Table S1. Mapping RETRO species to GEOS-Chem tracers

GEOS-Chem tracer	Applicable RETRO species	Global emissions (TgC/y)	Comments
ALK4	Butanes Pentanes Hexanes and Higher Alkanes	32.56	ALK4: >= C4 alkanes
ACET	Ketones	2.63	Acetone assumed to account for 75% of total ketones
MEK	Ketones	0.88	MEK: >= C4 ketones. MEK assumed to account for 25% of total ketones
ALD2	Other Alkanals	0.93	ALD2: acetaldehyde. ALD2 assumed to account for the total alkanal flux
PRPE	Propene	2.35	PRPE: >= C3 alkenes
C3H8	Propane	2.88	
CH2O	Methanal	0.46	
C2H6	Ethane	4.09	
C2H4	Ethene	5.55	
BENZ	Benzene	3.21	
TOLU	Toluene	5.60	
XYLE	Xylene and other aromatics	7.22	C_8 aromatic compounds including *o*-, *m*-, *p*-xylenes and ethylbenzene. "Other aromatics" assumed to be solely ethylbenzene.
C2H2	Ethyne	2.28	
HCOOH	Acids	0.49	HCOOH assumed to account for 25% of the total carboxylic acid flux [*Chebbi and Carlier,* 1996][1]
Total		71.14	

[1] Chebbi, A., and P. Carlier (1996), Carboxylic acids in the troposphere, occurrence, sources, and sinks: A review, *Atmos. Environ.*, 30(24), 4233-4249, doi: 10.1016/1352-2310(96)00102-1.

Table S2. *A posteriori* emission estimates over the contiguous US for the best-estimate (Opt1) and 24 sensitivity runs (Opt2 to Opt25).

		Benzene				Toluene				C_8 aromatics			
Opt number	Note	On-road	Non-road	US total	$J_{final}/J_{initial}$[1]	On-road	Non-road	US total	$J_{final}/J_{initial}$[1]	On-road	Non-road	US total	$J_{final}/J_{initial}$[1]
	A priori; NEI08; 2011; GgC	54	133	187		256	705	961		241	624	865	
1a	NEI08; 2011; Seasonal[2]: cold; DryDep[3]; GgC	22	72		0.29	86	101		0.2	161	117		0.87
1b	NEI08; 2011; Seasonal[2]: warm; DryDep[3]; GgC	48	67		0.62	124	97		0.34	246	298		0.91
1	Opt 1; GgC	70	139	209		210	198	408		407	415	822	
Sensitivity runs: deviation from the best-estimate Opt1													
2	NEI08; 2011; DryDep[3]; Annual[4]	-16%	18%	6%	0.41	-9%	23%	2%	0.27	-21%	7%	-7%	0.99
3	NEI08; 2011; Annual[4]	-11%	13%	5%	0.49	-9%	23%	2%	0.25	-22%	5%	-8%	0.99
4	NEI08; 2010; Annual[4]	-10%	-3%	-5%	0.65	-1%	8%	2%	0.28	-11%	4%	-3%	0.99
5	NEI08; 2010-2011; Annual[4]	-11%	5%	0%	0.57	-5%	13%	1%	0.27	-16%	5%	-5%	0.99
6	NEI08; 2011; Seasonal[2]; MLE[5]	-3%	-17%	-12%	0.52 (cold);0.78 (warm)	-7%	-67%	-28%	0.03 (cold);0.10 (warm)	-8%	-54%	-31%	0.41 (cold); 0.96 (warm)
7	NEI08; 2011; Annual[4]; MLE[5]	-33%	-4%	-14%	0.65	-25%	-61%	-37%	0.04	-32%	-68%	-51%	0.58
8	NEI08; 2011; Seasonal[2]; $\mathbf{S}_{mod}$ *2[6]	4%	10%	8%	0.72 (cold);0.70 (warm)	-1%	26%	8%	0.14 (cold);0.21 (warm)	4%	27%	16%	0.98 (cold); 0.86 (warm)
9	NEI08; 2011;	0%	24%	16%	0.58	-6%	36%	9%	0.17	-13%	35%	12%	0.99

	Annual[4]; $\mathbf{S}_{mod}$ *2[6]												
10	NEI08; 2011; Seasonal[2]; γ_{HO2} *2[7]	3%	-4%	-2%	0.38 (cold); 0.69 (warm)	-2%	-3%	-2%	0.18 (cold); 0.31 (warm)	-2%	-5%	-3%	0.84 (cold); 0.92 (warm)
11	NEI08; 2011; Annual[4]; γ_{HO2} *2[7]	-13%	10%	2%	0.52	-11%	17%	-1%	0.24	-22%	1%	-11%	0.98
12	NEI08; 2011; Seasonal[2]; E_{NOx} *0.6[8]	3%	-6%	-3%	0.36 (cold); 0.67 (warm)	0%	-13%	-5%	0.18 (cold); 0.31 (warm)	1%	-8%	-3%	0.85 (cold); 0.92 (warm)
13	NEI08; 2011; Annual[4]; E_{NOx} *0.6[8]	-13%	10%	2%	0.5	-8%	10%	-2%	0.24	-20%	1%	-10%	0.98
14	NEI08; 2011; Seasonal[2]; with Br chemsitry[9]	3%	-1%	0%	0.36 (cold); 0.65 (warm)	0%	0%	0%	0.19 (cold); 0.34 (warm)	-1%	0%	0%	0.86 (cold); 0.91 (warm)
15	NEI08; 2011; Annual[4]; with Br chemsitry[9]	-13%	13%	4%	0.49	-9%	23%	2%	0.25	-22%	5%	-8%	0.99
16	NEI08; 2011; Seasonal[2]; DryDep with relative uptake[10]	6%	57%	39%	0.3 (cold); 0.47 (warm)	8%	36%	18%	0.28 (cold); 0.47 (warm)	4%	17%	11%	0.92 (cold); 0.88 (warm)
17	NEI08; 2011; Annual[4]; DryDep with relative uptake[10]	-11%	70%	42%	0.37	-2%	61%	20%	0.36	-16%	16%	0%	0.99
18	NEI08; 2011; Seasonal[2]; PBL local[11]	10%	7%	8%	0.39 (cold); 0.62 (warm)	8%	0%	5%	0.23 (cold); 0.41 (warm)	-1%	4%	1%	0.87 (cold); 0.91 (warm)
19	NEI08; 2011; Annual[4]; PBL local[11]	-6%	21%	12%	0.5	-3%	23%	6%	0.3	-23%	7%	-8%	0.99

20	RETRO; 2011; Seasonal[2]; DryDep[3]			-14%	0.34 (cold); 0.35 (warm)			-2%	0.13 (cold); 0.08 (warm)		5%	0.43 (cold); 0.48 (warm)
21	RETRO; 2011; Annual[4]; DryDep[3]			-13%	0.35			-2%	0.11		3%	0.45
22	RETRO; 2011; Seasonal[2];			-12%	0.36 (cold); 0.34 (warm)			-2%	0.12 (cold); 0.08 (warm)		15%	0.47 (cold); 0.48 (warm)
23	RETRO; 2011; Annual[4]			-10%				-2%	0.13		15%	0.45
24	RETRO; 2010; Annual[4]			-8%				3%	0.13		25%	0.54
25	NEI08; 2011; 3-season[12]; DryDep[3]	-6%	-1%	-2%	0.24 (cold); 0.65 (warm); 0.37 (shoulder)	-12%	25%	1%	0.16 (cold); 0.26 (warm); 0.33 (shoulder)	-16%	21%	3%	0.66 (cold); 0.86 (warm); 0.98 (shoulder)
	Min[13]	-33%	-17%	-14%		-25%	-67%	-37%		-32%	-68%	-51%	
	Max[13]	10%	70%	42%		8%	61%	20%		4%	35%	25%	

[1] Cost function reduction; [2] Optimizations using warm season (April – September) or cold season (October – March) data only; see Section 4 and Table 3 for source combinations for different seasons; [3] With dry deposition as described in Section 2.3; [4] Aseasonal optimizations; [5] Estimation of $\mathbf{S}_a$ and $\mathbf{S}_\Sigma$ based on a maximum likelihood estimation approach; [6] Doubling the model error, which is the major part of the observational error covariance matrix $\mathbf{S}_\Sigma$; [7] doubling the reactive uptake coefficient for HO_2 on aqueous aerosols from 0.2 to 0.4; [8] Decreasing NO_x emissions over North America by 40%; [9] Including bromine chemistry; [10] Including reactive uptake during dry deposition of aromatics; [11] Using a local boundary layer mixing scheme; [12] Optimizations using 3-season inversion (cold= December - February; warm= June - August; shoulder= March - May + September - November). [13] Minimum or maximum deviation from Opt1 for the ensemble of sensitivity runs (Opt2 to Opt25).

Table S3. Emission correction factors from the best-estimate optimization based on the RETRO inventory.

	Benzene		Toluene		C_8 aromatics	
RETRO (Opt20)	U.S. emissions	Boundary condition	U.S. emissions	Boundary condition	U.S. emissions	Boundary condition
Cold	0.35	2.74	0.2	2.14	0.38	4.92
Warm	0.50	2.20	0.22	1	0.48	1

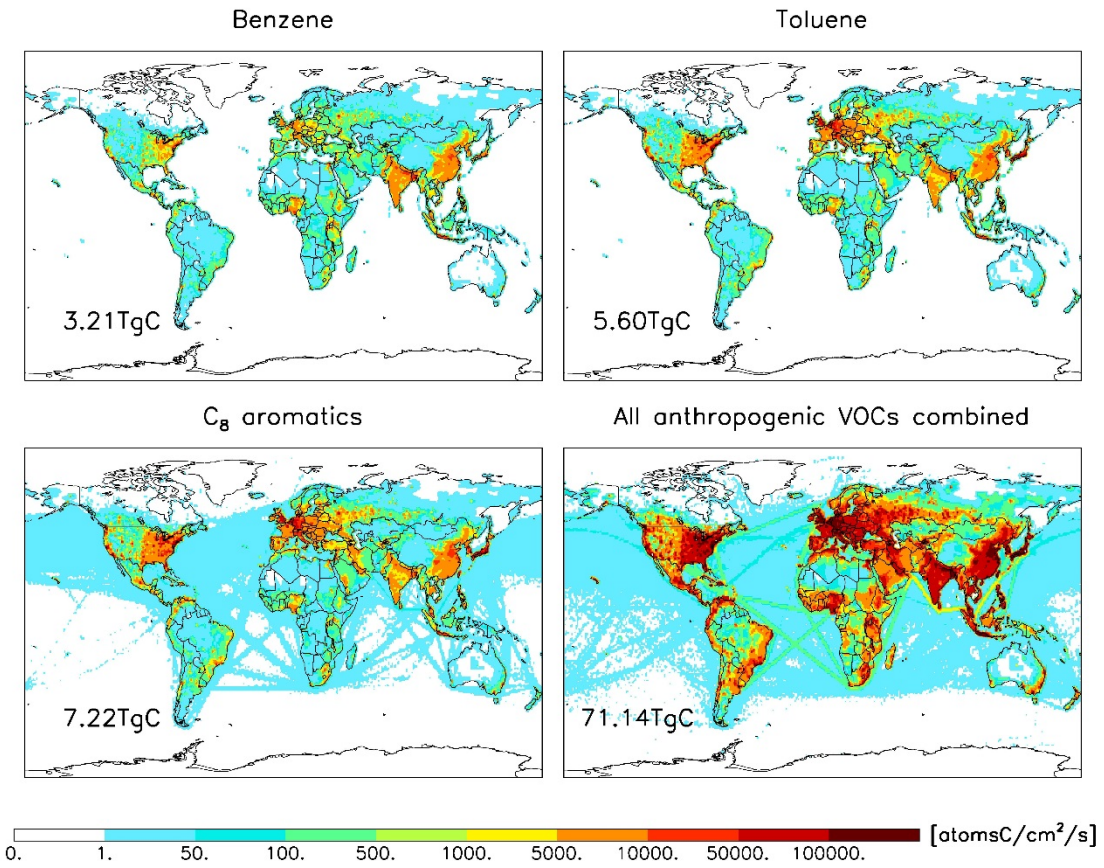

Figure S1. Annual emissions of C_6-C_8 aromatics and all anthropogenic VOCs combined (including aromatics) according to the year-2000 RETRO inventory as implemented in GEOS-Chem. Numbers insets show the aggregated annual flux for the corresponding GEOS-Chem tracers (see Table S1).

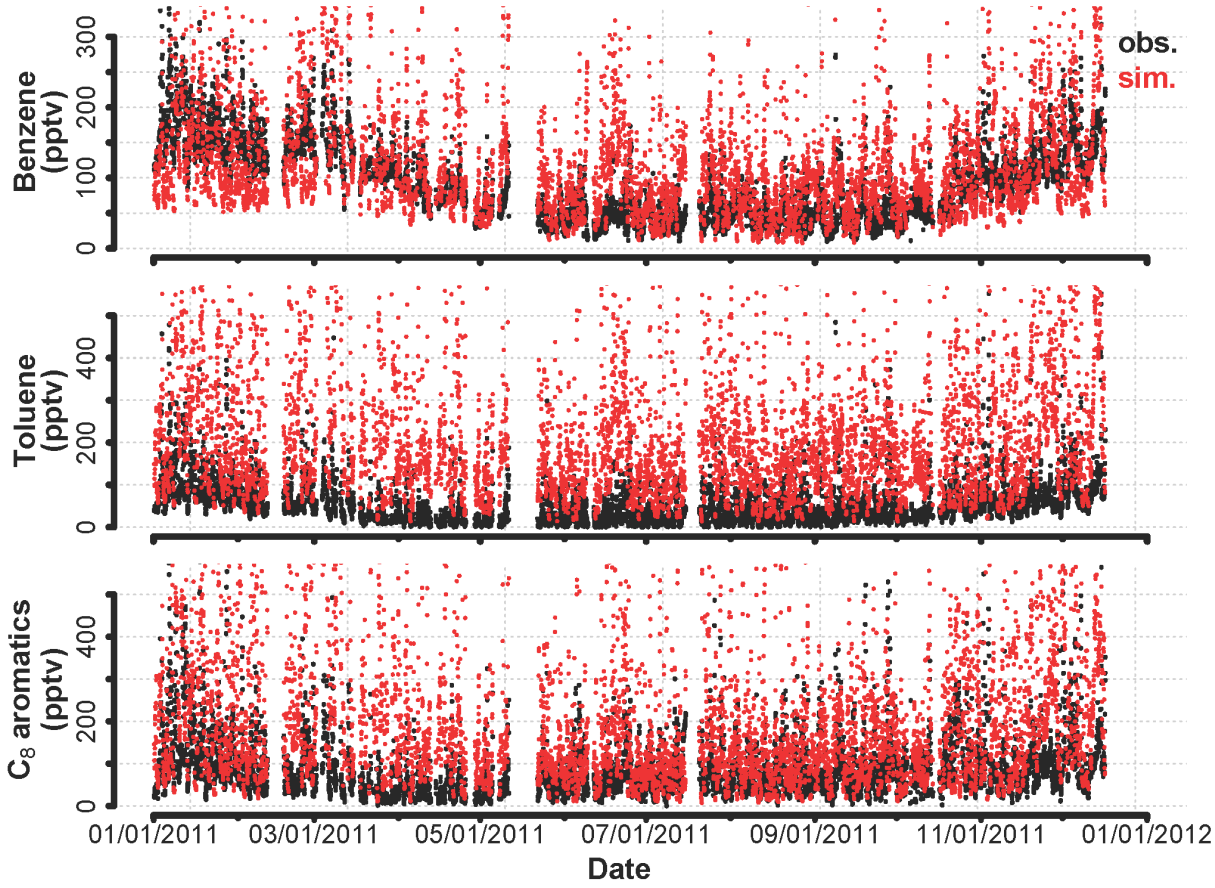

Figure S2. Annual cycle in benzene, toluene, and C_8 aromatics observed at the KCMP tall tower during 2011 (black). The GEOS-Chem *a priori* simulation based on the RETRO inventory is shown in red. All data points are 1-h means.

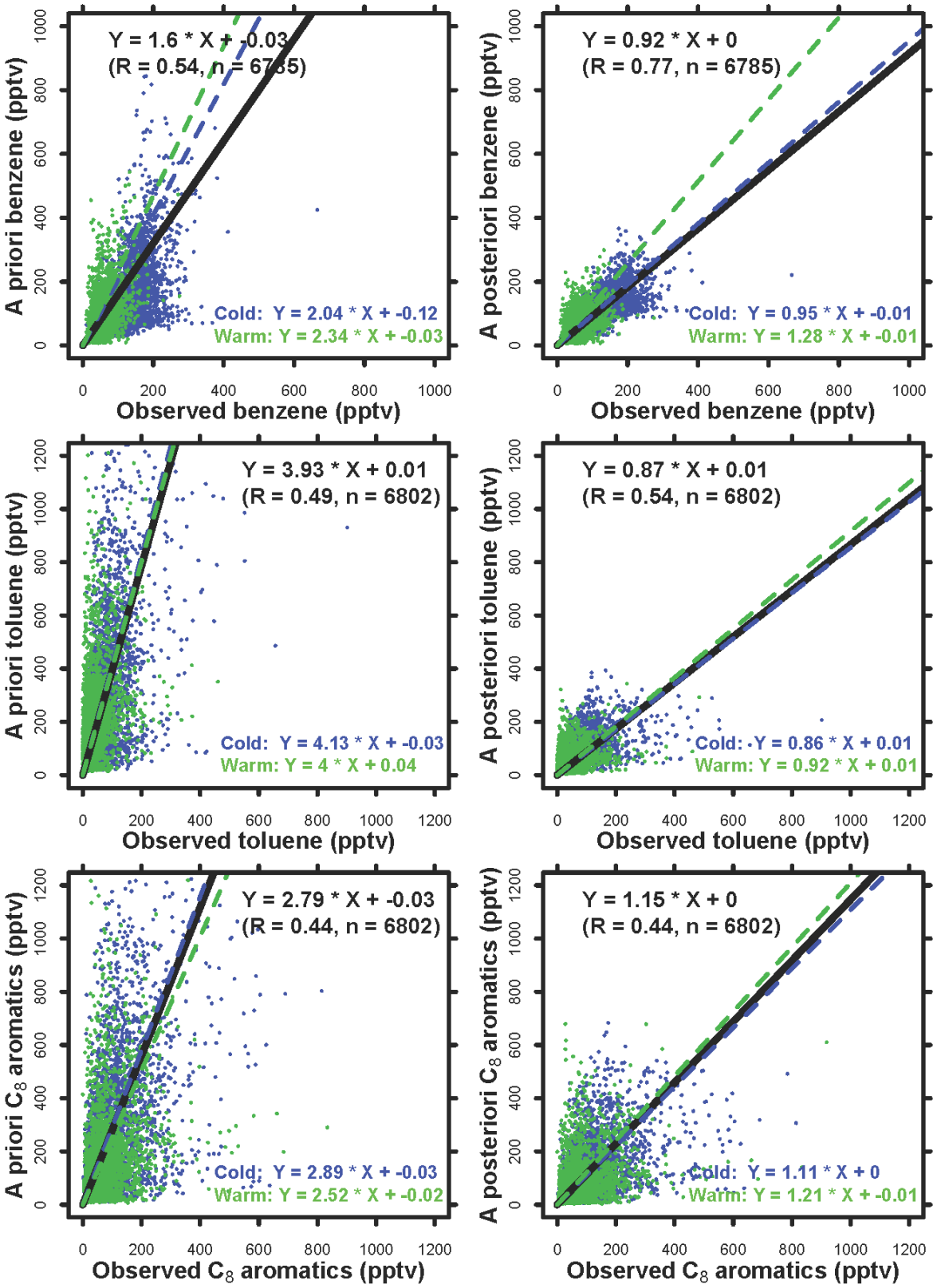
Y = 1.6 * X + -0.03
(R = 0.54, n = 6785)
Cold: Y = 2.04 * X + -0.12
Warm: Y = 2.34 * X + -0.03
A priori benzene (pptv)
Observed benzene (pptv)
Y = 0.92 * X + 0
(R = 0.77, n = 6785)
Cold: Y = 0.95 * X + -0.01
Warm: Y = 1.28 * X + -0.01
A posteriori benzene (pptv)
Observed benzene (pptv)
Y = 3.93 * X + 0.01
(R = 0.49, n = 6802)
Cold: Y = 4.13 * X + -0.03
Warm: Y = 4 * X + 0.04
A priori toluene (pptv)
Observed toluene (pptv)
Y = 0.87 * X + 0.01
(R = 0.54, n = 6802)
Cold: Y = 0.86 * X + 0.01
Warm: Y = 0.92 * X + 0.01
A posteriori toluene (pptv)
Observed toluene (pptv)
Y = 2.79 * X + -0.03
(R = 0.44, n = 6802)
Cold: Y = 2.89 * X + -0.03
Warm: Y = 2.52 * X + -0.02
A priori C8 aromatics (pptv)
Observed C8 aromatics (pptv)
Y = 1.15 * X + 0
(R = 0.44, n = 6802)
Cold: Y = 1.11 * X + 0
Warm: Y = 1.21 * X + -0.01
A posteriori C8 aromatics (pptv)
Observed C8 aromatics (pptv)

Figure S3: Atmospheric aromatic mixing ratios at the KCMP tall tower in 2011. The GEOS-Chem *a priori* (based on NEI08; left column) and best-estimate *a posteriori* (right column) simulations are compared to measured values at the tall tower in 2011, colored by warm and cold seasons. Solid and dashed lines show the corresponding major axis fits, with regression parameters given inset (95% confidence intervals for slopes are better than ± 6%, ± 9%, ± 10% for benzene, toluene, and C_8 aromatics, respectively). Data points are 1-h mean values.

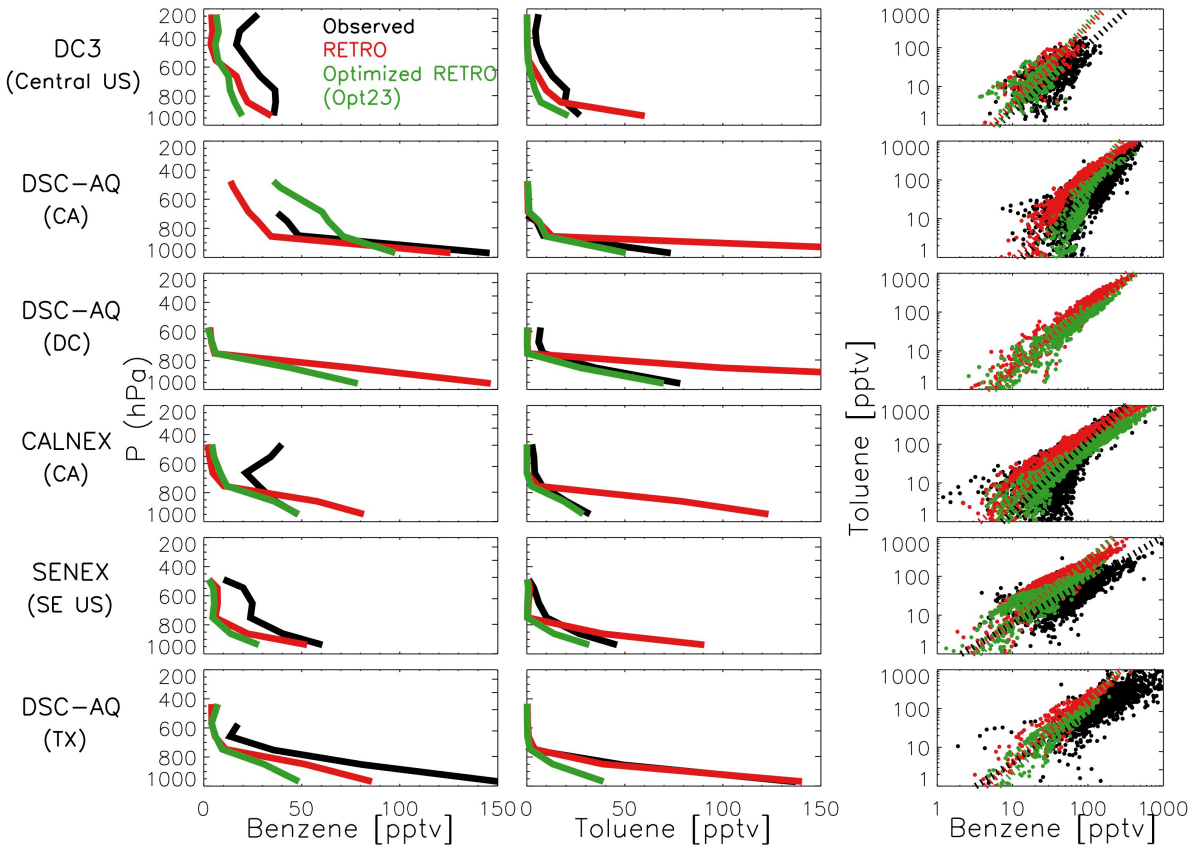

Figure S4. Vertical distribution of benzene and toluene over North America. Left and middle columns: median vertical profiles of benzene and toluene observed (black) and simulated by GEOS-Chem based on RETRO (red: base-case *a priori*; green: best-estimate optimization) during six recent aircraft campaigns. Right column: boundary layer (P > 800 hPa) benzene-toluene relationships for the same aircraft campaigns, with dashed lines showing the best fits from major axis regression. DSC-AQ (CA): DISCOVER-AQ California; DSC-AQ (DC): DISCOVER-AQ Baltimore-Washington DC; DSC-AQ (TX): DISCOVER-AQ Texas.

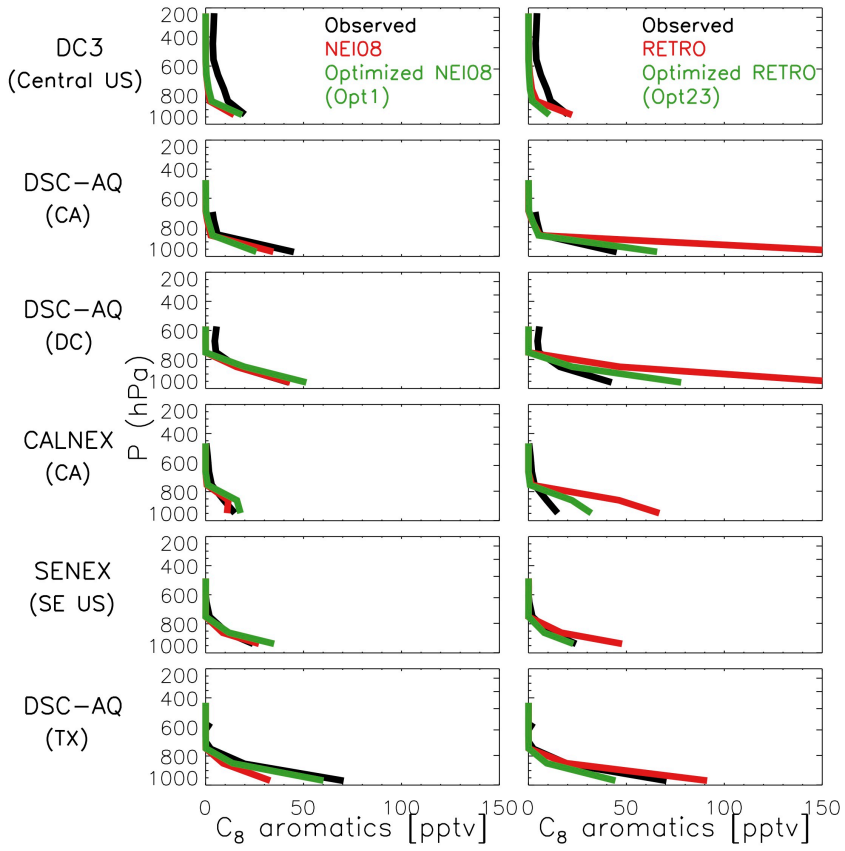

Figure S5. Vertical distribution of C_8 aromatics over North America. Shown are the median vertical profiles of C_8 aromatic observed (black) and simulated by GEOS-Chem based on NEI08 (left column) and RETRO (right column). The red lines show the *a priori* simulation and the green lines show the corresponding *a posteriori* simulation. DSC-AQ (CA): DISCOVER-AQ California; DSC-AQ (DC): DISCOVER-AQ Baltimore-Washington DC; DSC-AQ (TX): DISCOVER-AQ Texas.