Supplementary Materials

Improving the Performance of Pipeline Leak Detection Algorithms for the Mobile Monitoring of Methane Leaks

Tian Xia *, Julia Raneses and Stuart Batterman

Department of Environmental Health Sciences, School of Public Health, University of Michigan, Ann Arbor, MI, 48109, USA; jraneses@umich.edu (J.R.); stuartb@umich.edu (S.B.)
* Correspondence: xiatian@umich.edu

Table S1. Statistics of the 254 peaks detected by our algorithm with parameter set 4 setups on the second 10-day subset. Bolded data represents the median.

		C_B (ppm)	ΔC_{max} (ppm)	ΔC_{ave} (ppm)	ΔC_{med} (ppm)
Valid Sample Size		254	254	254	254
Average		2.03	1.26	0.44	0.31
Standard Deviation		0.05	3.28	0.79	0.45
Percentile	0.000	1.93	0.07	0.05	0.02
	0.250	1.99	0.14	0.12	0.12
	0.500	**2.02**	**0.26**	**0.19**	**0.18**
	0.750	2.08	0.72	0.35	0.31
	0.900	2.11	2.62	0.95	0.63
	0.980	2.14	10.74	3.28	1.51
	0.990	2.15	15.35	3.90	1.80
	0.999	2.17	29.44	6.68	4.66
	1.000	2.17	31.59	7.30	5.57

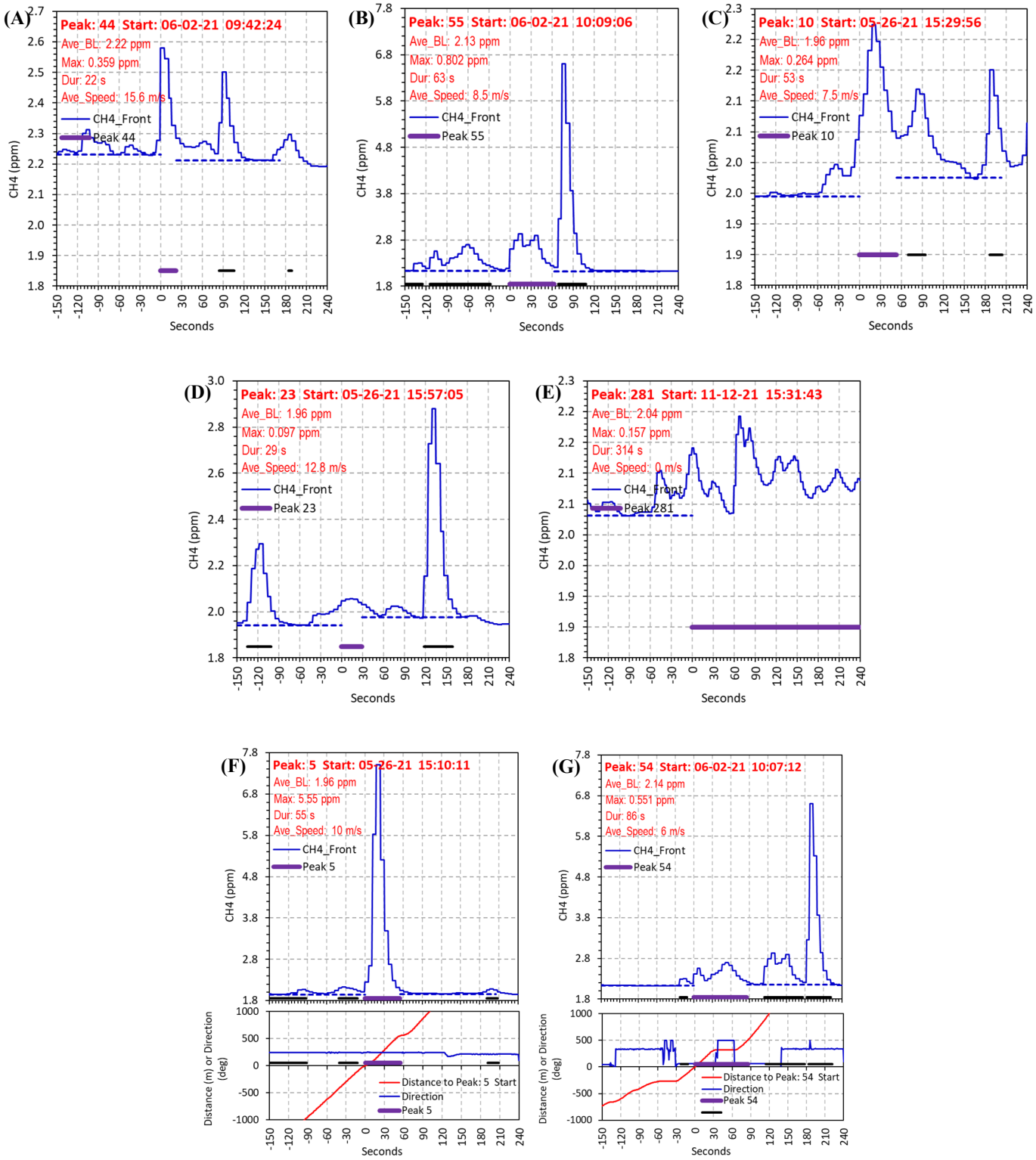
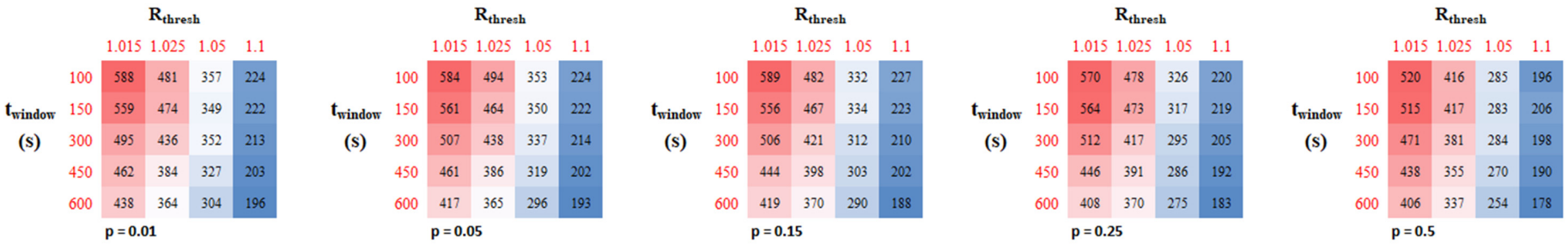


Figure S1. Graphical examples of flagged peaks. (A) Baseline (start) out of range (1.8 – 2.2 ppm); (B) consistently elevated observations in the baseline window (>50% of data in the baseline window as indicated by the blue dashed line); (C) large baseline shifts ($|C_{B,pre} - C_{B,post}| > 0.02$ ppm); (D) small peak ($\Delta C_{max} < 0.1$ ppm); (E) peak duration too long ($t_p > 240$ s); (F) wide peaks ($W > 500$ m); (G) peaks with stopped periods ($V_i < 0.05$ m/s exceeding 1 s during the peak, as indicated by 500 deg driving direction).

A) Number of Peaks after Pass 2



t_{window} (s) \ R_{thresh}	1.015	1.025	1.05	1.1
100	588	481	357	224
150	559	474	349	222
300	495	436	352	213
450	462	384	327	203
600	438	364	304	196

p = 0.01

t_{window} (s) \ R_{thresh}	1.015	1.025	1.05	1.1
100	584	494	353	224
150	561	464	350	222
300	507	438	337	214
450	461	386	319	202
600	417	365	296	193

p = 0.05

t_{window} (s) \ R_{thresh}	1.015	1.025	1.05	1.1
100	589	482	332	227
150	556	467	334	223
300	506	421	312	210
450	444	398	303	202
600	419	370	290	188

p = 0.15

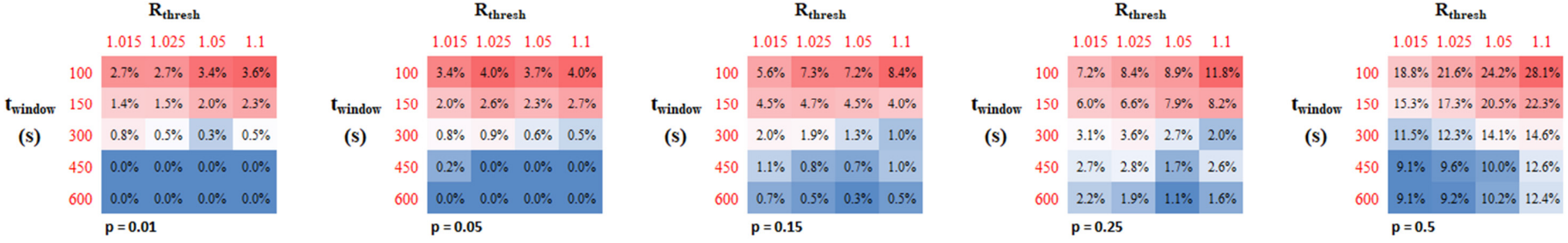
t_{window} (s) \ R_{thresh}	1.015	1.025	1.05	1.1
100	570	478	326	220
150	564	473	317	219
300	512	417	295	205
450	446	391	286	192
600	408	370	275	183

p = 0.25

t_{window} (s) \ R_{thresh}	1.015	1.025	1.05	1.1
100	520	416	285	196
150	515	417	283	206
300	471	381	284	198
450	438	355	270	190
600	406	337	254	178

p = 0.5

B) Baseline out of Range



t_{window} (s) \ R_{thresh}	1.015	1.025	1.05	1.1
100	2.7%	2.7%	3.4%	3.6%
150	1.4%	1.5%	2.0%	2.3%
300	0.8%	0.5%	0.3%	0.5%
450	0.0%	0.0%	0.0%	0.0%
600	0.0%	0.0%	0.0%	0.0%

p = 0.01

t_{window} (s) \ R_{thresh}	1.015	1.025	1.05	1.1
100	3.4%	4.0%	3.7%	4.0%
150	2.0%	2.6%	2.3%	2.7%
300	0.8%	0.9%	0.6%	0.5%
450	0.2%	0.0%	0.0%	0.0%
600	0.0%	0.0%	0.0%	0.0%

p = 0.05

t_{window} (s) \ R_{thresh}	1.015	1.025	1.05	1.1
100	5.6%	7.3%	7.2%	8.4%
150	4.5%	4.7%	4.5%	4.0%
300	2.0%	1.9%	1.3%	1.0%
450	1.1%	0.8%	0.7%	1.0%
600	0.7%	0.5%	0.3%	0.5%

p = 0.15

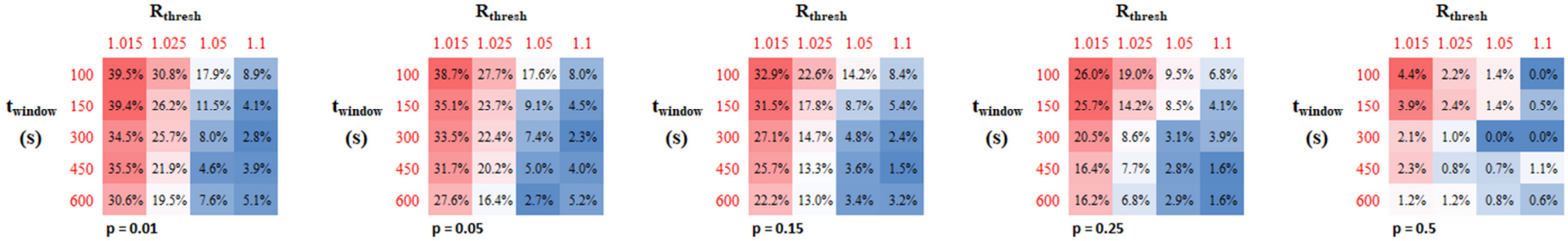
t_{window} (s) \ R_{thresh}	1.015	1.025	1.05	1.1
100	7.2%	8.4%	8.9%	11.8%
150	6.0%	6.6%	7.9%	8.2%
300	3.1%	3.6%	2.7%	2.0%
450	2.7%	2.8%	1.7%	2.6%
600	2.2%	1.9%	1.1%	1.6%

p = 0.25

t_{window} (s) \ R_{thresh}	1.015	1.025	1.05	1.1
100	18.8%	21.6%	24.2%	28.1%
150	15.3%	17.3%	20.5%	22.3%
300	11.5%	12.3%	14.1%	14.6%
450	9.1%	9.6%	10.0%	12.6%
600	9.1%	9.2%	10.2%	12.4%

p = 0.5

C) Consistently elevated observations in basline window



t_{window} (s) \ R_{thresh}	1.015	1.025	1.05	1.1
100	39.5%	30.8%	17.9%	8.9%
150	39.4%	26.2%	11.5%	4.1%
300	34.5%	25.7%	8.0%	2.8%
450	35.5%	21.9%	4.6%	3.9%
600	30.6%	19.5%	7.6%	5.1%

p = 0.01

t_{window} (s) \ R_{thresh}	1.015	1.025	1.05	1.1
100	38.7%	27.7%	17.6%	8.0%
150	35.1%	23.7%	9.1%	4.5%
300	33.5%	22.4%	7.4%	2.3%
450	31.7%	20.2%	5.0%	4.0%
600	27.6%	16.4%	2.7%	5.2%

p = 0.05

t_{window} (s) \ R_{thresh}	1.015	1.025	1.05	1.1
100	32.9%	22.6%	14.2%	8.4%
150	31.5%	17.8%	8.7%	5.4%
300	27.1%	14.7%	4.8%	2.4%
450	25.7%	13.3%	3.6%	1.5%
600	22.2%	13.0%	3.4%	3.2%

p = 0.15

t_{window} (s) \ R_{thresh}	1.015	1.025	1.05	1.1
100	26.0%	19.0%	9.5%	6.8%
150	25.7%	14.2%	8.5%	4.1%
300	20.5%	8.6%	3.1%	3.9%
450	16.4%	7.7%	2.8%	1.6%
600	16.2%	6.8%	2.9%	1.6%

p = 0.25

t_{window} (s) \ R_{thresh}	1.015	1.025	1.05	1.1
100	4.4%	2.2%	1.4%	0.0%
150	3.9%	2.4%	1.4%	0.5%
300	2.1%	1.0%	0.0%	0.0%
450	2.3%	0.8%	0.7%	1.1%
600	1.2%	1.2%	0.8%	0.6%

p = 0.5

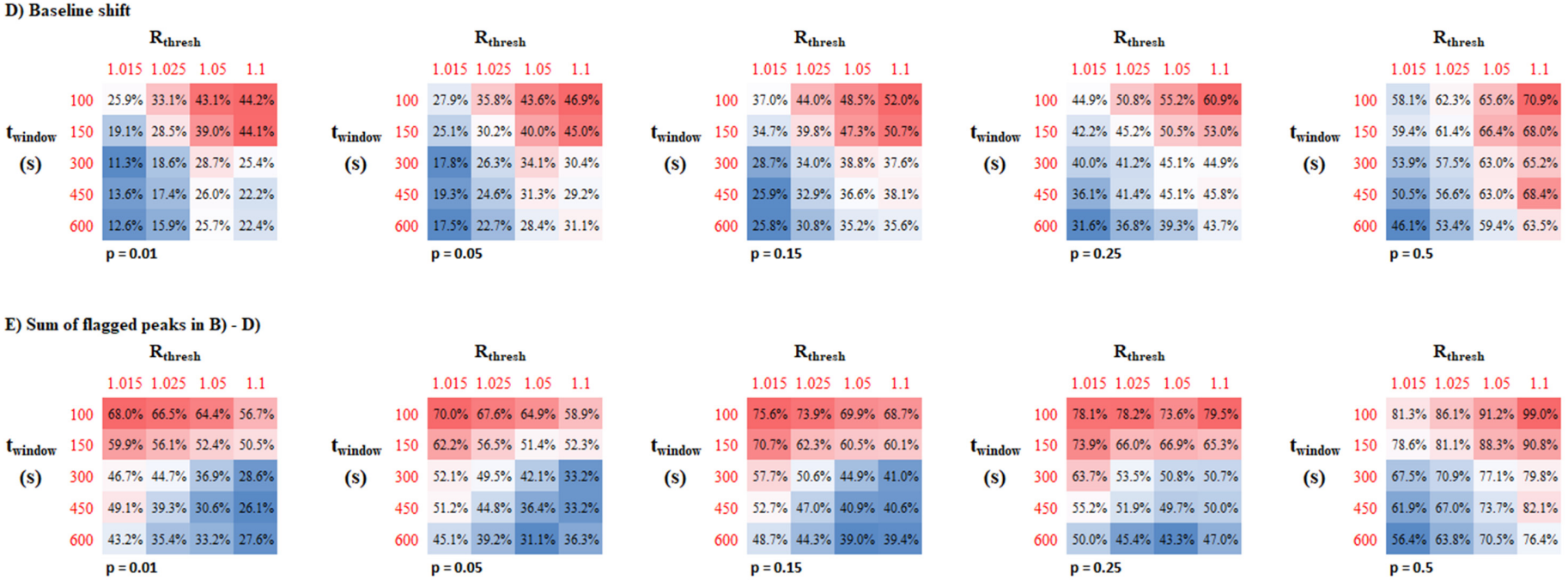


Figure S2. Heatmaps for the baseline sensitivity analyses.

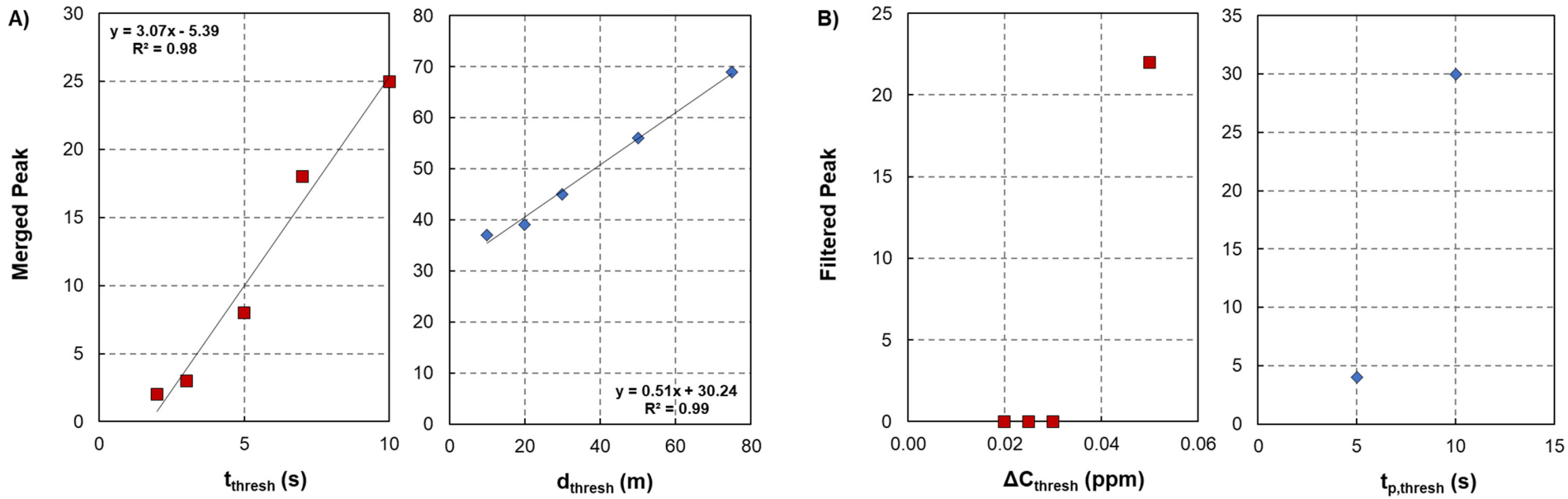


Figure S3. Number of A) merged peaks in Pass 3 and B) filtered out peaks in Pass 4 with different parameter setups.

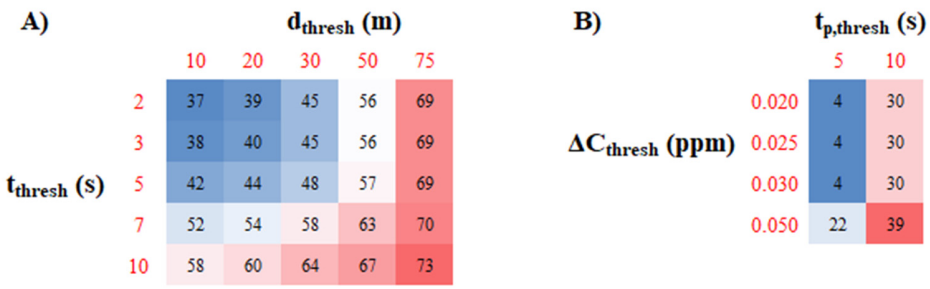
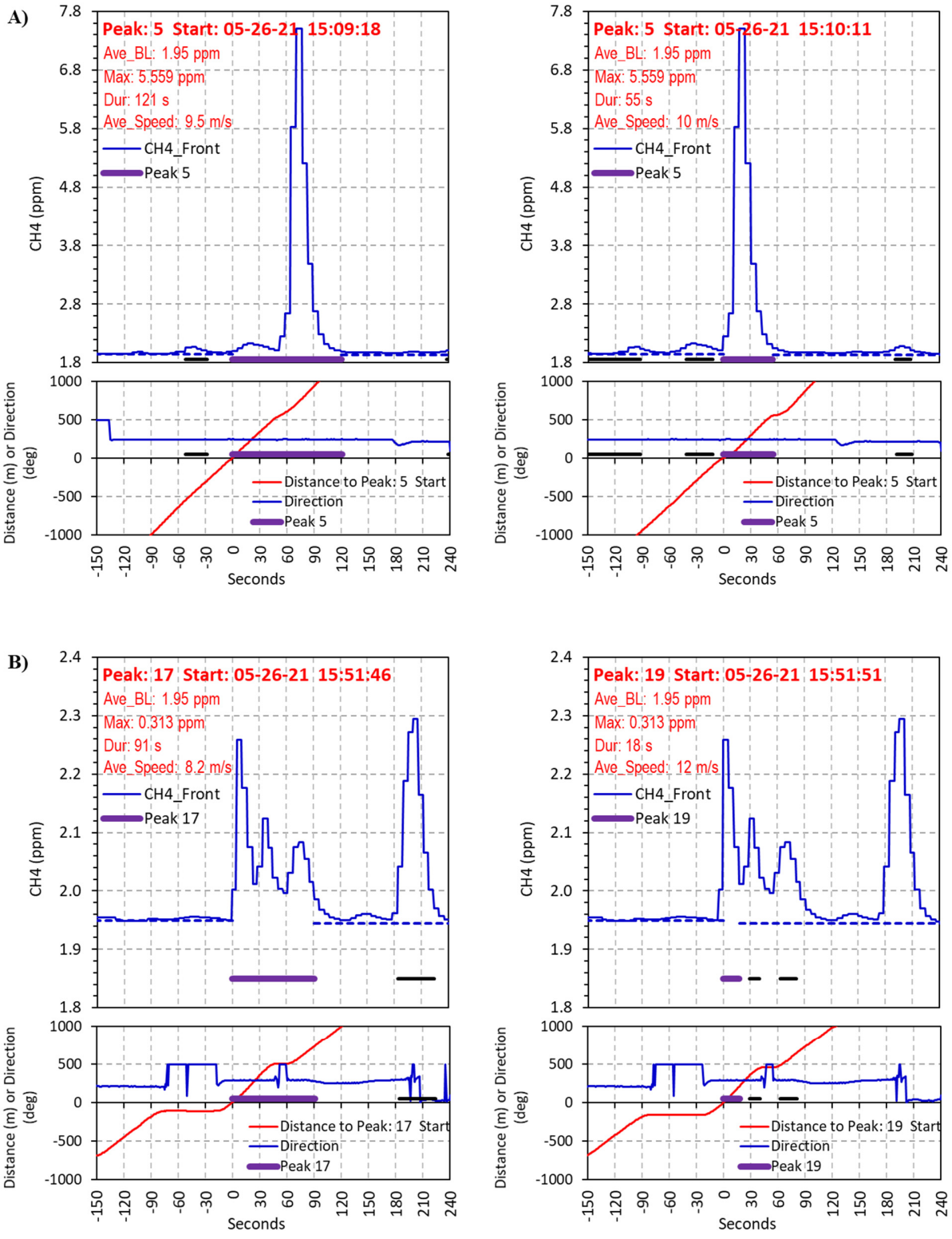


Figure S4. Heatmaps of the A) total merged peaks in Pass 3 with 25 t_{thresh} and d_{thresh} combinations and B) total filtered out peaks in Pass 4 with 8 $t_{p,thresh}$ and ΔC_{thresh} combinations.


A)
Peak: 5 Start: 05-26-21 15:09:18
Ave_BL: 1.95 ppm
Max: 5.559 ppm
Dur: 121 s
Ave_Speed: 9.5 m/s
CH4_Front
Peak 5
CH4 (ppm)
Distance (m) or Direction (deg)
Distance to Peak: 5 Start
Direction
Peak 5
Seconds
Peak: 5 Start: 05-26-21 15:10:11
Ave_BL: 1.95 ppm
Max: 5.559 ppm
Dur: 55 s
Ave_Speed: 10 m/s
CH4_Front
Peak 5
CH4 (ppm)
Distance (m) or Direction (deg)
Distance to Peak: 5 Start
Direction
Peak 5
Seconds
B)
Peak: 17 Start: 05-26-21 15:51:46
Ave_BL: 1.95 ppm
Max: 0.313 ppm
Dur: 91 s
Ave_Speed: 8.2 m/s
CH4_Front
Peak 17
CH4 (ppm)
Distance (m) or Direction (deg)
Distance to Peak: 17 Start
Direction
Peak 17
Seconds
Peak: 19 Start: 05-26-21 15:51:51
Ave_BL: 1.95 ppm
Max: 0.313 ppm
Dur: 18 s
Ave_Speed: 12 m/s
CH4_Front
Peak 19
CH4 (ppm)
Distance (m) or Direction (deg)
Distance to Peak: 19 Start
Direction
Peak 19
Seconds

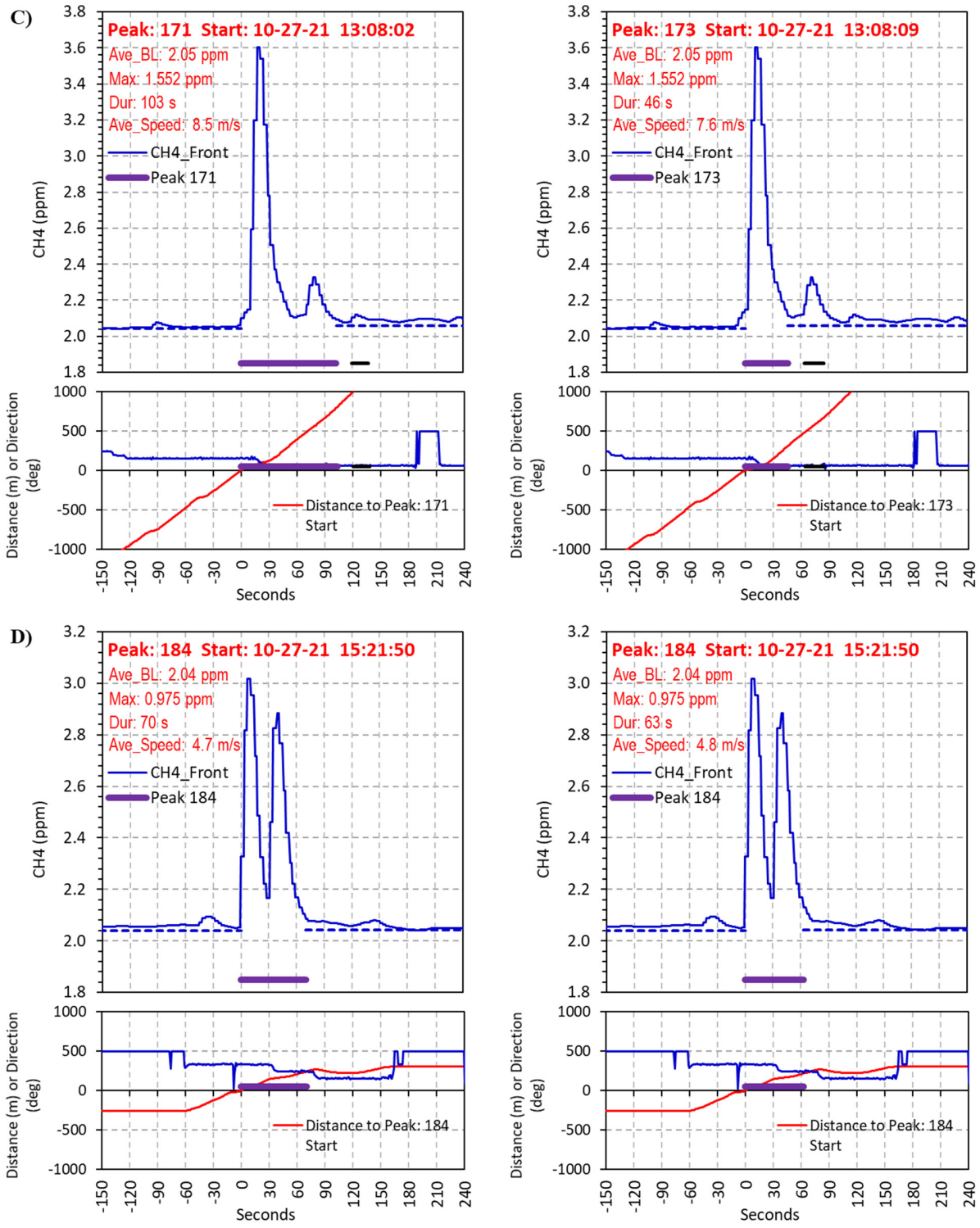


Figure S5. Comparison of selected peaks plots with R_{thresh} = 1.025 (left) and R_{thresh} = 1.05 (right). Fixed t_{window} = 450 s and p = 0.05 were applied. Solid bars under the trend plot indicate each peak detected, and dashed bars indicate pre- and post- peak baseline levels. The moving direction was set to 500 degrees if MPAL stopped.

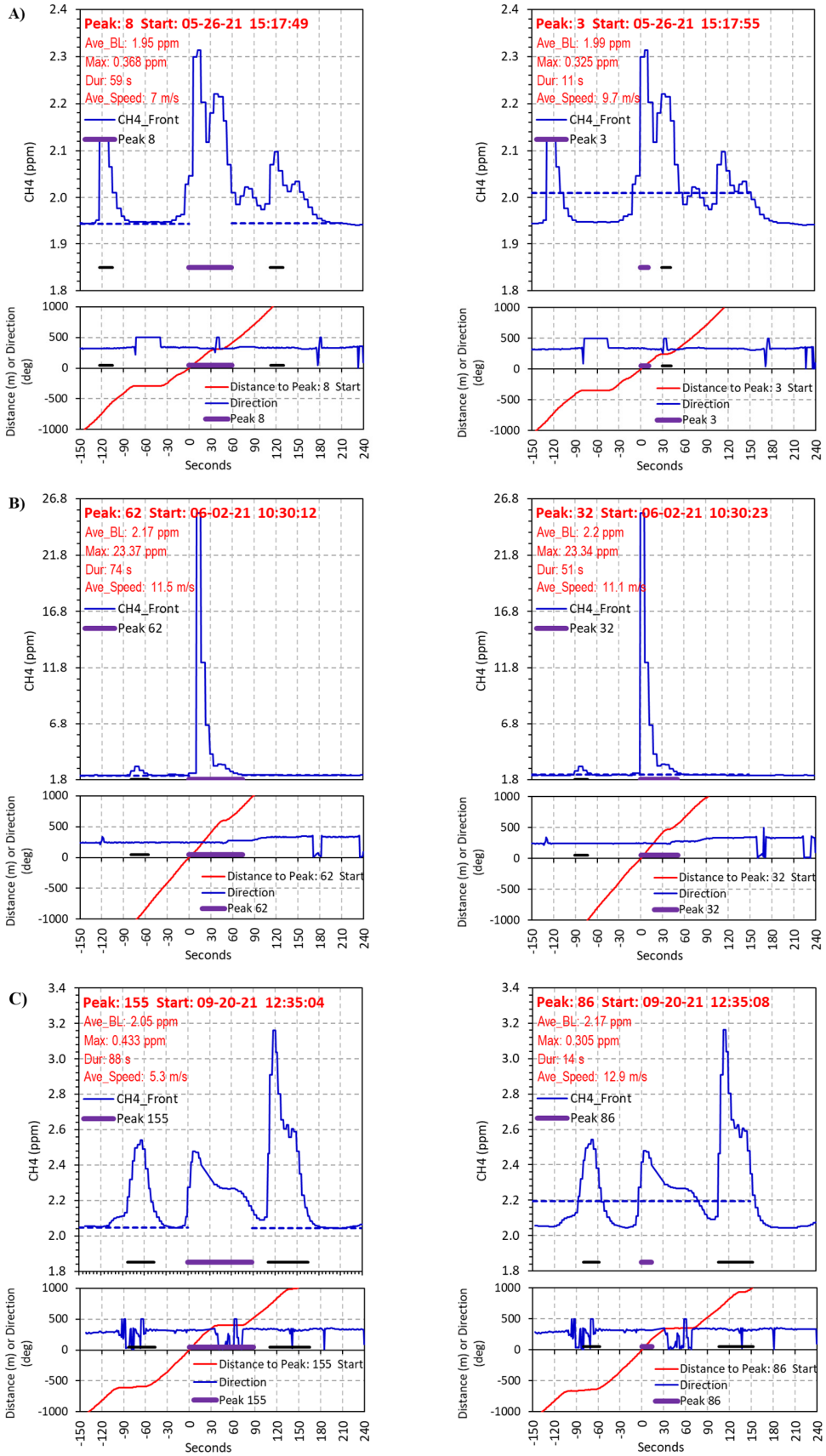


Figure S6. Comparison of selected peaks plots between parameter set 4 (left) and the parameters by Weller et al. [21] (right). Solid bars under the trend plot indicate each peak, dashed bars indicate peak baseline levels estimated using the corresponding method. The moving direction was set to 500 degrees if MPAL stopped.

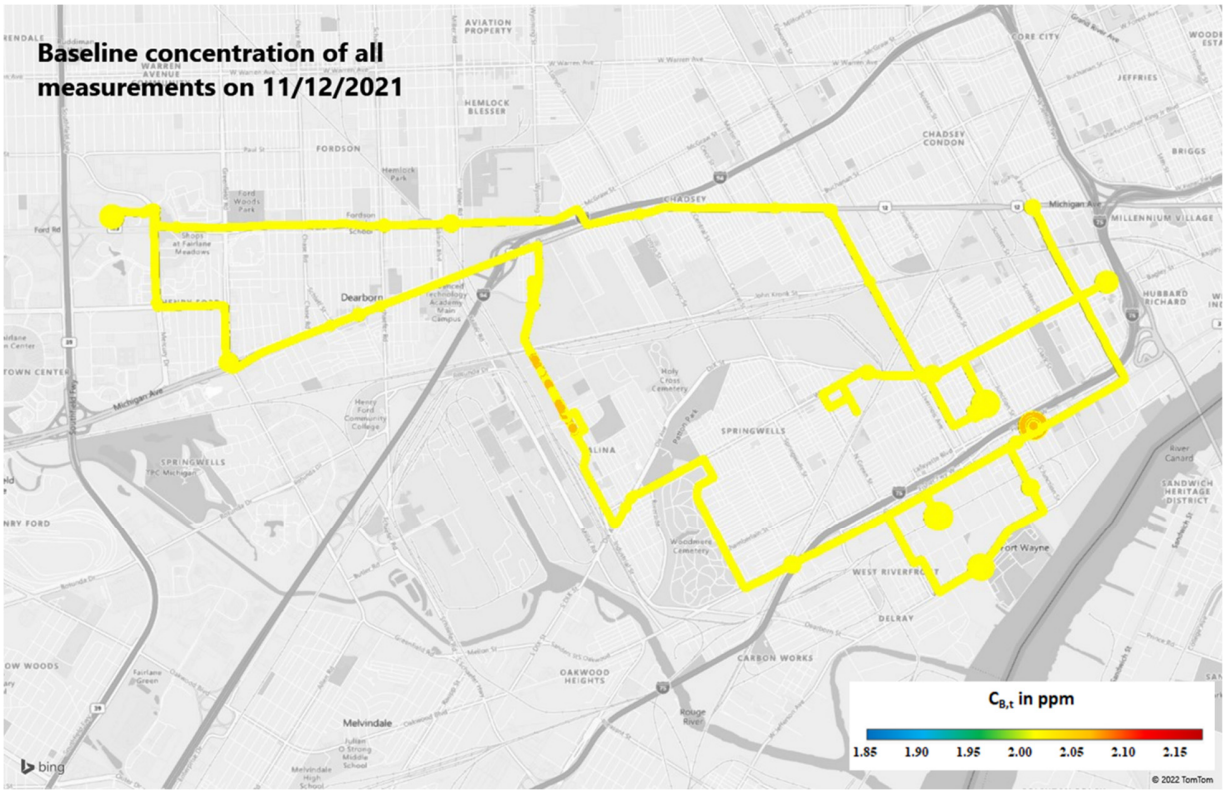


Figure S7. Map of $C_{B,t}$ for all measurements on 11/12/2021. The size of the dot indicates the number of measurements at the same location. No obvious spatial variations are found. Background CH_4 concentration was around 2 ppm and was only elevated by ~ 0.05 ppm at two spots.

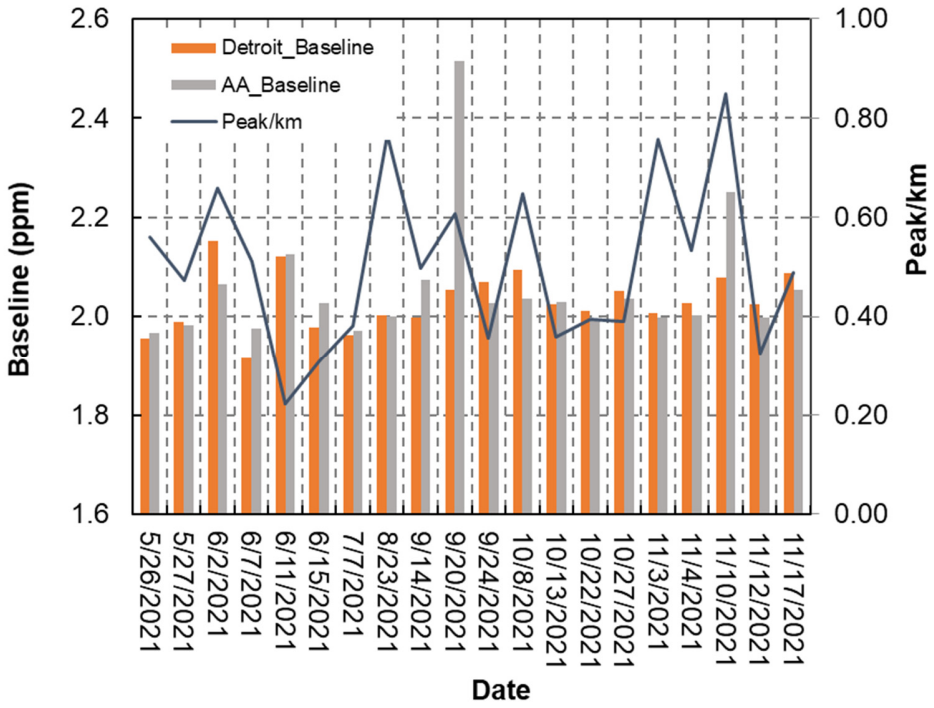


Figure S8. Median $C_{B,t}$ (columns) of all measurements on each of the analyzed 20 days, which showed temporal background variations. The number of peaks per kilometer traveled on each day (line) was relatively unaffected by the median baseline concentration.