Supplemental material for Predicting Ambient Aerosol Thermal Optical Reflectance (TOR) Measurements from Infrared Spectra: Organic Carbon

A. M. Dillner[1], S. Takahama[2]

[1]University of California - Davis, Davis, California

[2]Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland

Correspondence to: A. M. Dillner (amdillner@ucdavis.edu)

S1. Location of sampling sites

Figure S1. Aerosol sampling locations: Mesa Verde,CO (37.1984, -108.4907) 119 samples Jan - Dec 2011, Olympic, WA (48.0065, -122.9727) 120 samples Jan - Dec 2011, Phoenix, AZ (33.5038, -112.0958) 100 samples Jan - Dec 2011, Phoenix, AZ second sampler (33.5038, -112.0958) 99 samples Jan - Dec 2011, Proctor Maple Research Facility, VT (44.5284, -72.8688) 106 samples Jan - Dec 2011, Sac and Fox, KS (39.9791, -95.5682) 53 Samples Jan - Jun 2011, St. Marks, FL (30.0926, -84.1614) 108 Samples Jan - Dec 2011, Trapper Creek, AK (63.3153, -150.3156) 110 samples Jan - Dec 2011.

S2. Spectral types

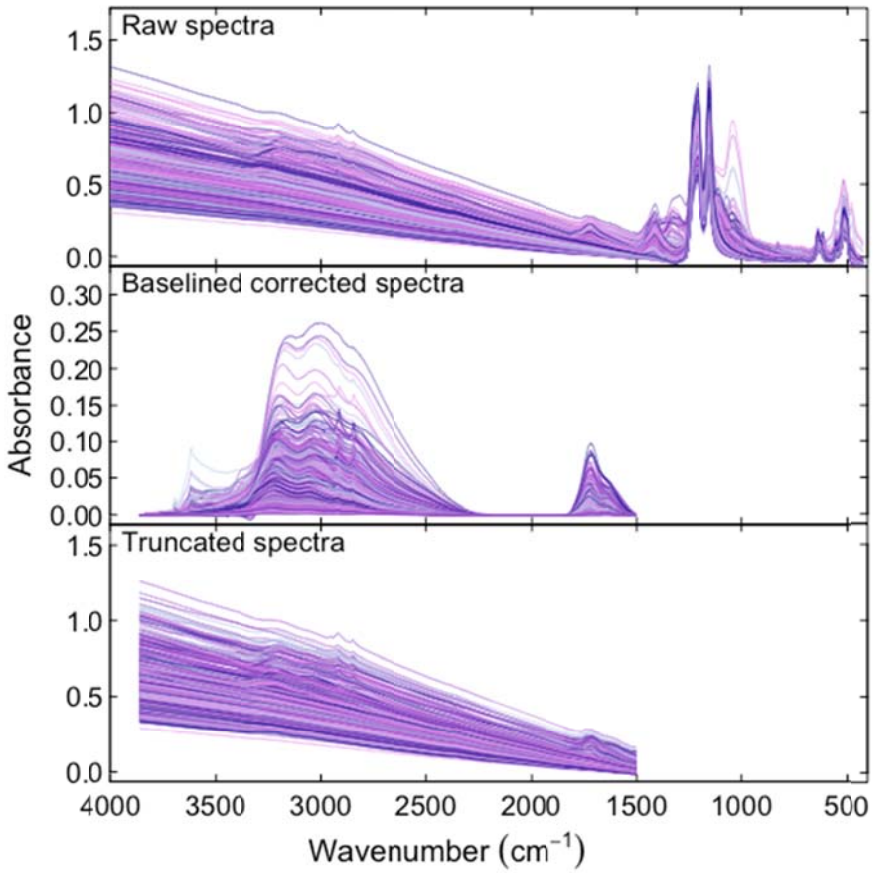

Figure S2. Spectra of 794 IMPROVE PTFE samples, shown as original absorbances ("raw" spectra) and after pretreatment ("baseline corrected" and "truncated" spectra) as described in Section 2.2. Each spectrum is differentiated by color.

S3. PLSR

The calibration model is selected on the basis of minimum root mean square error of prediction (RMSEP). *K*-fold "interleaved" cross validation is used. Figure S3 shows that the prediction accuracy is robust with respect to the value of *K* used in selection of our model. The number of components selected for the base case calibration model is 48 for the full spectra, 15 for the baseline corrected spectra, and 47 for the truncated spectra.

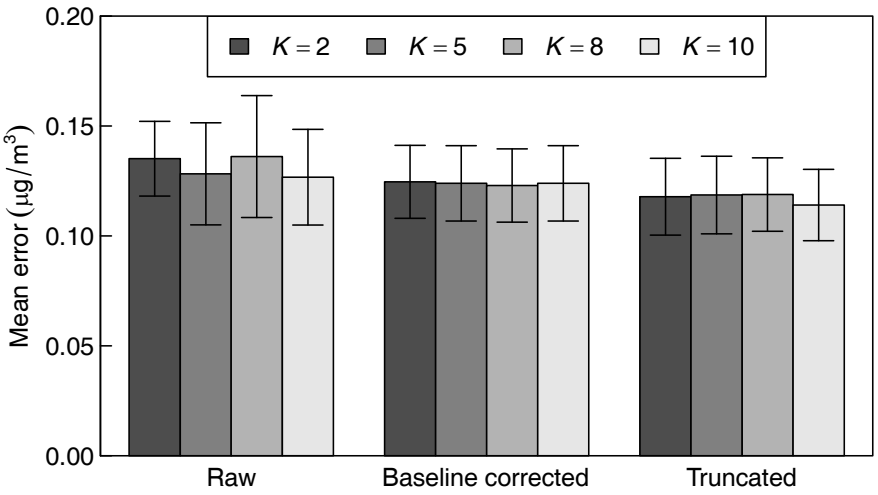

Figure S3. Mean prediction errors for models selected using varied number of segments in K-fold cross validation for each case of spectra pretreatment. Error bars indicate the 95% confidence interval on the means.

S4. Error distribution in uniform calibrations

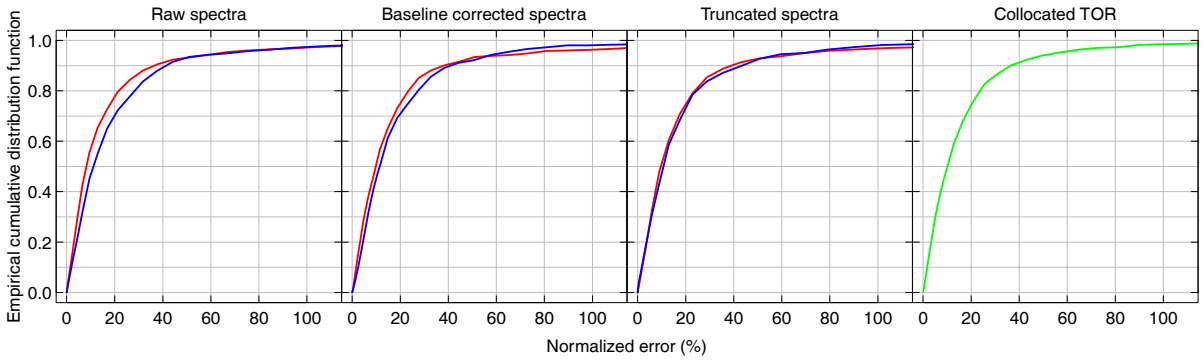

Figure S4. Distributions of normalized error for the base case and precision for collocated TOR samples. Calibration set is in red and the test set is in blue.

S5. Blank filters in calibration

The number of blanks in the calibration was varied from zero to 36 to evaluate their impact on MDL. The samples and blanks in the test set and the samples in the calibration set remained constant for each case. The MDL does not correlate with the number of blanks in the calibration set.

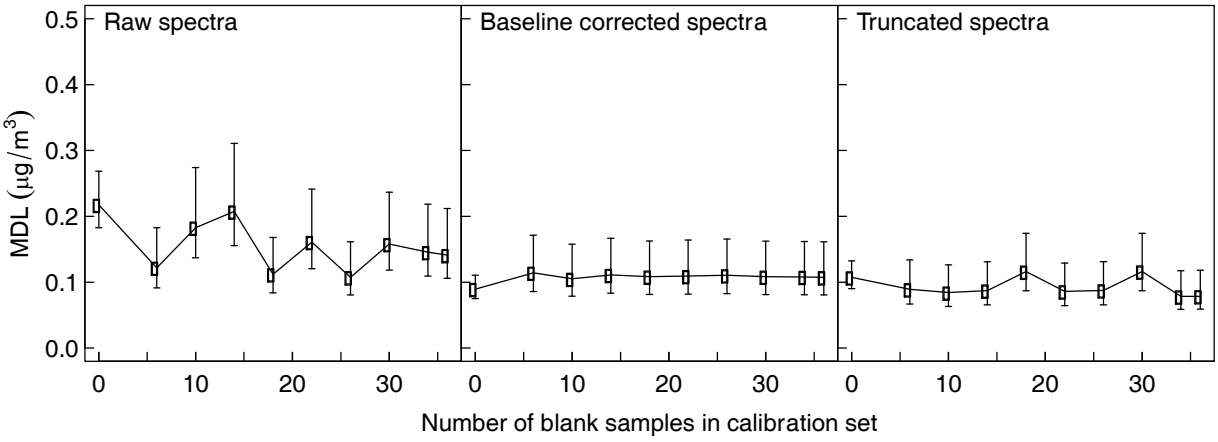

Figure S5. MDL calculated from the same 16 blank filters in the test set with the number of blanks in the calibration set ranging from zero to 36. Ambient samples in the calibration and test set are the same for each case. The

S6. OC/EC

In addition to the OC, OM/OC and ammonium/OC calibrations reported on in the paper, OC/EC calibrations are developed. Values of OC/EC have been used as an indicator of primary and secondary aerosol and so this parameter is evaluated for its role in the quality of the prediction of OC. TOR EC is reported by the IMPROVE network. Figure S5 shows calibrations based on samples ordered by OC/EC for uniform and non-uniform cases. Figure S6 shows the results for each site.

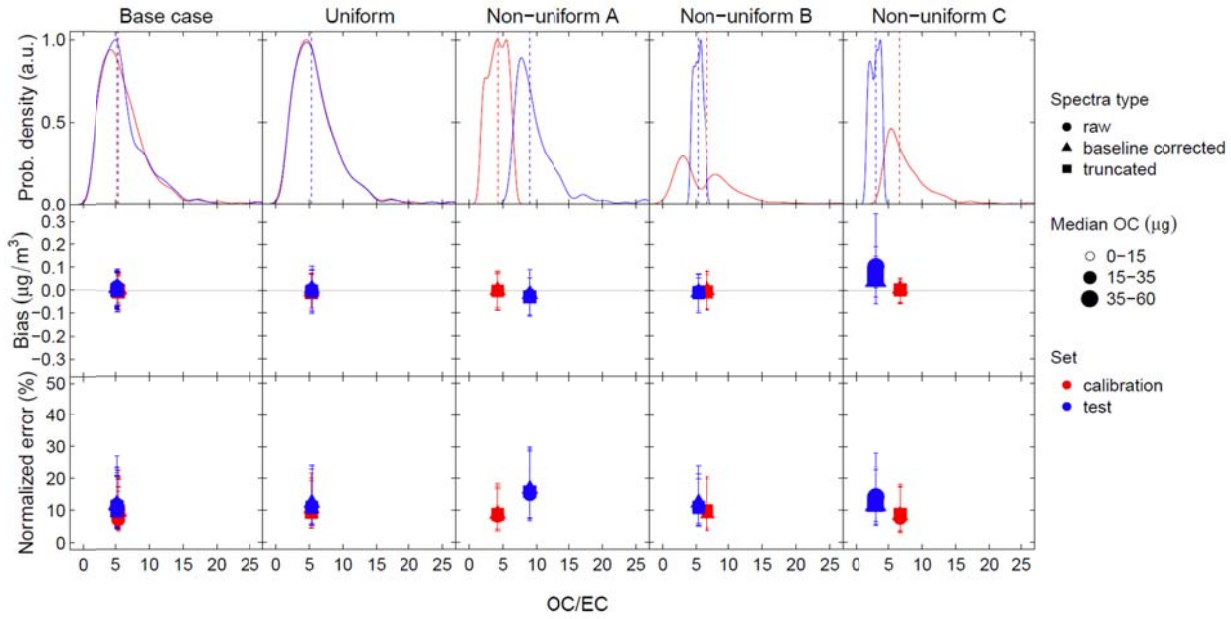

Figure S6. The probability density distribution of OC/EC and bias and normalized error (with the interquartile range shown by error bars) in the calibration (red) and test (blue) sets for five calibration cases: the Base case, the Uniform OC case and three Non-uniform OC/EC cases. Vertical lines are the median of the OC/EC distributions.

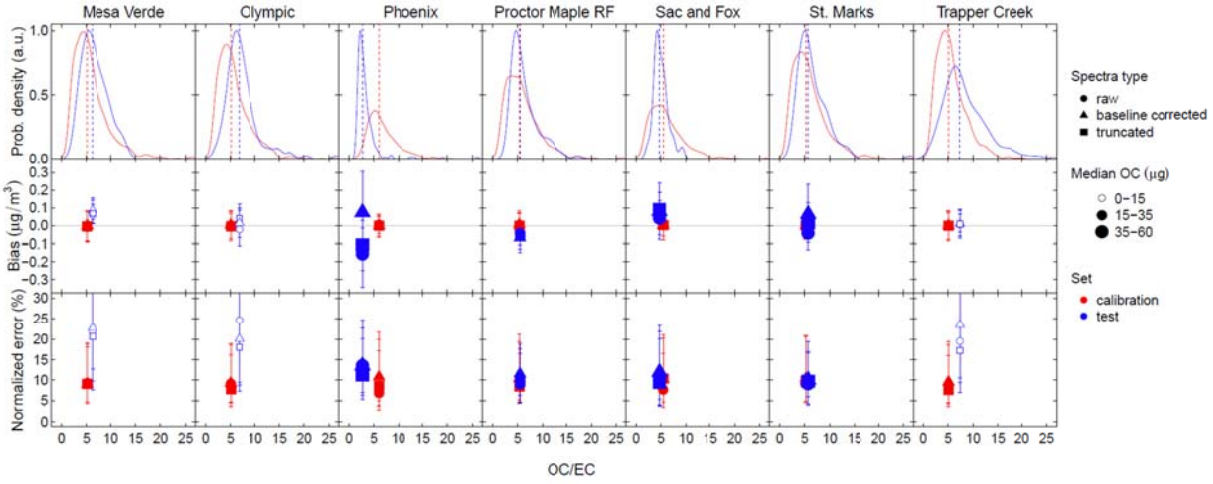

Figure S7. The distribution of OC/EC in the calibration (red) and test (blue) sets, and the bias and normalized error (with the interquartile range shown by error bars) for calibrations developed for each site. Each calibration has all samples in the calibration set except for the site to be predicted. Vertical lines are the median of the OC/EC distributions.

S7. Probability density functions of OC, OM/OC, OC/EC and Ammonium/OC for the test and calibration sets for all calibrations developed.

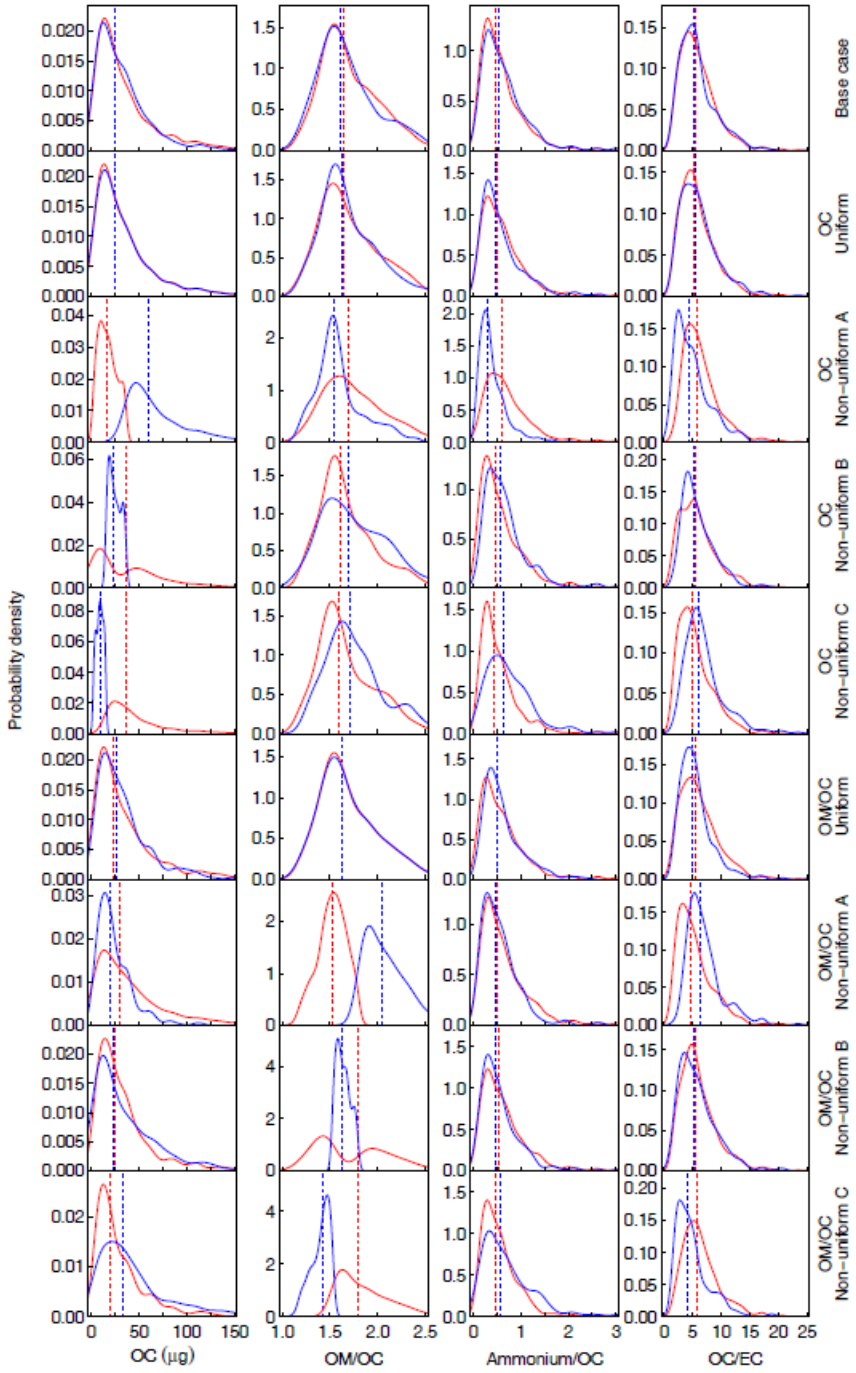

Figure S8. Distribution of OC, OM/OC, Ammonium/OC and OCEC for various calibrations (indicated on right-hand y-axis).

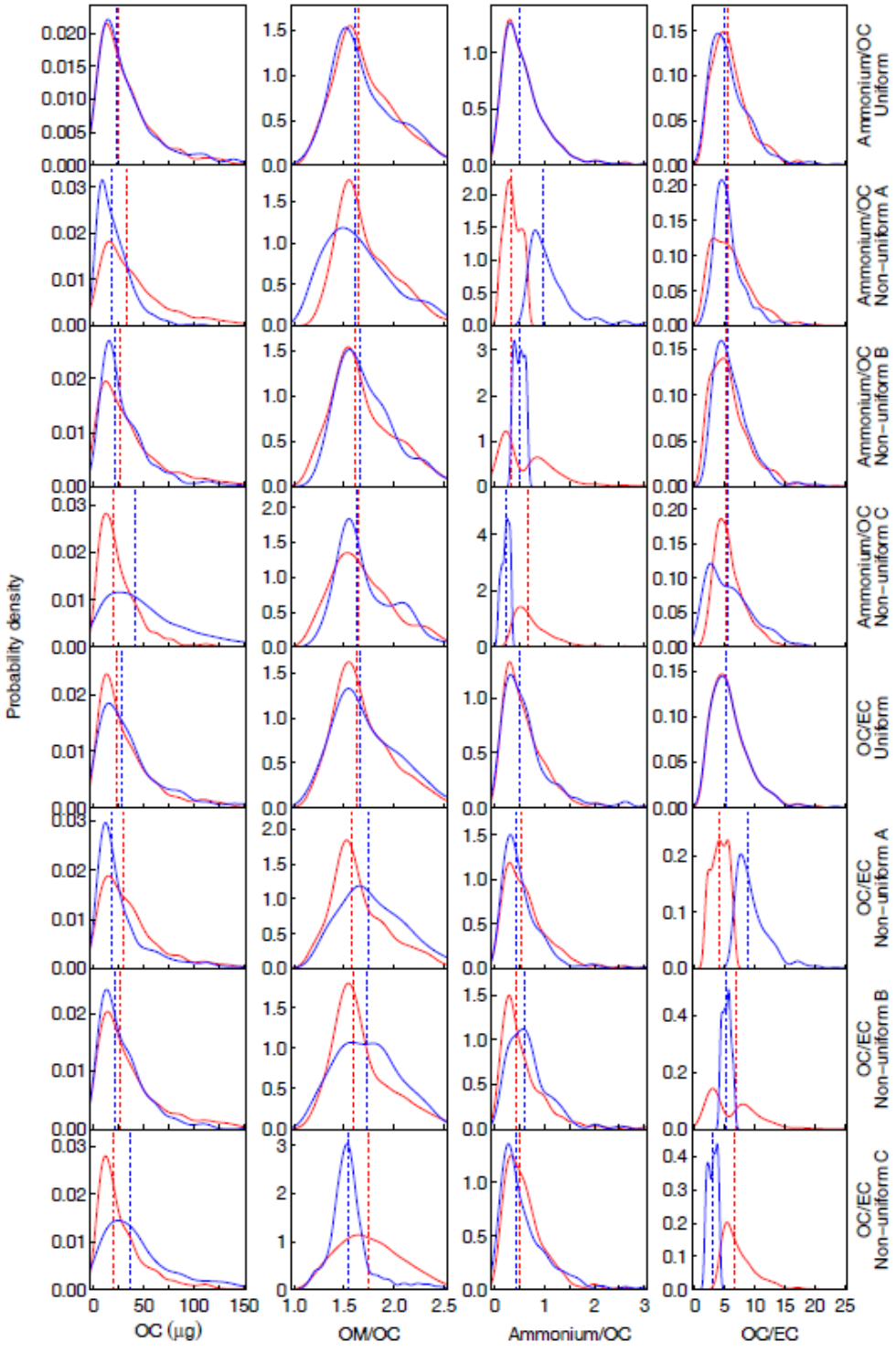

Figure S9. Distribution of OC, OM/OC, Ammonium/OC and OCEC for various calibrations (indicated on right-hand y-axis).

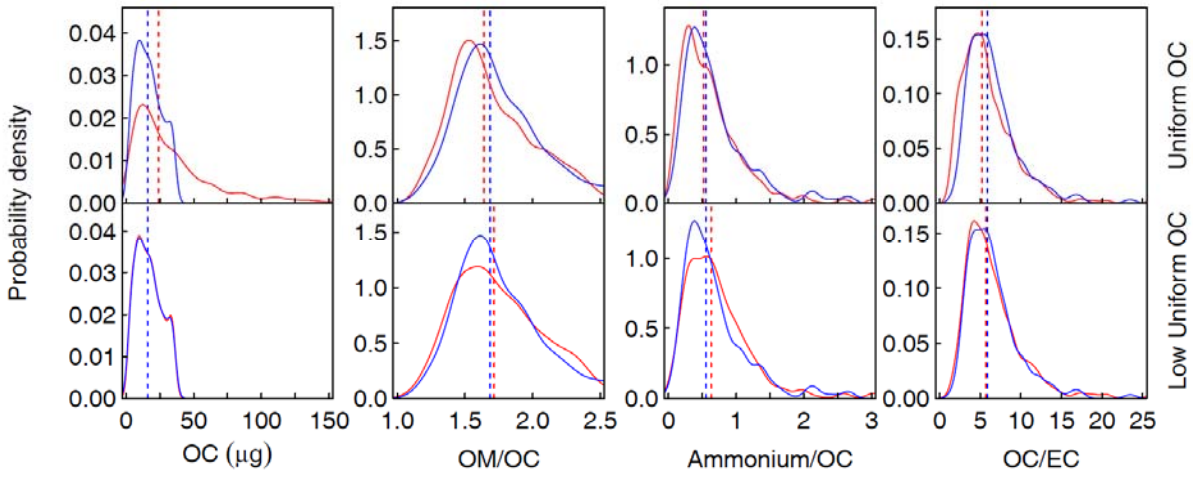

Figure S10. Distribution of OC, OM/OC, Ammonium/OC and OCEC for Uniform OC and Low Uniform OC (indicated on right-hand y-axis).

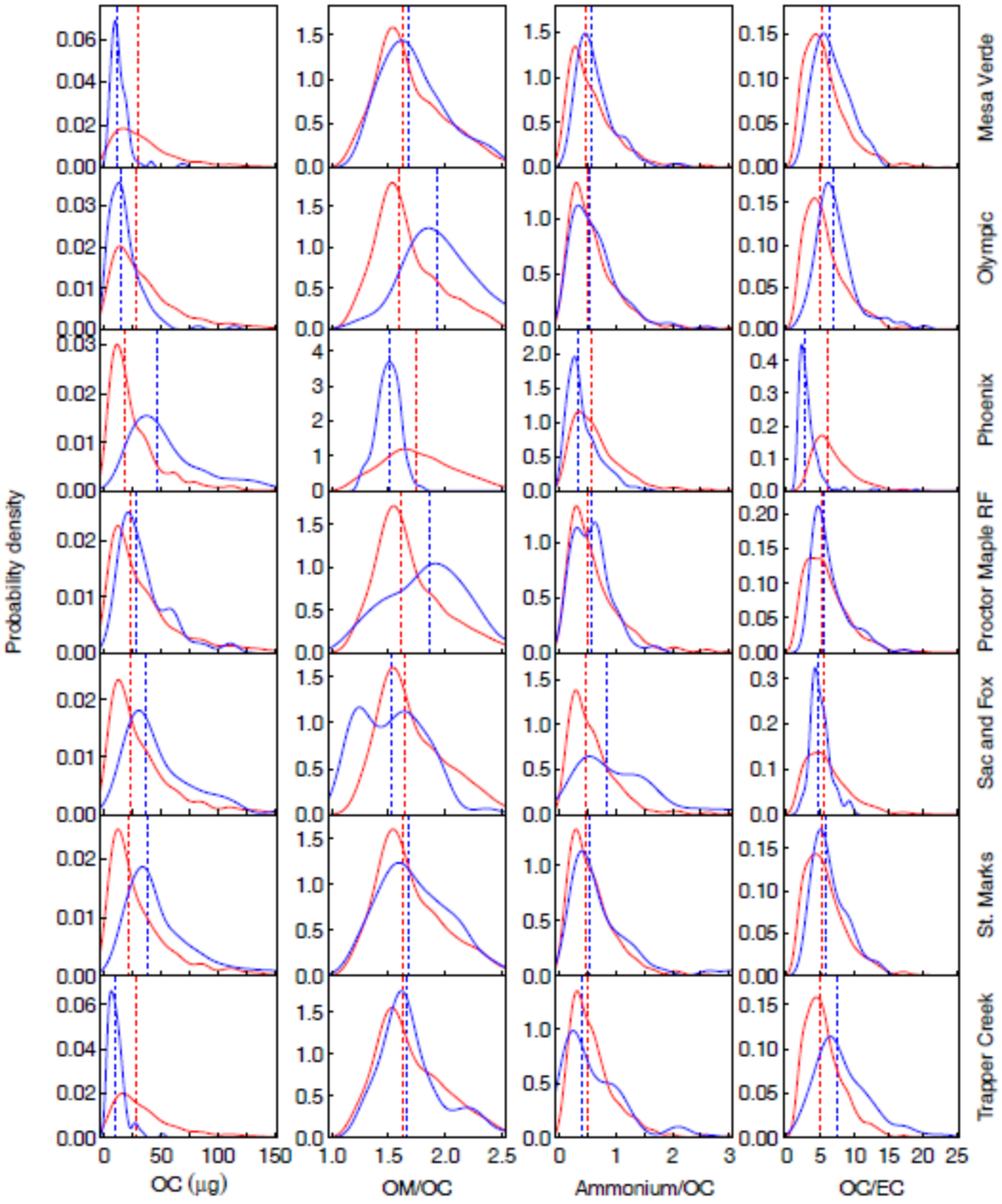

Figure S11. Distribution of OC, OM/OC, Ammonium/OC and OCEC for each site (indicated on right-hand y-axis).

S8. Residuals

PLSR, as with many other regression models, assumes that the reference measurement contains no or relatively small error, and that the magnitude of errors are the same across samples (homoscedastic) and furthermore uncorrelated. While violation of these assumptions preclude drawing robust conclusions for statistical inference, application of PLSR for multivariate calibration and prediction is common when errors in both *x* and *y* variables are present, and in heteroscedastic form (Martens and Geladi, 2004). As the basis for our calibration are analytical measurements, heteroscedastic errors which scale with concentration can be expected. Faber and Kowalski (Faber and Kowalski, 1997) present theoretical expressions for prediction intervals due to heteroscedastic errors, while Schreyer et al. (2002) and Khayamian (2007) empirically find that predictions from PLSR seem to be robust with respect to such errors.

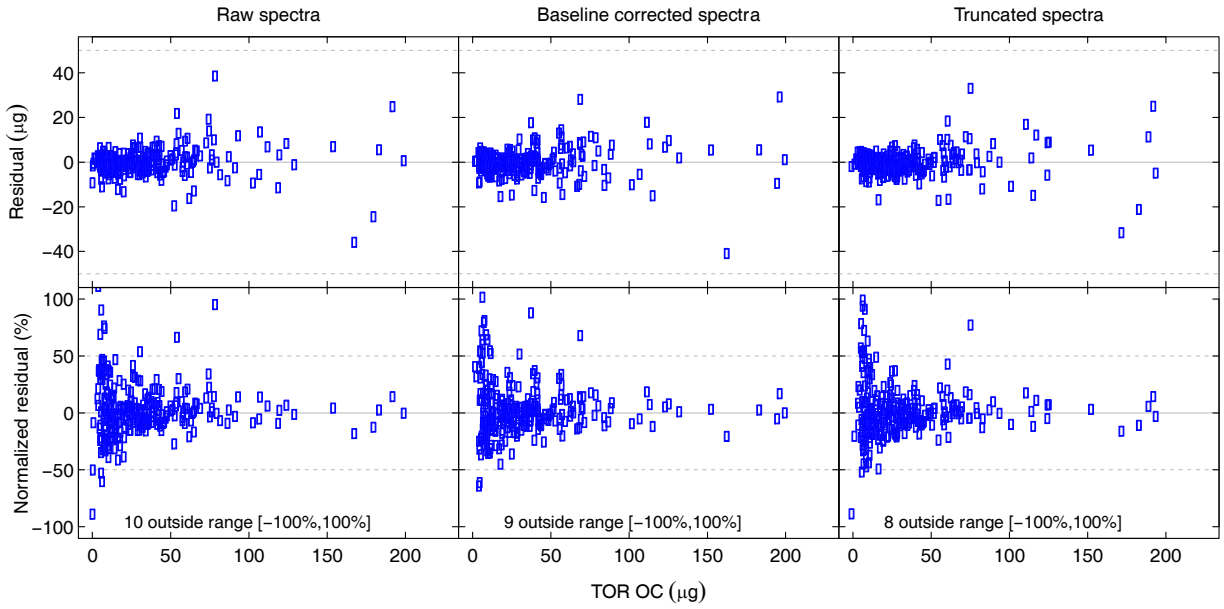

Figure S12. Residuals for base case test set predictions using raw, baseline corrected, and truncated spectra.

We examine how such errors are manifested in residuals from the PLSR regression for our study for the base case (Fig. S1s). The main body of the manuscript focuses on central tendencies of the residual distribution by the metrics of median bias and error, but we consider additional aspects of the residual population in this section. We observe that the residuals are indeed heteroscedastic, increasing with

higher concentrations of observed and predicted OC. This structure can lead to residuals which are approximately symmetric but long-tailed with respect to a normal distribution. Possible strategies to meet model assumptions of residual normality include variable scaling, smoothing, wavelength selection, and local (i.e., piecewise) regression (Geladi et al., 1999). To work directly with OC concentrations in measured units, we have avoided variable scaling, but examine the residuals and corresponding prediction errors when a subset of wavelengths are used (as described in Section 2.2), or the regression is localized over a smaller range of concentrations (Section 3.4). Comparing residuals for the Uniform OC and Low Uniform OC calibration and test samples, we find that the residuals are approximately symmetric for all spectra types and concentration regimes (Figure S13). Normality of each residual distribution is evaluated by examining the *p*-value of a chi-square test. Residuals for the Uniform OC case are generally long-tailed with respect to a normal distribution (except for the calibration set predicted by the truncated spectra) and approximately normal for most of the Low Uniform OC sets (except for test set predictions with the baseline corrected spectra), but largely due to the range in concentrations addressed by the calibration model rather than inherent differences in its development. If the Uniform OC calibration model is evaluated only for the Low Uniform OC test samples, the residuals are normally distributed for all spectra types (*p*-values range between 0.2 and 0.9 for the three spectra types). Regardless of the normality (or lack thereof) in the residual distribution, we find no major differences in the mean and dispersion of prediction errors.

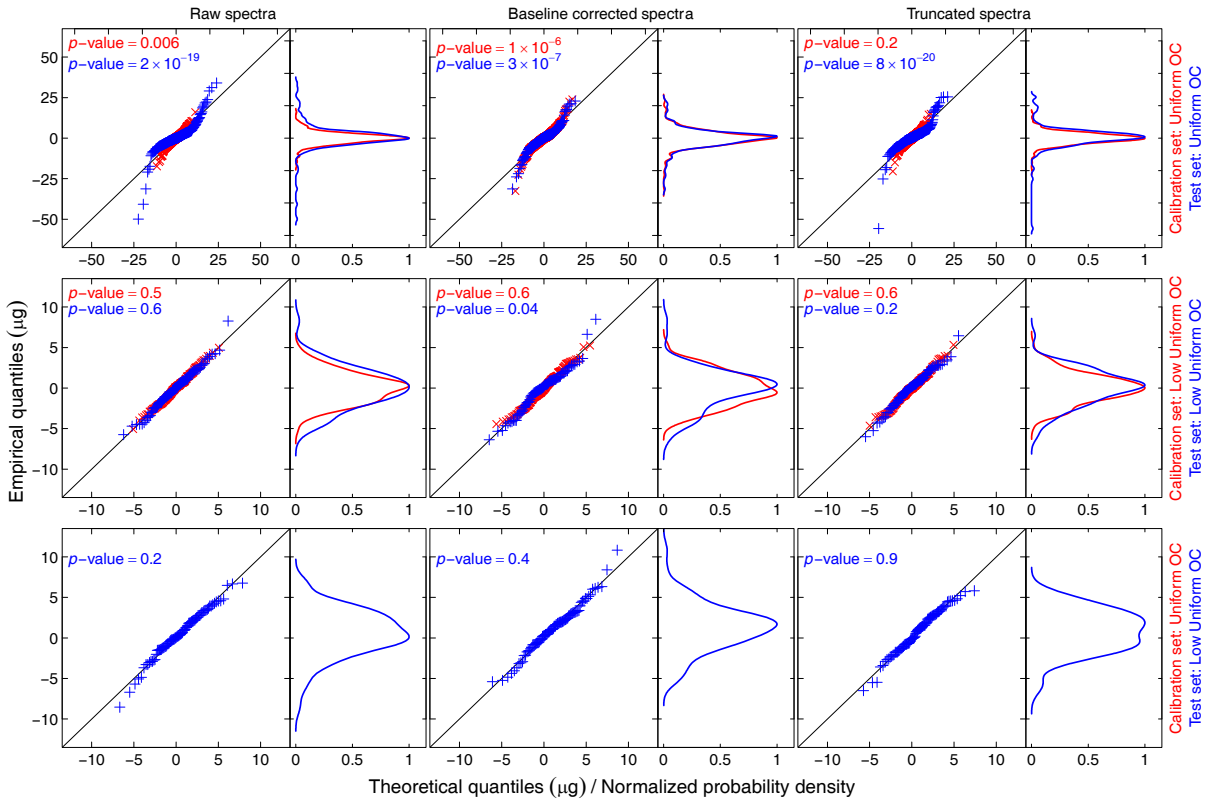

Figure S13. Density distributions of regression residuals, and corresponding comparison of empirical and theoretical quantiles of a normal distribution. Red symbols correspond to calibration set samples and blue symbols correspond to test set samples. *P*-values are calculated from the chi-squared test for normality. Residuals for the calibration set a in the last row are identical to those in the first row and are therefore not shown.

References

Faber, K., and Kowalski, B. R.: Propagation of measurement errors for the validation of predictions obtained by principal component regression and partial least squares, Journal of Chemometrics, 11, 181-238, 10.1002/(SICI)1099-128X(199705)11:3<181::AID-CEM459>3.0.CO;2-7, 1997.

Geladi, P., Hadjiiski, L., and Hopke, P.: Multiple regression for environmental data: nonlinearities and prediction bias, Chemometrics and Intelligent Laboratory Systems, 47, 165-173, 10.1016/S0169-7439(98)00204-4, 1999.

Khayamian, T.: Robustness of PARAFAC and N-PLS regression models in relation to homoscedastic and heteroscedastic noise, Chemometrics and Intelligent Laboratory Systems, 88, 35-40, 10.1016/j.chemolab.2006.10.006, 2007.

Martens, H., and Geladi, P.: Multivariate Calibration, in: Encyclopedia of Statistical Sciences, John Wiley & Sons, Inc., 2004.

Schreyer, S. K., Bidinosti, M., and Wentzell, P. D.: Application of Maximum Likelihood Principal Components Regression to Fluorescence Emission Spectra, Applied Spectroscopy, 56, 789-796, 10.1366/000370202760076857, 2002.