Supplement of Atmos. Meas. Tech., 11, 291–313, 2018
https://doi.org/10.5194/amt-11-291-2018-supplement

Atmospheric Measurement Techniques
Open Access
EGU

Supplement of

A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring

Naomi Zimmerman et al.

Correspondence to: R. Subramanian (subu@cmu.edu)

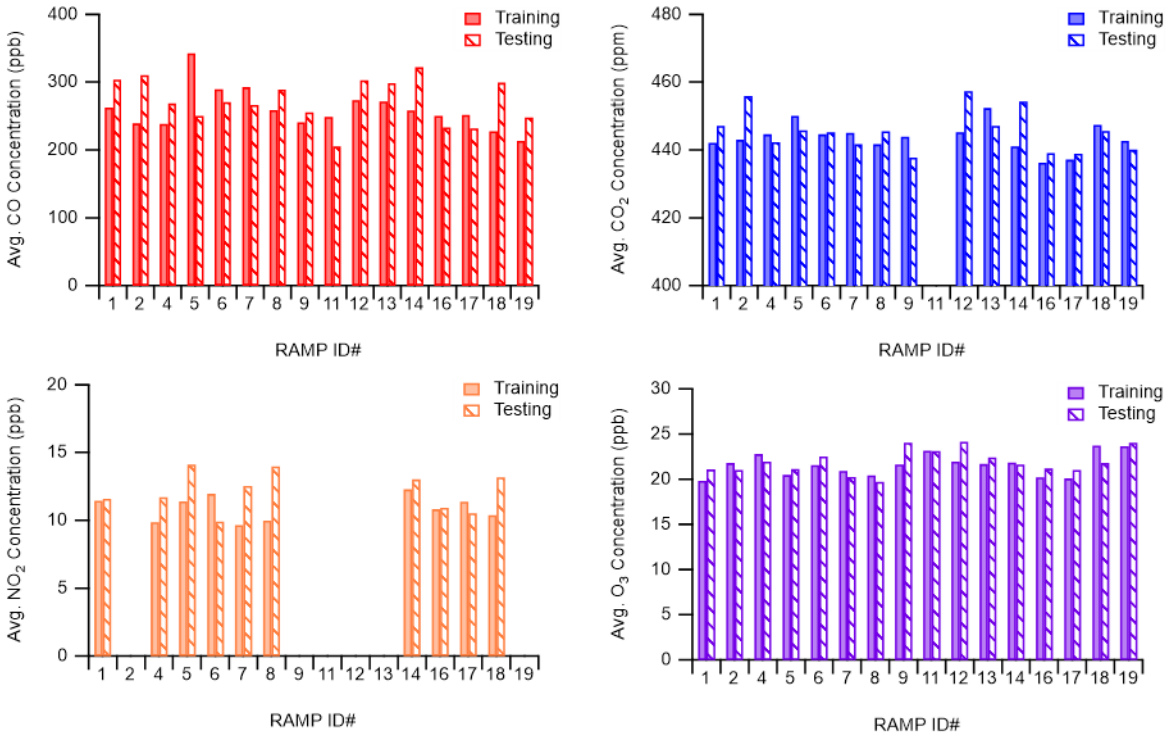

Figure S1. Average reference monitor concentrations during training and testing windows for each RAMP

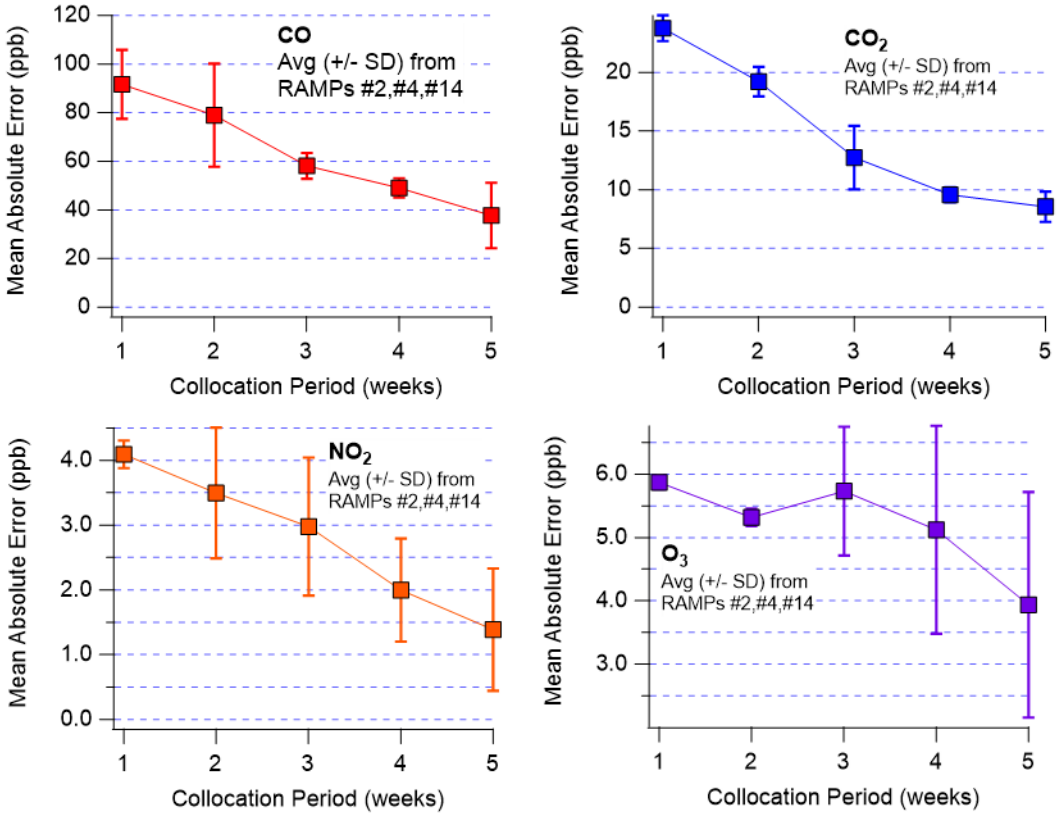

Figure S2: Choosing collocation length for training. Optimization assessed on a subset of three RAMPs (RAMP #2, #4, and #14). Ultimately a 4-week collocation period was chosen as being the best across all four species.

Table S1: Change in model performance if one consecutive 4-week colocation at the beginning of the study is conducted vs spacing out training data throughout the study in 8 half week increments.

Impact of using one consecutive four week training window vs. distributed half week colocations				
Metric	**CO**	**CO_2**	**NO_2**	**O_3**
RMSE	+11.9 ppb	+3.2 ppm	+0.4 ppb	+1.6 ppb
MAE	+12.1 ppb	+1.8 ppm	+0.4 ppb	+1.6 ppb
Pearson r	-0.01	-0.03	-0.08	-0.05

Table S2. Metrics used for comparing sensor data. M indicates a value measured by one of the sensors participating in the experiment and O indicates the observations from the reference measurements.

Statistic	Abbrev.	Formula	Characteristics
Mean Bias Error	MBE	$\mathrm{MBE} = \overline{\mathrm{M}} - \overline{\mathrm{O}}$	Estimation of the magnitude of differences (bias) between sensors estimation and reference values averaged over the whole sampling period
Mean Absolute Error	MAE	$\mathrm{MAE} = \frac{1}{\mathrm{n}}\sum_{i=1}^{n}\lvert M_i - O_i\rvert$	• Indicates the average of the magnitude of the errors. • Sensitive to outliers.
Pearson Correlation Coefficient	r	$\mathrm{r} = \frac{\sum_{i=1}^{n}(M_i - \bar{M})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(M_i - \bar{M})^2(O_i - \bar{O})^2}}$	Measures the strength and the direction of a linear relationship between two variables.
Root Mean Square Error	RMSE	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(M_i - O_i)^2}$	• Magnitude of the error and retains the variable's unit • Sensitive to extreme values and to outliers
Centred Root Mean Square Error	CRMSE	$CRMSE = \sqrt{RMSE^2 - MBE^2}$	• RMSE corrected for bias • Measure of random error

Plots of Goodness of Fit (Results from Training, Figures S3-S6)

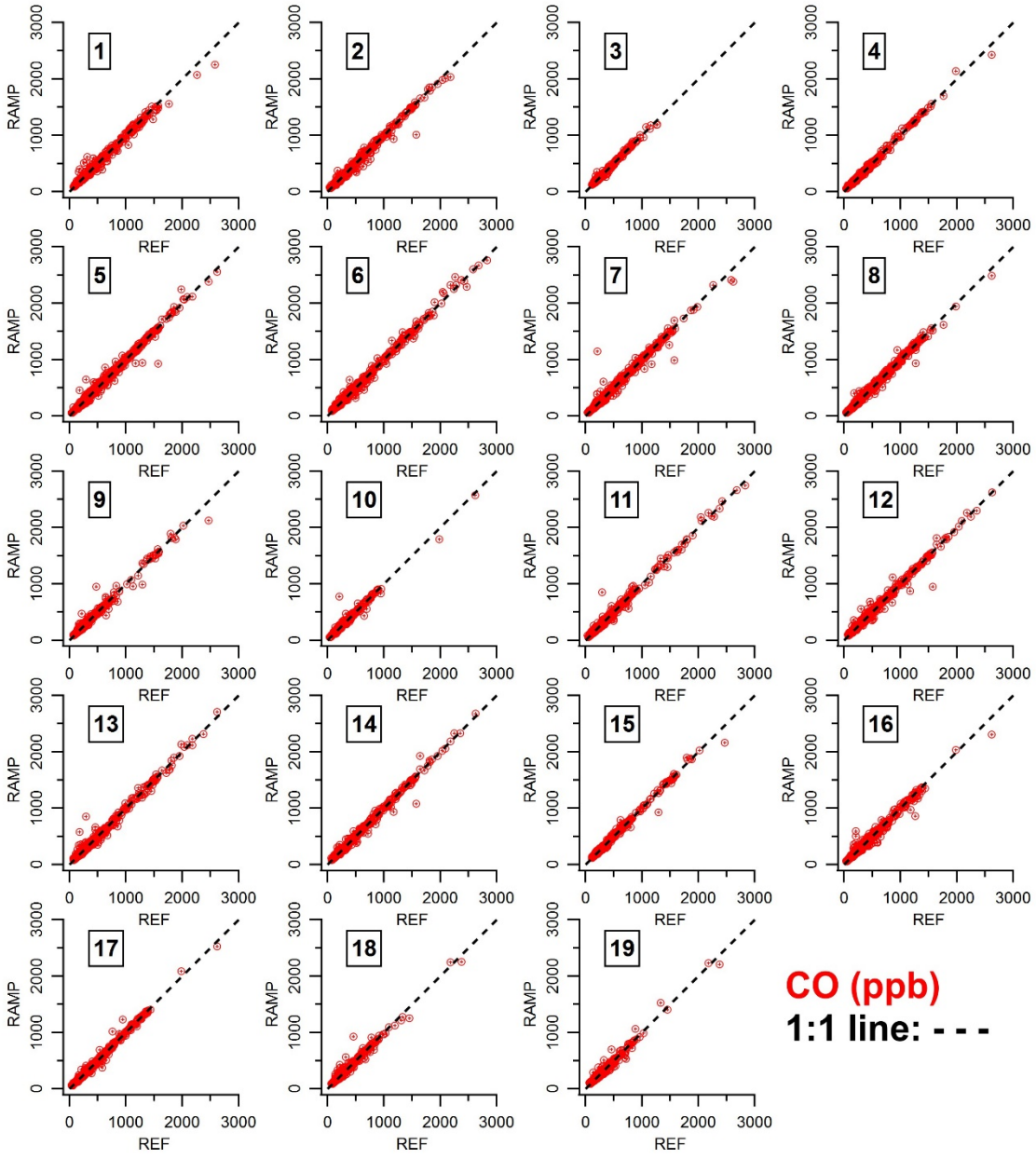

Figure S3: Goodness of fit on training data across all 19 RAMPs for CO using random forests.

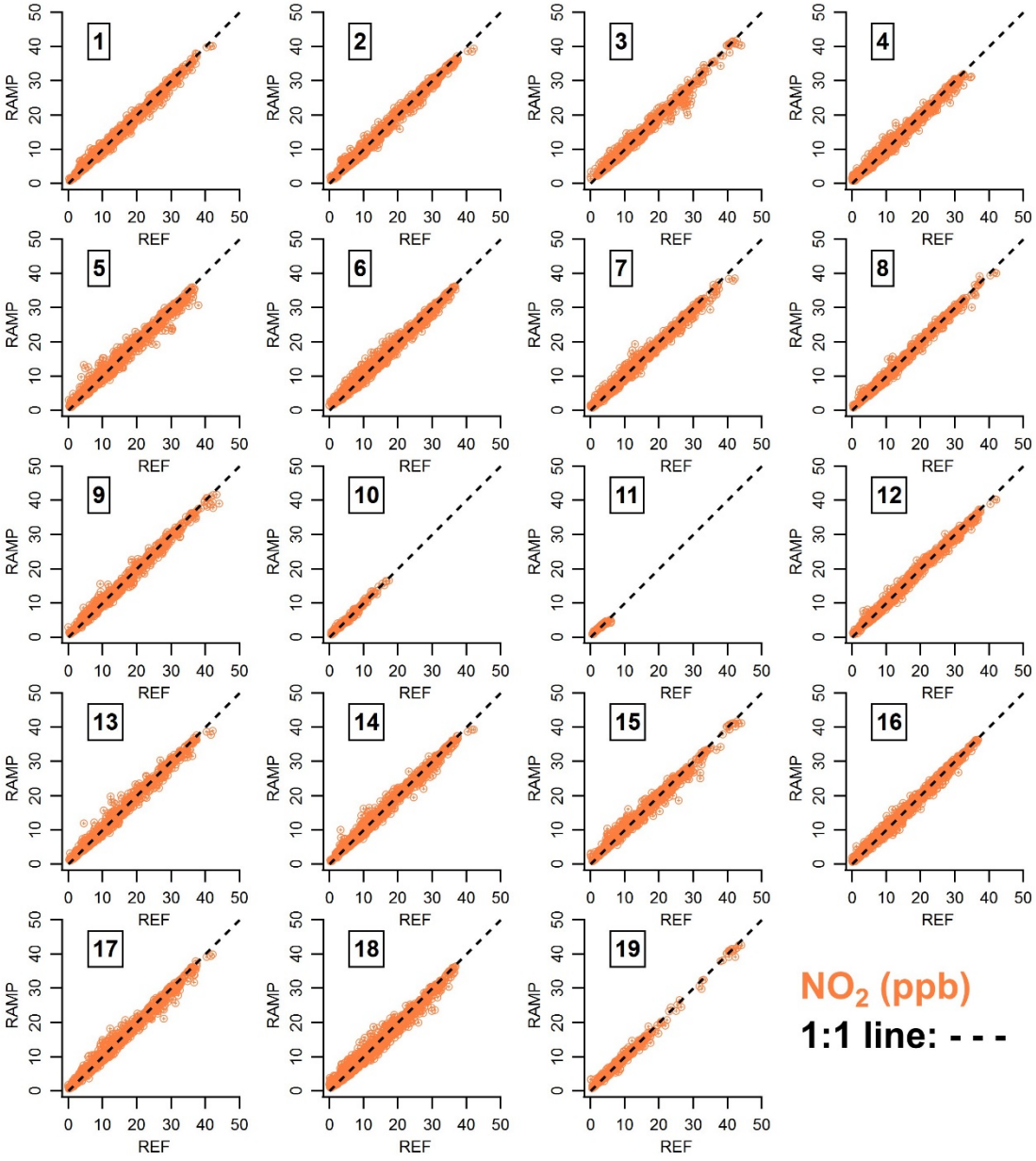

Figure S4: Goodness of fit on training data across all 16 RAMPs for NO_2 using random forests.

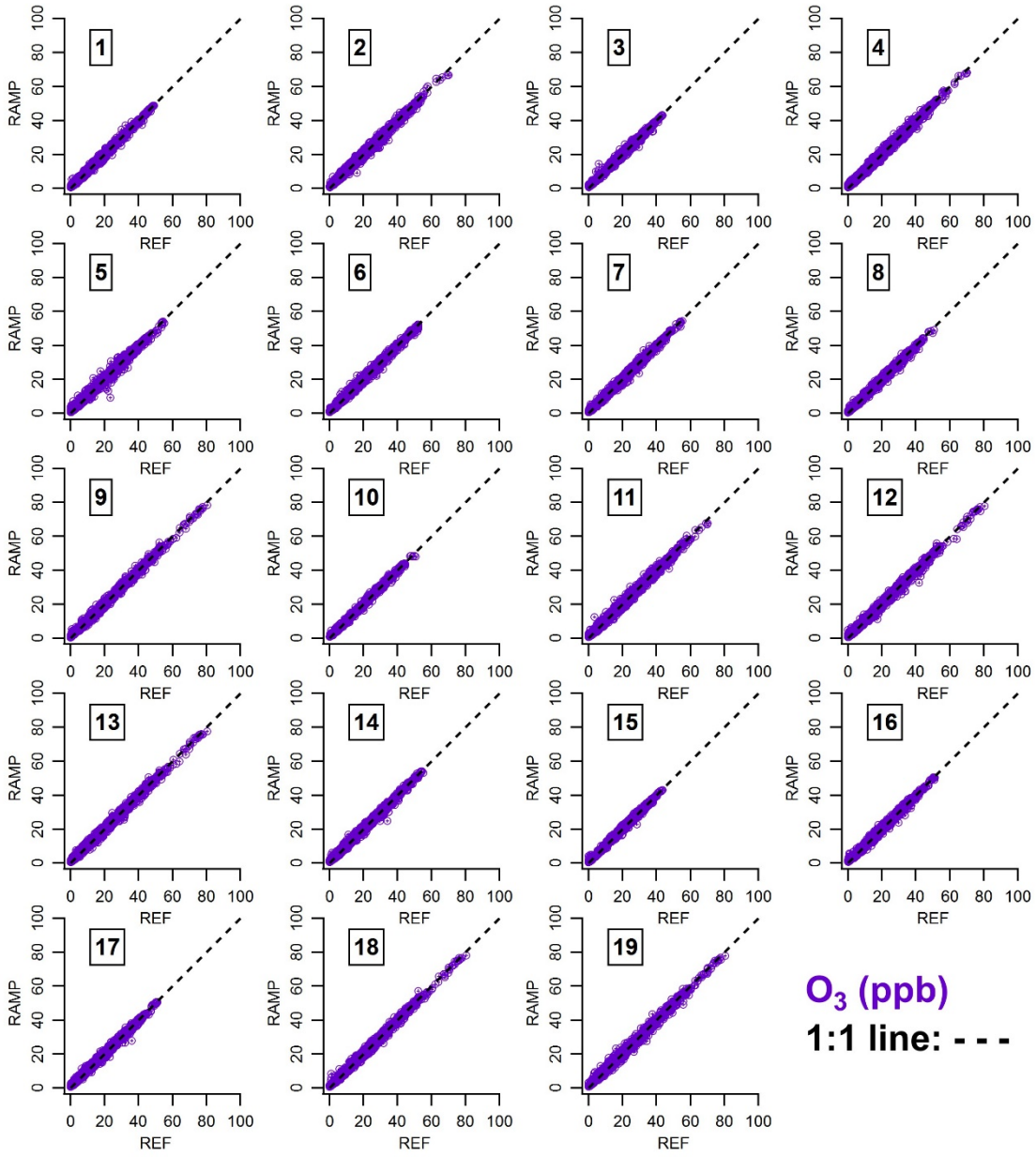

Figure S5: Goodness of fit on training data across all 19 RAMPs for O_3 using random forests.

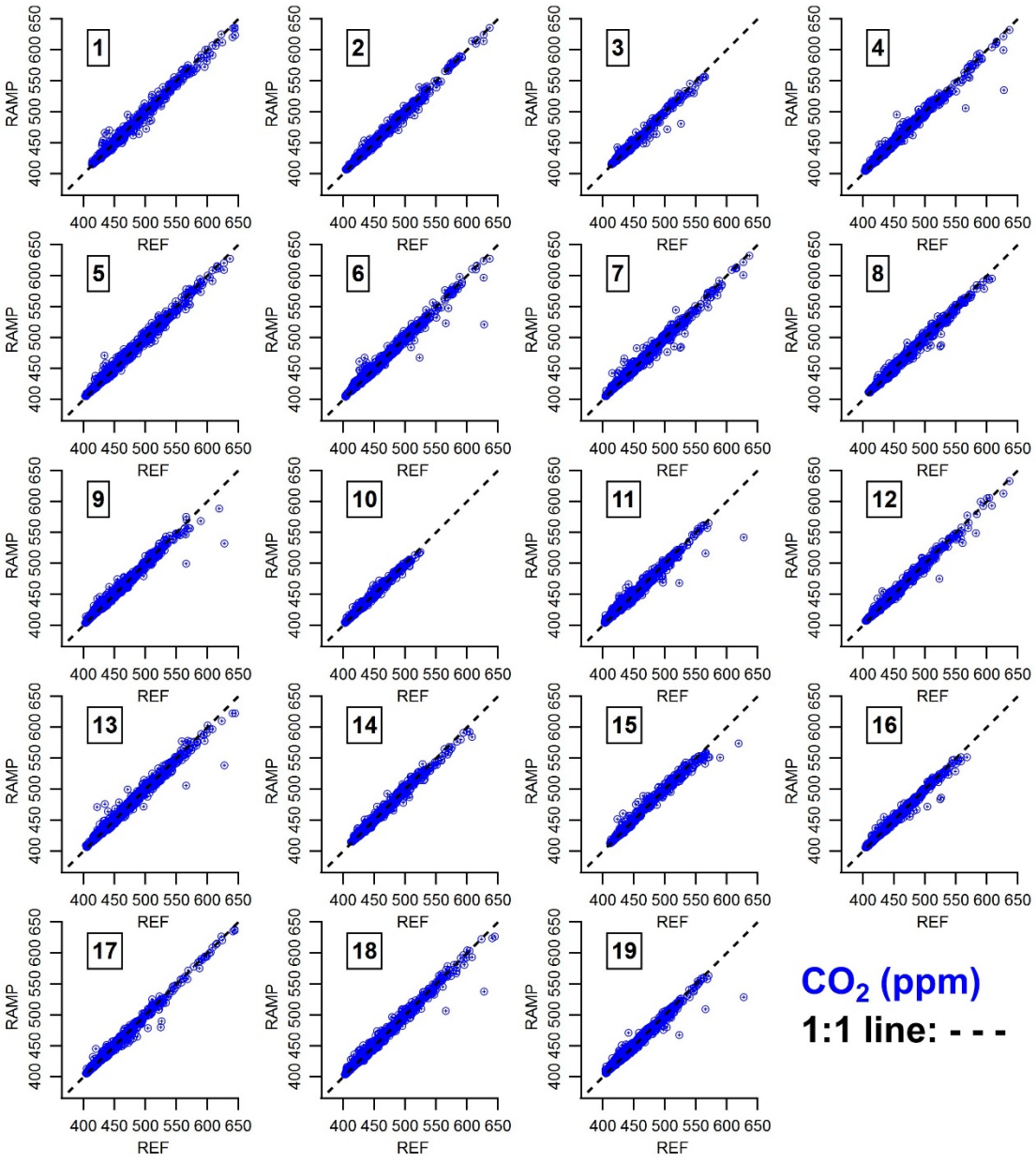

Figure S6: Goodness of fit on training data across all 19 RAMPs for CO_2 using random forests.

Plots of Random Forest Performance by RAMP (Results from Testing, Figures S7-S10)

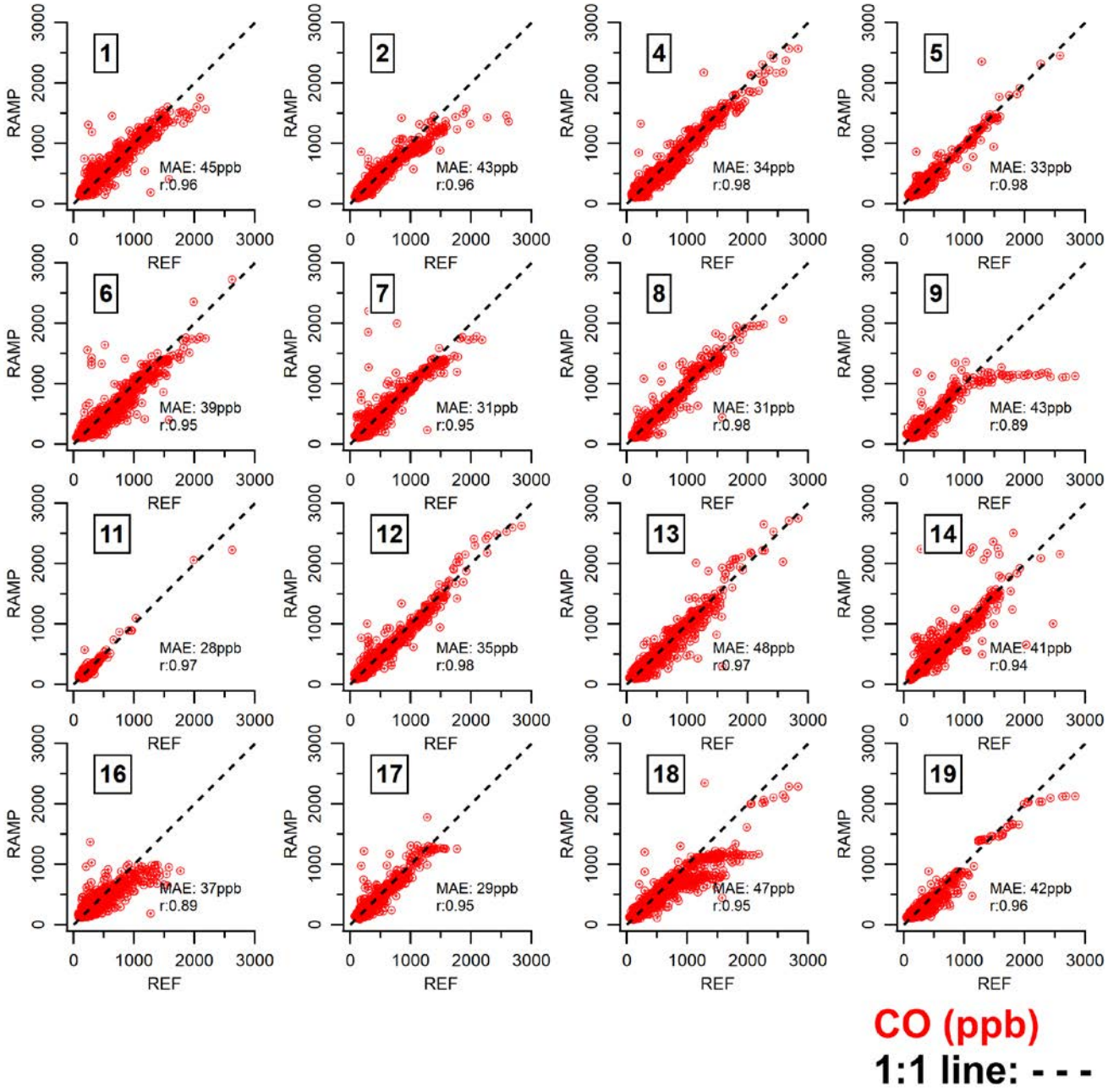

Figure S7: Model performance on testing data across 16 RAMPs for CO using random forests.

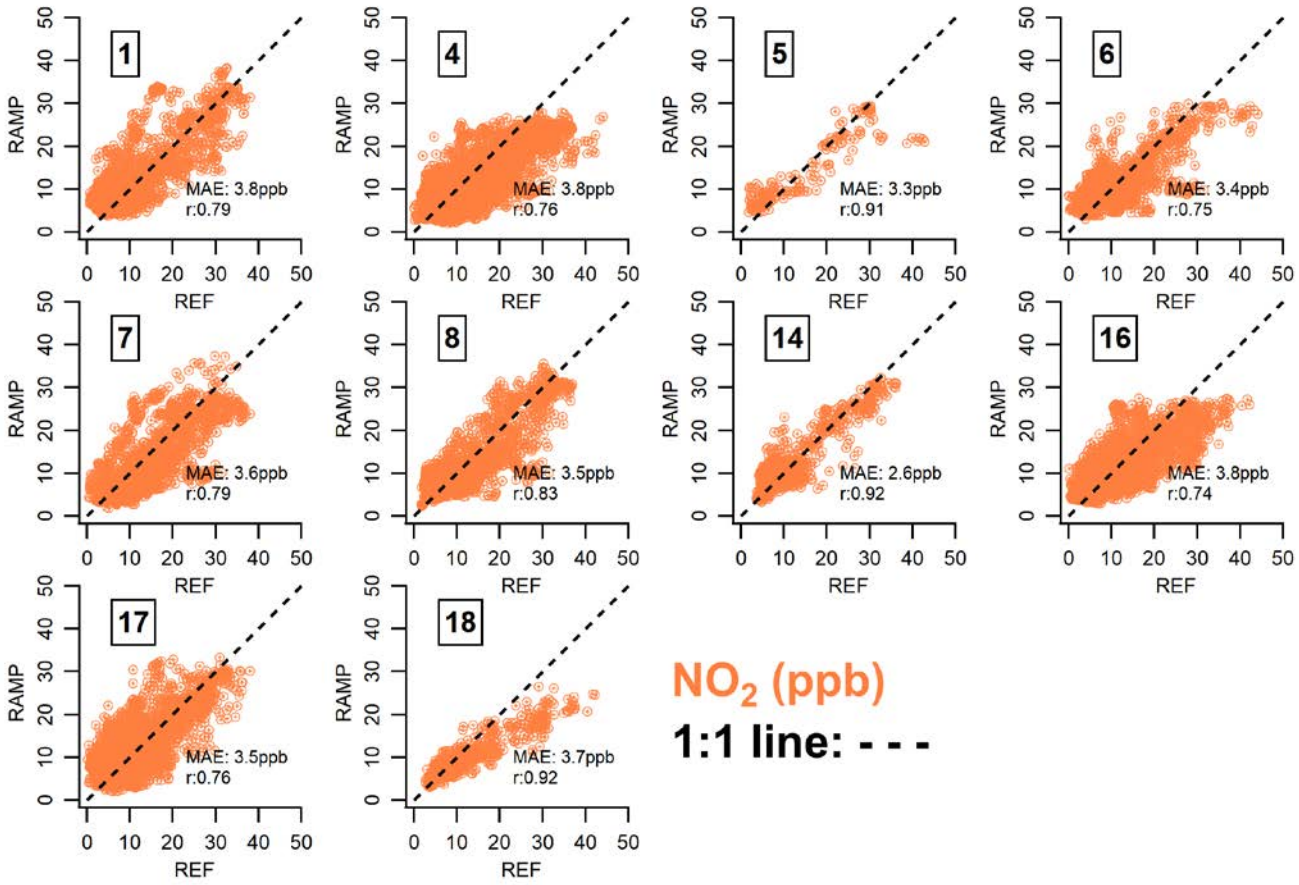

Figure S8: Model performance on testing data across 10 RAMPs for NO_2 using random forests.

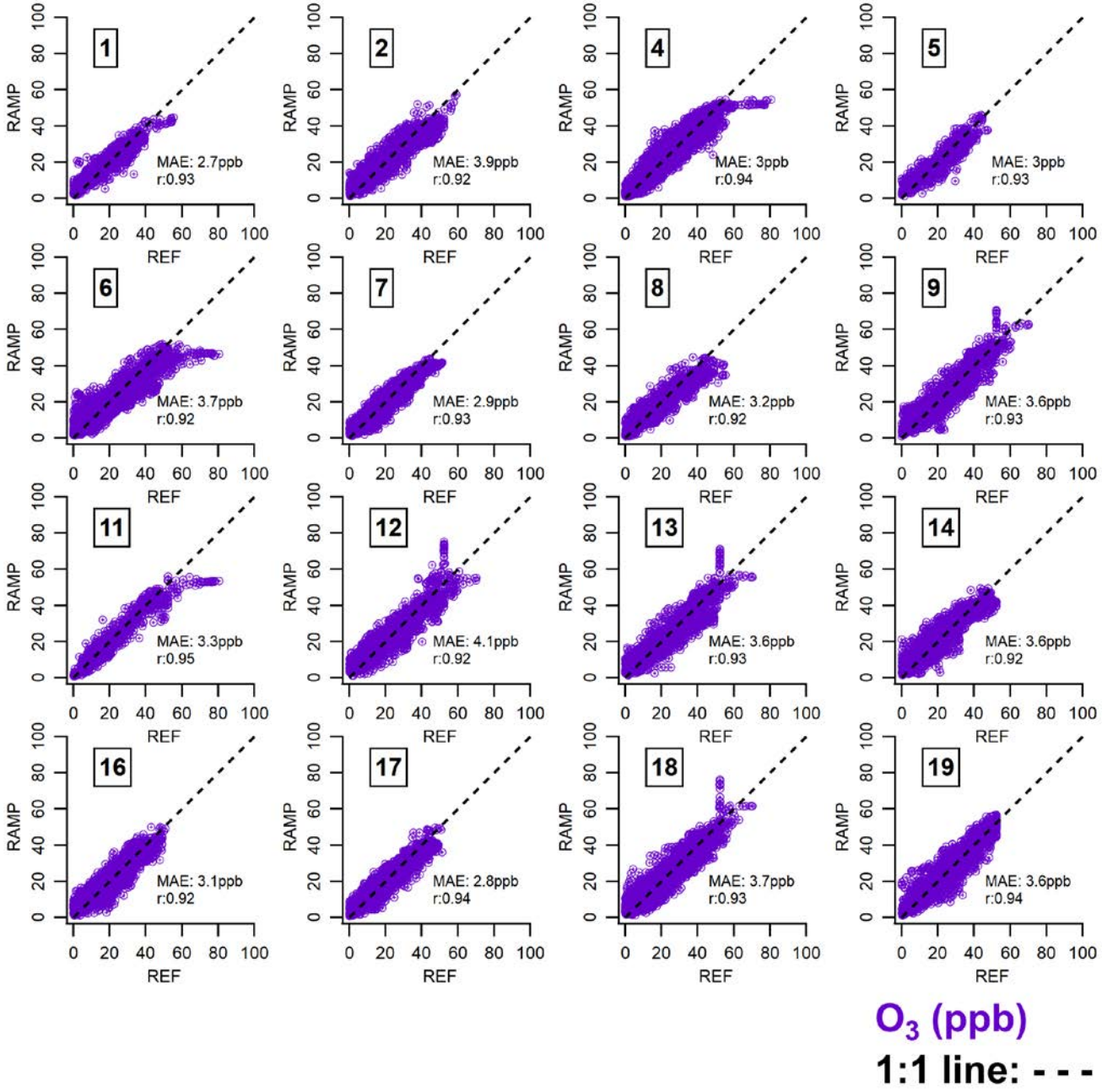

Figure S9: Model performance on testing data across 16 RAMPs for O_3 using random forests.

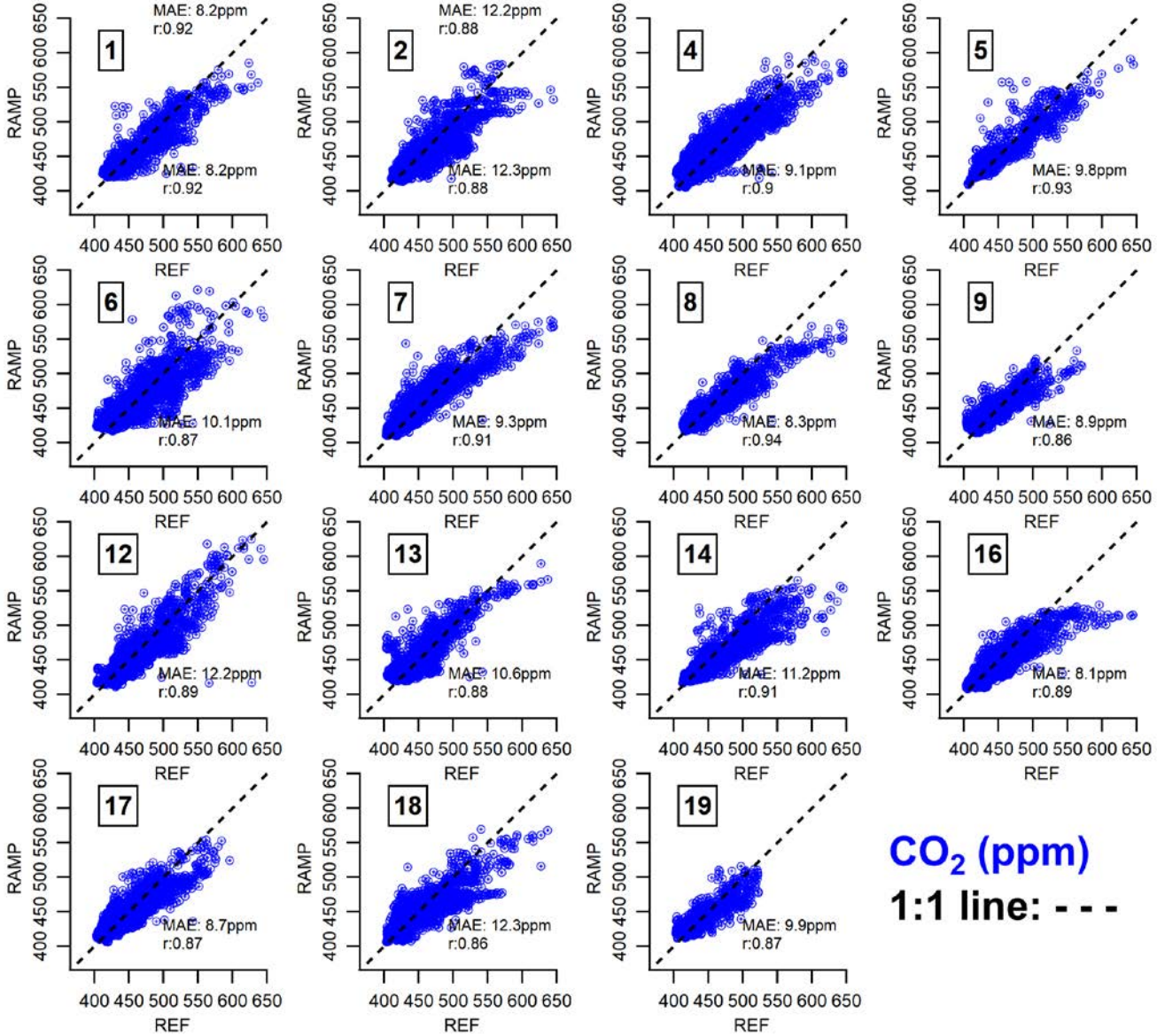

Figure S10: Model performance on testing data across 15 RAMPs for CO_2 using random forests.

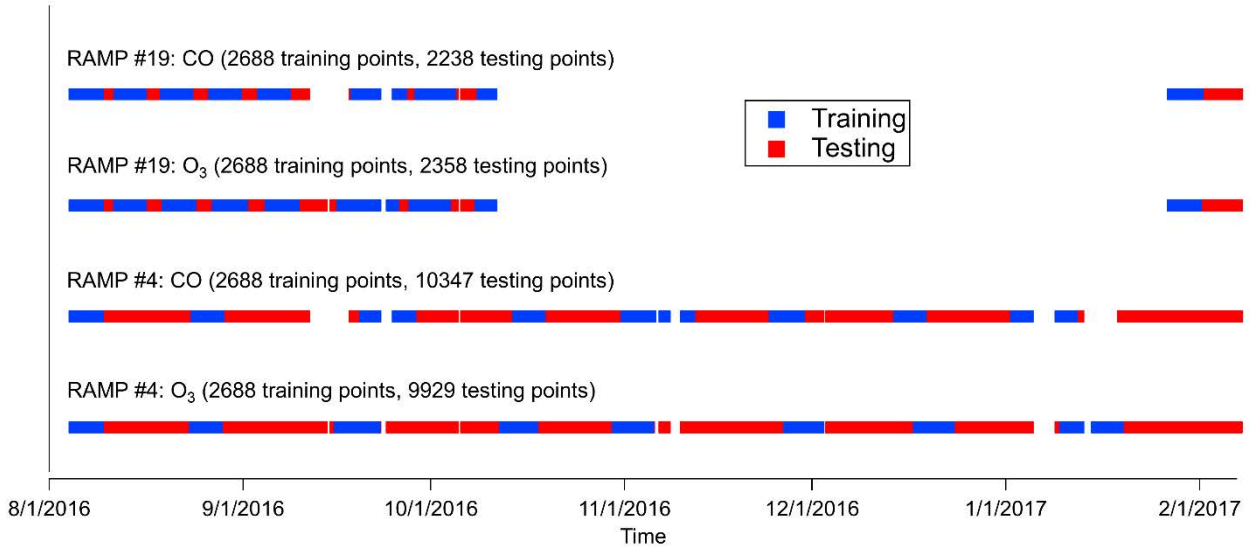

Figure S11: Examples of training and testing periods for two RAMPs. The training window (4 weeks) is the same for each sensor; remaining data are used for model testing. RAMP #4 had the longest testing window (15 weeks), while RAMP #19 had one of the shorter testing windows (under 4 weeks.) Only two sensors from each RAMP are shown for clarity.

US EPA Air Sensor Guidebook Precision and Bias Estimators:

Precision:

The precision estimator (CV) is the upper bound of the 90% confidence interval on the coefficient of variation.

$$\mathbf{CV} = \sqrt{\frac{\mathbf{n} \cdot \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{d_i^2} - (\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{d_i})^2}{\mathbf{n(n-1)}}} \cdot \sqrt{\frac{\mathbf{n-1}}{\boldsymbol{\chi}^2_{\mathbf{(0.1,n-1)}}}}$$

Where n is the number of data points, $\chi^2_{(0.1,n-1)}$ is the 10th percentile of a chi-squared distribution with n-1 degrees of freedom, and d_i is equal to:

$$\mathbf{d_i} = \frac{\mathbf{RAMP - reference}}{\mathbf{reference}} \cdot \mathbf{100\%}$$

Bias:

The bias estimator is the upper bound of the 95% confidence interval on the mean absolute value of the percent difference between the RAMPs and the reference monitor.

$$\mathbf{|Bias|} = \mathbf{AB} + \mathbf{t_{0.95,n-1}} \cdot \frac{\mathbf{AS}}{\sqrt{\mathbf{n}}}$$

Where $t_{0.95,n-1}$ is the 95th quartile of a t-distibution with n-1 degrees of freedom and AB is the mean of the absolute values of the d_i's.

$$\mathbf{AB} = \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{|d_i|}$$

And AS is the standard deviation of the absolute value of the d_i's:

$$\mathbf{AS} = \sqrt{\frac{\mathbf{n} \cdot \sum_{\mathbf{i=1}}^{\mathbf{n}} |\boldsymbol{d_i}|^2 - (\sum_{\mathbf{i=1}}^{\mathbf{n}} |\boldsymbol{d_i}|)^2}{\mathbf{n(n-1)}}}$$