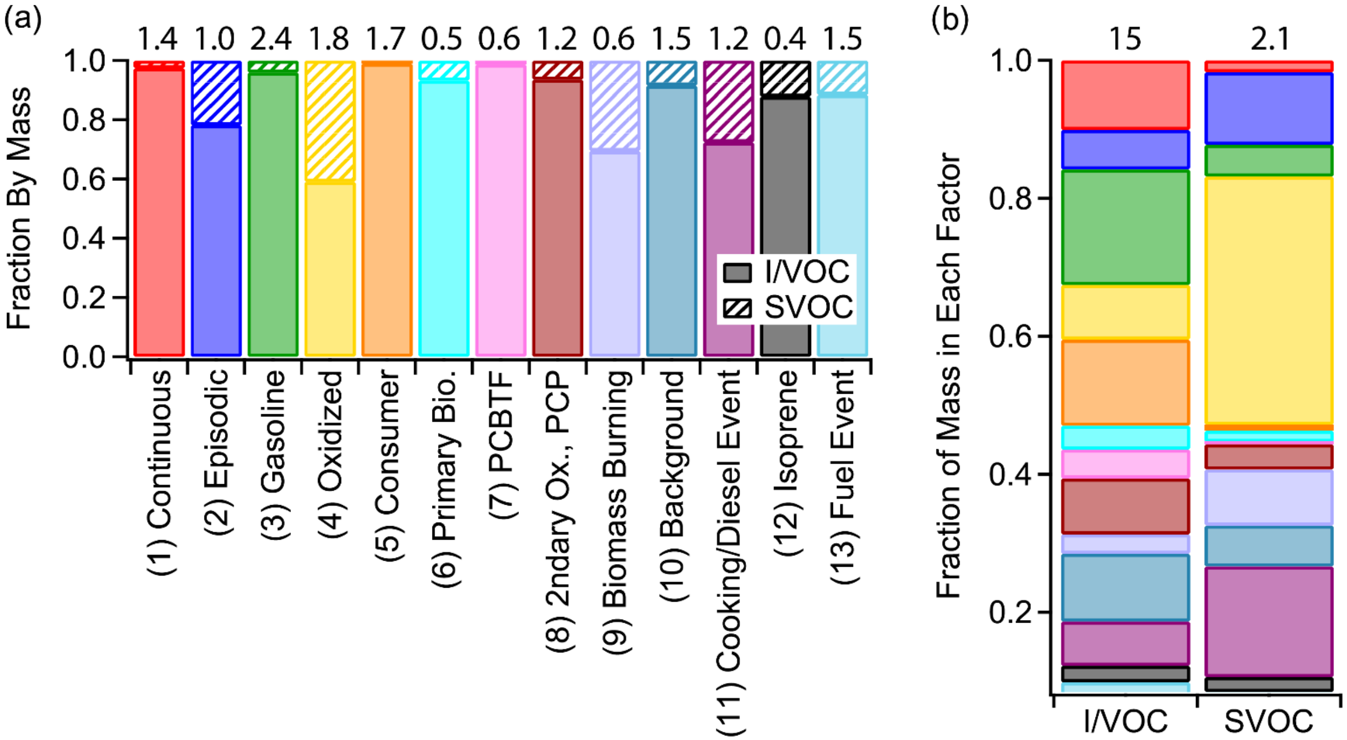


Figure S6. (a) Fraction of average mass from each cTAG measurement channel in each factor. (b) Fraction of the average total mass of compounds measured on each channel found in each factor. The numbers at the top represent the average total mass for (a) that factor or (b) that class of compounds in µg m^{-3}.

S4 Ratio of isoprene to benzene in gasoline exhaust

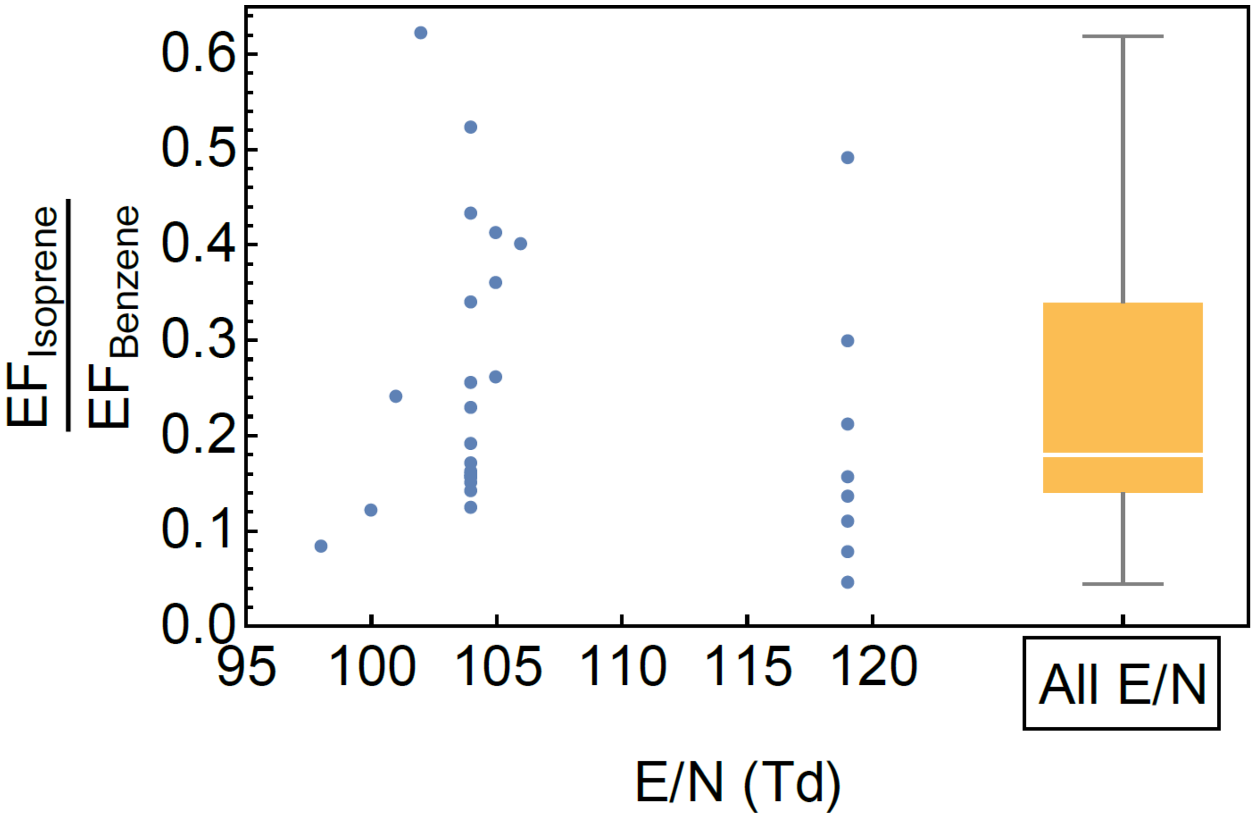


Figure S7. The ratio of isoprene to benzene emission factors (EF) (mg (kg Fuel)$^{-1}$) for cold-start emissions for a wide range of vehicle classes, showing a mean value of 0.18. Observed ratios are sorted by the E/N (electric field density ratio) used in measurements made by a quadrupole proton transfer reaction mass spectrometer. Vehicles span model year 1990 - 2014 and Pre-LEV to SULEV emissions categories. There is no trend with E/N, suggesting a minimal effect of substituted cyclohexanes on isoprene measurements (Gueneron et al., 2015).

References

Gueneron, M., Erickson, M. H., VanderSchelden, G. S., and Jobson, B. T.: PTR-MS fragmentation patterns of gasoline hydrocarbons, International Journal of Mass Spectrometry, 379, 97–109, https://doi.org/10.1016/j.ijms.2015.01.001, 2015.

Ulbrich, I. M., Canagaratna, M. R., Zhang, Q., Worsnop, D. R., and Jimenez, J. L.: Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data, Atmos. Chem. Phys., 9, 2891–2918, https://doi.org/10.5194/acp-9-2891-2009, 2009.