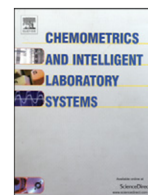
Chemometrics and Intelligent Laboratory Systems 118 (2012) 239–245

Contents lists available at SciVerse ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab



An efficient method for signal improvement in near-infrared spectroscopy measurements during calibration and validation processes

David Abookasis [a,*], Jerome J. Workman [b,c,d]

[a] Department of Electrical and Electronics Engineering, Ariel University Center of Samaria, Ariel 44837, Israel
[b] Unity Scientific, Brookfield, CT 06804, USA
[c] Liberty University, Lynchburg, VA 24502, USA
[d] National University, La Jolla, CA 92037, USA

ARTICLE INFO

Article history:
Received 4 April 2012
Received in revised form 31 May 2012
Accepted 31 May 2012
Available online 7 June 2012

Keywords:
Spectroscopy
PLS analysis
Spectra binning
Correlation
Fourier transform spectroscopy

ABSTRACT

Standard spectroscopic practice involves the measurements of replicate spectra (designated as s) for each sample where these replicates are co-added in order to reduce random noise by a factor of $\sqrt{s}$. However, when systematic or structured noise are present, due to instrument or sample upset conditions, or subject motion, this practice tends to degrade and distort spectral bands. When co-adding multiple replicate measurements for single samples, such distortion tends to cause biased calibration coefficients and larger prediction errors, resulting in loss of analytical accuracy. In this work a simple automated procedure is presented and aimed to circumvent the above mentioned concern. Multiple replicate spectra of a sample are correlated with the median of the entire set of replicate spectra and then ranked by similarity based on the correlation of each spectrum to this 'reference' median spectrum. A tunable 'binning size' parameter is chosen by dividing the set of ranked, correlated replicate spectra into sub-spectral groups. The highest correlation spectra then co-added with the median to yield what is termed here as a single 'ideal' spectrum. These steps are repeated for each set of sample measurements and performed for both calibration and validation data sets before modeling or prediction. Results from experiments show a substantial decrease in both standard errors of prediction and bias in comparison to the classical replicate spectra co-averaging approach highlights the effectiveness of the proposed method.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Near-Infrared spectroscopy (NIRS) is a well-known and established technique widely applied in both noninvasive and invasive modes to measure the biochemical composition and morphological properties of different biological samples, such as urine, serum, and whole blood glucose [1–4]. Other NIRS application used for pathology purposes including measurement of tissues, such as cancer, brain, skin, prostate, etc. have also been reported [5–8]. NIRS combined with a multivariate regression analysis approach, such as partial least-squares (PLS) regression [9,10], can produce precise and informative diagnostic information [11–14]. Consequentially, various NIRS modalities and advanced processing techniques have been developed by multiple scientific research groups over the years, each with its own merits and limitations [2,15,16]. One of the popular forms of NIRS is based on a Michelson interferometer known as a Fourier transform spectrometer (FTS) [17–19]. In order to minimize noise during sample measurements occurring from scan-to-scan by the moveable mirror in FTS, multiple single-scans spectra are summed (i.e., co-added) together into a single averaged spectrum. In this technique the noise can be suppressed in proportion to the square root of the number of replicate spectra. Since this simple summation operation constitutes an ensemble averaged over many different scans, having both "poor" quality and "good" quality spectral measurements (see an example in Fig. 1), an artifact spectral shape can be created, having no basis in the true spectral characteristics of real samples. Therefore, it is essential to reduce or eliminate the effect of such conditions in order to build the optimum predictive model. In addition, in some applications the scanning time is a critical issue, and prolonged scanning times seriously impede the spectral quality.

In NIRS measurements we broadly categorize two basic types of noise: random (white) noise and systematic (system or subject) noise. During algorithm development two basic assumptions regarding the characteristic of measurement noise types were conducted. First, the random noise, which follows a $\sqrt{s}$ noise reduction phenomenon, is uncorrelated to analyte signal (where s is the number of replicate scans or spectral measurements co-added for each sample). Second, systematic variation in a measurement system or sample presentation (e.g., changes in temperature, flow, sample scattering, instrument vibration, etc.) can be correlated or uncorrelated to the analyte signal. Therefore, an improved signal-to-noise ratio can be obtained by reducing the impact of structured

* Corresponding author.
E-mail address: davida@ariel.ac.il (D. Abookasis).

0169-7439/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2012.05.016

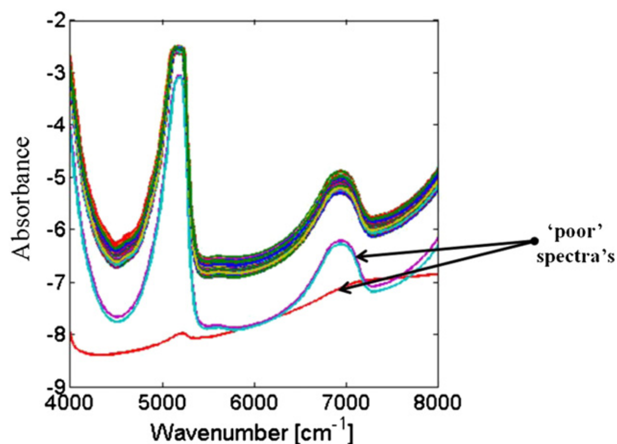


Fig. 1. Example of ~460 replicate Intensity spectra for a single sample, demonstrating a range of "poor" and "good" quality measurements. If one uses conventional co-averaging over the entire set of replicate scans, an artificial or synthetic spectral shape is created having no basis in true spectral characteristics of the sample.

noise by using a method to select out "poor" quality spectra, while selecting "good" quality spectra for co-addition. One such statistically appropriate method to accomplish this is to compute the statistical median spectrum from a set of replicate scans; to select those replicate spectra closest in feature shape to the median spectrum for each sample; and then co-add these selected replicates into the final measurement spectrum. Different methods have been suggested by several research groups to cope with the influence of systematic and structured noise during spectroscopic measurements [20–28]. A common feature of these methods is the designing appropriate mathematical filters and the use of pre-processing techniques to reduce the noise variance structure prior to replicate scan averaging. However, even when these methods are applied successfully, low signal-to-noise signals can still seriously affect the final analytical performance. We note here that the proper de-noising of signals obtained from replicate scans is essential.

In this work, we report on a new preprocessing hybrid algorithm which is applied at the single scan-level, referred to as Ideal Spectrum Adaptive Filtering (ISAF), which was created to circumvent the above mentioned structured noise concerns. The proposed hybrid algorithm makes use of the combination of spectral correlation, spectral binning, and "Lucky Imaging" concepts developed for astronomy [29–31]. In lucky imaging, a technique which has revolutionized astronomical imaging using land based telescopes, short-exposure photographs least affected by telescope temperature and atmospheric refraction variations are shifted to a common center point on an image plane, and then co-added into a single average image [32–34] based on signal-to-noise characteristics; the best images are selected from a series of images and then co-added to result in much higher resolution and quality images. Land based telescopes have been improved by three orders of magnitude in image quality. The ISAF algorithm is intended to mimic "Lucky Imaging" and includes five main steps for each physical sample as follows: First, the median of the entire set of replicate spectra is calculated (this is accomplished as the median of each absorbance at each wavenumber). Second, a cross-correlation between each individual replicate spectrum compared to the calculated median is computed (each replicate has a correlation value as compared to the median spectrum). Third, individual replicate scans (spectra) are ranked by correlation to the median. Fourth, the ranked replicate sets are divided into sub-spectral groups based on a specific classification criterion. Finally, the highest correlation spectra of the first group are selected and co-added with the computed median to result in one averaged 'ideal' spectrum per sample. These steps are then repeated for each set of individual samples for the calibration, and prediction process. This ISAF approach was tested on blood sample spectra with a clinical range of physiologic blood glucose and hematocrit levels.

2. Materials and methods

2.1. Instrumentation

We use a Fourier transform near-infrared (FT-NIR) spectrometer work at the spectral range of 4000–8000 cm^{-1} with a resolution of 32 cm^{-1}. It consists of a uniform blackbody light source (~1035 °C), beam-splitter and compensator plates, fixed and moveable corner-cube retro-reflectors, and an InGaAs PIN photodiode detector. The active mirror is continuously scanning in rate of 7.67 Hz. The blood mixture and background (Saline, NaCl 0.9%) samples were pumped through a flow cell with a nominal pathlength of 1 mm, controlled at a set-point of 34 °C during measurements.

2.2. Procedures

Basic blood components of plasma and red blood cells were blended to produce multiple sample matrices to provide a diverse spectral population of samples. Each blood sample containing native glucose and a specific hematocrit (HCT) level is spiked with a certain glucose concentration, using the same volume dilution, at least 10 minutes before the NIR measurement. This sample is then gently mixed during that time to ensure uniform dispersion of glucose and RBCs within the sample. Approximately 5.0 mL of blood sample is placed in a 30 mL conical tube in which 4.5 mL is used to complete a draw for each NIR measurement and another 0.5 mL is used for measuring HCT. Reference hematocrit values are measured using the HemoCue glucose analyzer, while whole blood and plasma glucose values on two identical YSI glucose analyzers. The glucose values collected from the YSI and HemoCue are then analyzed and compared to spectra and predicted concentration values (mg/dL) obtained from the FT-NIR instruments.

2.3. NIR measurements

Background spectra of saline solution were collected before and after blood sample measurement for 30 seconds to account for variation in instrument noise and drift profiles. On the other hand, blood measurements were obtained for 60 seconds. Glucose samples were in the physiologic concentration range 30–500 mg/dL with varied hematocrit level range from 24% to 48%. These samples were measured using variations in temperature, flow rate and flowcell pathlength normally experienced during routine measurement operations, as outlined in Table 1. Therefore, an increased number of outliers were expected in addition to those generated from instrument measurement variation, and variation in the sample composition. The sample set was represented by 579 samples at nominal conditions; 162 samples at flow rate variations (equally divided as minimum and maximum); 162 samples at pathlength variation, and 157 samples at varied temperature ranges. Approximately one third of the samples were randomly selected and measured with each of the three instruments (~353 each). This experimental design represented the greatest possible variation in sampling and measurement effects which still remained within the measurement system

Table 1
Background variation obtained during experiments.

Type of variation	Value
Low flow	0.45 ml/min
Nominal flow	0.50 ml/min
High flow	0.55 ml/min
Small pathlength	0.95 mm
Nominal pathlength	1.00 mm
Large pathlength	1.05 mm
Low temperature	15 °C
Nominal temperature	25 °C
High temperature	35 °C

Download English Version:

https://daneshyari.com/en/article/7563665

Download Persian Version:

https://daneshyari.com/article/7563665

Daneshyari.com