Supplementary Information: Portable, low-cost samplers for distributed sampling of reactive gases

James F. Hurley[1], Ksenia Onufrieva[2,3], Alejandra Caceres[4], Deborah F. McGlynn[1], Mary E. Tovillo[1], Roger Schürch[2], Suzanne Pinar[2], Gabriel Isaacman-VanWertz[1]

[1]The Charles E. Via, Jr. Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA 24061, USA
[2]Department of Entomology, Virginia Tech, Blacksburg, VA 24061, USA
[3]Center for Advanced Innovation in Agriculture, Virginia Tech, Blacksburg, VA 24061, USA
[4]Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061, USA

Correspondence to: Gabriel Isaacman-VanWertz (ivw@vt.edu)

S1 Sampler Precision and Reproducibility Data

As described in Sections 2.3.1 and 3.1, two sets of samplers, Set A (placed on lab bench, n=5) and Set B (placed beside a closed window, n=7), were run to assess sampler precision and detect any differences for a given analyte between the two sampling sets. After a Jarque-Bera Test was run to confirm normality, a two-tailed t-test, homoscedastic (two-sample equal variance) at 95% confidence was run in Excel to obtain a p-value. A p-value of less than 0.05 means the two sampling sets have a statistically significant difference for that analyte, and these compounds are marked with an asterisk * in the table. Table S1 shows that 20 of the 31 compounds were statistically different between the two sets. The table presents analytes in order of decreasing volatility, and in general, except for the latest eluting analytes, the Bench Set (Set A) tended to have higher concentrations. Strong intra-set precision (low % RSD) allows small differences in analyte concentration (e.g., 10-15%) between sets to become statistically significant.

Table S1. The mean and % relative standard deviation (%RSD) are given for the two sampling sets. Also included are the % increase in the Bench Set (Set A, n=5) versus the Window Set (Set B, n=7). Statistically different analytes have a p-value < 0.05 and are marked with an asterisk*.

Compound	Set A (Bench) (mean ± % RSD)	Set B (Window) (mean ± % RSD)	% Increase on Bench	p-Value
Benzene	1.028E8 ± 4.8%	9.635E7 ± 9.2%	6.7	0.174
1-Butanol	3.833E8 ± 7.9	2.485E8 ± 6.4	54.2	0.000 *
3-Methyl Hexane	1.040E7 ± 6.2	9.851E6 ± 3.9	5.5	0.095
Heptane	2.725E7 ± 6.2	2.524E7 ± 5.0	7.9	0.039 *
tert-Butyl Acetate	1.469E7 ± 5.6	1.351E7 ± 4.8	8.8	0.018 *
MIK (Methyl Isobutyl Ketone)	3.956E6 ± 4.5	3.588E6 ± 4.4	10.2	0.004 *
Propylene Glycol	2.930E7 ± 9.2	2.381E7 ± 5.8	23.1	0.001 *
Toluene	8.330E8 ± 6.8	8.036E8 ± 6.3	3.6	0.369
Octane	1.588E7 ± 4.9	1.422E7 ± 5.6	11.7	0.005 *
Tetrachloroethylene	2.160E7 ± 5.5	2.340E7 ± 6.8	-7.7	0.060
Butyl Acetate	2.356E7 ± 6.2	2.104E7 ± 3.6	11.9	0.003 *
Furfural	1.542E7 ± 5.1	1.330E7 ± 3.0	15.9	0.000 *
Chlorobenzene	1.022E7 ± 6.1	9.220E6 ± 2.8	10.9	0.003 *
Maleic Anhydride	7.758E7 ± 6.0	7.050E7 ± 4.4	10.0	0.010 *
Ethylbenzene	1.199E8 ± 11.1	1.069E8 ± 3.8	12.2	0.034 *
p-, o-Xylene	2.931E8 ± 10.3	2.665E8 ± 4.5	10.0	0.058
Phenylethyne	2.882E7 ± 3.4	2.603E7 ± 6.2	10.8	0.006 *
o-Xylene	1.187E8 ± 15.9	1.062 E8 ± 5.7	11.8	0.125
Nonane	8.958E6 ± 7.0	8.248E6 ± 6.2	8.6	0.055
2BE (2-Butoxy- ethanol)	3.070E6 ± 14.7	3.259E6 ± 22.6	-5.8	0.624
Butyrolactone	2.528E6 ± 6.6	2.226E6 ± 2.3	13.5	0.001 *
α-Pinene	3.510E7 ± 6.8	3.005E7 ± 4.4	16.8	0.001 *
TMB (Trimethyl-benzene)	7.510E7 ± 5.9	6.735E7 ± 3.9	11.5	0.003 *
2EH (2-Ethyl-hexanol)	2.054E7 ± 29.0	1.906E7 ± 24.1	7.7	0.637
Limonene	1.188E6 ± 8.9	1.080E6 ± 11.7	9.9	0.154
o-Cymene	1.428E7 ± 6.4	1.247E7 ± 4.7	14.5	0.002 *
Undecane	1.002E6 ± 6.6	9.449E5 ± 5.6	6.1	0.125
Dodecane	2.081E8 ± 28.5	3.119E8 ± 24.1	-33.3	0.028 *
1P2P (1-Phenoxy-2-Propanol)	3.595E8 ± 6.1	5.705E8 ± 5.0	-37.0	0.000 *
Tridecane	5.923E7 ± 7.5	1.011E8 ± 5.5	-41.4	0.000 *
Phthalic Anhydride	6.008E7 ± 4.2	5.328E7 ± 3.9	12.7	0.000 *

35

S2 Ozone Scrubbing Data

As discussed in Sections 2.3.2 and 3.2, ozone scrubbing experiments were conducted in both indoor and outdoor settings to confirm that inexpensive scrubbers attached to the sampler boxes could preserve susceptible analytes. Analytes for this validation experiment were chosen if their match strengths (identification) were strong and for their environmental relevance.
40 Selected analytes were readily classified into two "families", the "terpene" family and the "BTEX" family, respectively representing high and negligible reactivity with ozone. Sulcatone (not detected in the indoor samples) is an unsaturated ketone found in air fresheners and cleaning products and was grouped with the terpene family as its match strength was high and its ozone rate constant (k_{O3}) is comparable to true terpenes. Table S2 gives information about the selected compounds.

45 **Table S2. Information about the chosen analytes for the Ozone Scrubbing Experiment, most significantly the ozone reaction rate coefficient (k_{O3}).**

Compound	Family	Formula	k_{O3} ($cm^3molec^{-1}s^{-1}$)	Notes
Benzene	BTEX	C_6H_6	7×10^{-23}	
Toluene	BTEX	C_7H_8	1.5×10^{-22}	
Ethylbenzene	BTEX	C_8H_{10}	1.7×10^{-22}	
p-Xylene	BTEX	C_8H_{10}	4×10^{-22}	Elutes with *meta* isomer
m-Xylene	BTEX	C_8H_{10}	6×10^{-22}	Elutes with *para* isomer
o-Xylene	BTEX	C_8H_{10}	7×10^{-22}	
α-Pinene	Terpenes	$C_{10}H_{16}$	1.07×10^{-16}	
Camphene	Terpenes	$C_{10}H_{16}$	4.5×10^{-19}	
Sulcatone	Terpenes	$C_8H_{14}O$	2.6×10^{-16}	Not a terpene
β-Pinene	Terpenes	$C_{10}H_{16}$	2.35×10^{-17}	
3-Carene	Terpenes	$C_{10}H_{16}$	3.7×10^{-17}	
Limonene	Terpenes	$C_{10}H_{16}$	2.5×10^{-16}	

Rate constants for the terpenes were found in the NIST database (Manion et al., 2015). Rate constants for the BTEX compounds were taken/estimated from other sources (Pate et al., 1976; Tseng et al., 2009). The *para* and *meta* isomers of xylene are
50 difficult to resolve, and their peak areas were integrated together, using a rate constant of 4×10^{-22} cm^3molec^{-1} s^{-1} for Figure 3. Rate constants for BTEX are less well-documented than those of terpenes and, due to their far slower kinetics, the BTEX compounds serve well as a control in this experiment.

The indoor and outdoor data for the analytes a given in Table S3. Mean values are given with their % RSD, and the ratio of unscrubbed/scrubbed is given with an absolute uncertainty. The latter is determined from propagating the error (the % RSD)
55 for the two mean values used in the quotient to get the ratio. A p-value of less than 0.05 (determined by a t-test as described in

Section 2.3.1 and Supplemental Section S1 for the sampler precision experiment) means the scrubbed and unscrubbed values are significantly different, with such values marked by an asterisk in the table.

Table S3. Data for the ozone scrubbing experiments, giving mean values for the scrubbed (n=5) and unscrubbed (n=5) sampling sets in both indoor and outdoor settings. A p-value < 0.05 means there is a significant difference between scrubbed and unscrubbed samples. Such values are denoted with an asterisk (*).

Compound	With Scrubber (Mean ± %RSD)	W/O Scrubber (Mean ± %RSD)	Ratio (Unscrb./Scrb.) ± Uncertainty	p-Value
Indoors				
Benzene	1.80E8 ± 30.1%	1.28E8 ± 24.4%	0.710 ± 0.275	0.099
Toluene	1.93E9 ± 6.4	1.99E9 ± 2.3	1.034 ± 0.070	0.295
Ethylbenzene	1.37E8 ± 9.7	1.43E8 ± 9.0	1.044 ± 0.138	0.493
m-, *p-*Xylene	4.14E8 ± 10.4	4.35E8 ± 9.8	1.050 ± 0.150	0.466
*o-*Xylene	1.32E8 ± 13.8	1.37E8 ± 9.4	1.037 ± 0.173	0.635
α-Pinene	3.86E8 ± 7.5	3.60E8 ± 2.3	0.931 ± 0.073	0.084
Camphene	3.82E7 ± 10.7	4.51E7 ± 5.5	1.182 ± 0.143	0.012 *
β-Pinene	1.96E7 ± 22.7	1.44E7 ± 13.1	0.736 ± 0.193	0.044 *
3-Carene	1.92E7 ± 8.5	1.35E7 ± 7.2	0.705 ± 0.078	0.000 *
Limonene	3.55E7 ± 13.7	1.80E7 ± 11.9	0.506 ± 0.09	0.000 *
Outdoors				
Benzene	1.57E8 ± 16.2	1.76E8 ± 29.0	1.120 ± 0.372	0.481
Toluene	2.47E9 ± 2.6	2.52E9 ± 7.1	1.022 ± 0.077	0.542
Ethylbenzene	6.24E7 ± 5.3	6.12E7 ± 10.7	0.980 ± 0.117	0.708
m-, *p-*Xylene	1.71E8 ± 5.7	1.67E8 ± 10.1	0.974 ± 0.113	0.628
*o-*Xylene	5.76E7 ± 7.6	5.34E7 ± 9.1	0.927 ± 0.109	0.186
α-Pinene	4.84E7 ± 5.0	1.39E7 ± 21.3	0.288 ± 0.063	0.000 *
Camphene	3.18E6 ± 5.0	2.24E6 ± 12.9	0.705 ± 0.097	0.000 *
Sulcatone	1.20E7 ± 13.5	6.93E5 ± 33.1	0.058 ± 0.021	0.000 *
β-Pinene	1.15E6 ± 17.1	3.69E5 ± 16.0	0.320 ± 0.075	0.000 *
3-Carene	3.96E6 ± 10.4	7.32E5 ± 18.4	0.185 ± 0.039	0.000 *
Limonene	8.48E7 ± 51.4	3.04E6 ± 37.1	0.036 ± 0.023	0.003 *

No BTEX compounds show statistically significant differences between scrubbed and unscrubbed samples, either indoors or outdoors. For the terpenes, except for α-pinene indoors, all show significant differences. In all these cases except camphene indoors, the scrubbing results in higher measured concentrations of terpenes, with the ratios pronouncedly smaller outdoors, as anticipated. Camphene has the slowest ozone kinetics among terpenes by over two orders of magnitude, which manifests as the highest ratio amongst terpenes, though this does not satisfactorily explain camphene's anomalous indoor result. Camphene meets expectations in the outdoor experiment in that its ratio is the highest but still well below 1. α-Pinene gives somewhat

higher ratios than expected given its comparatively fast kinetics. Altogether, the ozone scrubbers appear effective and the
70 relation between the unscrubbed/scrubbed ratio and the ozone rate constant is very strong.

S3. Times Series Data

Discerning temporal (day-to-day) differences in analyte concentration requires high precision (i.e., low % RSD) in the sampler
set. This time series experiment utilized fewer samplers (n=3), so this was a good test of whether such high precision can be
achieved and day-to-day variability detected. Table S4 gives the Mean Daily % RSD (i.e., the mean of the 9 daily % RSD for
75 that compound) for the 13 analytes and the Aggregate % RSD. Aggregate % RSD is calculated as:

$$Aggregate\ \%\ RSD = 100\ x\ \left(\frac{SD\ of\ daily\ means}{Mean\ of\ daily\ means} \right)$$

Table S4. Comparison of analyte Mean Daily % RSD versus aggregate % RSD for the Time Series Experiment.

Analyte	Mean Daily % RSD	Aggregate %RSD
Benzene	13.6	20.7
Toluene	7.9	81.7
Ethylbenzene	5.3	41.5
p,m-Xylene	5.4	32.1
o-Xylene	6.2	30.8
Trichloroethylene	21.0	33.6
Tetrachloroethylene	9.2	74.8
Furfural	6.3	51.8
α-Pinene	10.2	66.3
Sulcatone	9.0	88.0
β-Pinene	7.1	28.5
3-Carene	7.2	56.3
Limonene	7.4	84.8
Mean ±SD	**8.9 ± 4.1**	**53.1 ± 22.9**

80

It is obvious that temporal variability (as measured in the Aggregate % RSD) is consistently greater than the Mean Daily %
RSD. Benzene and trichloroethylene show both comparatively poor precision and low temporal variability, so reliably
detecting slight day-to-day variation in these two analytes may not be possible. For the other 11 analytes, the differences in %
85 RSD are substantial and temporal differences should be readily measurable. These experiments show that with a small sample
set (n=3), the intersampler precision (low Mean Daily % RSD) observed in other experiments is maintained and day-to-day
differences in common indoor air compounds can be recorded.

S4 Optimization of Transect Mapping

As described in Section 3.4 on the text, there are trade-offs when deciding how many transect angles to use for transect
90   mapping. In our study, two angles were used, so the sampling grid consisted of a 400 m² (20m x 20m) square grid of 25 equally sized square grid cells. The total number of samples collected in this approach, n, equals twice the number of transects on each angle ($n = 10$ in 5 x 5 example discussed in the main text), and the number of grid cells (i.e., resolution, R) is the square of the number of transects in each angle, or $R = \left(\frac{n}{2}\right)^2 = \frac{1}{4}n^2$, where n is meaningful only when divisible by two. Each cell is defined as containing one transect in each angle, and for every point inside the cell, the nearest transect in each angle is the one
95   contained in the cell. Emissions that have two sources (e.g., decane and dodecane in our experiment) yield a phantom image artefact where the one transect crossing one source meets one transect crossing the other source to yield an apparent hotspot where none exists. Including additional angles mitigates the phantom image artefact; there exists no artefact so long as the number of sources of a particular emission is less than the number of angles. We exam here the addition of a third angle as a potential approach to optimization and allowing unique identification of two sources of a given emission.

100   Transects at three evenly spaced angles, 0°, 60° and 120°, form a hexagonal sampling grid with equally sized triangular grid cells (Fig. S1). In this case, the total number of samples is three times the number of transects in each angle, and the number of grid cells grows as $R = \frac{1}{6}n^2 - \frac{\frac{n}{3}\bmod 2}{2}$ with cells defined the same as in the case of the square grid and n meaningful only when divisible by 3. The latter term in the equation, equal to 0 or 1, accounts for differences in the shape of the grid formed by odd versus even numbers of transects in each angle and is negligible for most n, so $R \approx \frac{1}{6}n^2$. A grid with this shape can
105   uniquely locate two emissions sources, while a third source may generate a single phantom image.
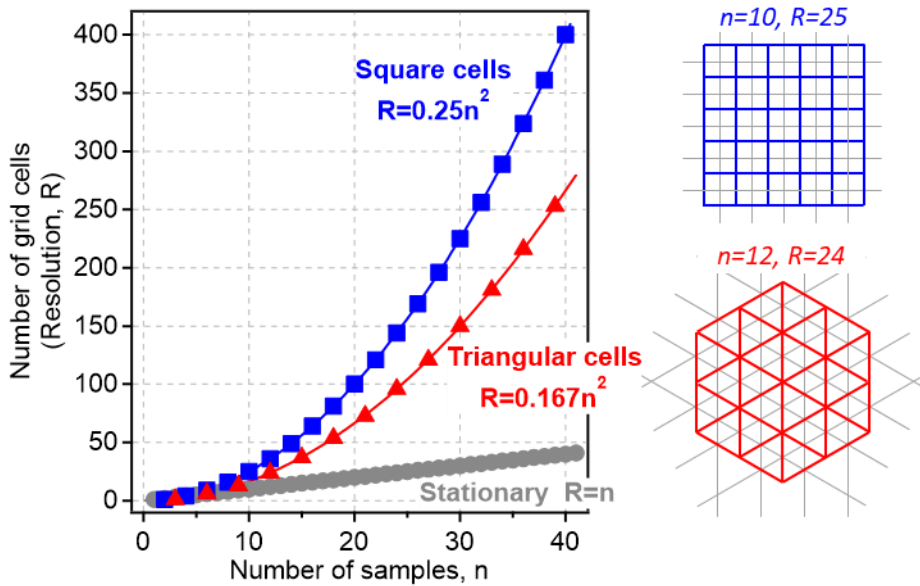
Figure S1: A plot of resolution (number of grid cells) versus total number of samples collected for stationary sampling and transect mapping with either two or three angles (square and triangular grid cells, respectively).

Figure S1 shows the exponential nature of the increase in resolution with the transect mapping approach versus the linear resolution increase with stationary sampling. As an example, a grid of 25 equal cells can be resolved with 25 samples using a stationary strategy with very high accuracy (no phantom images), 10 samples using a square grid with phantom images in the case of multiple emissions sources, or 12 samples using a triangular/hexagonal grid with no phantom images for doubled sources, and reduced images for higher numbers of sources. These advantages grow for higher numbers of samples. Considered from the other direction, if 24 samples are used, two-angle sampling provides a resolution of 144 square grid cells, three-angle sampling provides a resolution of 96 triangular grid cells, and stationary sampling provides a resolution of 24 grid cells. However, each reduction in resolution comes with a reduction in artefacts. Consequently, significant work could be conducted to optimize transect mapping for researchers' particular needs and constraints, but the general approach is shown to work well in the main text.

References

Manion, J. A., Huie, R. E., Levin, R. D., Burgess Jr., D. R., Orkin, V. L., Tsang, W., McGivern, W. S., Hudgens, J. W., Knyazev, V. D., Atkinson, D. B., Chai, E., Tereza, A. M., Lin, C.-Y., Allison, T. C., Mallard, W. G., Westley, F., Herron, J.

130   T., Hampson, R. F. and Frizzell, D. H.: IST Chemical Kinetics Database (Web Version), Natl. Inst. Stand. Technol. Gaithersburg, Maryl. [online] Available from: http://kinetics.nist.gov/, 2015.

Pate, C. T., Atkinson, R. and Pitts, J. N.: The gas phase reaction of o3 with a series of aromatic hydrocarbons, J. Environ. Sci. Heal. Part A Environ. Sci. Eng., 11(1), 1–10, doi:10.1080/10934527609385750, 1976.

Tseng, K. H., Wang, J. L., Kuo, P. H. and Tsuang, B. J.: Using a photochemical index to discuss the ozone formation and

135   estimate hydroxyl concentration at downwind area, Aerosol Air Qual. Res., 9(4), 441–452, doi:10.4209/aaqr.2009.05.0026, 2009.