



Supplement of

Improving the quantification of peak concentrations for air quality sensors via data weighting

Caroline Frischmon et al.

Correspondence to: Caroline Frischmon (caroline.frischmon@colorado.edu)

The copyright of individual parts of the supplement might differ from the article licence.

Table S1. Sensor information

Sensor	Target Pollutant	Manufacturer	Approximate Cost (USD 2024)
Alphasense CO-B4	CO	Alphasense	80
Figaro 2600	tVOC	Figaro	15
Figaro 2602	tVOC	Figaro	15
Figaro 2611	CH4	Figaro	15
MQ4	CH4	Hanwei	5

Table S2. Summary statistics for CO calibration models, where RF is random forest and MLR is multiple linear regression. Piece indicates a piecewise weighting function with a percentile cut off listed subsequently.

Model	Training R2	Testing R2	R2	Training RMSE	Testing RMSE	RMSE	Training MBE	Testing MBE	MBE
RF unweighted	0.82	0.78	0.81	0.075	0.092	0.080	0.000	0.012	0.004
RF Sigmoid	0.79	0.73	0.77	0.082	0.102	0.088	0.031	0.043	0.035
RF Sigmoid z=0.5	0.76	0.71	0.74	0.086	0.106	0.093	0.040	0.052	0.044
RF Sigmoid z=1	0.73	0.68	0.70	0.092	0.111	0.099	0.050	0.061	0.054
RF Sigmoid z=1.5	0.68	0.63	0.65	0.101	0.119	0.107	0.063	0.074	0.067
RF Sigmoid z=2	0.63	0.59	0.60	0.108	0.125	0.114	0.073	0.083	0.077
RF Sigmoid z=2.5	0.57	0.53	0.54	0.117	0.134	0.123	0.085	0.095	0.088
RF Sigmoid z=3	0.43	0.38	0.40	0.135	0.154	0.141	0.105	0.119	0.109
RF piece 75	0.76	0.72	0.74	0.087	0.104	0.093	0.036	0.045	0.040
RF piece 80	0.77	0.72	0.75	0.084	0.103	0.090	0.031	0.040	0.035

RF piece 85	0.78	0.72	0.76	0.084	0.103	0.090	0.029	0.038	0.033
RF piece 90	0.79	0.74	0.77	0.081	0.100	0.087	0.024	0.034	0.027
RF piece 95	0.80	0.76	0.79	0.079	0.096	0.083	0.015	0.027	0.018
RF piece 99	0.82	0.78	0.81	0.076	0.092	0.080	0.004	0.017	0.007
RF piece 99.5	0.82	0.78	0.81	0.075	0.091	0.080	0.003	0.016	0.007
RF piece 99.9	0.82	0.79	0.81	0.075	0.090	0.079	0.001	0.014	0.005
MLR unweighted	0.74	0.77	0.75	0.090	0.095	0.091	0.000	0.000	0.003
MLR Sigmoid	0.71	0.74	0.71	0.096	0.101	0.098	0.032	0.033	0.035
MLR Sigmoid z=.5	0.68	0.72	0.68	0.100	0.104	0.102	0.043	0.043	0.046
MLR Sigmoid z=1	0.64	0.68	0.64	0.107	0.111	0.109	0.055	0.056	0.058
MLR Sigmoid z=1.5	0.57	0.63	0.58	0.116	0.119	0.118	0.070	0.071	0.073
MLR Sigmoid z=2	0.47	0.54	0.48	0.129	0.132	0.131	0.088	0.090	0.091
MLR Sigmoid z=2.5	0.31	0.41	0.32	0.147	0.151	0.149	0.110	0.114	0.113
MLR Sigmoid z=3	0.07	0.20	0.08	0.172	0.176	0.174	0.138	0.143	0.141
MLR piece 75	0.68	0.71	0.68	0.101	0.105	0.103	0.045	0.046	0.048
MLR piece 80	0.68	0.71	0.68	0.100	0.105	0.103	0.044	0.045	0.047
MLR piece 85	0.68	0.71	0.68	0.100	0.105	0.102	0.043	0.043	0.046
MLR piece 90	0.69	0.71	0.69	0.099	0.105	0.101	0.037	0.036	0.040
MLR piece 95	0.70	0.71	0.70	0.097	0.105	0.099	0.026	0.025	0.029
MLR piece 99	0.72	0.74	0.73	0.094	0.101	0.094	0.011	0.009	0.014

MLR piece 99.5	0.73	0.74	0.74	0.092	0.099	0.093	0.008	0.006	0.011
MLR piece 99.9	0.74	0.76	0.75	0.090	0.096	0.091	0.003	0.003	0.006

Table S3. Summary statistics for CH₄ calibration models, where RF is random forest and MLR is multiple linear regression. Piece indicates a piecewise weighting function with a percentile cut off listed subsequently.

Model	Training R2	Testing R2	R2	Training RMSE	Testing RMSE	RMSE	Training MBE	Testing MBE	MBE
RF unweighted	0.83	0.67	0.79	0.149	0.221	0.169	0.000	-0.060	- 0.006
RF Sigmoid	0.86	0.72	0.8	0.137	0.203	0.163	0.025	-0.034	0.021
RF Sigmoid z=0.5	0.85	0.73	0.79	0.141	0.201	0.166	0.036	-0.023	0.032
RF Sigmoid z=1	0.83	0.74	0.78	0.148	0.197	0.172	0.047	-0.010	0.045
RF Sigmoid z=1.5	0.8	0.73	0.75	0.160	0.200	0.182	0.063	0.006	0.061
RF Sigmoid z=2	0.77	0.72	0.72	0.174	0.205	0.195	0.080	0.019	0.077
RF Sigmoid z=2.5	0.7	0.69	0.65	0.196	0.214	0.215	0.101	0.041	0.100
RF Sigmoid z=3	0.59	0.63	0.54	0.232	0.235	0.247	0.134	0.076	0.134
RF piece 75	0.84	0.72	0.79	0.144	0.206	0.167	0.027	-0.026	0.023
RF piece 80	0.83	0.72	0.78	0.147	0.206	0.170	0.031	-0.023	0.028
RF piece 85	0.83	0.72	0.78	0.149	0.204	0.172	0.035	-0.020	0.031
RF piece 90	0.82	0.73	0.77	0.151	0.202	0.174	0.038	-0.021	0.034
RF piece 95	0.82	0.74	0.77	0.154	0.198	0.176	0.039	-0.024	0.036

RF piece 99	0.86	0.71	0.8	0.137	0.207	0.165	0.021	-0.040	0.016
RF piece 99.5	0.86	0.72	0.81	0.134	0.205	0.161	0.014	-0.042	0.009
RF piece 99.9	0.86	0.7	0.81	0.133	0.213	0.161	0.007	-0.052	0.002
MLR unweighted	0.75	0.69	0.73	0.182	0.215	0.190	0.000	-0.048	- 0.006
MLR Sigmoid	0.71	0.65	0.7	0.196	0.230	0.200	0.031	-0.058	0.024
MLR Sigmoid z=.5	0.66	0.56	0.66	0.210	0.256	0.214	0.042	-0.073	0.036
MLR Sigmoid z=1	0.56	0.35	0.56	0.238	0.311	0.243	0.056	-0.100	0.050
MLR Sigmoid z=1.5	0.36	-0.12	0.35	0.287	0.408	0.295	0.073	-0.141	0.067
MLR Sigmoid z=2	-0.01	-1.03	- 0.05	0.362	0.551	0.376	0.096	-0.193	0.090
MLR Sigmoid z=2.5	-0.64	-2.54	- 0.74	0.461	0.728	0.483	0.133	-0.244	0.127
MLR Sigmoid z=3	-1.59	-4.56	-1.8	0.580	0.912	0.613	0.203	-0.265	0.197
MLR piece 75	0.7	0.64	0.69	0.198	0.233	0.203	0.036	-0.065	0.030
MLR piece 80	0.69	0.63	0.68	0.201	0.236	0.206	0.045	-0.059	0.038
MLR piece 85	0.68	0.62	0.67	0.205	0.239	0.210	0.054	-0.052	0.047
MLR piece 90	0.65	0.59	0.65	0.212	0.247	0.216	0.061	-0.046	0.054
MLR piece 95	0.62	0.56	0.63	0.221	0.256	0.224	0.065	-0.038	0.058
MLR piece 99	0.62	0.54	0.61	0.223	0.261	0.228	0.048	-0.020	0.042
MLR piece 99.5	0.63	0.54	0.62	0.218	0.262	0.226	0.037	-0.016	0.031
MLR piece 99.9	0.71	0.64	0.7	0.194	0.232	0.202	0.019	-0.032	0.013

Table S4. Summary statistics for TVOC calibration models, where RF is random forest and MLR is multiple linear regression. Piece indicates a piecewise weighting function with a percentile cut off listed subsequently.

Model	Training R2	Testing R2	R2	Training RMSE	Testing RMSE	RMSE	Training MBE	Testing MBE	MBE
RF unweighted	0.72	0.64	0.69	52.13	42.57	52.60	-0.09	-0.35	5.75
RF Sigmoid	0.69	0.64	0.63	54.17	42.95	56.85	13.18	9.73	19.54
RF Sigmoid z=0.5	0.67	0.62	0.6	55.99	43.72	59.28	17.87	12.81	24.33
RF Sigmoid z=1	0.66	0.58	0.56	57.18	46.46	62.26	22.83	17.38	29.51
RF Sigmoid z=1.5	0.59	0.49	0.47	62.73	50.74	68.60	29.24	21.54	36.00
RF Sigmoid z=2	0.49	0.39	0.34	69.76	55.59	76.14	36.78	27.31	43.61
RF Sigmoid z=2.5	0.31	0.22	0.12	81.72	63.16	88.31	46.27	33.35	52.96
RF Sigmoid z=3	0.09	0.02	- 0.12	93.72	70.71	99.56	61.68	41.36	67.33
RF piece 75	0.67	0.61	0.61	56.72	44.81	58.66	14.76	10.78	21.03
RF piece 80	0.68	0.6	0.62	55.20	45.23	58.24	15.46	11.00	21.59
RF piece 85	0.64	0.59	0.58	58.71	45.56	60.88	16.83	11.18	22.51
RF piece 90	0.64	0.6	0.57	59.18	45.30	61.90	17.05	9.96	22.63
RF piece 95	0.68	0.65	0.6	55.20	42.47	59.69	14.47	6.68	20.42
RF piece 99	0.72	0.66	0.63	51.58	41.79	57.41	6.86	2.48	13.40
RF piece 99.5	0.74	0.65	0.65	50.09	42.26	55.42	4.56	1.53	11.19
RF piece 99.9	0.74	0.63	0.68	49.56	43.32	53.53	1.77	0.38	8.18

MLR unweighted	0.57	0.61	0.58	64.40	44.68	61.24	0.00	1.20	5.50
MLR Sigmoid	0.5	0.56	0.47	69.65	47.19	68.21	11.79	16.15	17.29
MLR Sigmoid z=.5	0.43	0.51	0.4	73.79	49.99	72.74	14.06	20.10	19.56
MLR Sigmoid z=1	0.33	0.42	0.28	80.26	54.43	79.56	15.50	23.65	21.00
MLR Sigmoid z=1.5	0.15	0.26	0.09	90.29	61.51	89.85	15.53	26.37	21.03
MLR Sigmoid z=2	-0.16	-0.05	- 0.26	105.70	72.95	105.40	13.85	28.35	19.35
MLR Sigmoid z=2.5	-0.7	-0.59	- 0.84	127.76	90.10	127.55	10.86	30.70	16.36
MLR Sigmoid z=3	-1.51	-1.47	- 1.73	155.35	112.11	155.22	8.20	35.72	13.70
MLR piece 75	0.51	0.55	0.49	68.36	48.12	66.80	13.38	17.47	18.88
MLR piece 80	0.49	0.52	0.47	69.73	49.25	68.44	15.00	20.10	20.49
MLR piece 85	0.47	0.49	0.44	71.63	50.95	70.63	16.97	23.35	22.47
MLR piece 90	0.42	0.45	0.38	74.54	52.85	73.86	19.30	26.52	24.80
MLR piece 95	0.35	0.45	0.3	79.24	52.76	78.89	21.69	25.48	27.19
MLR piece 99	0.26	0.52	0.2	84.33	49.51	84.04	18.75	15.06	24.25
MLR piece 99.5	0.31	0.56	0.26	81.47	47.35	80.94	15.21	10.13	20.71
MLR piece 99.9	0.48	0.62	0.46	70.50	44.18	68.88	7.10	6.70	12.60

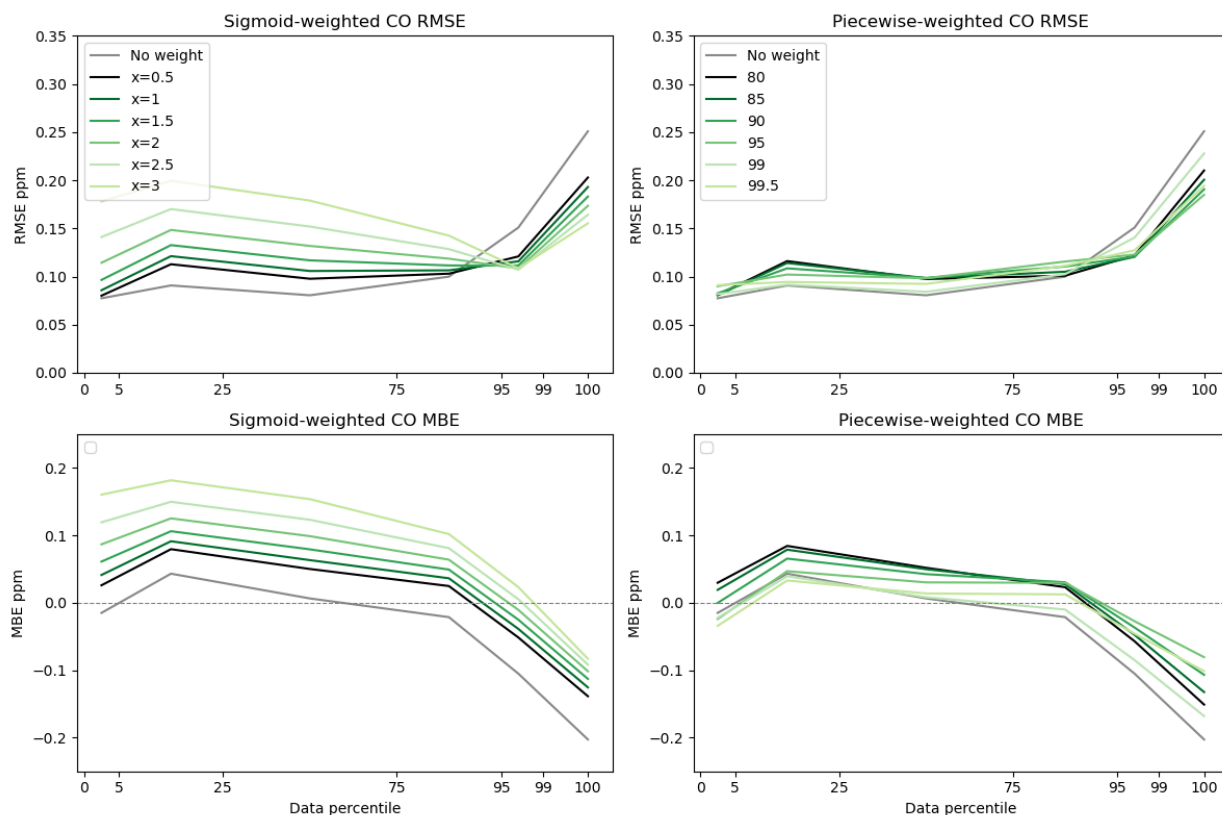


Figure S1. CO sensitivity to weighting parameters for MLR models. RMSE and MBE are displayed as a function of data percentile for unweighted data as well as for piecewise and sigmoidal weights. Concentrations of CO range between 0.04 and 1.84 ppm. Dashed lines indicate MLR fits whereas filled lines represent RF fits. Lighter colors indicate increased offsets/percentile set points for weighting distributions.

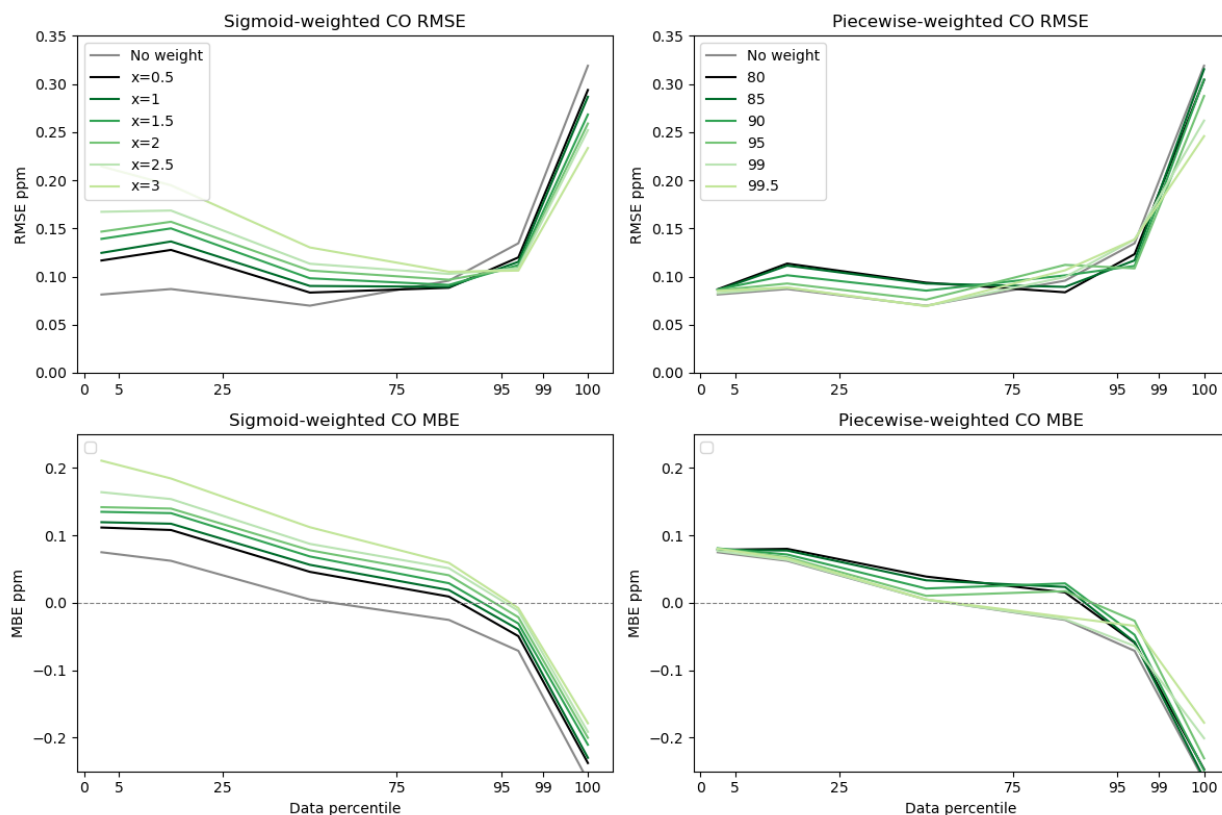


Figure S2. CO sensitivity to weighting parameters for RF models. RMSE and MBE are displayed as a function of data percentile for unweighted data as well as for piecewise and sigmoidal weights. Concentrations of CO range between 0.04 and 1.84 ppm. Dashed lines indicate MLR fits whereas filled lines represent RF fits. Lighter colors indicate increased offsets/percentile set points for weighting distributions.

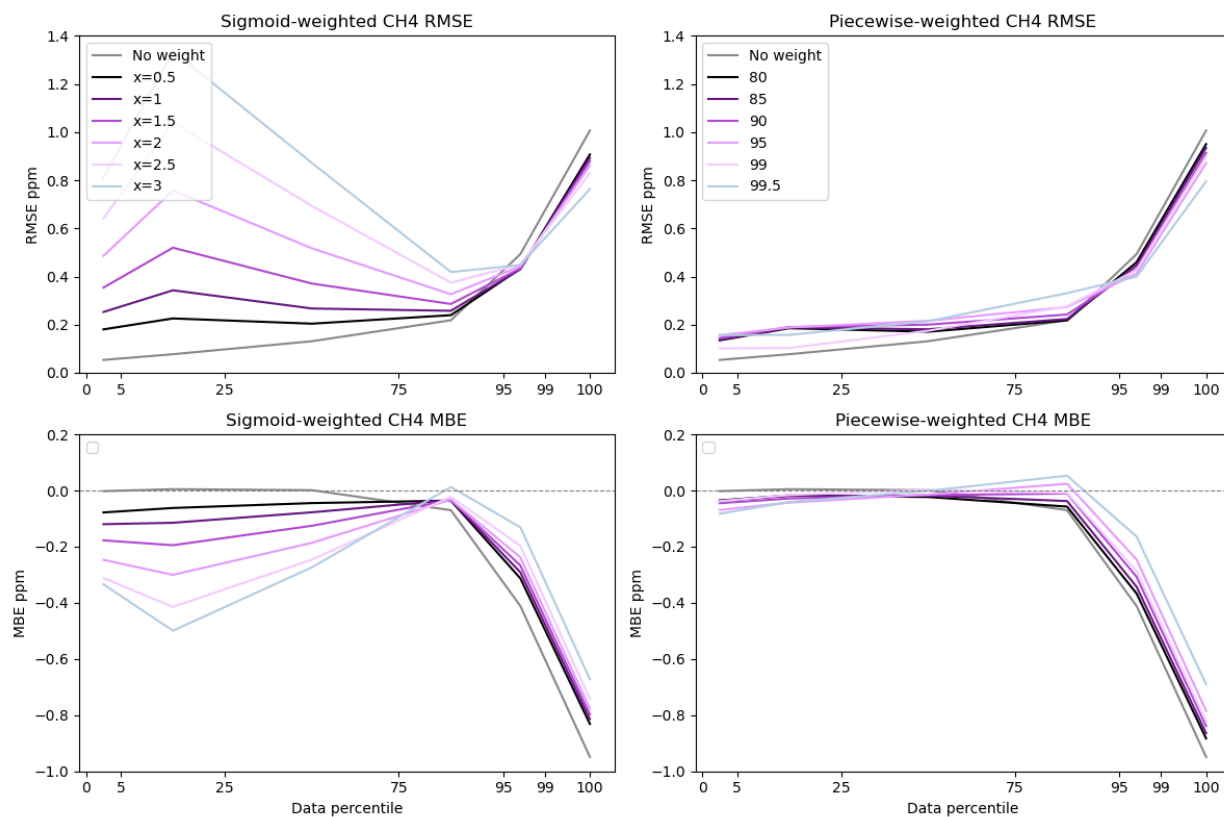


Figure S3. CH₄ sensitivity to weighting parameters for MLR models. RMSE and MBE are displayed as a function of data percentile for unweighted data as well as for piecewise and sigmoidal weights. Concentrations of CH₄ range between 1.96 and 8.75 ppm. Lighter colors indicate increased offsets/percentile set points for weighting distributions.

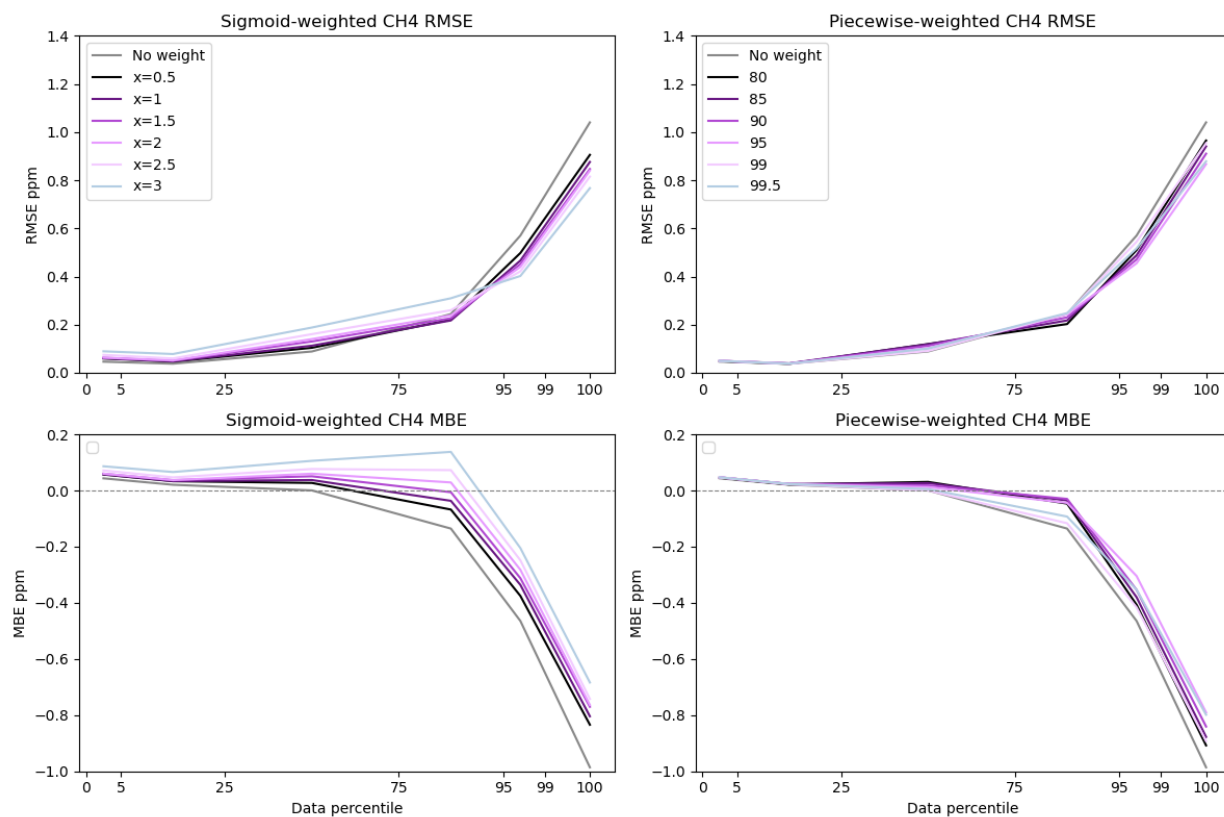


Figure S4. CH₄ sensitivity to weighting parameters for RF models. RMSE and MBE are displayed as a function of data percentile for unweighted data as well as for piecewise and sigmoidal weights. Concentrations of CH₄ range between 1.96 and 8.75 ppm. Lighter colors indicate increased offsets/percentile set points for weighting distributions.

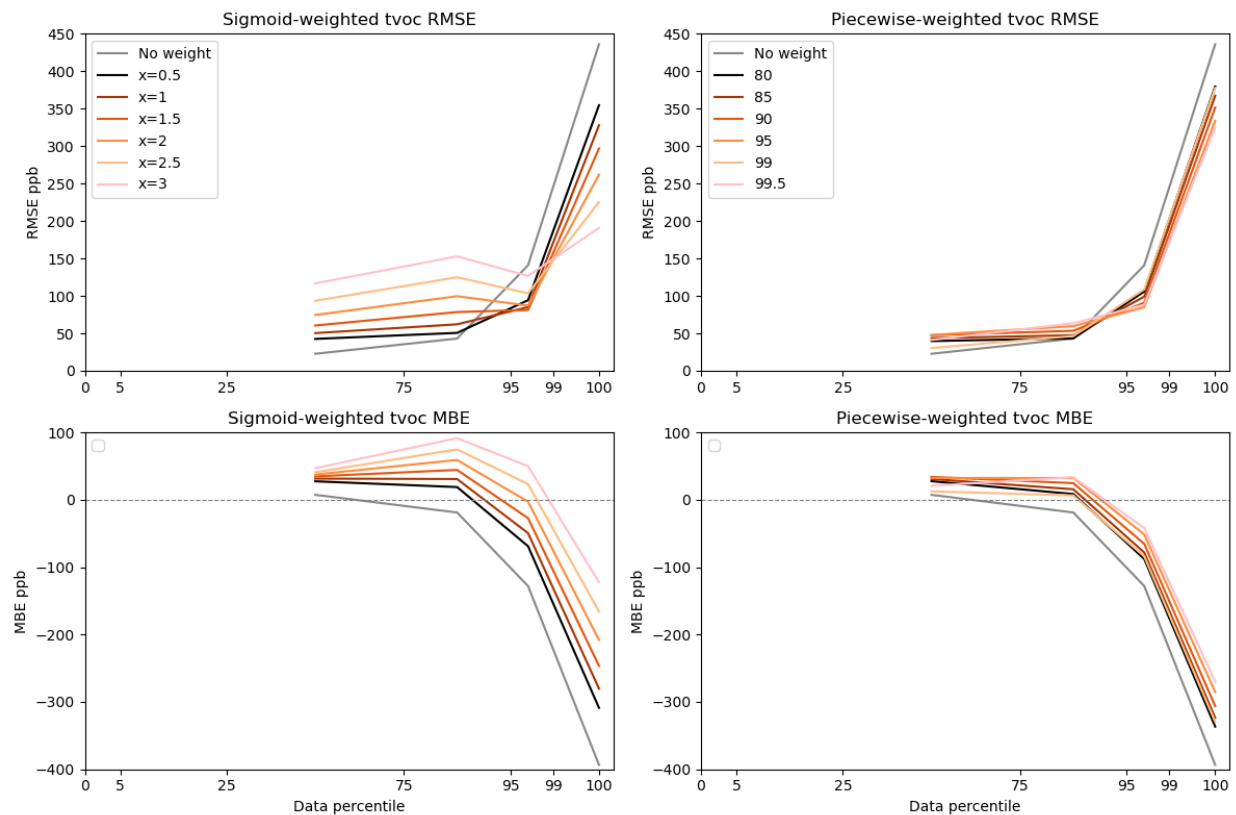


Figure S5. TVOC sensitivity to weighting parameters for MLR models. RMSE and MBE are displayed as a function of data percentile for unweighted data as well as for piecewise and sigmoidal weights. TVOC Concentrations ranged from below the detection limit of the reference instrument (15 ppb) to 2980 ppb. Lighter colors indicate increased offsets/percentile set points for weighting distributions.

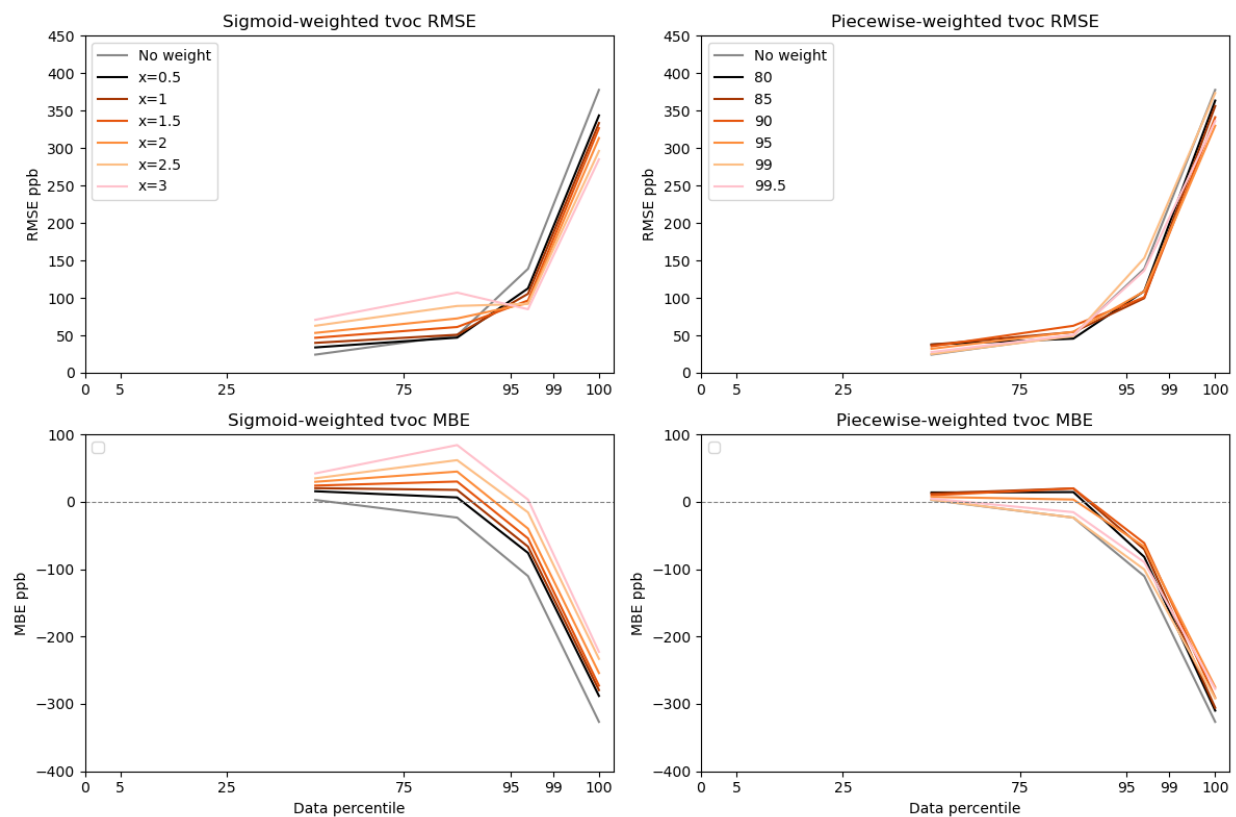


Figure S6. TVOC sensitivity to weighting parameters for RF models. RMSE and MBE are displayed as a function of data percentile for unweighted data as well as for piecewise and sigmoidal weights. TVOC Concentrations ranged from below the detection limit of the reference instrument (15 ppb) to 2980 ppb. Lighter colors indicate increased offsets/percentile set points for weighting distributions.