Design of experiments, chemometrics, and Raman spectroscopy for the quantification of hydroxylammonium, nitrate, and nitric acid

Luke R. Sadergaski[1,*], *Travis J. Hager*[2], *Hunter B. Andrews*[1]

[1]Oak Ridge National Laboratory, 1 Bethel Valley Rd., Oak Ridge, TN, 37830, USA

[2]University of Missouri, Department of Chemistry, 125 Chemistry Building Columbia, MO, 65211, USA

*sadergaskilr@ornl.gov

This Supporting Information Contains:

Pages: 9

Tables: 5

Figures: 7

EXTENDED METHODS

Tukey pair-wise statistics

The Tukey-Kramer method was used for the pairwise comparison of RMSEPs of each PLSR model assuming the null hypothesis H_0: $\mu_i = \mu_j$. At the level $\alpha = 0.05$, a type I error is made 5% of the time (i.e., H_0 is falsely rejected). When making multiple comparisons, the critical t value must be adjusted to account for α inflation and avoid false significant results. Therefore, the critical t value was adjusted to avoid making a type I error.

Model comparisons were carried out using a three-step process: (1) the sample means $\hat{y}_i$ were ranked in order from largest to smallest, (2) the largest sample mean was compared with the smallest mean until the largest was compared with the second largest, and (3) the second largest mean was compared with the smallest mean until the second largest was compared with the third largest. This pairwise comparison continued until every combination was tested.

TABLES

Table S1. Solution compositions selected by D-optimal design for generating a training set to build regression models

ID	HAN	HNO_3	Space type	Build type
1	0.43	0.0	Edge	Model

2	0.0	0.85	Edge	Model
3	0.315	1.0	Edge	Model
4	0.50	0.645	Interior	Model
5	0.0	0.0	Vertex	Model
6	0.215	0.425	Interior	Model

Table S2. D-optimal validation set comprising lack-of-fit points

ID	HAN	HNO_3	Space type	Build type
1	0.085	0.40	Interior	Lack of fit
2	0.265	1.0	Edge	Lack of fit
3	0.22	0.255	Interior	Lack of fit
4	0.1325	0.835	Interior	Lack of fit
5	0.135	0.0	Edge	Lack of fit
6	0.35	0.43	Interior	Lack of fit
7	0.50	0.0	Vertex	Lack of fit
8	0.0	0.67	Edge	Lack of fit
9	0.345	0.737	Interior	Lack of fit
10	0.50	0.69	Edge	Lack of fit

Table S3. Order of all possible HA^+ comparisons based on the magnitude of the mean ($\hat{y}_i$) prediction values

Comparisons	A	B
1	U-S	D-optimal
2	U-S	U-DD
3	U-S	None
4	NP	D-optimal
5	NP	U-DD
6	U-DD	D-optimal

Abbreviations: user-defined design (U-DD), user selected (U-S), and raw data (NP)

Table S4. Order of all possible H^+ comparisons based on the magnitude of the mean ($\hat{y}_i$) prediction values

Comparison #	A	B
1	U-S	U-DD
2	U-S	D-optimal
3	U-S	None
4	NP	U-DD
5	NP	D-optimal
6	D-optimal	U-DD

Abbreviations: user-defined design (U-DD), user selected (User), and raw data (NP)

Table S5. Order of all possible NO_3^- comparisons based on the magnitude of the mean ($\hat{y}_i$) prediction values. Abbreviations: user-defined design (U-DD), user selected (U-S) and raw data (NP).

Comparison #	A	B
1	U-S	U-DD
2	U-S	D-optimal
3	U-S	NP
4	NP	U-DD
5	NP	D-optimal
6	D-optimal	U-DD

Abbreviations: user-defined design (U-DD), user selected (U-S) and raw data (NP)

FIGURES

Figure S1. General-purpose Raman probe and probe/cuvette holder made by Spectra solutions Inc.

Photograph courtesy of Luke R. Sadergaski.

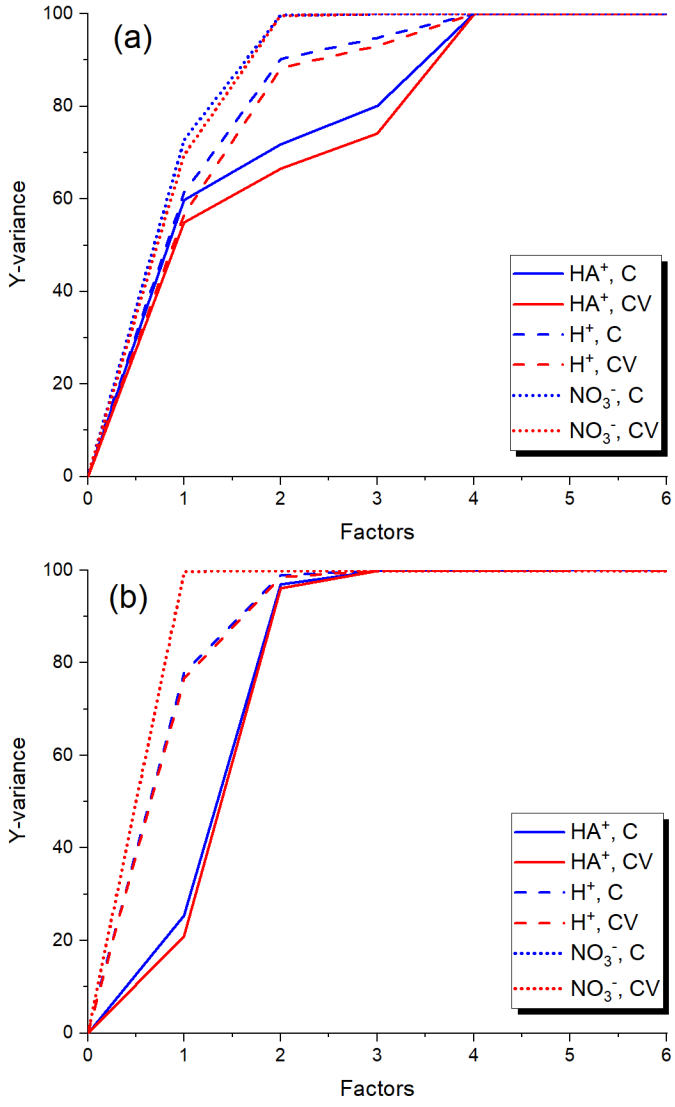
(a)
Y-variance
Factors
HA+, C
HA+, CV
H+, C
H+, CV
NO3-, C
NO3-, CV
(b)
Y-variance
Factors
HA+, C
HA+, CV
H+, C
H+, CV
NO3-, C
NO3-, CV

Figure S2. Explained *Y*-variance of the calibration (C) and cross validation (CV) vs. the number of factors describing hydroxylammonium (HA^+), H^+, and NO_3^- concentrations for PLSR models built (a) without preprocessing and (b) with the optimal strategy selected using D-optimal design.

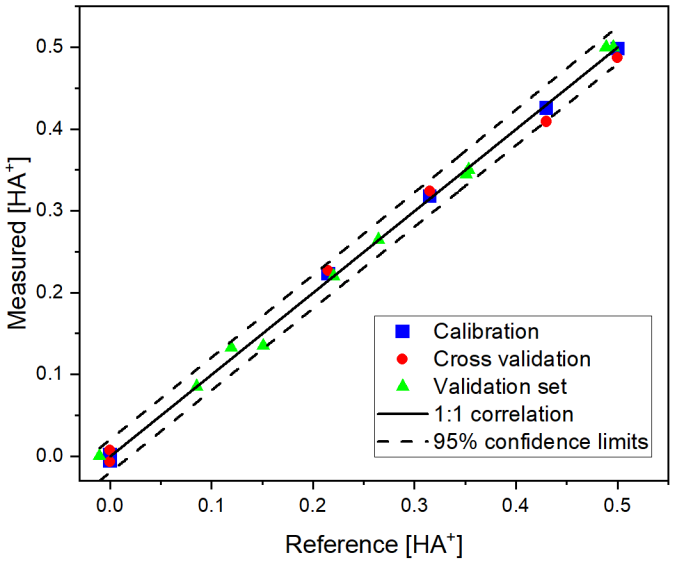

Figure S3. Parity plots for PLSR D-optimal model hydroxylammonium (HA^+) results. Blue squares represent the calibration set, red circles represent the cross validation, green triangles represent the predictions on the validation set, and the line represents a notional 1:1 correlation (slope = 1) with 95% confidence limits of the linear regression between measured and reference values.

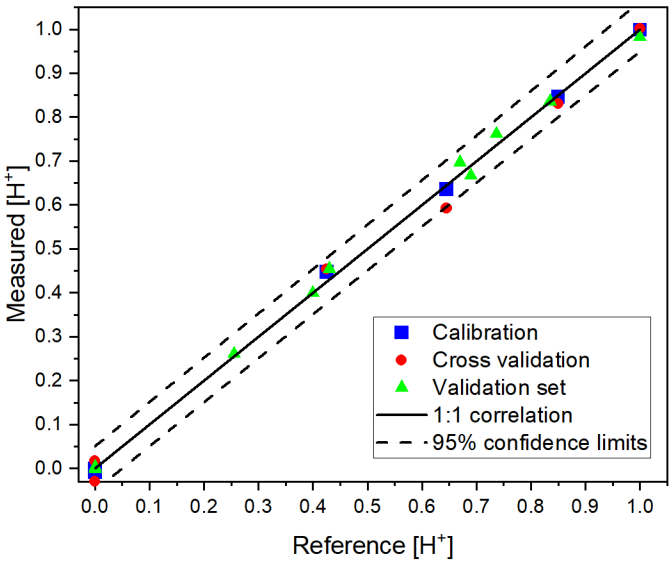

Figure S4. Parity plots for PLSR D-optimal model free acid (H^+) results. Blue squares represent the calibration set, red circles represent the cross validation, green triangles represent the predictions on the validation set, and the line represents a notional 1:1 correlation (slope = 1) with 95% confidence limits of the linear regression between measured and reference values.

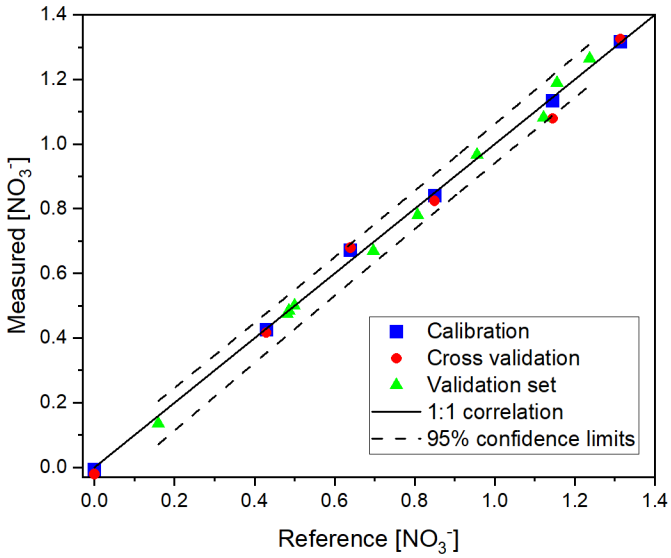

Figure S5. Parity plots for PLSR D-optimal model nitrate (NO_3^-) results. Blue squares represent the calibration set, red circles represent the cross validation, green triangles represent the predictions on the validation set, and the line represents a notional 1:1 correlation (slope = 1) with 95% confidence limits of the linear regression between measured and reference values.

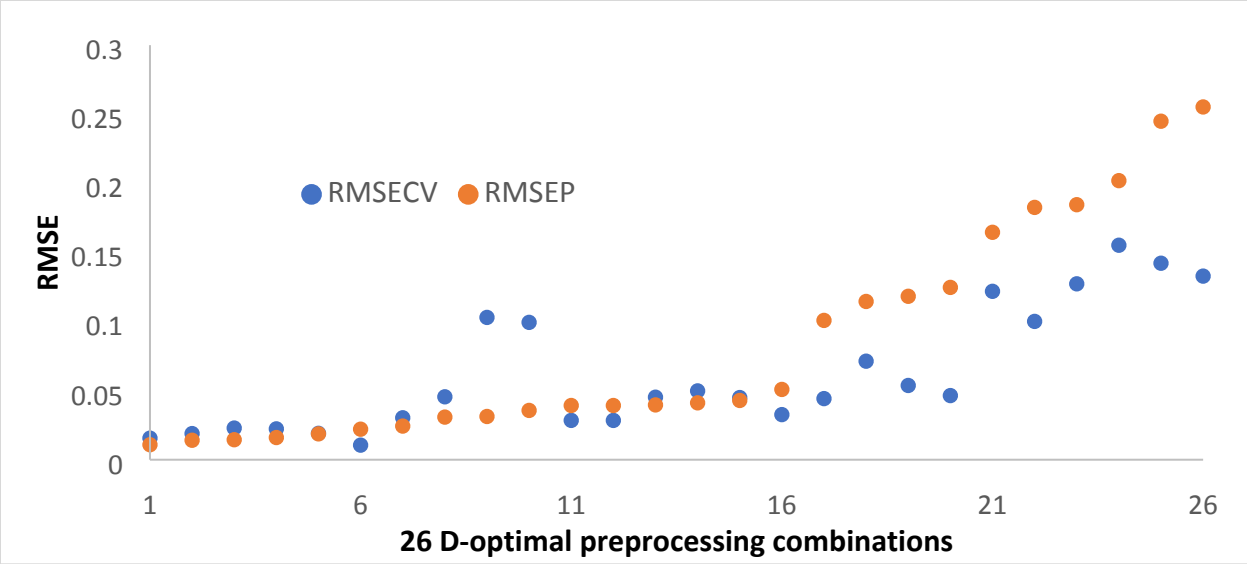

Figure S6. RMSECV and RMSEP for HA^+ concentration in order of smallest to largest RMSEP values for PLSR models built using every D-optimal preprocessing strategy.

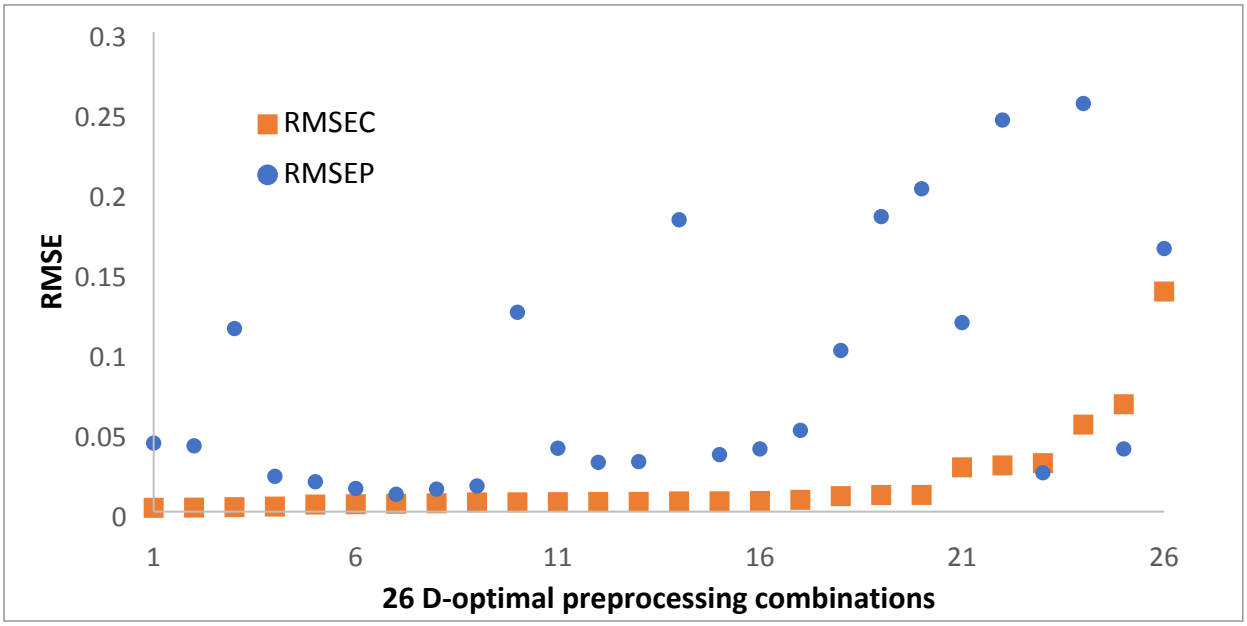

Figure S7. RMSEC and RMSEP for HA^+ concentration in order of smallest to largest RMSEC values for PLSR models built using every D-optimal preprocessing strategy.