

BASELINE REMOVAL IN LIBS AND FTIR SPECTROSCOPY: OPTIMIZATION TECHNIQUES. S. Giguere¹, C. Carey¹, M. D. Dyar², T. F. Boucher¹, M. Parente³, T. J. Tague, Jr.⁴, and S. Mahadevan¹. ¹School of Computer Science, Univ. of Massachusetts, Amherst, MA 01003, (sgiguere@cs.umass.edu), ²Mount Holyoke College, Dept. of Astronomy, South Hadley, MA 01075, ³Dept. of Electrical and Computer Engineering, Univ. of Massachusetts, Amherst MA 01003, ⁴Bruker Optics, Inc., 19 Fortune Dr., Billerica, MA 01821, USA.

Introduction: The task of proper baseline or continuum removal is common to nearly all types of spectroscopy. Its goal is to remove any portion of a signal that is irrelevant to features of interest while preserving any predictive information. Despite its importance, spectroscopists typically employ default parameters or use commercially-available software supplied with their instruments for the baseline removal task.

There are two current applications in planetary science where baseline removal may be critical to obtaining quantitative results. The first is laser-induced breakdown spectroscopy (LIBS), currently deployed as part of the ChemCam instrument on Mars Science Laboratory [1,2]. The LIBS signal used for measuring the chemical composition of surface materials on Mars is complicated by a Bremsstrahlung continuum and ion-electron recombination processes that are not relevant to the atomic emission signal of interest. The ChemCam team removes this continuum via decomposition into a set of cubic spline undecimated wavelet scales in which local minima or convex hulls are found. A spline function is then interpolated through the different minima [3]. The second important application with a need for careful baseline removal is Fourier transform infrared spectroscopy (FTIR) reflectance spectroscopy. For example, the Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) experienced challenges from spectral artifacts due to residual atmospheric contributions and detector-based effects as the flight hardware aged for which careful noise suppression techniques were developed [4,5].

Because the baseline signal in these two types of spectroscopy comes from very different phenomena, they provide a useful basis for comparison of baseline removal techniques. We here treat the issue of finding the best baseline removal procedure as an optimization problem in which the method used and its parameters are adjusted to maximize the predictive accuracy of the resulting models.

Background: Baseline removal techniques fall into three general categories: methods that smooth spectra, those that fit a function to the baseline, and models that project the spectra onto a basis that captures the structure of the baseline.

Smoothing Methods are often single-pass techniques that output a smoothed version of the original spectrum [6,7]. As a result, these methods can be used

directly for baseline correction methods, although smoothing is often used as a sub-procedure in more complex correction methods.

Fitting Methods optimize a loss function to learn the baseline. Because the baseline is either always below or above the peaks in the spectrum, these methods account for this asymmetry by considering points on the peaks as less important than points on the baseline. One approach uses an asymmetric weighting scheme, as in Asymmetric Least Squares (ALS) [8]; others use peak detection to assign a weight to each band that controls how much that band's intensity value influences the fit. Due to error in peak detection, methods using the fitting approach tend to rely on iteration, where estimates in an iteration are improved by building on results from prior iterations.

Projection Methods compute a baseline by projecting the original spectrum onto a basis that captures the structure of the baseline but not the peaks. Basis functions may be learned in an unsupervised way (Empirical Mode Decomposition, or EMD, and Ensemble-EMD, or EEMD) [9], chosen beforehand ([10], which uses P-splines), or computed using frequency-analyzing transformations (Orthogonal Basis uses SVD, wavelet methods, etc.) [11].

In this project, we evaluate three algorithms based on Whittaker smoothing (ALS, airPLS, and FABC), two that include iterative thresholding based on mean and variance (FABC, Dietrich), and two that involve fitting polynomial functions (Kajfosz-Kwiatk, Parente), as described in [4] and listed in **Table 1**.

Samples: This study uses two data sets for the optimization problem. The LIBS data set includes 93 randomly-chosen geological samples for which spectra

Table 1. Baseline Correction Methods Tested

Name	Acronym	AP	Ref.
Asymmetric Least Squares	ALS	2	[13]
Adaptive Iteratively Re-weighted Penalized Least Squares	airPLS	1	[8]
Fully Automatic Baseline Correction	FABC	2	[14]
Iterative Thresholding	IT	2	[15]
Kajfosz-Kwiatk	K-K	2	[16]
Hermite polynomial	HP	1	[4,5]

AP – number of adjustable parameters

were acquired at six locations with six laser shots each in the LIBS lab at Mount Holyoke. The FTIR reflectance data include 18,880 terrestrial soil samples available from the U.S. Department of Agriculture's Natural Resources Conservation Service (NRCS) at (<http://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm>). We arbitrarily chose variables to test prediction accuracy of the various baseline removal algorithms: nine major elements in wt.% oxide for LIBS, which ranges from 0 to 100, and carbon for FTIR, reported in wt% NCS, which ranges from 0 to 57.34.

Computational Methods: To quantify the effectiveness of the baseline removal methods in Table 1, we compared the root-mean-square error (RMSE) and relative-RMSE (the ratio of RMSE to response mean) of the calibration curves produced by each method. For this, we created a 4-step *pipeline*: (1) the raw LIBS or FTIR data are input, (2) a baseline correction method is randomly selected with random parameter values (e.g., the smoothness parameter in Whittaker smoothing methods) drawn uniformly from appropriate ranges, (3) cross validation (CV) is performed with the processed spectra and a 10-component partial least squares (PLS) model, and (4) the mean RMSE across CV-folds and the chosen method and parameters are recorded. After running the pipeline repeatedly for 24 hours, the lowest RMSE for each method is reported.

Table 2. Comparison of Average Oxide Results (LIBS)

	RMSE	Relative RMSE	N*
ALS	1.961	0.1840	21748
airPLS	2.042	0.1916	9571
FABC	1.917	0.1797	4009
IT	1.897	0.1779	23016
K-K	2.148	0.2015	1371
HP	2.144	0.2010	11
No Correction	2.040	0.1913	-

*N = the number of hits to each method.

LIBS Results: The LIBS results given in Table 2 illustrate that while baseline correction often improves model accuracy, the wrong choice of correction method or parameters can degrade performance. Reductions in mean square errors relative to uncorrected data were observed after applying the FABC and IT methods, while applying ALS or airPLS had little effect on model accuracy. Both the K-K and HP methods increased the RMSE. As a caveat, we note that due to computational issues, few parameters were tested for the HP method, possibly explaining its poor performance.

FTIR Results: Table 3 contains the performance for each correction method applied to FTIR spectra. In contrast to the LIBS results, the best performing correction method was ALS. Improvements were observed after applying each method except for HP and FABC, which had little effect on RMSE. These results

suggest that ALS is good candidate for use with FTIR spectra. It is unsurprising that LIBS and FTIR, in which fundamentally dissimilar mechanisms give rise to baselines, should behave differently as a function of baseline removal algorithm. We have yet to evaluate if different baseline removal protocols would be favored for different spectral regions or variables within the same types of spectra, i.e. the H peak in FTIR or minor elements in LIBS.

Table 3. Comparison of Carbon Results (FTIR)

	RMSE	Relative RMSE	N*
ALS	2.9119	0.8543	1541
airPLS	3.2681	0.9588	4576
FABC	3.3721	0.9893	913
IT	3.3336	0.9780	6747
K-K	3.0844	0.9049	137
HP	4.1861	1.2281	6
No Correction	3.3712	0.9890	-

*N = the number of hits to each method.

Future Work: We have tested only six algorithms covering only the fitting/smoothing methods for baseline removal. Future comparisons will include projection-based methods, such as orthogonal basis (OB) methods [11] and Fuzzy Optimal Associative Memory (FOAM) techniques [18]. Another promising direction is to examine non-quadratic loss functions when using least squares based methods [19]. We will continue to implement other techniques and optimize them for specific applications. It is possible not only that each type of spectroscopy will have a preferred correction method, but that specific methods will be best suited to particular applications as well. As computing power continues to expand exponentially, it is feasible that future baseline removal will involve on-the-fly optimization of the types tested here.

Acknowledgments: Research supported by NASA MFR program grant NNX15AC82G.

References: [1] Wiens R. C. et al. (2012) *Space Sci. Revs.*, 170, 167-227. [2] Maurice S. et al. (2012) *Space Sci. Rev.*, 170, 95-166. [3] Wiens R. C. (2013) *Spectrochim. Acta B*, 82, 1-27. [4] Parente M. (2008) *LPSC XXXIX*, Abstract #2528. [5] Parente M. (2010) Ph.D. thesis, Stanford Univ. [6] Ruchstuhl A. F. et al. (2012) *Atmos. Meas. Tech.*, 5, 2613-2624. [7] Bao Q. et al. (2012) *J. Mag. Reson.*, 218, 35-43. [8] Zhang Z.-M. et al. (2010) *Analyst*, 135, 1138-1146. [9] Mariyappa N. et al. (2014) *Med. Eng. & Phys.*, 36, 1266-1276. [10] de Rooi J. J. et al. (2013) *Chemom. Intell. Lab. Systems*, 117, 56-60. [11] Wang Z. et al. (2014) *Anal. Chem.*, 86, 9050-9057. [12] Carey C. et al. (2015) this meeting. [13] Eilers P. H. C. and Boelens H. F. M. (2005) Leiden Univ. Med. Centre Rep. <http://www.science.uva.nl/~hboelens/>. [14] Cobas J. C. et al. (2006) *J. Mag. Reson.*, 183, 145-151. [15] Dietrich W. et al. (1991) *J. Mag. Reson.*, 91, 1-11. [16] Kajfosz J. and Kwiatek W. M. (1987) *Nucl. Instrum. Methods Phys. Res.*, 22, 78-81. [17] Dyar M. D. (2015) this meeting. [18] Harrington, P. d. B. (2014) *Anal. Chem.*, 86, 4883-4892. [19] Mazet V. et al. (2005) *Chemom. Intell. Lab. Systems*, 76, 121-133.