

Table S1: *Lasso* linear regression models for full-scale WTP HAA(fp) and THM(fp) data.

Coefficient	Full-scale THM(fp)		Coefficient	Full-scale HAA(fp)	
	Excitation wavelength (nm)	Emission wavelength (nm)		Excitation wavelength (nm)	Emission wavelength (nm)
996.6	435	480	1026.7	393	410
510.8	408	437	98.4	399	597
503.7	390	604	86.8	264	374
442.7	366	610	60.3	246	570
382.5	402	610	50.3	402	604
23.4	243	540	43.7	255	384
21.8	<i>ind.</i> ^a	<i>ind.</i> ^a	31.1	240	520
15.8	243	563	22	252	391
0.7	<i>int.</i> ^b	<i>int.</i> ^b	20.6	243	520
			10.8	249	557
			3.1	249	604
			3.1	243	377
			-0.3	246	620
			-3	<i>int.</i>	<i>int.</i>
			-7.8	240	312
			-41	249	594
			-49.9	255	290
			-127	261	309
			-218.3	252	293

^aformation potential indicator variable (1=formation potential, 0=instantaneous)

^bintercept

Table S2: *Lasso* linear regression models for bench-scale HAAfp and THMfp data.

	Bench-scale THMfp			Bench-scale HAAfp	
Coefficient	Excitation wavelength (nm)	Emission wavelength (nm)	Coefficient	Excitation wavelength (nm)	Emission wavelength (nm)
109.1	438	567	351.8	423	453
65.6	<i>int.</i> ^a	<i>int.</i>	127.8	405	590
42.6	258	610	62.3	<i>int.</i> ^a	<i>int.</i>
36	402	594	55.3	429	466
9.4	243	607	21	243	540
8.8	246	557	8.5	246	553
8.1	246	319	6.4	246	607
6.6	252	322	3.2	573	620
6	258	309	3.1	249	607
5.8	243	319	1.1	429	583
4.2	255	319	0.4	384	620
1.1	243	287	-0.5	249	570
0.004	240	414	-2.5	240	381
-1.2	258	293	-3	240	447
-1.4	261	290	-4.2	240	312
-1.4	240	563	-6.3	261	387
-1.9	243	358	-6.4	243	325
-2.3	264	287	-7.4	255	530
-3	267	284	-8	252	345
-4.6	252	358	-13.5	243	397
-7.2	243	430	-13.5	270	560
-7.7	243	303	-14.1	258	358
-8.9	246	277	-17.7	243	358
-9.1	246	600			
-35.7	258	284			

^aintercept

Table S3: FRI linear regression models for each of the four data sets.

Variable	Coefficients			
	Full-scale THM(fp)	Full-scale HAA(fp)	Bench- scale THMfp	Bench- scale HAAfp
Formation potential indicator ^a	51.981	6.836	<i>n/a</i>	<i>n/a</i>
Intercept	-11.273	-5.393	66.041	90.886
Region I	-4.003	1.142	0.039	-0.265
Region II	2.574	4.896	-0.649	-1.202
Region III	-0.008	-1.031	-0.092	-0.689
Region IV	-0.532	-1.054	-0.003	-0.074
Region V	0.148	0.323	0.041	0.17

^a 1=formation potential, 0=instantaneous

Table S4: Principal components linear regression models for each of the four data sets.

Variable	Coefficients			
	Full-scale THM(fp)	Full-scale HAA(fp)	Bench- scale THMfp	Bench- scale HAAfp
Formation potential indicator ^a	44.6	-0.9	<i>n/a</i>	<i>n/a</i>
Intercept	25.15	43.4	75.94	79.17
PC 1	0.15	0.15	-0.07	-0.08
PC 2	-0.19	0.06	-0.31	-0.84
PC 3	-0.12	-0.08	-0.27	-1.1
PC 4	-0.34	0.28	-0.19	0.61
PC 5	-0.53	0.24	-0.08	-0.32
PC 6	-0.54	0.11	0.53	
PC 7	-0.13	-0.18	0.09	
PC 8		0.13	0.83	
PC 9		0.59	-0.05	
PC 10			-0.38	
PC 11			0.36	
PC 12			0.22	
PC 13			0.75	
PC 14			-0.65	
PC 15			0.2	
PC 16			0.71	
PC 17			-0.55	
PC 18			0.16	
PC 19			-0.34	
PC 20			-0.32	
PC 21			0.64	

^a 1=formation potential, 0=instantaneous

Table S5: Supervised principal components linear regression models for each of the four data sets.

Variable	Coefficients			
	Full-scale THM(fp)	Full-scale HAA(fp)	Bench- scale THMfp	Bench- scale HAAfp
Formation potential indicator ^a	39.7	-2	<i>n/a</i>	<i>n/a</i>
Intercept	28.6	44.2	75.9	79.2
PC 1	86.2	155.3	-13.4	-5.7
PC 2	-169.6	39.8		-48.4
PC 3		-108.9		2.4
PC 4		75.7		-9.5
PC 5		784.3		-9
PC 6		-492.5		0.3
PC 7		35.8		-32.6
PC 8		-145.4		-20
PC 9				30.7

^a 1=formation potential, 0=instantaneous

Table S6: Spectral features accounting for 95% of relative influence in boosted regression tree ensembles fitted to full-scale HAA(fp) and THM(fp) data (listed in decreasing order by relative influence).

<u>Full-scale THM(fp)</u>		<u>Full-scale HAA(fp)</u>	
Excitation wavelength (nm)	Emission wavelength (nm)	Excitation wavelength (nm)	Emission wavelength (nm)
405	433	267	381
243	358	333	577
240	417	276	604
240	427	246	516
558	587	393	407
348	583	246	453
414	533	261	546
243	309	381	580
591	614	264	567
246	322	354	594
534	573	252	600
243	284	342	358
243	410	255	567
435	476	372	600
		264	617
		411	557
		318	607
		408	423
		252	290
		246	567
		258	364
		435	543

Table S7: Spectral features accounting for 95% of relative influence in boosted regression tree ensembles fitted to bench-scale HAAfp data (listed in decreasing order by relative influence).

Ex. (nm)	Em. (nm)	Ex. (nm)	Em. (nm)	Ex. (nm)	Em. (nm)	Ex. (nm)	Em. (nm)	Ex. (nm)	Em. (nm)	Ex. (nm)	Em. (nm)
249	342	396	617	270	557	414	440	243	607	255	309
243	355	243	536	432	466	249	335	255	361	255	332
444	470	315	329	246	587	258	530	246	520	240	553
252	342	246	557	243	567	246	433	246	319	582	617
255	526	393	600	447	496	249	543	297	306	318	335
438	463	381	614	450	557	267	303	441	480	408	583
252	293	243	394	243	384	246	476	249	590	264	322
258	523	240	417	246	309	255	557	255	345	249	553
246	387	240	410	261	384	261	348	240	377	441	553
573	617	408	430	252	303	252	355	243	296	261	303
240	355	267	355	279	355	243	387	246	533	261	332
402	414	300	610	549	597	246	332	249	490	306	316
402	423	240	526	243	600	243	503	417	573	435	557
396	410	258	300	240	533	420	450	249	486	240	368
582	604	249	540	258	610	240	351	420	570	246	280
246	604	240	563	258	526	372	600	264	319	249	417
249	604	393	614	525	587	240	342	261	607	258	280
258	335	240	496	516	553	243	570	240	604	387	423
402	617	297	610	243	319	249	287	261	342	261	284
240	500	252	583	258	306	582	614	240	335	252	335
444	560	570	604	246	486	252	553	246	526	243	427
405	587	240	610	249	309	264	520	246	364	243	513
360	614	255	594	444	476	243	280	387	617	246	306
366	614	558	583	243	293	240	580	240	513	240	540
438	536	240	319	240	443	243	443	594	614	249	296
420	453	426	453	384	597	243	533	240	490		
408	553	243	303	240	516	399	577	252	617		
270	617	246	329	240	309	426	577	243	614		
246	355	240	420	243	420	249	319	252	374		

Table S8: Spectral features accounting for 95% of relative influence in boosted regression tree ensembles fitted to bench-scale THMfp data (listed in decreasing order by relative influence).

Ex. (nm)	Em. (nm)	Ex. (nm)	Em. (nm)	Ex. (nm)	Em. (nm)	Ex. (nm)	Em. (nm)	Ex. (nm)	Em. (nm)	Ex. (nm)	Em. (nm)
402	590	249	325	240	567	258	300	255	355	264	300
246	600	246	546	246	348	402	417	240	443	321	335
435	583	423	597	381	594	441	553	243	546	276	325
240	560	246	316	405	433	261	309	243	587	246	550
243	530	249	607	240	312	321	338	240	423	243	480
243	355	252	604	240	355	384	604	444	493	276	604
411	617	243	516	243	433	240	394	243	407	252	300
246	322	438	463	240	450	429	536	285	364	255	338
243	604	267	617	429	453	258	580	240	433	240	381
396	410	240	296	249	573	426	460	279	335	273	342
240	325	240	351	324	342	249	306	240	570	444	553
240	284	246	577	255	587	402	617	246	280	249	610
252	355	243	364	246	342	243	280	252	570	252	322
252	325	252	319	279	423	243	351	258	371	246	590
246	329	258	322	549	617	240	503	246	325	246	364
243	573	249	597	240	510	435	480	267	309	252	312
258	607	246	290	240	594	243	506	246	338	270	332
240	437	249	355	243	600	438	553	249	287	588	600
246	553	417	440	249	371	261	300	249	543	273	293
249	567	444	470	243	306	246	358	255	348	255	312
246	300	249	580	252	371	282	351	423	466	252	573
243	284	240	540	435	553	276	338	264	342	258	348
252	607	438	466	264	583	243	543	243	614	393	614
240	516	267	342	426	466	243	338	405	583	252	583
240	526	258	309	444	496	252	536	240	563	408	433
243	316	240	533	246	430	435	466	243	594	384	607
246	597	243	557	240	358	243	540	240	410	246	540
243	300	258	567	312	319	240	513	243	570	240	543
423	456	246	526	243	290	240	506	267	577	576	594
255	567	255	290	246	523	249	374	384	617	252	597
243	427	258	293	252	394	258	306	261	348	255	316
252	309	243	583	246	274	243	322	429	577	243	391
240	557	438	583	249	384	240	290	435	460	288	303
249	594	252	617	246	460	240	384	243	397	285	306
243	550	240	573	423	580	240	364	261	345	249	342
261	604	240	348	255	306	549	597	432	563	261	594
249	604	525	577	246	567	249	560	240	377	240	306
279	427	246	557	246	617	438	473	294	312	282	345
240	546	255	563	300	309	249	290	246	303	249	617
411	594	243	617	258	594	288	342	423	573	318	332
267	329	249	557	558	583	387	617	552	583	243	387
246	536	426	580	240	536	273	604	240	486	246	361
249	284	477	540	285	332	396	617	246	573	438	543
315	322	396	607	252	332	243	597	441	570		
252	614	243	536	255	329	273	361	255	319		
438	536	414	433	246	587	246	563	252	287		