Supplementary Information for Spectroscopic Quantification of Target Species in a Complex Mixture using Blind Source Separation and Partial Least-Squares Regression: A Case Study on Hanford Waste

Stefani Kocevska, Giovanni Maria Maggioni, Ronald W. Rousseau, and Martha A. Grover*

School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

E-mail: martha.grover@chbe.gatech.edu

June 28, 2021

1 Training Data Set

Concentrations associated with the training data set, $\mathbf{X}_{tr}$, are given in Table S1.

2 Simulated Data

The spectra are simulated by multiplying each unimolar reference spectrum by the desired concentration, and adding the contributions of each species to create a mixture with a given

Table S1: Concentrations of the target species in the training data set, including 12 repetitions of the central point.

Mixture Index	NO_3^- [mol kg^{-1}]	NO_2^- [mol kg^{-1}]	SO_4^{2-} [mol kg^{-1}]	CO_3^{2-} [mol kg^{-1}]	H_2O [mol kg^{-1}]
1	0.00	0.00	0.00	0.00	55.51
2	0.00	0.00	0.00	0.61	55.51
3	0.00	0.00	0.19	0.61	55.51
4	0.00	1.26	0.19	0.61	55.51
5	1.49	1.26	0.19	0.61	55.51
6	0.00	0.00	0.19	0.00	55.51
7	0.00	1.26	0.19	0.00	55.51
8	1.50	1.26	0.19	0.00	55.51
9	0.00	1.27	0.00	0.00	55.51
10	0.00	1.27	0.00	0.61	55.51
11	1.50	1.27	0.00	0.61	55.51
12	1.48	0.00	0.00	0.00	55.51
13	1.48	0.00	0.19	0.00	55.51
14	1.48	0.00	0.19	0.61	55.51
15	1.48	0.00	0.00	0.61	55.51
16	1.47	1.25	0.00	0.00	55.51
17	0.00	0.63	0.10	0.31	55.51
18	1.49	0.63	0.10	0.31	55.51
19	0.75	0.00	0.10	0.31	55.51
20	0.75	1.27	0.10	0.31	55.51
21	0.74	0.62	0.00	0.31	55.51
22	0.74	0.62	0.19	0.31	55.51
23	0.74	0.63	0.10	0.00	55.51
24	0.74	0.63	0.10	0.61	55.51
25	0.75	0.63	0.10	0.31	55.51
26	0.74	0.63	0.10	0.31	55.51
27	0.75	0.63	0.09	0.31	55.51
28	0.75	0.63	0.10	0.31	55.51
29	0.75	0.63	0.10	0.31	55.51
30	0.75	0.63	0.10	0.31	55.51
31	0.75	0.63	0.10	0.31	55.51
32	0.75	0.63	0.10	0.31	55.51
33	0.75	0.63	0.10	0.31	55.51
34	0.75	0.63	0.09	0.30	55.51
35	0.75	0.63	0.09	0.31	55.51
36	0.75	0.63	0.10	0.31	55.51

composition. Next, to simulate the presence of noise in the experimental measurements, a predetermined level of white, Gaussian noise is added to the spectra. For example, to simulate the test set from Table 1, using the eight reference spectra from Figure 1, one can evaluate:

$$\mathbf{X}_{ts,s} = \sum_{k=1}^{n_K} \mathbf{C}_{ts,s}^{k} \mathbf{L}^{k} + \boldsymbol{\eta} \tag{1}$$

where the subscript s is used to distinguish the simulated data, the superscript k represents each species, $\mathbf{C}_{ts,s} \in \mathbb{R}^{n_N \times n_K}$ represents the concentrations, $\mathbf{L} \in \mathbb{R}^{n_K \times n_L}$ is the library of the reference spectra and $\boldsymbol{\eta} \in \mathbb{R}^{n_K \times n_L}$ represents the noise, which is assumed to be additive to the signal. The level of added noise was based on the measurement noise associated with the experimental system, represented by the variance of the noise, σ_η^2. This value was used to calculate the signal-to-noise ratio (SNR). The local SNR (labeled with subscript l) for measurement i at a specific wavenumber j can be calculated as:

$$SNR_l^{ij} = \frac{X_{ij}^2}{\sigma_\eta^2} \tag{2}$$

The average SNR (labeled with subscript avg) for measurement i can be calculated as:

$$SNR_{avg}^{i} = \frac{1}{n_L} \sum_{j=1}^{n_L} SNR_l^{ij} \tag{3}$$

All SNR values were converted to decibel (dB), defined as $10 \log_{10}(SNR)$, for better resolution at both lower and higher levels. In this work, the SNR values for Raman and IR were calculated from experiments and similar levels were used in the simulated data, so that the simulations resemble the conditions of real experiments. The same approach was used to simulate the training data, $\mathbf{X}_{tr,s}$, using the experimental design from Table S1.

2.1 Raman Spectroscopy

The simulated Raman spectra for $\mathbf{X}_{ts,s}$ and $\mathbf{X}_{tr,s}$ are shown in Figures S1(a) and S2(a). Note that noise that corresponds to a *SNR* value of 40 dB was added to the simulated data. The algorithm identified seven independent components, which were compared against the reference spectra of the five target species. Figure S3 shows the correlation results. While sources for nitrate, nitrite, sulfate and carbonate were identified correctly, water was not identified as a unique source; instead the peak of water is present in the source which correlates highly with nitrite. In order to prevent any of the remaining sources (non-targets) to be identified as water in the correlation matrix, a cut-off value of 0.95 was imposed. The intensity of water is low in Raman, which could have contributed to the discrepancy.

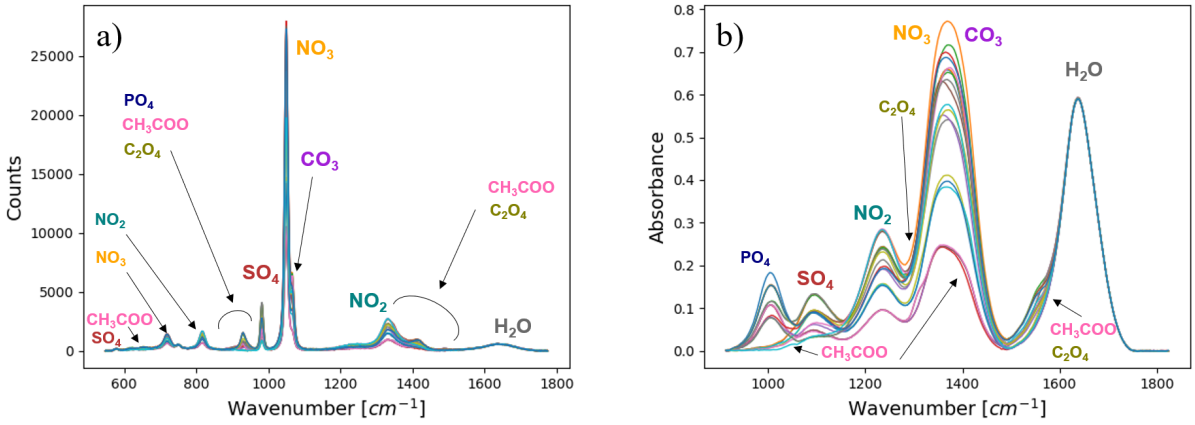

Figure S1: Simulated test data set mixtures for a) Raman spectroscopy with SNR_{avg} = 40 dB and b) ATR-FTIR spectroscopy with SNR_{avg} = 55 dB. Concentrations for the 21 mixtures are given in Table 1.

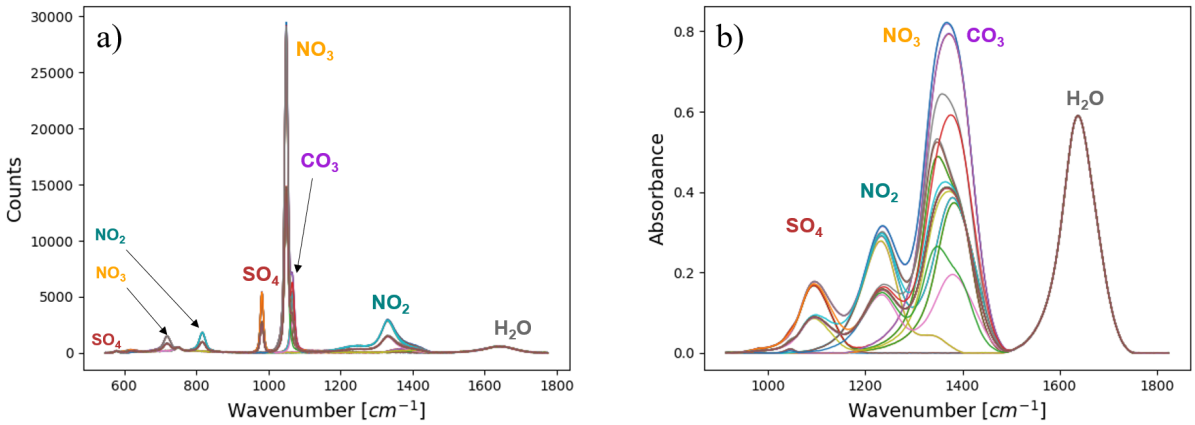

Figure S2: Simulated training data set mixtures for a) Raman spectroscopy with SNR_{avg} = 40 dB and b) ATR-FTIR spectroscopy with SNR_{avg} = 55 dB. The concentrations that match each measurement are shown in the Supplementary Information as Table S1.

The sources from Figure S3 that did not correlate with the target reference spectra, were classified as the non-target species. Their contributions to the mixtures in $\mathbf{X}_{ts,s}$ were subtracted, as shown in Figure S4. The resulting preprocessed spectra for $\mathbf{X}_{ts,s}$, shown in red, only include the signals of the target species.

The preprocessed test data for $\mathbf{X}_{ts,s}$ were used in Step 4 of the framework, to calculate the concentrations of the target species using a PLSR model. The predicted concentrations for the original and preprocessed data (using the same PLSR model) are compared to the real values in Figure S5. The parity plots show that generally, the prediction errors improved for the preprocessed data set, but the outcome of both PLSR and BSS-preprocessed PLSR also depends on the data treatment of the spectra, such as the order of derivation and window length for the Savitzky-Golay filter. A summary of the percent errors for each point in the original and preprocessed data sets is shown in Tables S2, S3 and S4, for different values for the order of derivation and Savitzky-Golay smoothing window length.

Table S2: Absolute values of the percent errors for the simulated Raman data set using a 0th derivative in the PLSR model. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species.

	PLSR only					BSS-PLSR				
Mixture Index	Nitrate (%)	Nitrite (%)	Sulfate (%)	Carbonate (%)	Water (%)	Nitrate (%)	Nitrite (%)	Sulfate (%)	Carbonate (%)	Water (%)
1	1.7	6.8	7.2	20.2	0.0	1.9	1.2	4.8	22.3	0.0
2	12.7	8.2	0.4	4.2	0.0	0.2	3.4	5.4	21.8	0.0
3	38.4	13.9	49.1	4.2	0.0	0.1	3.1	2.3	15.6	0.0
4	7.3	4.5	94.7	23.4	0.0	1.0	2.7	20.4	49.9	0.0
5	35.3	8.5	127.1	35.9	0.0	4.3	4.1	28.4	66.6	0.0
6	26.6	39.0	100.5	34.2	0.0	1.2	10.3	14.1	66.0	0.0
7	29.4	39.5	115.7	5.2	0.0	4.5	11.3	23.6	38.3	0.0
8	15.2	16.7	29.9	9.4	0.0	2.1	1.7	1.6	4.0	0.0
9	1.3	13.1	5.9	6.7	0.0	1.1	6.0	2.7	5.8	0.0
10	39.6	9.4	24.0	72.2	0.0	1.6	0.3	1.4	17.5	0.0
11	11.3	1.3	79.7	66.0	0.0	0.1	3.4	6.4	9.8	0.0
12	32.3	7.7	99.2	2.7	0.0	2.5	5.0	5.3	14.3	0.0
13	18.9	5.6	176.0	7.8	0.0	1.7	3.4	18.5	14.5	0.0
14	14.6	7.2	42.3	1.1	0.0	2.8	2.1	3.5	7.8	0.0
15	0.8	0.8	53.0	22.0	0.0	4.0	0.6	10.4	1.4	0.0
16	27.2	7.7	94.1	45.7	0.0	7.0	3.8	0.2	24.2	0.0
17	17.8	4.8	25.6	4.8	0.0	4.2	3.3	2.9	15.3	0.0
18	16.6	5.8	178.5	1.4	0.0	4.2	1.2	22.9	7.1	0.0
19	29.5	3.7	79.8	26.2	0.0	1.0	1.3	3.0	17.0	0.0
20	0.2	0.0	11.2	14.7	0.0	1.0	2.7	12.9	17.2	0.0
21	31.6	11.2	100.7	0.5	0.0	3.0	2.5	5.8	14.0	0.0
Average	**19.4**	**10.3**	**71.2**	**19.4**	**0.0**	**2.4**	**3.5**	**9.4**	**21.4**	**0.0**

Table S3: Absolute values of the percent errors for the simulated Raman data set using a 1st derivative in the PLSR model, with a Savitzky-Golay filter with a second order polynomial and window length size of 7. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species.

	PLSR only					BSS-PLSR				
Mixture Index	Nitrate (%)	Nitrite (%)	Sulfate (%)	Carbonate (%)	Water (%)	Nitrate (%)	Nitrite (%)	Sulfate (%)	Carbonate (%)	Water (%)
1	5.1	5.8	9.3	12.2	0.0	4.0	4.0	5.1	11.3	0.0
2	1.9	8.0	20.8	22.0	0.0	5.7	2.9	7.0	19.5	0.0
3	9.9	16.0	58.0	46.6	0.0	3.2	3.7	9.4	35.4	0.0
4	10.9	3.2	3.2	136.7	0.0	8.2	1.5	46.5	159.4	0.0
5	20.4	5.9	43.7	129.7	0.0	8.6	5.2	28.6	157.8	0.0
6	8.3	13.9	80.2	87.8	0.0	1.2	0.5	1.6	118.2	0.0
7	7.1	9.7	76.7	99.6	0.0	0.2	4.6	6.5	128.9	0.0
8	5.9	9.7	40.8	22.5	0.0	10.3	1.3	8.6	12.0	0.0
9	0.6	5.5	3.1	8.7	0.0	1.5	6.6	4.9	9.2	0.0
10	13.5	11.1	32.2	8.0	0.0	1.4	2.2	6.5	2.2	0.0
11	3.7	5.5	46.0	0.7	0.0	0.2	0.0	2.3	17.7	0.0
12	4.2	16.0	69.0	24.9	0.0	7.9	4.8	13.3	12.9	0.0
13	2.4	7.5	84.2	21.4	0.0	4.5	0.3	63.1	15.2	0.0
14	3.5	3.5	40.2	18.5	0.0	2.6	5.2	1.2	10.6	0.0
15	5.0	4.1	14.2	17.8	0.0	6.1	1.7	17.8	25.3	0.0
16	3.6	15.3	68.7	7.5	0.0	8.5	4.5	12.2	15.5	0.0
17	10.4	7.6	20.0	28.9	0.0	3.0	2.7	17.1	26.4	0.0
18	6.0	9.0	73.7	8.3	0.0	1.3	2.1	65.5	0.3	0.0
19	10.6	7.4	51.4	31.2	0.0	0.5	0.6	8.1	18.6	0.0
20	4.0	6.1	42.7	10.5	0.0	3.8	5.1	57.0	9.8	0.0
21	6.0	15.9	69.5	37.3	0.0	6.5	0.1	9.4	22.1	0.0
Average	**6.8**	**8.9**	**45.1**	**37.2**	**0.0**	**4.2**	**2.8**	**18.6**	**39.4**	**0.0**

Table S4: Absolute values of the percent errors for the simulated Raman data set using a 1st derivative in the PLSR model, with a Savitzky-Golay filter with a second order polynomial and window length size of 27. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species.

	PLSR only					BSS-PLSR				
Mixture Index	Nitrate (%)	Nitrite (%)	Sulfate (%)	Carbonate (%)	Water (%)	Nitrate (%)	Nitrite (%)	Sulfate (%)	Carbonate (%)	Water (%)
1	10.5	5.2	8.8	2.5	0.0	1.4	4.5	16.2	12.7	0.0
2	25.0	12.7	94.2	111.1	0.0	0.4	1.6	9.6	2.1	0.0
3	118.6	36.0	258.5	344.1	0.0	2.3	2.0	13.7	0.3	0.0
4	77.6	43.3	122.3	220.0	0.0	8.9	7.8	22.2	34.8	0.0
5	160.6	53.5	340.1	654.3	0.0	21.4	9.4	59.4	72.2	0.0
6	92.6	80.6	392.3	598.3	0.0	1.5	9.7	67.4	13.8	0.0
7	99.0	78.1	397.4	542.4	0.0	5.6	6.9	58.3	14.2	0.0
8	50.6	36.6	137.2	188.2	0.0	4.8	3.0	6.2	27.4	0.0
9	24.1	6.5	2.0	44.2	0.0	6.5	2.0	7.7	42.1	0.0
10	109.3	19.8	213.1	315.3	0.0	1.7	10.2	10.7	33.3	0.0
11	99.6	42.0	101.4	179.7	0.0	4.7	1.3	0.3	18.5	0.0
12	116.7	33.0	387.4	329.8	0.0	1.3	3.4	37.8	3.6	0.0
13	68.5	18.9	661.9	151.8	0.0	0.3	5.0	98.5	18.2	0.0
14	59.1	26.1	201.1	219.9	0.0	0.1	1.3	12.3	0.8	0.0
15	22.2	13.0	43.2	90.9	0.0	1.6	2.0	31.2	0.7	0.0
16	107.6	33.8	389.9	567.1	0.0	10.3	0.7	25.4	74.0	0.0
17	52.3	6.0	229.3	142.1	0.0	0.9	4.3	52.9	14.9	0.0
18	69.7	18.8	689.0	237.4	0.0	1.5	2.8	23.6	15.1	0.0
19	119.3	29.3	228.8	284.9	0.0	2.1	2.0	16.3	12.0	0.0
20	5.8	6.1	15.7	7.6	0.0	2.9	2.4	66.5	8.6	0.0
21	121.2	47.0	390.8	454.0	0.0	0.5	1.6	13.6	25.9	0.0
Average	**76.7**	**30.8**	**252.6**	**270.7**	**0.0**	**3.8**	**4.0**	**30.9**	**21.2**	**0.0**

2.2 ATR-FTIR Spectroscopy

Similar simulation analyses were performed for ATR-FTIR, for a *SNR* value of 55 dB, based on the level of noise present in experimental IR data. While the IR spectra are less noisy compared to Raman, they feature more significant overlap among the target and non-target species, which complicates the identification of BSS components. For example, the carbonate and nitrate overlap with each other almost completely, but also with acetate, and to a lower extent with oxalate and nitrite. At the same time, the signals of both the target and non-target species are strong, which could help with identification, compared to Raman, where the signals of phosphate, water, oxalate and acetate are weaker and therefore harder to identify. The test and training data sets are plotted in Figures S1(b) and S2(b). The preprocessing procedure identified eight independent sources, shown in Figure S6. The sources that correlated with nitrate, sulfate and water preserved their shapes completely, while carbonate and nitrite show minor deviations. Even with the deviations, the correlation coefficients are relatively high, as shown in Figure S6.

Next, the test data were preprocessed to remove the signals associated with the non-target sources, shown in Figure S7. The parity plot that compares the predicted concentrations for the original and preprocessed data shows improvements for the preprocessed mixtures (Figure S8). A summary of the percent errors for each point in the original and preprocessed data sets is shown in Table S5.

Table S5: Absolute values of the percent errors for the simulated IR data set using a 1st derivative in the PLSR model, with a Savitzky-Golay filter with a second order polynomial and window length size of 5. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species.

	PLSR only					BSS-PLSR				
Mixture Index	Nitrate (%)	Nitrite (%)	Sulfate (%)	Carbonate (%)	Water (%)	Nitrate (%)	Nitrite (%)	Sulfate (%)	Carbonate (%)	Water (%)
1	1.0	8.3	9.3	2.5	0.0	4.6	6.1	9.1	7.5	0.0
2	10.1	6.5	8.6	38.8	0.0	0.7	6.9	11.4	3.5	0.0
3	24.0	36.3	183.2	49.9	0.0	9.2	3.0	11.0	25.2	0.0
4	108.2	129.2	551.4	94.1	0.0	2.2	30.8	27.8	74.1	0.0
5	91.4	121.7	544.0	13.2	0.0	21.6	20.9	25.7	98.2	0.0
6	81.1	98.7	502.3	45.3	0.0	2.2	26.7	24.8	83.7	0.0
7	82.2	91.4	538.6	55.2	0.0	2.1	18.7	30.5	67.0	0.0
8	39.8	30.8	169.2	27.4	0.0	12.5	1.1	4.2	4.8	0.0
9	1.6	9.3	11.3	12.5	0.0	5.2	10.3	13.4	11.2	0.0
10	52.1	9.5	8.5	115.6	0.0	0.1	17.2	9.7	32.2	0.0
11	134.3	120.6	506.2	44.6	0.0	7.9	10.3	17.6	8.2	0.0
12	41.5	45.1	382.3	38.2	0.0	12.6	0.4	22.0	17.1	0.0
13	22.0	28.0	661.8	25.4	0.0	3.9	6.0	73.0	6.9	0.0
14	17.7	42.8	191.0	33.3	0.0	3.8	12.7	16.7	10.8	0.0
15	41.6	41.1	321.5	28.0	0.0	5.0	10.2	36.4	21.3	0.0
16	40.3	46.1	386.5	72.2	0.0	11.5	0.8	21.7	36.8	0.0
17	21.4	0.9	5.1	35.9	0.0	1.1	7.6	31.8	7.4	0.0
18	17.0	30.8	637.1	36.8	0.0	0.0	7.6	60.5	6.4	0.0
19	63.3	63.3	391.6	15.4	0.0	0.3	6.1	19.8	10.9	0.0
20	2.5	7.3	15.3	3.1	0.0	3.2	2.3	54.5	5.2	0.0
21	27.1	81.9	387.7	65.8	0.0	1.4	7.8	17.3	11.2	0.0
Average	**43.8**	**50.0**	**305.4**	**40.6**	**0.0**	**5.3**	**10.2**	**25.7**	**26.2**	**0.0**

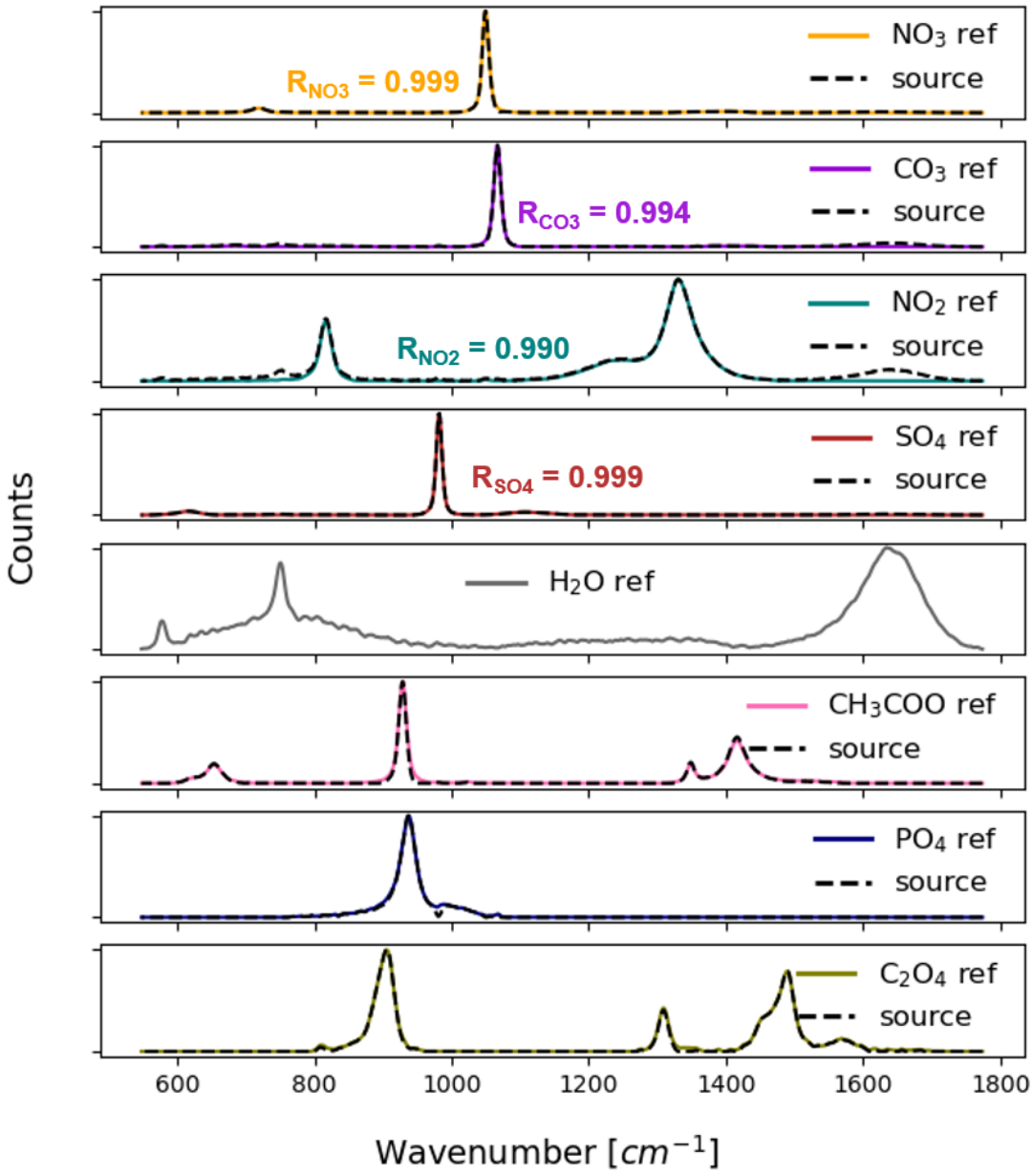

Figure S3: Reference spectra-sources correlation: the first five rows show the normalized reference spectra of the target species, while the last three rows show the normalized reference spectra of the non-target species. The normalized Raman sources identified by the blind source separation algorithm are plotted in black (dashed). The BSS algorithm did not identify a unique source for water.

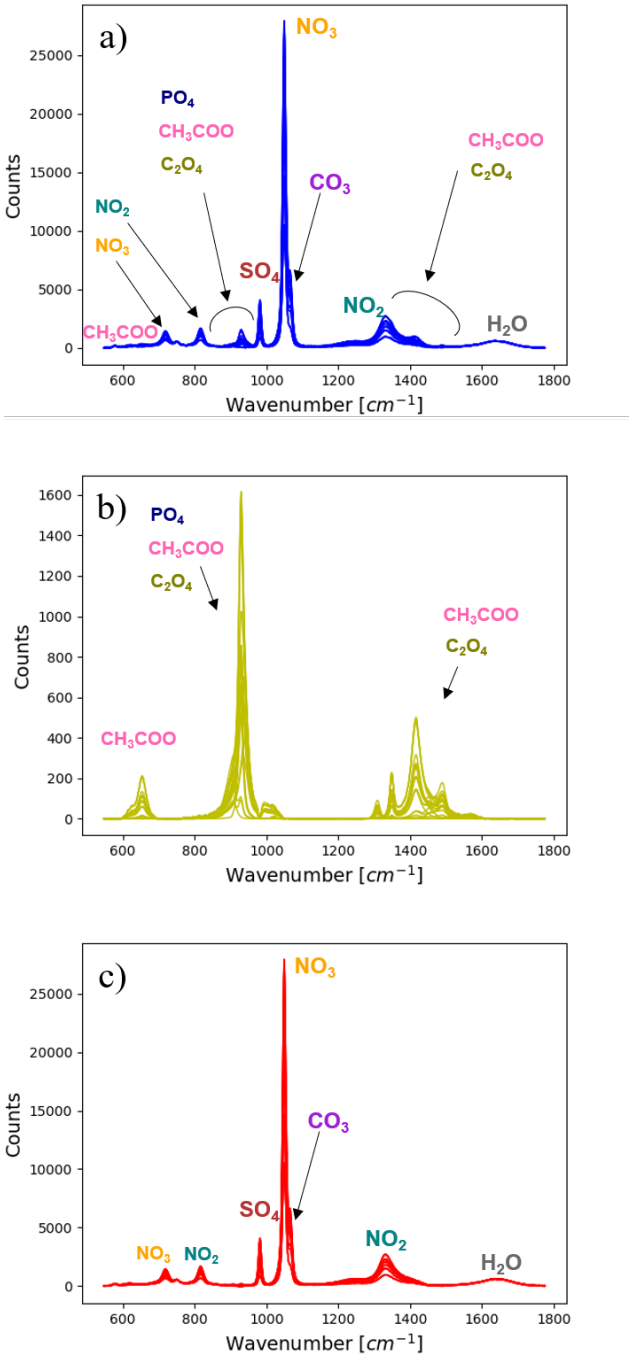

Figure S4: Removing the signals of the non-target species for Raman: a) original test set mixture data (blue), containing both target and non-target species, b) signals of the non-target species (yellow) that are subtracted to generate the preprocessed spectra (red) in c). The intensity of the yellow signals is lower relative to the spectra of the targets.

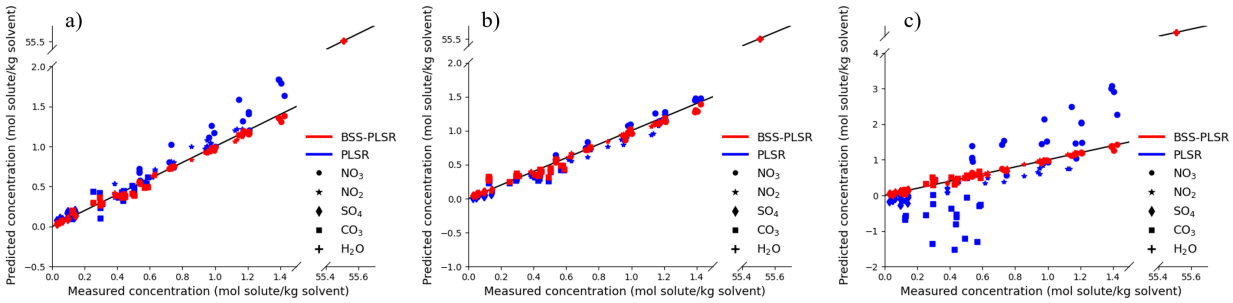

Figure S5: Parity plot comparing the Raman predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for Raman: a) 0th derivative spectra; b) 1st derivative spectra using a Savitzky-Golay filter with a second order polynomial and window length size of 7; and c) 1st derivative spectra using a Savitzky-Golay filter with a second order polynomial and window length size of 27.

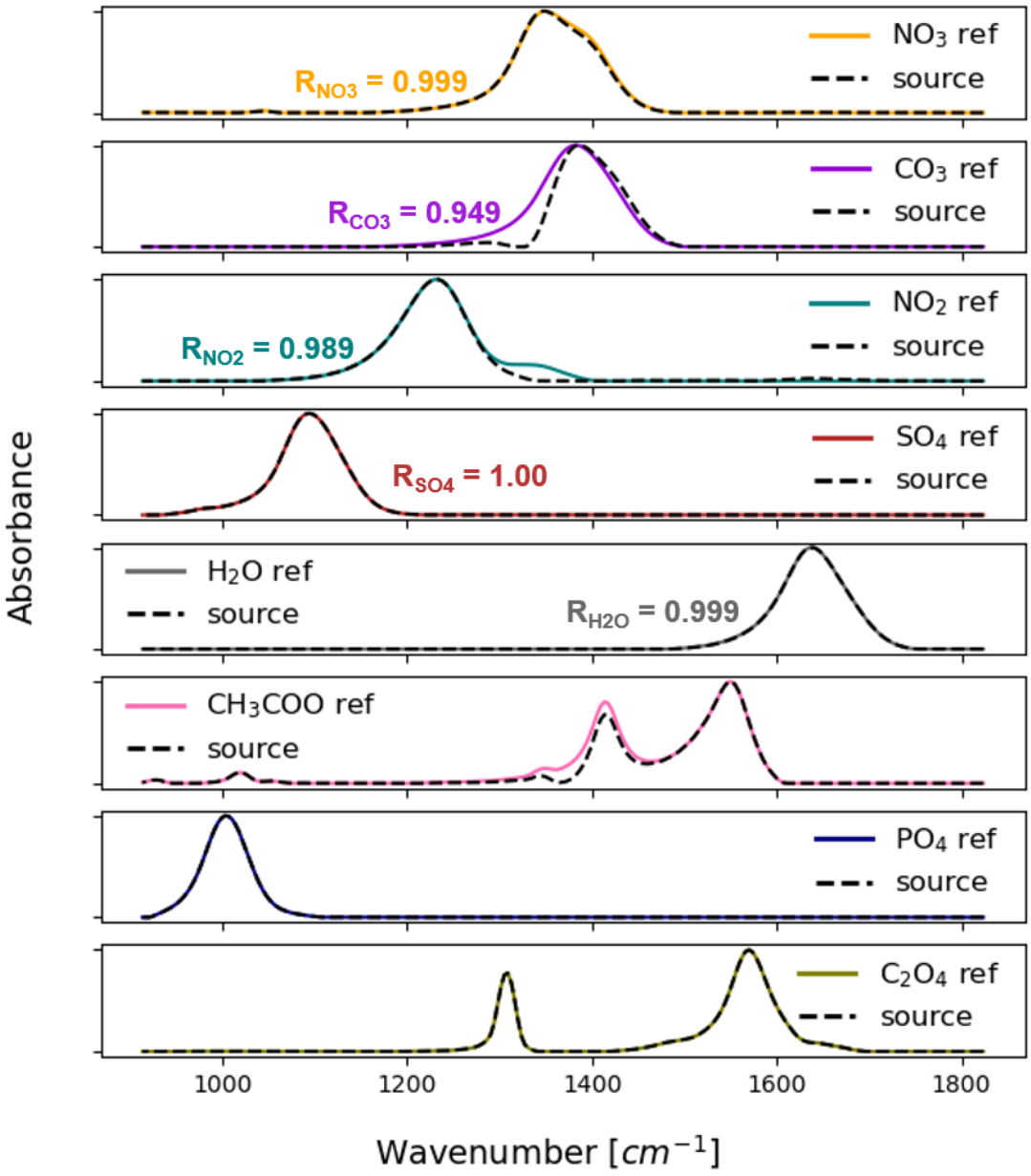

Figure S6: Reference spectra-sources correlation: the first five rows show the normalized reference spectra of the target species, while the last three rows show the normalized reference spectra of the non-target species. The normalized IR sources identified by the blind source separation algorithm are plotted in black (dashed).

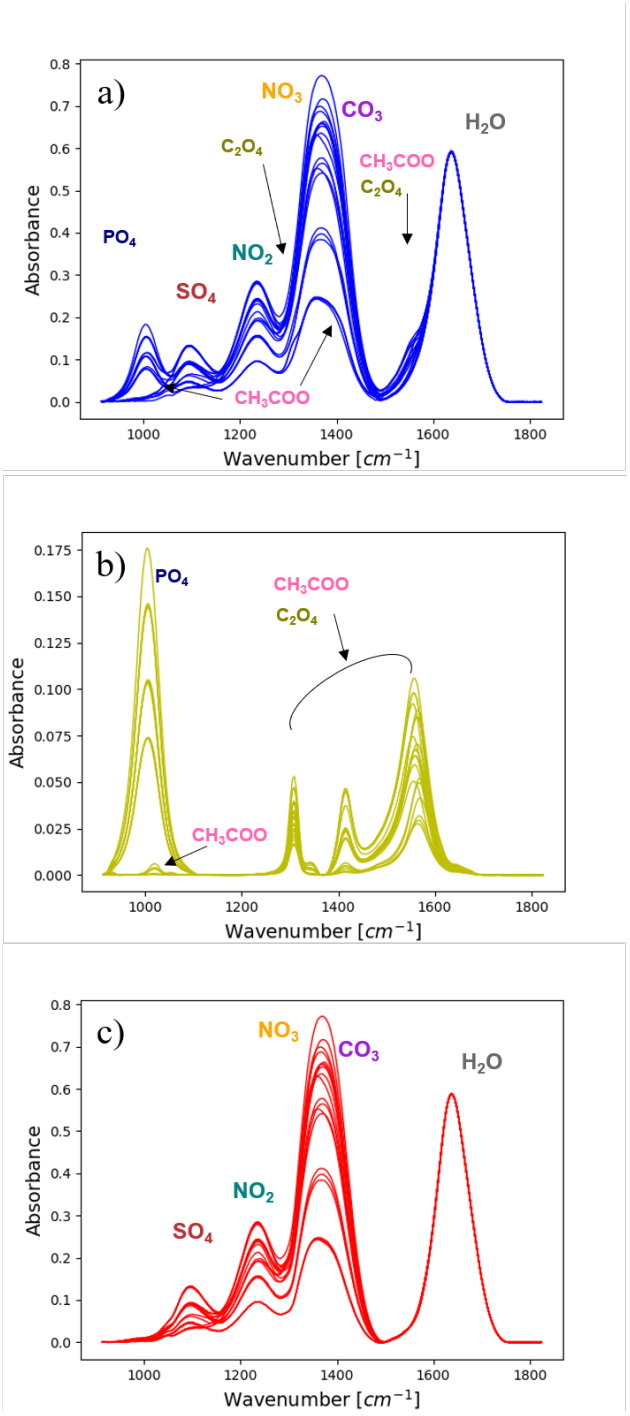

Figure S7: Removing the signals of the non-target species for IR: a) original mixture data (blue), containing both target and non-target species, b) signals of the non-target species (yellow) that are subtracted to generate the preprocessed spectra (red) in c). The intensity of the yellow signals is lower relative to the spectra of the targets.

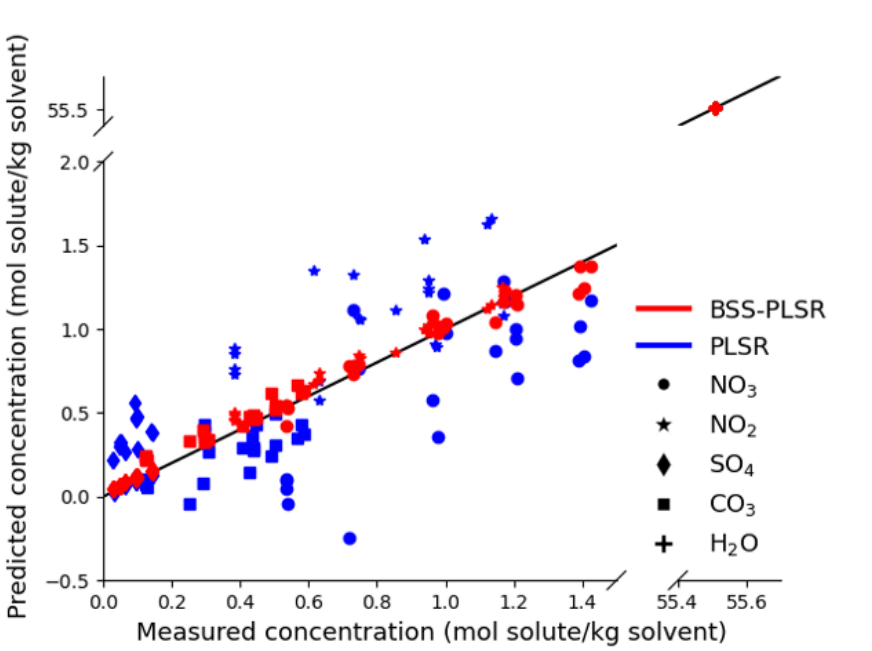

Figure S8: Parity plot comparing the predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for IR.

3 Experimental Data

3.1 Raman Spectroscopy

3.1.1 Water not identified as a source in Raman

The resulting reference spectra-sources correlation is shown in Figure S9.

3.1.2 Percent errors for the experimental Raman data

The percent errors from the experimental Raman data are given in Table S6.

Table S6: Absolute values of the percent errors for the experimental Raman data set using a 1st derivative in the PLSR model, with a Savitzky-Golay filter with a second order polynomial and window length size of 7. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species.

	PLSR only					BSS-PLSR				
Mixture Index	Nitrate (%)	Nitrite (%)	Sulfate (%)	Carbonate (%)	Water (%)	Nitrate (%)	Nitrite (%)	Sulfate (%)	Carbonate (%)	Water (%)
1	0.9	2.4	11.2	28.9	0.0	7.3	2.6	20.8	5.2	0.0
2	47.5	0.6	74.3	203.6	0.0	4.0	3.4	11.1	19.1	0.0
3	46.7	4.0	32.8	239.8	0.0	15.9	1.9	90.1	6.4	0.0
4	2.0	10.3	257.5	64.9	0.0	14.1	0.1	107.4	92.3	0.0
5	134.9	31.0	19.2	955.7	0.0	0.2	9.3	68.6	26.4	0.0
6	130.2	16.5	4.2	879.3	0.0	8.5	0.2	34.0	158.8	0.0
7	124.8	15.8	47.3	889.7	0.0	11.3	0.6	76.0	95.5	0.0
8	58.4	1.5	32.0	292.8	0.0	19.5	5.3	71.8	14.3	0.0
9	3.4	1.8	2.6	9.6	0.0	0.9	1.5	0.3	33.1	0.0
10	168.5	4.0	231.5	861.1	0.0	6.1	5.6	25.4	150.7	0.0
11	9.9	24.9	241.8	91.9	0.0	14.6	10.7	80.7	35.5	0.0
12	107.6	5.3	18.9	531.3	0.0	15.2	8.8	196.1	62.8	0.0
13	64.1	7.2	158.8	286.4	0.0	11.7	0.1	343.0	43.8	0.0
14	50.7	5.2	9.1	376.8	0.0	14.3	3.7	134.8	48.9	0.0
15	6.3	5.0	174.0	37.9	0.0	15.6	2.8	144.3	33.1	0.0
16	104.5	11.6	37.9	1018.8	0.0	18.6	2.3	179.5	94.5	0.0
17	80.3	6.7	244.7	254.3	0.0	0.4	10.2	18.2	48.4	0.0
18	65.0	1.8	135.5	345.5	0.0	11.2	5.4	349.4	23.1	0.0
19	85.2	16.6	60.1	376.6	0.0	11.6	4.4	110.0	30.1	0.0
20	0.2	4.6	36.3	6.6	0.0	4.8	4.8	3.1	8.8	0.0
21	104.8	16.6	46.2	680.4	0.0	16.1	5.6	153.7	68.1	0.0
Average	**66.5**	**9.2**	**89.3**	**401.5**	**0.0**	**10.6**	**4.3**	**105.6**	**52.3**	**0.0**

3.1.3 Parity plot for 0th derivative spectra for Raman

The parity plot for PLSR and BSS-PLSR predictions on 0th order spectra is shown in Figure S10.

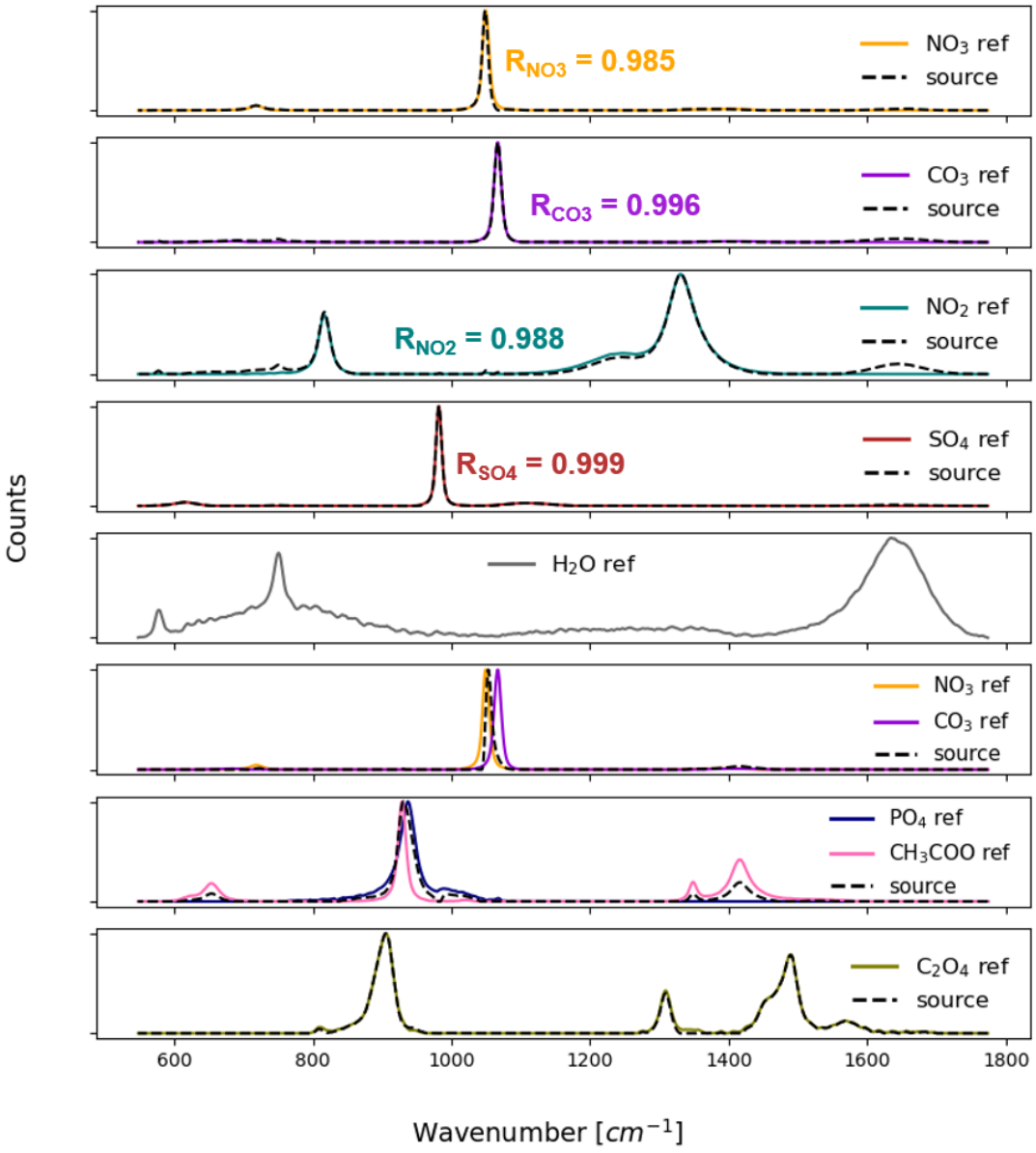

Figure S9: Reference spectra-sources correlation: the first five rows show the normalized reference spectra of the target species, while the last two rows show the normalized reference spectra of the non-target species. The normalized Raman sources identified by the blind source separation algorithm are plotted in black (dashed). The BSS algorithm did not identify water as a unique source.

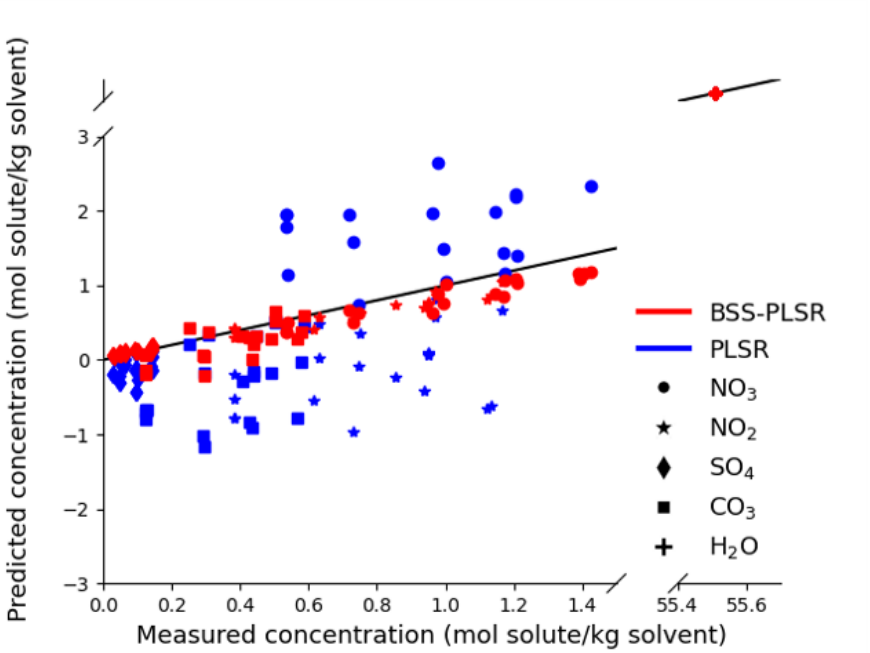

Figure S10: Parity plot comparing the Raman predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for Raman experimental data, using 0th derivative spectra.

3.1.4 Parity plot for Raman, prior to re-classification of the nitrate shift source as a target species

The parity plot for PLSR and BSS-PLSR predictions prior to reclassifying the nitrate blue shift source as a target species is shown in Figure S11.

3.2 FTIR Spectroscopy

3.2.1 Percent errors for the experimental IR data

The percent errors from the experimental IR data are given in Table S7.

3.2.2 Parity plot for 0th derivative spectra for IR

The parity plot for PLSR and BSS-PLSR predictions on 0th order spectra is shown in Figure S12.

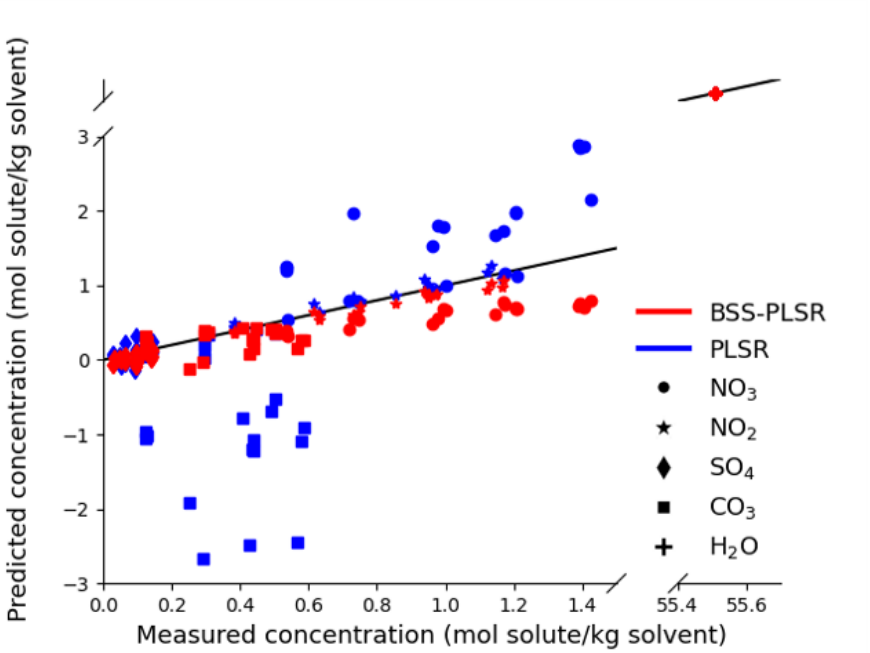

Figure S11: Parity plot comparing the Raman predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for Raman experimental data, while the nitrate shift source was classified as a non-target species.

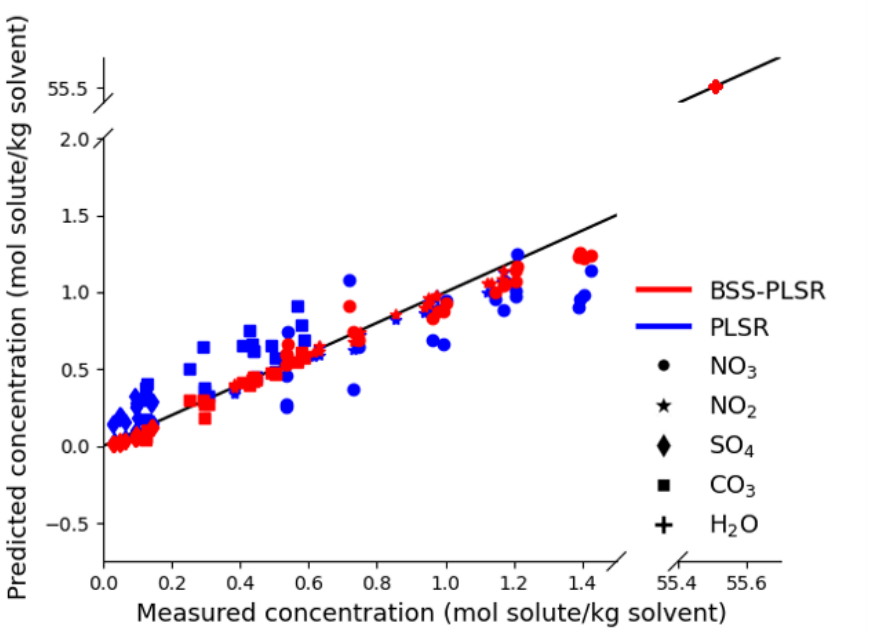

Figure S12: Parity plot comparing the IR predictions made by PLSR on the original test data (blue points) to those made by PLSR on the BSS-preprocessed test data (red points) for IR experimental data, using 0th derivative spectra.

Table S7: Absolute values of the percent errors for the experimental IR data set using a 1st derivative in the PLSR model, with a Savitzky-Golay filter with a second order polynomial and window length size of 5. Results for the original data are shown on the left, while results using BSS-preprocessing are shown on the right side. The last row shows the average percent error for each species.

	PLSR only					BSS-PLSR				
Mixture Index	Nitrate (%)	Nitrite (%)	Sulfate (%)	Carbonate (%)	Water (%)	Nitrate (%)	Nitrite (%)	Sulfate (%)	Carbonate (%)	Water (%)
1	7.1	5.4	5.4	17.6	0.0	12.9	13.9	2.9	7.5	0.0
2	13.9	19.1	12.7	37.3	0.0	14.1	18.4	1.5	9.8	0.0
3	26.7	36.4	38.4	91.9	0.0	9.8	16.7	13.2	5.7	0.0
4	99.9	151.5	65.1	263.2	0.0	2.8	13.7	3.6	72.3	0.0
5	89.3	117.1	86.3	358.4	0.0	7.5	18.8	4.1	58.5	0.0
6	48.6	67.0	115.6	409.1	0.0	17.2	39.1	2.6	66.2	0.0
7	29.4	78.4	123.7	368.2	0.0	43.7	41.7	7.0	89.1	0.0
8	27.6	40.1	37.4	126.4	0.0	17.4	20.4	5.6	1.1	0.0
9	25.6	14.8	10.0	12.3	0.0	34.7	27.5	1.0	18.7	0.0
10	19.9	39.4	21.6	50.3	0.0	11.4	23.8	10.4	49.5	0.0
11	133.1	162.4	58.5	190.2	0.0	15.5	6.6	9.9	24.1	0.0
12	19.8	53.6	80.2	110.6	0.0	13.3	5.7	12.8	12.0	0.0
13	15.2	29.2	179.3	60.9	0.0	12.8	13.0	60.0	6.7	0.0
14	9.1	38.0	39.4	77.8	0.0	13.3	13.1	6.2	7.7	0.0
15	31.3	57.3	30.4	80.6	0.0	1.7	7.9	3.4	20.9	0.0
16	30.9	43.5	77.8	253.6	0.0	9.1	15.3	11.7	7.9	0.0
17	9.2	26.2	37.0	28.1	0.0	5.5	21.2	9.2	3.6	0.0
18	22.3	22.3	155.8	92.2	0.0	4.7	18.3	34.1	0.2	0.0
19	61.6	75.5	66.3	135.7	0.0	7.7	11.6	17.1	4.9	0.0
20	9.9	7.6	27.0	0.0	0.0	17.0	15.0	14.4	12.0	0.0
21	22.6	90.0	70.6	149.2	0.0	12.1	12.5	3.7	13.2	0.0
Average	**35.9**	**55.9**	**63.7**	**138.8**	**0.0**	**13.5**	**17.8**	**11.2**	**23.4**	**0.0**

82