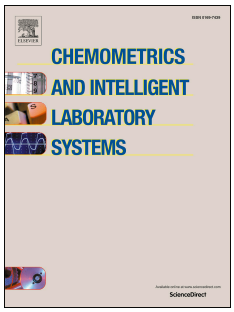
Accepted Manuscript



Leveraging multiple linear regression for wavelength selection

Tony Lemos, John H. Kalivas

PII: S0169-7439(17)30373-8

DOI: 10.1016/j.chemolab.2017.07.011

Reference: CHEMOM 3474

To appear in: *Chemometrics and Intelligent Laboratory Systems*

Received Date: 31 May 2017

Revised Date: 18 July 2017

Accepted Date: 24 July 2017

Please cite this article as: T. Lemos, J.H. Kalivas, Leveraging multiple linear regression for wavelength selection, *Chemometrics and Intelligent Laboratory Systems* (2017), doi: 10.1016/j.chemolab.2017.07.011.

Leveraging Multiple Linear Regression for Wavelength Selection

Tony Lemos[a], John H. Kalivas[a*]

[a]*Department of Chemistry, Idaho State University, Pocatello, ID 83209, USA*

*Corresponding Author.
Email Address: kalijohn@isu.edu

ABSTRACT

Wavelength selection is often used for multivariate calibration methods to lower prediction error for the calibrated sample properties. As a result, there are a plethora of wavelength selection methods to select from; all with unique advantages and disadvantages. All wavelength selection methods involve a range of tuning parameters making the methods cumbersome or complex and hence, difficult to work with. The goal of this study is to provide a simple process to select wavelengths for multivariate calibration methods while trying to standardize values of the five adjustable algorithm tuning parameters across data sets. The proposed method uses multiple linear regression (MLR) as an indicator to which wavelengths should be used further to form a multivariate calibration model by some processes such as partial least squares (PLS). From a collection of MLR models formed from randomly selected wavelengths, those models within a thresholds of the bias indicator root mean square error of calibration (RMSEC) and variance indicator model vector L_2 norm are evaluated to ascertain the most frequently selected wavelengths. Portions of the most frequent wavelengths are retained and used to produce a calibration model by PLS. This proposed wavelength selection method is compared to PLS models based on full spectra. Several near infrared data sets are evaluated showing that PLS models based on MLR selected wavelengths provide improved prediction errors. Of the five