Analytica Chimica Acta 681 (2010) 63–70

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Wavelet analysis used for spectral background removal in the determination of glucose from near-infrared single-beam spectra

Boyong Wan [1], Gary W. Small *

Department of Chemistry & Optical Science and Technology Center, University of Iowa, Iowa City, IA 52242, United States

ARTICLE INFO

Article history:
Received 26 April 2010
Received in revised form 23 August 2010
Accepted 14 September 2010
Available online 22 September 2010

Keywords:
Near-infrared
Wavelet analysis
Multivariate calibration
Glucose
Partial least-squares

ABSTRACT

Wavelet analysis is developed as a preprocessing tool for use in removing background information from near-infrared (near-IR) single-beam spectra before the construction of multivariate calibration models. Three data sets collected with three different near-IR spectrometers are investigated that involve the determination of physiological levels of glucose (1–30 mM) in a simulated biological matrix containing alanine, ascorbate, lactate, triacetin, and urea in phosphate buffer. A factorial design is employed to optimize the specific wavelet function used and the level of decomposition applied, in addition to the spectral range and number of latent variables associated with a partial least-squares calibration model. The prediction performance of the computed models is studied with separate data acquired after the collection of the calibration spectra. This evaluation includes one data set collected over a period of more than 6 months. Preprocessing with wavelet analysis is also compared to the calculation of second-derivative spectra. Over the three data sets evaluated, wavelet analysis is observed to produce better-performing calibration models, with improvements in concentration predictions on the order of 30% being realized relative to models based on either second-derivative spectra or spectra preprocessed with simple additive and multiplicative scaling correction. This methodology allows the construction of stable calibrations directly with single-beam spectra, thereby eliminating the need for the collection of a separate background or reference spectrum.

1. Introduction

Near-infrared (near-IR) spectroscopy has been studied for its potential use in the determination of clinically relevant analytes in biological matrixes [1–4], as well as in direct noninvasive measurements in tissue [5–7]. In combination with multivariate calibration techniques such as partial least-squares (PLS) regression, near-IR spectroscopy can provide quick quantitative determinations of species such as blood glucose with minimal sample preparation.

One of the main impediments to the practical use of this approach is the instability of multivariate calibration models over time. This instability of calibration can be due to changes in the instrumental response, instrumental drift, or sample-based variation. The existence of such varying background information will not only worsen the precision of concentration predictions, but also complicate the multivariate models, which can lead to models with many terms that may be over-fit to the calibration data.

Removal of background information from spectra is typically performed by taking the ratio to a suitable reference or blank spectrum, followed by additional preprocessing to remove any residual background effects (e.g., baseline variation). Collection of a closely matching reference spectrum is complicated in aqueous-based samples, however, by the extreme temperature sensitivity of the water absorbance. Even slight temperature mismatches between aqueous blank and sample spectra produce baseline artifacts when the ratio is taken. Furthermore, if the spectral measurement is made directly in tissue, there is no convenient way to obtain a good matrix-matched blank.

To overcome these complications, our laboratory has explored the development of near-infrared calibrations directly with the single-beam sample spectra. This approach requires no collection of a reference background. Through the use of suitable preprocessing strategies, good calibrations have been developed [8–10], although maintaining calibration stability over a long timeframe remains a challenge.

In the work reported here, wavelet analysis is explored as a way to remove background signatures from single-beam spectra and thereby allow stable calibration models to be formed. Wavelet analysis has found applications such as peak resolution, denoising, data compression, and baseline correction in analytical chemistry

* Corresponding author. Tel.: +1 319 335 3214; fax: +1 319 353 1115.
E-mail address: gary-small@uiowa.edu (G.W. Small).
[1] Present address: Bristol-Myers Squibb Co., 1 Squibb Dr., New Brunswick, NJ 08903, United States.

0003-2670/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2010.09.022

[11]. In near-IR spectroscopy, wavelet analysis has been used in conjunction with multivariate calibration for a number of applications [12–16], although not for use in building calibrations with single-beam spectra.

In this research, wavelet analysis is applied to near-IR single-beam spectra from three different instruments to suppress non-constant background information. Following this preprocessing step, PLS calibration models are developed. These models are evaluated for their predictive ability when applied to data collected shortly after the calibration, as well as data collected over a time span of approximately 6 months. The performance of the wavelet preprocessing step is compared to second-derivative preprocessing, another common background suppression technique.

2. Experimental

2.1. Instrumentation and data collection protocols

Near-infrared data from three different instruments were used in this work. The first instrument, termed instrument A, was a Digilab FTS-60A Fourier transform (FT) spectrometer (Varian, Inc., Randolph, MA). Instrument B was a Nicolet Nexus 670 FT spectrometer (Thermo Electron Corp., Madison, WI) and instrument C was a Thermo Nicolet 6700 FT spectrometer (Thermo Electron Corp.). Each instrument was equipped with a tungsten–halogen source, calcium fluoride beam splitter, and liquid nitrogen-cooled indium antimonide (InSb) detector. A K-band optical interference filter (Barr Associates, Westford, MA) was placed in the optical path to isolate the 5000–4000 cm^{-1} region.

For all instruments, samples were contained in the same demountable liquid transmission cell with a 20-mm diameter circular aperture (Wilmad-Labglass, Buena, NJ), sapphire windows (Meller Optics, Inc., Providence, RI), and a 1.5 mm path-length. Sample temperatures were controlled to the range of 36.9–37.1 °C by use of a refrigerated temperature bath (Model 9100, Fisher Scientific, Pittsburgh, PA) to circulate water through an integrated jacket in the sample cell. The temperature of the sample cell was monitored with a type-T thermocouple probe and a digital thermocouple meter (Omega Engineering, Inc., Stamford, CT). Due to the extreme temperature sensitivity of the background water absorbance, some temperature-induced variance was still present in the acquired spectral data.

In the data collection process, for each sample, three replicate interferograms were measured consecutively without removing the sample from the spectrometer. To minimize correlations between concentration and time during the measurement process, samples were run in a random order with respect to the component concentrations.

With spectrometers A and C, double-sided interferograms containing 8192 points and based on 256 co-added scans were sampled at every zero-crossing of the HeNe reference laser. With spectrometer B, double-sided interferograms containing 4096 points and based on 256 co-added scans were sampled at every two zero-crossings of the laser. All interferograms were Fourier processed with triangular apodization, Mertz phase correction, and one level of zero-filling to produce single beam spectra with a point spacing of about 1.9 cm^{-1}. The region of 5000–4000 cm^{-1} was extracted from all spectra and used for subsequent calculations.

2.2. Reagents

Two sets of samples were used in this work. In both cases, the sample matrix consisted of α-D-glucose (analyte), sodium L-lactate, urea, sodium ascorbate, β-alanine, and triacetin in phosphate buffer. This sample matrix did not mimic a specific biological sample, but rather included components whose structural types are commonly found in such samples. As shown in Fig. 1, these six components have highly overlapped absorption peaks in the near-IR combination region of 5000–4000 cm^{-1}, thereby providing a good assessment regarding the ability of a calibration methodology to extract the glucose signature from single-beam spectra corresponding to a complex sample matrix.

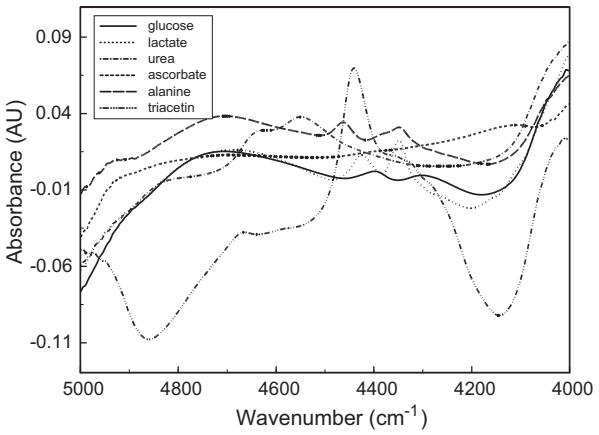

Fig. 1. Pure-component absorbance spectra of the six components obtained by taking the ratio of single-beam spectra of each constituent to a spectrum of phosphate buffer. The trace of glucose (solid line) corresponding to a concentration of 150 mM clearly shows the two C–H combination bands located at 4300 and 4400 cm^{-1} and the O–H combination band located at 4700 cm^{-1}. The other five components, lactate, urea, ascorbate, alanine and triacetin, were prepared at concentration levels of 400, 200, 60, 150 and 250 mM, respectively. Negative absorbance values arise due to differences in water concentration between the sample and reference spectra used in the absorbance calculation.

The first sample set consisted of 80 samples. The solvent used was a pH 6.86, 0.025 M phosphate buffer prepared with distilled water using 0.025 M KH_2PO_4 and 0.025 M Na_2HPO_4. As a preservative, 5-fluoruracil (0.1%, w/w) was added to the buffer. All samples were prepared by weighing appropriate amounts of α-D-glucose (ACS reagent, Aldrich Chemical Co., Inc., Milwaukee, WI), sodium L-lactate (Aldrich), urea (minimum 99.5%, Sigma Chemical Co., St. Louis, MO), sodium ascorbate (>99%, Aldrich), β-alanine (>99%, Aldrich), and triacetin (approximately 99%, Sigma), and diluting with the phosphate buffer to 50 mL.

The concentration of each component for each sample was randomly assigned across the 80 samples. The concentration ranges for glucose, lactate, urea, ascorbate, alanine and triacetin were 0.68–34.27, 1.18–33.34, 1.20–30.23, 1.18–22.55, 1.75–38.32 and 0.46–27.95 mM, respectively, and correlation coefficients computed between the component concentrations ranged from −0.18 to 0.11, confirming that each species was independent. Avoiding correlations among the component concentrations in the calibration samples is important to ensure that the calibration model is based on the unique spectral signature of the analyte relative to the background matrix. This requirement holds for the samples used to compute the calibration model but does not apply to samples to which the computed model would be subsequently applied to obtain predicted concentrations.

Each of the 80 samples was measured with spectrometers A and B as described above. Before collecting near-IR spectra, glucose concentrations were confirmed with a YSI 2300 STAT PLUS clinical analyzer (YSI, Inc., Yellow Springs, OH).

The second sample set consisted of 89 samples composed of separately designed groups of 64 calibration samples and 25 prediction samples. These samples were measured with spectrometer C. Reagents for the six solutes were the same as described above. The solvent was a 0.1 M phosphate buffer at pH 7.4 prepared by titrating NaH_2PO_4 (ACS reagent, Fisher Scientific) with 50% (w/w)