Supplement of Atmos. Meas. Tech., 12, 903–920, 2019
https://doi.org/10.5194/amt-12-903-2019-supplement

Atmospheric Measurement Techniques
Open Access
EGU

Supplement of

Development of a general calibration model and long-term performance evaluation of low-cost sensors for air pollutant gas monitoring

Carl Malings et al.

Correspondence to: Carl Malings (cmalings@andrew.cmu.edu)

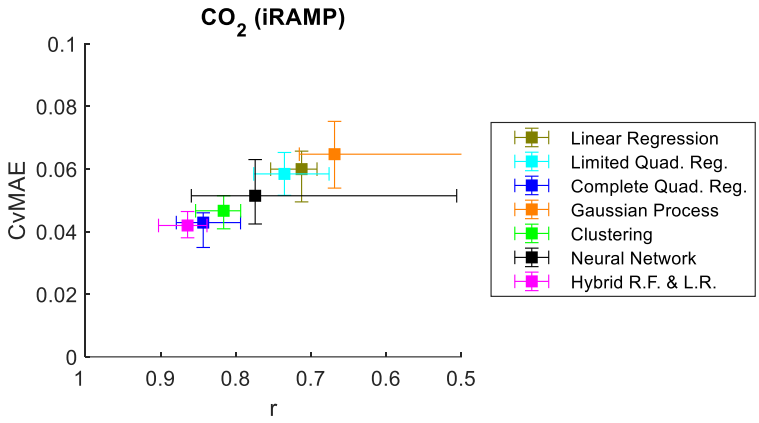


Figure S1. Results corresponding to Fig. 2 for CO_2.

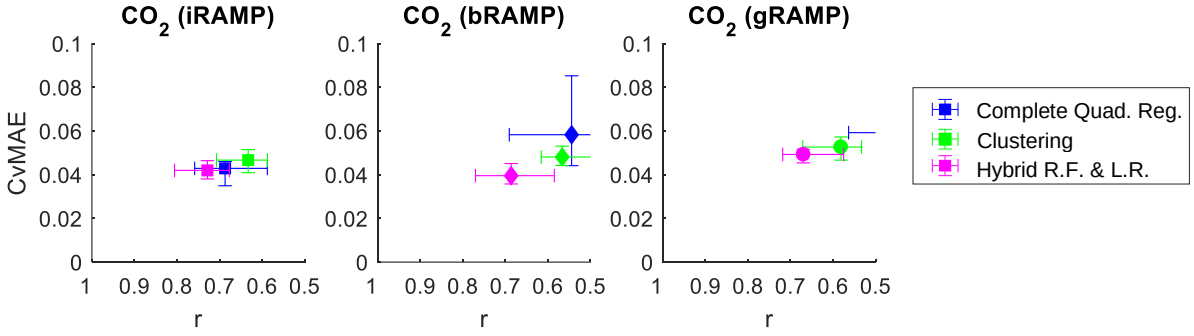


Figure S2. Results corresponding to Fig. 3 for CO_2.

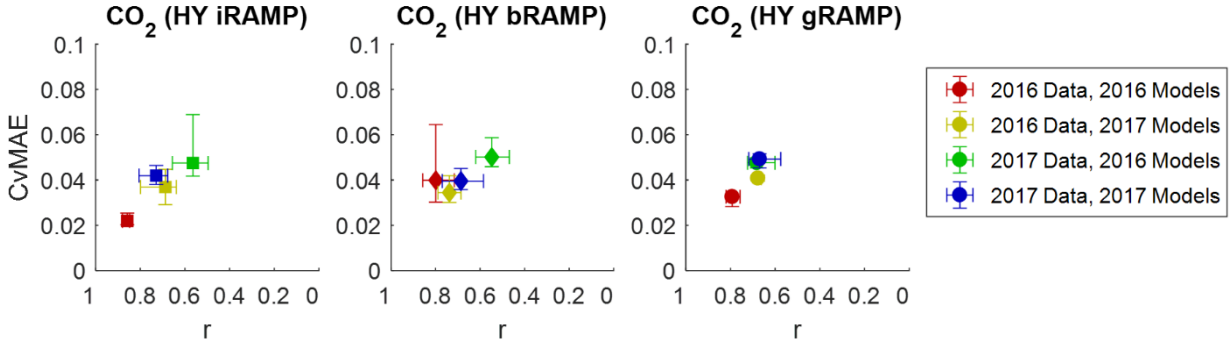


Figure S3. Results corresponding to Fig. 6 for CO_2.

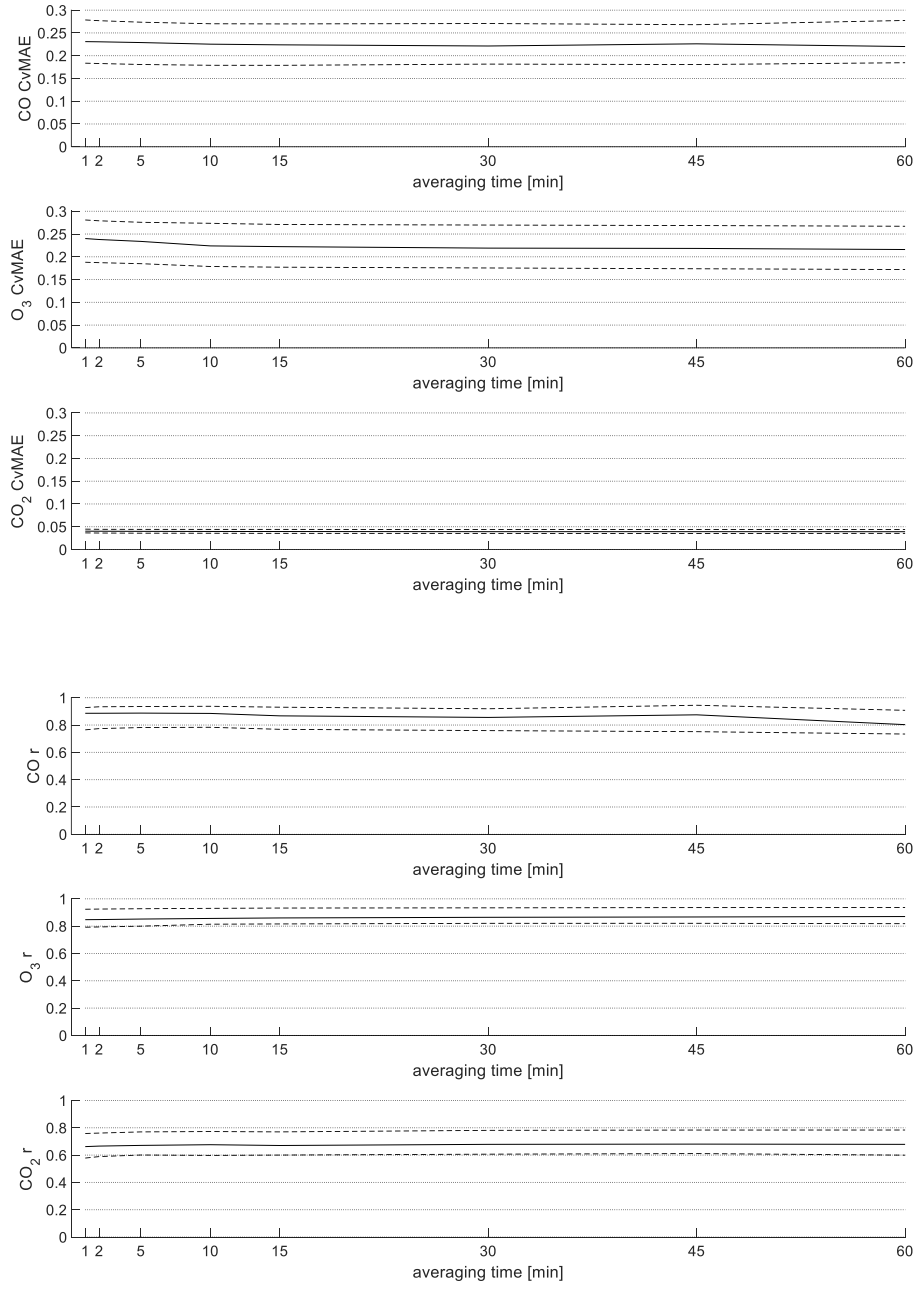
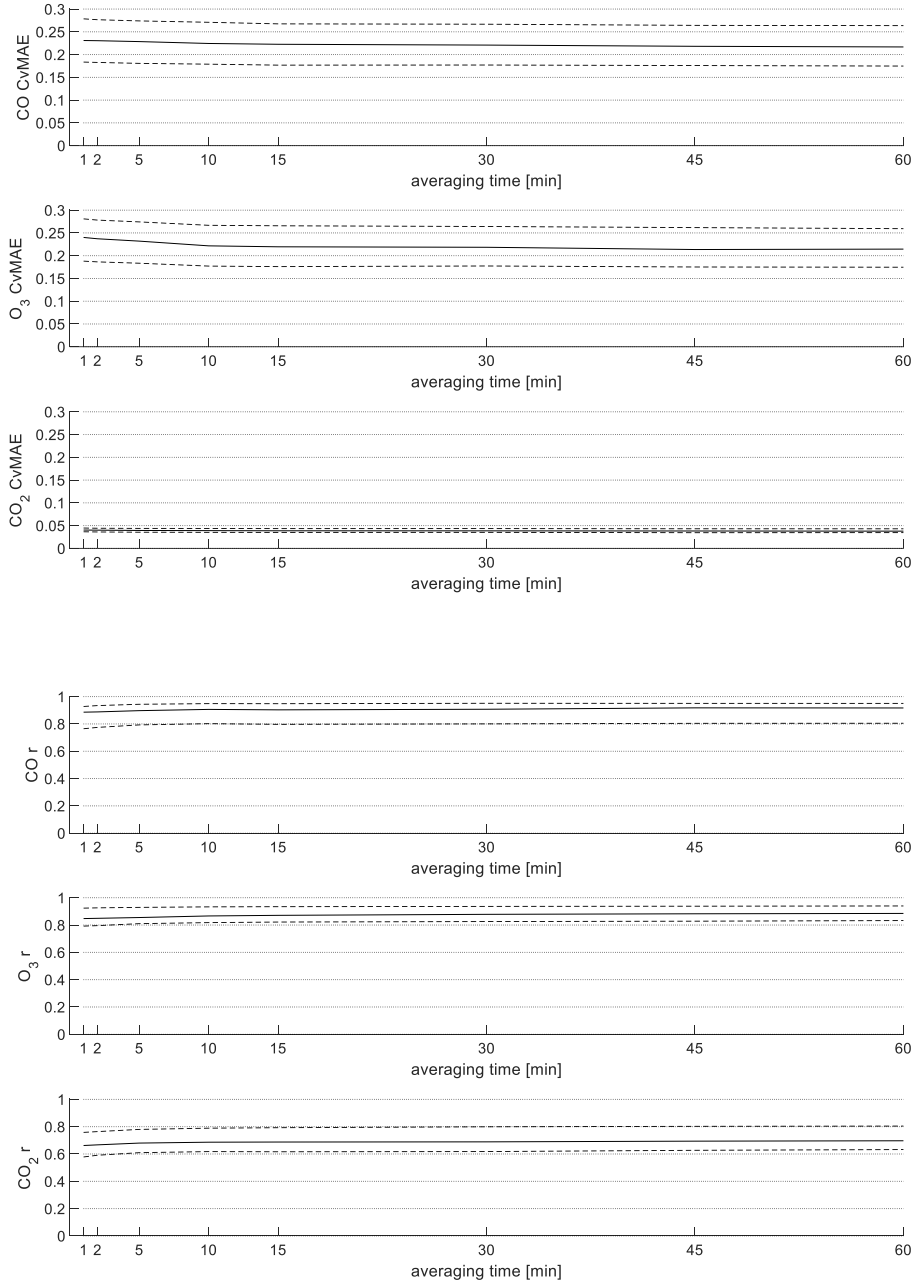


Figure S4. An evaluation of the performance of the calibration algorithms as a function of the averaging period applied to the raw RAMP data. All models are trained using data collected at the CMU site in 2017. Performance is also evaluated at the CMU site in 2017. Solid lines indicate median performance across RAMPs, dashed lines indicate 25th and 75th percentiles of performance. For CO, the gRAMP LQR model is used; for O_3 the gRAMP HY model is used; for CO_2 the gRAMP HY model is used. Note that all models were originally developed using data averaged at 15 minutes. Results are presented for CvMAE and Pearson r, for averaging times ranging from 1 minute up to 1 hour. These results indicate that the performance of the calibration approaches are fairly stable

for data averaged over periods ranging up to 1 hour. At longer averaging periods, the use of time-averaged environmental variables (such as temperature and relative humidity) in the calibration model appears to reduce performance.



5 **Figure S5. An evaluation of the performance of the calibration algorithms as a function of the averaging period; in contrast to the previous figure, this presents the results when averaging is performed after calibration, rather than before. In terms of CvMAE, performance improves as averaging time increases. In terms of Pearson r, results can be worse with longer averaging, due to the**

reduction in the number of points used to evaluate correlation (since there are fewer time periods overall to compare) and to the reduction in the variability (although accuracy is improving as averaging time increases, the variability in the data are also being reduced, and so correlation is decreasing).

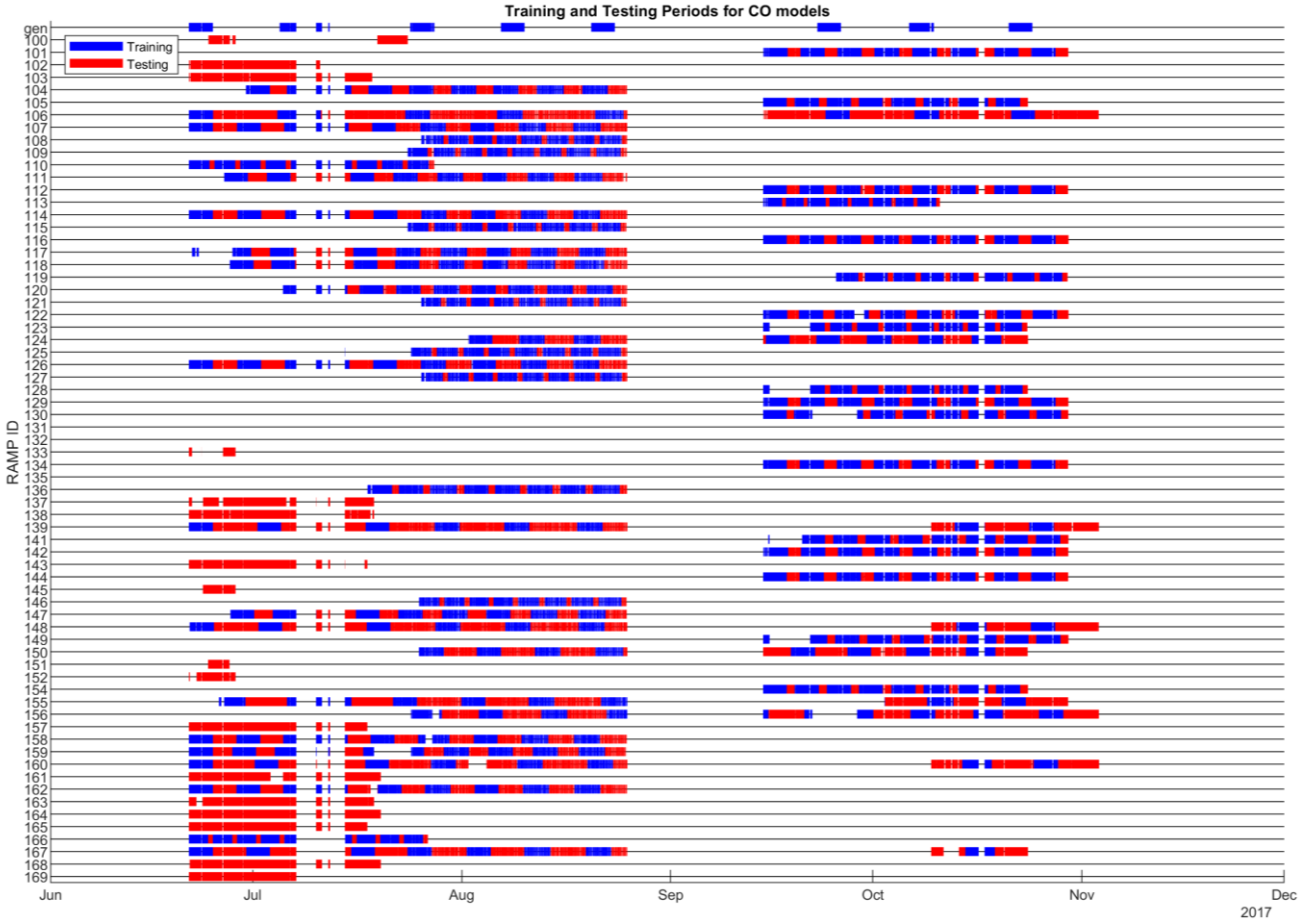


Figure S6. Description of the training and testing periods used for CO models. Blue bars indicate periods used for training data, while red bars denote periods set aside for testing. Time divisions for individual RAMPs (with numeric IDs) are presented corresponding to data used for iRAMP and bRAMP models. Divisions for the "gen" RAMP indicate the training data periods used for gRAMP models, derived from the median of data from the training set of RAMPs collected during these periods; testing data for gRAMP models is drawn from RAMPs which are not part of the training set of RAMPs.

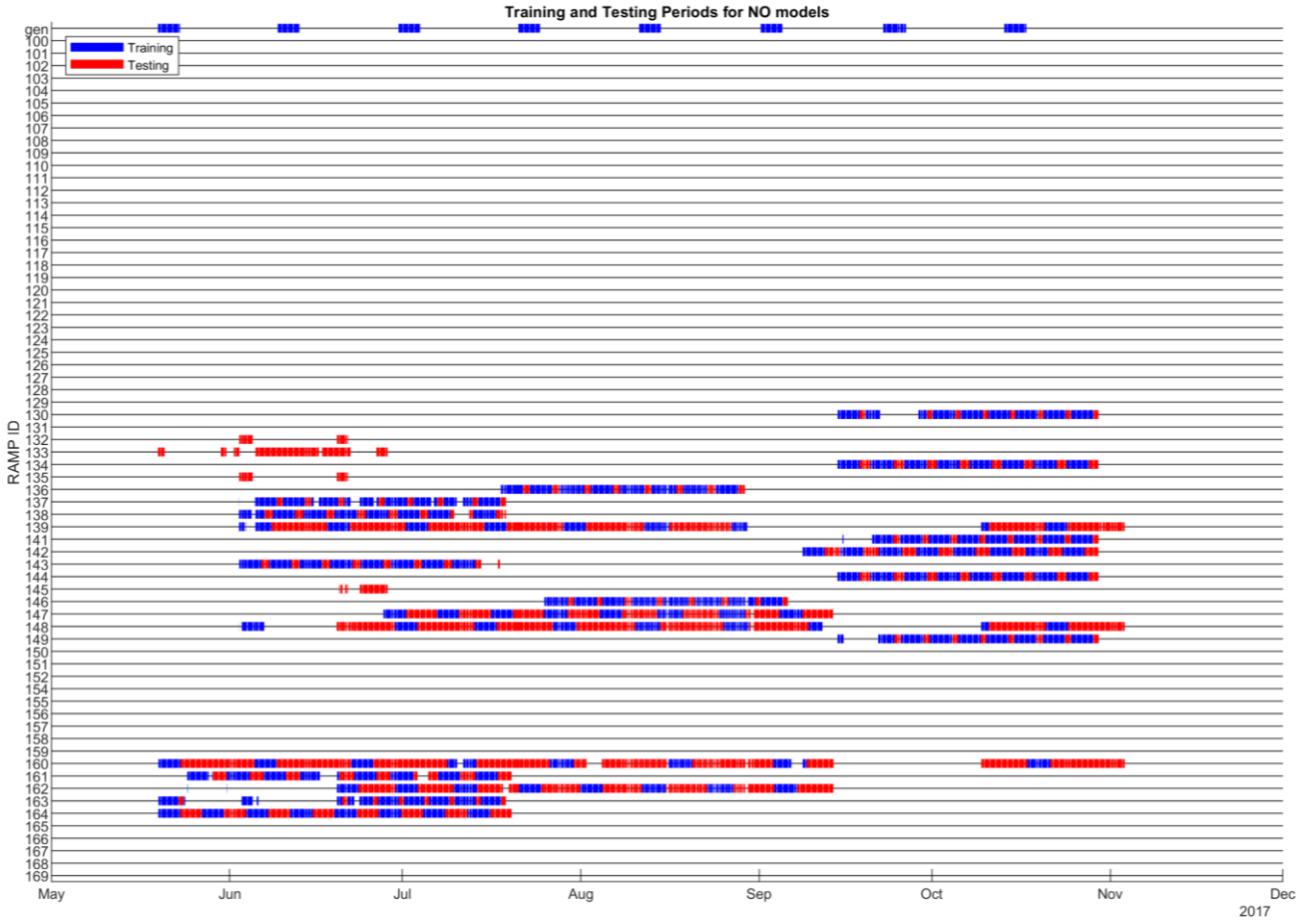


Figure S7. Description of the training and testing periods used for NO models. Blue bars indicate periods used for training data, while red bars denote periods set aside for testing. Time divisions for individual RAMPs (with numeric IDs) are presented corresponding to data used for iRAMP and bRAMP models. Divisions for the "gen" RAMP indicate the training data periods used for gRAMP models, derived from the median of data from the training set of RAMPs collected during these periods; testing data for gRAMP models is drawn from RAMPs which are not part of the training set of RAMPs.

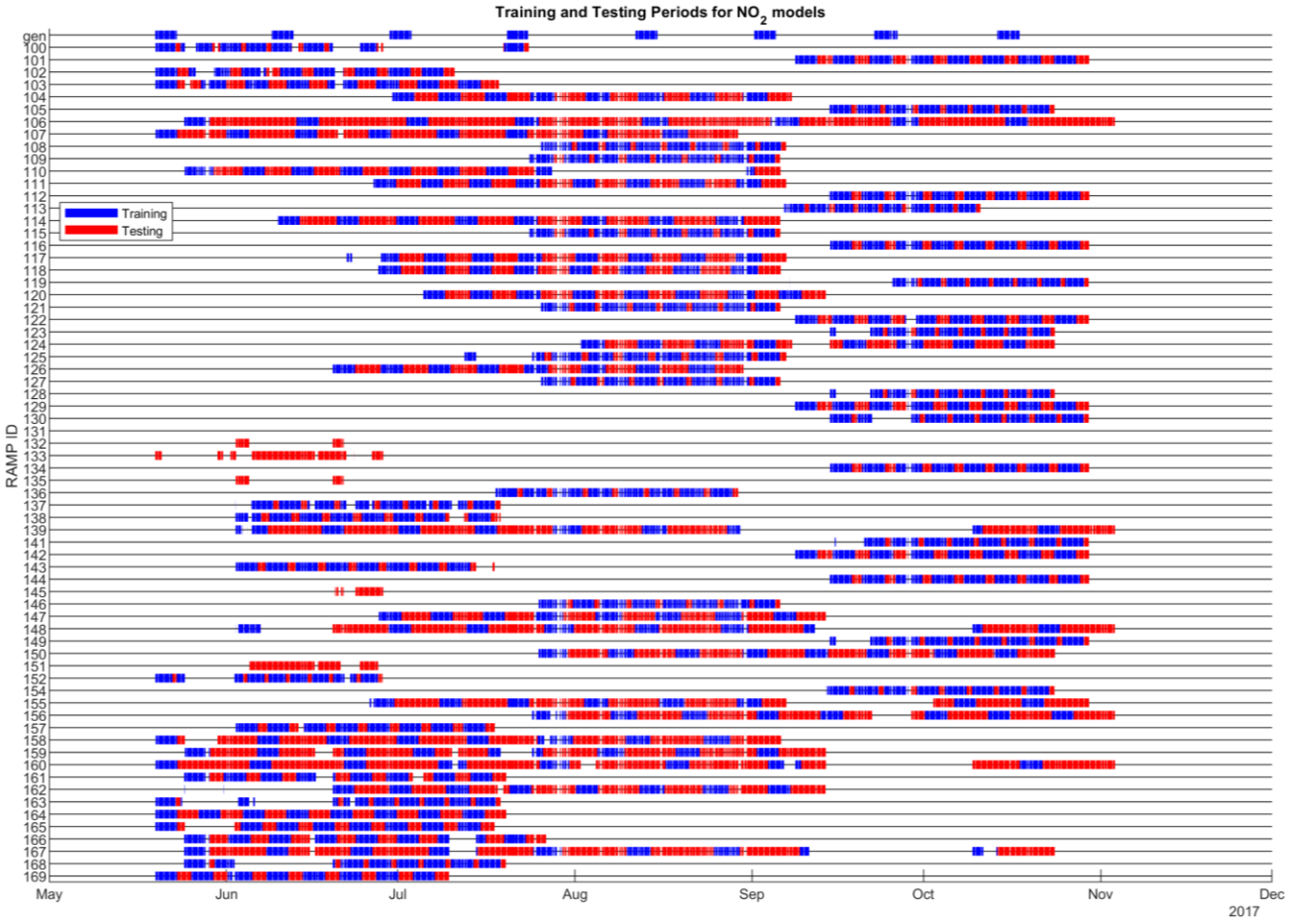


Figure S8. Description of the training and testing periods used for NO_2 models. Blue bars indicate periods used for training data, while red bars denote periods set aside for testing. Time divisions for individual RAMPs (with numeric IDs) are presented corresponding to data used for iRAMP and bRAMP models. Divisions for the "gen" RAMP indicate the training data periods used for gRAMP models, derived from the median of data from the training set of RAMPs collected during these periods; testing data for gRAMP models is drawn from RAMPs which are not part of the training set of RAMPs.

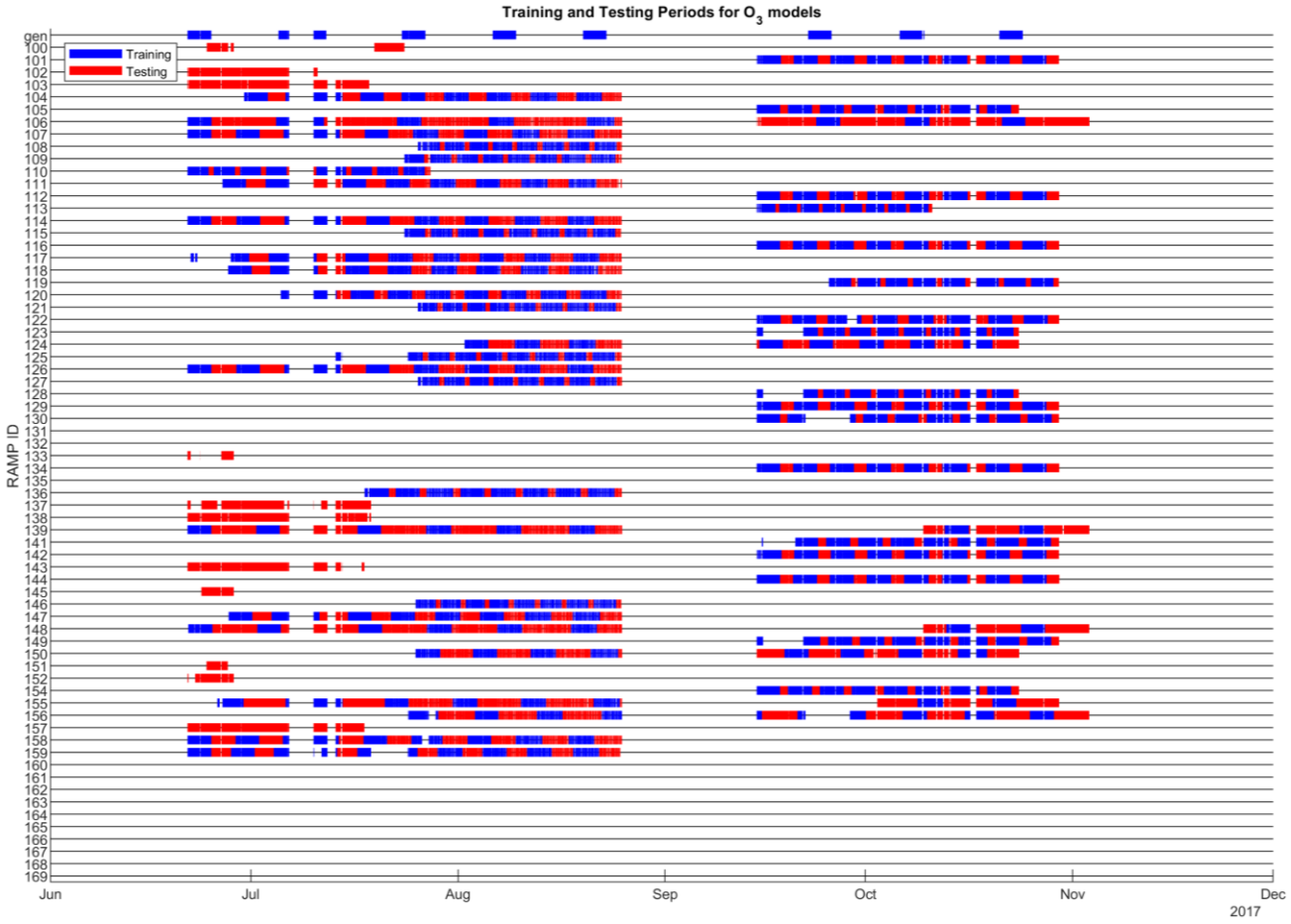


Figure S9. Description of the training and testing periods used for O_3 models. Blue bars indicate periods used for training data, while red bars denote periods set aside for testing. Time divisions for individual RAMPs (with numeric IDs) are presented corresponding to data used for iRAMP and bRAMP models. Divisions for the "gen" RAMP indicate the training data periods used for gRAMP models, derived from the median of data from the training set of RAMPs collected during these periods; testing data for gRAMP models is drawn from RAMPs which are not part of the training set of RAMPs.

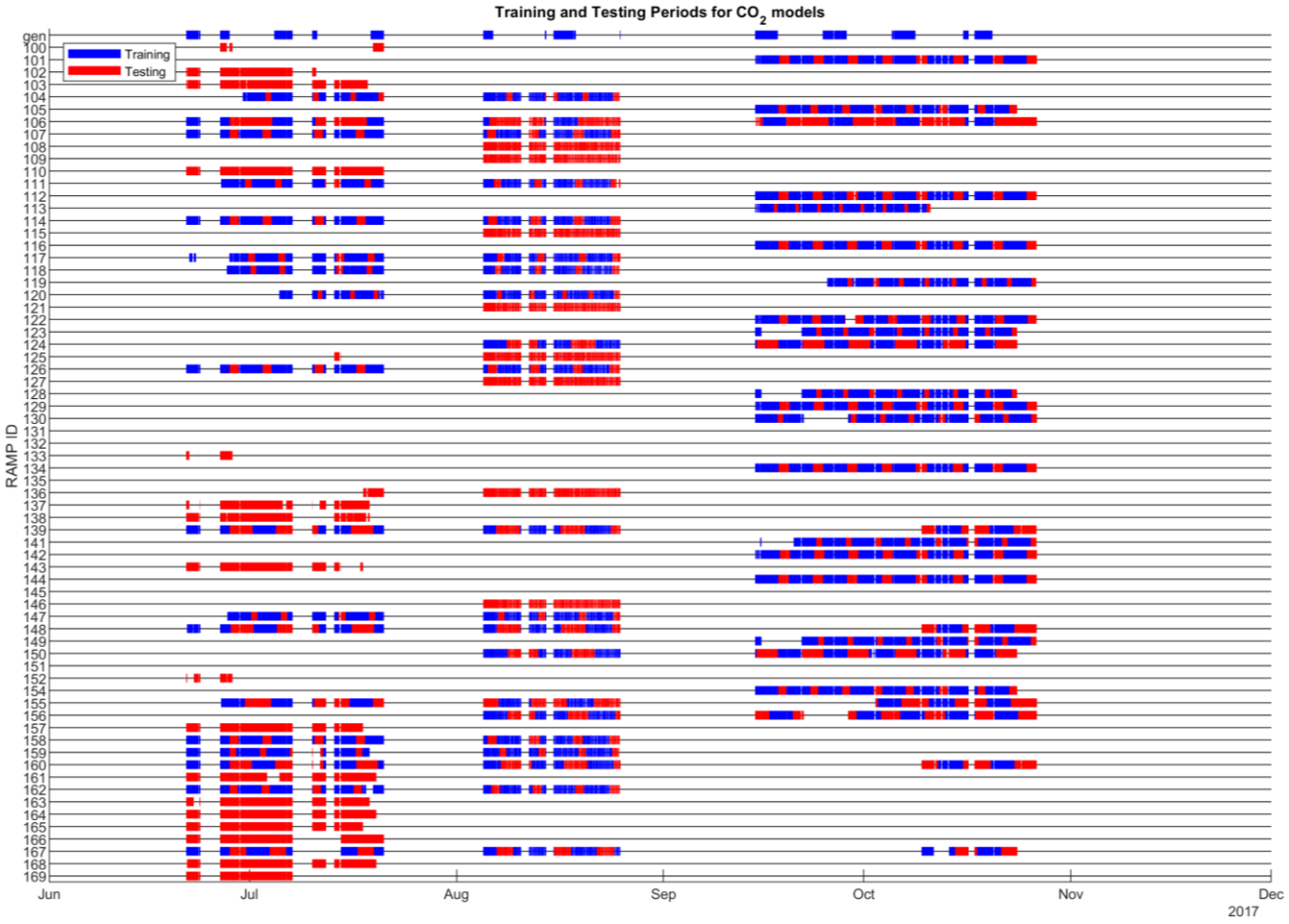


Figure S10. Description of the training and testing periods used for CO_2 models. Blue bars indicate periods used for training data, while red bars denote periods set aside for testing. Time divisions for individual RAMPs (with numeric IDs) are presented corresponding to data used for iRAMP and bRAMP models. Divisions for the "gen" RAMP indicate the training data periods used for gRAMP models, derived from the median of data from the training set of RAMPs collected during these periods; testing data for gRAMP models is drawn from RAMPs which are not part of the training set of RAMPs.

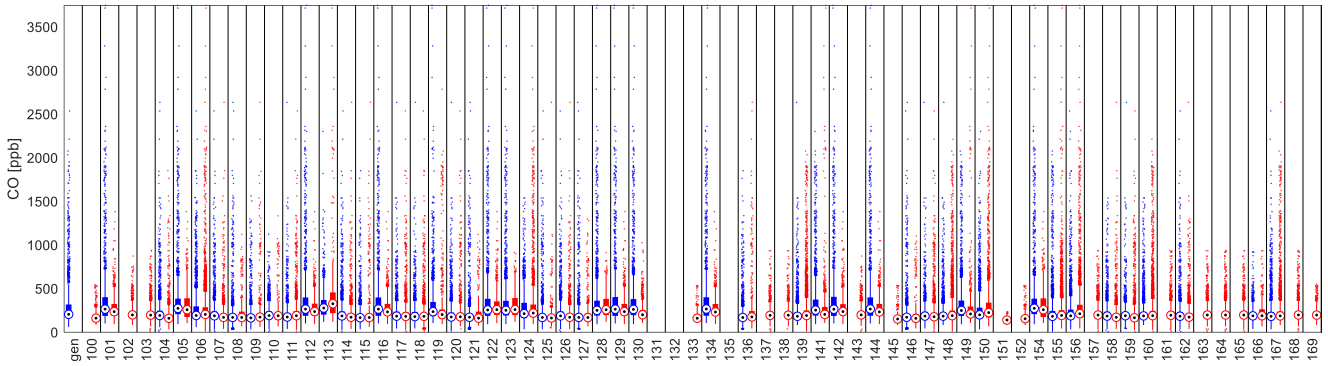


Figure S11. Depiction of the range of CO concentrations experienced during training and testing. Blue boxplots indicate training ranges, while red boxplots denote testing ranges. Dots with circles indicate the midpoint, thicker bars indicate the interquartile range, thinner bars show 1st and 99th percentiles, and colored dots depict outliers. The horizontal axis shows the RAMP ID number (or "gen", which depicts the concentration range used for training gRAMP models).

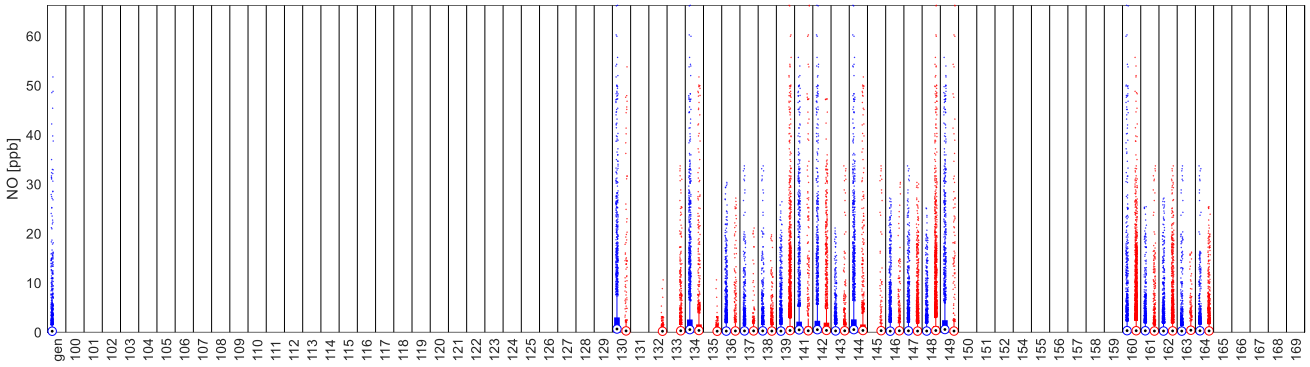


Figure S12. Depiction of the range of NO concentrations experienced during training and testing. Blue boxplots indicate training ranges, while red boxplots denote testing ranges. Dots with circles indicate the midpoint, thicker bars indicate the interquartile range, thinner bars show 1st and 99th percentiles, and colored dots depict outliers. The horizontal axis shows the RAMP ID number (or "gen", which depicts the concentration range used for training gRAMP models).

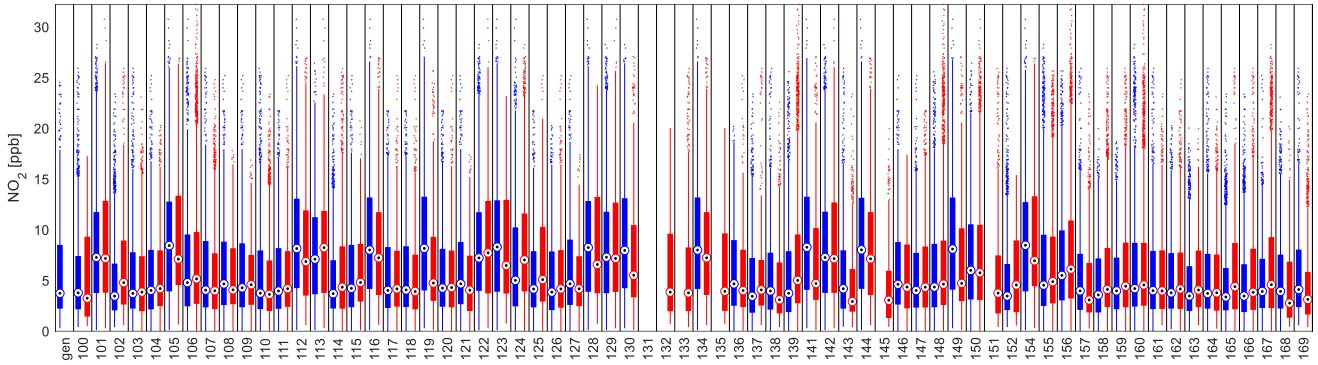


Figure S13. Depiction of the range of NO_2 concentrations experienced during training and testing. Blue boxplots indicate training ranges, while red boxplots denote testing ranges. Dots with circles indicate the midpoint, thicker bars indicate the interquartile range, thinner bars show 1st and 99th percentiles, and colored dots depict outliers. The horizontal axis shows the RAMP ID number (or "gen", which depicts the concentration range used for training gRAMP models).

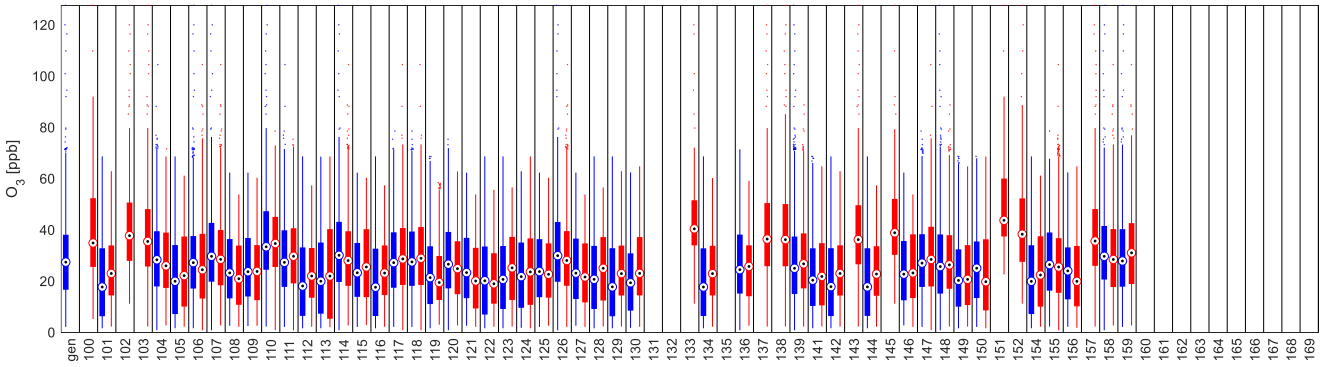


Figure S14. Depiction of the range of O_3 concentrations experienced during training and testing. Blue boxplots indicate training ranges, while red boxplots denote testing ranges. Dots with circles indicate the midpoint, thicker bars indicate the interquartile range, thinner bars show 1st and 99th percentiles, and colored dots depict outliers. The horizontal axis shows the RAMP ID number (or "gen", which depicts the concentration range used for training gRAMP models).

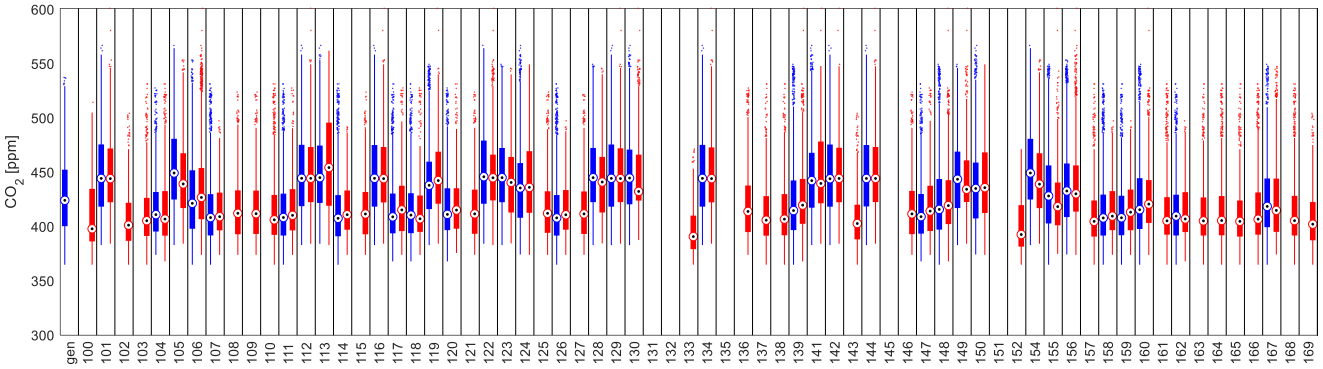


Figure S15. Depiction of the range of CO_2 concentrations experienced during training and testing. Blue boxplots indicate training ranges, while red boxplots denote testing ranges. Dots with circles indicate the midpoint, thicker bars indicate the interquartile range, thinner bars show 1st and 99th percentiles, and colored dots depict outliers. The horizontal axis shows the RAMP ID number (or "gen", which depicts the concentration range used for training gRAMP models).

Table S1: Results corresponding to Table 4 for CO_2.

Gas	Training Period				Testing Period			
	Duration	Concentration			Duration	Concentration		
	[days]	[ppm]			[days]	[ppm]		
	Range	*lower range*	*average range*	*upper range*	*Range*	*lower range*	*average range*	*upper range*
CO_2	21 - 28	365-384	413-454	528-567	2 - 50	365-388	399-458	471-601

Table S2: Results corresponding to Table 5 for CO_2.

Gas	Model		Testing Performance							
	Type	#	Slope		r^2		MAE [ppm]		Bias [ppm]	
			Avg.	*SD*	*Avg.*	*SD*	*Avg.*	*SD*	*Avg.*	*SD*
CO_2	LR	38	0.74	0.23	0.21	0.09	24	5	0.6	6.1
	LQR	38	0.62	0.22	0.23	0.12	25	5	1.0	8.1
	CQR	38	0.74	0.20	0.47	0.16	18	3	-1.0	4.8
	CL	38	0.76	0.13	0.43	0.13	20	3	1.4	4.4
	NN	38	0.47	1.53	0.28	0.25	23	7	-2.1	6.5
	HY	38	0.79	0.26	0.53	0.15	19	4	3.2	6.0

5

Table S3: Results corresponding to Table 5 for bRAMP models.

Gas	Model Type	#	Testing Performance							
			Slope		r^2		MAE [ppb]		Bias [ppb]	
			Avg.	*SD*	*Avg.*	*SD*	*Avg.*	*SD*	*Avg.*	*SD*
CO	LR	1	0.68	0.28	0.66	0.23	80	38	54	138
	LQR	1	0.75	0.25	0.71	0.21	59	28	56	119
	CQR	4	0.62	0.29	0.58	0.29	143	337	156	568
	CL	4	0.83	0.31	0.64	0.23	66	29	63	78
	NN	4	1.09	0.58	0.46	0.20	81	31	42	90
	HY	4	0.74	0.26	0.69	0.24	63	38	98	90
NO	LR	1	0.60	0.67	0.09	0.09	8.3	20.6	8.9	19.6
	LQR	1	0.69	0.53	0.14	0.14	3.8	7.2	1.9	6.5
	CQR	2	0.55	0.44	0.19	0.17	8.0	23.2	5.4	19.2
	CL	2	0.38	0.35	0.09	0.13	3.1	1.7	1.1	1.4
	NN	2	1.00	0.57	0.21	0.16	2.2	1.1	0.2	2.0
	HY	2	0.58	0.31	0.26	0.15	2.5	1.8	1.4	2.2
NO_2	LR	1	0.75	0.42	0.14	0.10	4.7	9.6	-0.2	7.0
	LQR	1	0.64	0.27	0.18	0.12	3.5	0.9	-1.4	1.8
	CQR	4	0.44	0.35	0.25	0.19	4.2	2.1	2.1	4.4
	CL	4	0.58	0.29	0.21	0.14	3.5	0.6	2.1	2.9
	NN	4	0.86	0.37	0.33	0.18	3.1	0.7	0.5	2.4
	HY	4	0.78	0.30	0.32	0.19	3.2	1.2	1.4	2.5
O_3	LR	1	0.76	0.24	0.70	0.23	10.5	24.8	4.7	21.5
	LQR	1	0.85	0.22	0.72	0.24	6.1	3.1	-3.1	8.9
	CQR	2	0.75	0.27	0.65	0.27	9.8	17.1	2.9	15.5
	CL	2	0.91	0.30	0.50	0.18	8.7	2.1	-1.7	8.2
	NN	2	0.90	0.53	0.65	0.26	7.0	3.9	-2.6	7.4
	HY	2	1.06	0.21	0.75	0.13	5.8	1.8	0.3	6.5
CO_2	LR	1	0.65	0.40	0.18	0.14	23	5	13	19
	LQR	1	0.41	0.32	0.16	0.16	27	9	15	25
	CQR	4	0.43	0.27	0.31	0.21	58	171	17	148
	CL	4	0.70	0.18	0.32	0.15	21	4	9	17
	NN	4	0.79	0.53	0.29	0.18	31	32	-9	46
	HY	4	0.95	0.27	0.47	0.16	18	3	12	17

Table S4: Results corresponding to Table 5 for gRAMP models.

Gas	Model Type	#	Testing Performance							
			Slope		r^2		MAE [ppb]		Bias [ppb]	
			Avg.	*SD*	*Avg.*	*SD*	*Avg.*	*SD*	*Avg.*	*SD*
CO	LR	1	1.03	0.24	0.80	0.11	68	12	26	109
	LQR	1	0.90	0.08	0.85	0.09	56	8	6	93
	CQR	1	0.69	0.18	0.66	0.20	106	41	60	95
	CL	1	1.02	0.19	0.72	0.10	80	12	21	57
	NN	1	0.67	0.22	0.51	0.17	134	59	88	111
	HY	1	0.75	0.11	0.61	0.11	110	41	75	54
NO	LR	1	1.51	0.92	0.07	0.03	3.8	1.8	0.1	1.8
	LQR	1	0.67	0.34	0.08	0.04	7.1	9.6	3.5	8.5
	CQR	1	0.15	0.09	0.06	0.04	5.9	6.4	2.7	6.3
	CL	1	0.43	0.13	0.13	0.03	3.1	1.0	-0.6	0.5
	NN	1	0.49	0.23	0.22	0.12	4.1	3.0	2.3	4.9
	HY	1	0.40	0.22	0.17	0.08	13.2	22.7	9.9	20.1
NO_2	LR	1	1.07	0.40	0.14	0.05	3.9	0.7	-1.2	1.5
	LQR	1	0.86	0.31	0.18	0.07	3.8	0.7	-1.1	1.9
	CQR	1	0.67	0.21	0.30	0.09	3.5	0.4	-1.0	2.7
	CL	1	0.67	0.19	0.26	0.12	3.6	0.5	-0.1	2.4
	NN	1	0.88	0.21	0.34	0.15	3.3	0.4	-0.3	2.9
	HY	1	0.76	0.27	0.30	0.17	3.4	0.5	0.2	2.6
O_3	LR	1	0.89	0.27	0.72	0.22	6.4	2.8	2.2	4.9
	LQR	1	0.77	0.29	0.66	0.24	7.5	4.1	4.4	6.8
	CQR	1	0.76	0.28	0.67	0.25	7.2	4.2	3.5	6.5
	CL	1	0.91	0.13	0.45	0.19	8.9	1.8	-0.8	5.7
	NN	1	0.92	0.18	0.73	0.18	5.9	2.4	1.7	5.1
	HY	1	1.00	0.12	0.73	0.12	5.9	1.5	0.9	2.9
CO_2	LR	1	0.63	0.14	0.21	0.06	26	4	-2	12
	LQR	1	0.55	0.14	0.20	0.06	27	4	-6	12
	CQR	1	0.39	0.11	0.22	0.11	27	5	-9	15
	CL	1	0.71	0.15	0.37	0.13	23	2	3	16
	NN	1	0.30	0.25	0.15	0.12	38	25	-12	19
	HY	1	0.80	0.16	0.43	0.11	21	2	4	15