

# Application of Multivariate Calibration Techniques to Quantitative Analysis of Bandpass-Filtered Fourier Transform Infrared Interferogram Data

MUTUA J. MATTU,\* GARY W. SMALL,† and MARK A. ARNOLD

*Center for Intelligent Chemical Instrumentation, Department of Chemistry, Clipping Laboratories, Ohio University, Athens, Ohio 45701-2979 (M.J.M., G.W.S.); and Department of Chemistry, University of Iowa, Iowa City, IA 52242 (M.A.A.)*

Multivariate calibration models are developed that allow quantitative analysis of short segments of Fourier transform infrared (FT-IR) interferogram data. Before the interferogram segments are submitted to partial least-squares (PLS) regression analysis, a bandpass digital filter is applied to isolate a narrow range of frequencies that correspond to an absorption band of the target analyte. This adds frequency selectivity to the analysis, thereby overcoming the principal obstacle to the direct use of interferogram data for quantitative analysis. With the optimization of the frequency response function of the filter, as well as the position and length of the interferogram segment employed, calibration models are developed that compare well with those computed with conventional absorbance spectra. This methodology is demonstrated by developing calibration models for determining glucose in an aqueous buffer matrix over the physiologically relevant concentration range of 1–20 mM. Through the use of a time-domain filter designed to isolate the modulated interferogram frequencies corresponding to the glucose C–H combination band at 4400 cm<sup>-1</sup>, a three-factor PLS calibration model is computed on the basis of interferogram points 601–850. This model is characterized by standard errors of calibration (SEC) and prediction (SEP) of 0.3311 and 0.6950 mM, respectively. The best model obtained in a thorough analysis of the corresponding absorbance spectra was also based on three PLS factors. This model was characterized by values of SEC and SEP of 0.2396 and 0.6115, respectively. In addition to achieving similar calibration and prediction results to the spectral-based model, the interferogram-based method has the advantage of requiring no background measurement of the sample matrix. Furthermore, since the analysis is based on only a 250-point segment of the interferogram, a reduction in the instrumentation and data collection requirements is realized.

Index Headings: Fourier transform infrared; Near-infrared; Interferogram; Glucose.

## INTRODUCTION

Fourier transform infrared (FT-IR) spectroscopy is receiving increased attention for its potential use in dedicated monitoring applications such as environmental remote sensing, chemical process monitoring, and clinical measurements. In each of these applications, a premium is placed on whether rugged, low-cost spectrometers can be designed that possess sufficient optical performance to allow the analytical determination to be made reliably. One approach to improving the ruggedness of an FT-IR spectrometer is to simplify the design specifications of the interferometer by reducing the resolution requirements of the collected spectra. This improvement reduces

the optical retardation required in the interferometer and can make possible the use of lower cost and potentially more rugged interferometer designs. This strategy is limited, however, by the requirements of the Fourier transform calculation used to convert the collected interferogram data to spectra. Extremely short interferograms produce highly distorted, low-resolution spectra that are difficult to use for quantitative analyses.

Recent work in our laboratories has focused on the development of methodology to overcome this limitation of the Fourier transform. Building upon several studies that demonstrated that quantitative information could be extracted directly from interferogram data,<sup>1–3</sup> we recently introduced a novel method for the direct quantitative analysis of short interferogram segments.<sup>4</sup> In this approach, the spectral selectivity normally obtained through the use of the Fourier transform is provided by the application of bandpass digital filters directly to the interferogram data. By tuning the filter bandpass to the modulated interferogram frequency associated with the absorption band of a target analyte and by selecting the appropriate interferogram segment to use, one can directly isolate selective analyte information from the interferogram segment. This procedure also eliminates the need for a background interferogram or spectrum for use in the data processing. In and of itself, eliminating the need for a background measurement can provide significant benefits in many dedicated monitoring applications such as environmental remote sensing in which it is extremely difficult to make a representative infrared background measurement without the analyte species present.

In our initial work, the successful use of a univariate calibration approach with bandpass-filtered interferogram segments of benzene and nitrobenzene of varying concentrations in carbon disulfide was reported. These preliminary studies were based on the analysis of a single analyte in a sample matrix containing no overlapping spectral bands due to the solvent or other interferences. In this paper, the methodology is extended to a significantly more challenging analysis, the determination of physiological levels of glucose in an aqueous buffer matrix. Here, the glucose spectral bands are weak and are heavily overlapped with those of the strongly absorbing aqueous matrix.

Due to this spectral overlap, digital filtering alone cannot provide sufficient selectivity to allow a univariate calibration procedure to be performed. In the work reported here, this problem is successfully overcome through the use of a multivariate calibration procedure based on partial least-squares (PLS) regression.

Received 28 August 1996; accepted 18 February 1997.

\* Present address: Instrumentation Metrics, Inc., 2085 Technology Circle, Suite 102, Tempe, AZ 85284.

† Author to whom correspondence should be sent.

## EXPERIMENTAL

**Apparatus and Reagents.** The samples used in this work consisted of aqueous glucose solutions in phosphate buffer. A portion of this data set has been used in a previous study based on the analysis of absorbance spectra.<sup>5</sup> The sample preparation and data collection will be briefly summarized here.

Thirty standard glucose solutions were prepared by diluting a stock glucose solution with a pH 7.3 phosphate buffer. The solutions spanned the physiologically relevant concentration range of 1.25–19.19 mM. The stock solution was prepared by dissolving the appropriate amount of dried reagent-grade glucose powder in the buffer, which contained 0.1 M phosphate and 0.483 g/L 5-fluorouracil. The 5-fluorouracil was added as a preservative. All solutions were prepared with reagent-grade water obtained by passing house-distilled water through a Milli-Q three-stage water purification system (Millipore, Inc., Bedford, MA).

Interferograms were collected with a Nicolet 740 FT-IR spectrometer equipped with a 150-W tungsten-halogen source, CaF<sub>2</sub> beamsplitter, and cryogenically cooled InSb detector. The detected light was restricted to the region of 5000–4000 cm<sup>-1</sup> through the use of an optical interference filter (Barr Associates, Westford, MA). This region contains several glucose spectral bands and has proven useful in previous work that employed absorbance spectra.<sup>5</sup> Samples were contained in an Infracil quartz transmission cell with a pathlength of 1 mm. Sample temperatures were controlled through the use of a water-jacketed cell holder and refrigerated temperature bath.

**Procedures.** During the data collection, double-sided interferograms with 16,384 points were collected on the basis of 256 coadded scans. Points were sampled at every zero-crossing of the HeNe reference laser, resulting in a maximum spectral frequency of 15,798.57 cm<sup>-1</sup>. In a procedure to obtain spectra for use in generating comparison results to the interferogram-based analysis, these interferograms were triangularly apodized and Fourier transformed to produce spectra with a point spacing of 1.9 cm<sup>-1</sup>. Mertz phase correction was applied to the spectra by use of a phase array based on 200 points on each side of the interferogram centerburst. The Fourier processing was performed with the data collection software resident on the Nicolet 620 computer controlling the spectrometer.

The data collection spanned six days and was designed for the study of how changes in sample temperature affect the glucose spectra. The data for 24 of the glucose solutions were collected at 37 °C, while for five solutions, interferograms were collected at 32, 33, 35, 36, 37, 39, and 41 °C. For the last glucose solution, interferograms were collected at 32, 33, 35, 37, 39, and 41 °C. Two to four replicate interferograms (each consisting of 256 coadded scans) were collected consecutively for each combination of glucose solution and temperature. In addition, five of the concentration/temperature combinations were repeated on a different day. This produced a total of 70 experiments [24 + 5 + (5 × 7) + (1 × 6)] and 206 interferograms. Interferograms corresponding to a given sample temperature were collected as a group, although, within the group, glucose solutions were run in

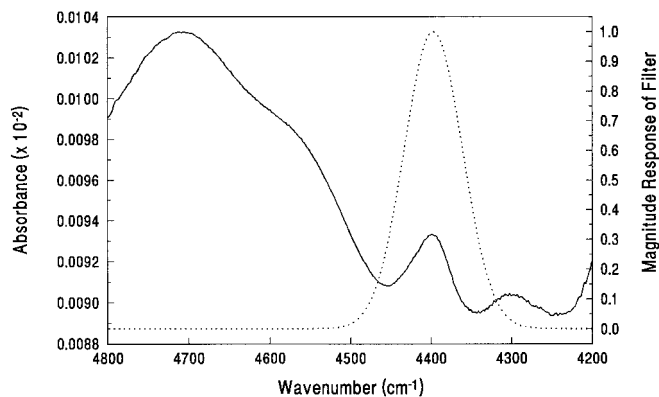


FIG. 1. Near infrared absorbance spectra for 11.0 mM glucose (solid line) and a filter frequency response function centered on the glucose absorption band at 4400 cm<sup>-1</sup> (dotted line). The FWHH of the frequency response is 84.9 cm<sup>-1</sup>.

a random order with respect to concentration. The sample temperatures were randomized with respect to time.

Throughout each day, interferograms of the pure phosphate buffer were collected, and the corresponding single-beam spectra were computed for subsequent use in the calculation of absorbance spectra of the glucose solutions. All measurements of the buffer were performed at 37 °C. The same Fourier processing steps described above were performed for these spectra. The glucose absorbance spectra computed with these background spectra thus had temperature-induced artifacts for the 40 experiments out of 70 in which the sample temperature was not 37 °C. These artifacts will be characterized further below.

The collected interferograms and computed single-beam spectra were transferred to a Silicon Graphics 4D/460 computer operating under Irix (Version 4.0.5, Silicon Graphics, Mountain View, CA). The remaining calculations reported here were performed on this computer with software written in FORTRAN 77. Subroutines used to perform Fourier transforms and multiple linear regression analysis were obtained from the IMSL library (IMSL, Inc., Houston, TX).

## RESULTS AND DISCUSSION

**Characterization of Collected Data.** The NIR spectrum of water is dominated by large absorption bands with peak maxima at 6876, 5267, and 3800 cm<sup>-1</sup>. For the quantitative analysis of glucose to be possible, glucose spectral bands that can be isolated from the strong absorption bands of water must be identified. The solid line in Fig. 1 is an absorbance spectrum of a 11.0-mM glucose solution. The spectrum was referenced to a phosphate buffer spectrum collected at approximately the same time. The plotted spectral region of 4800–4200 cm<sup>-1</sup> contains three glucose absorption bands centered at 4700, 4400, and 4300 cm<sup>-1</sup>. The band at 4400 cm<sup>-1</sup> was the selected target for this work because it is located near the minimum between the water bands, which results in relatively high optical throughput and a corresponding maximum signal-to-noise ratio. There are also relatively smaller baseline variations in the vicinity of the 4400-cm<sup>-1</sup> glucose band, because the water absorbance spec-

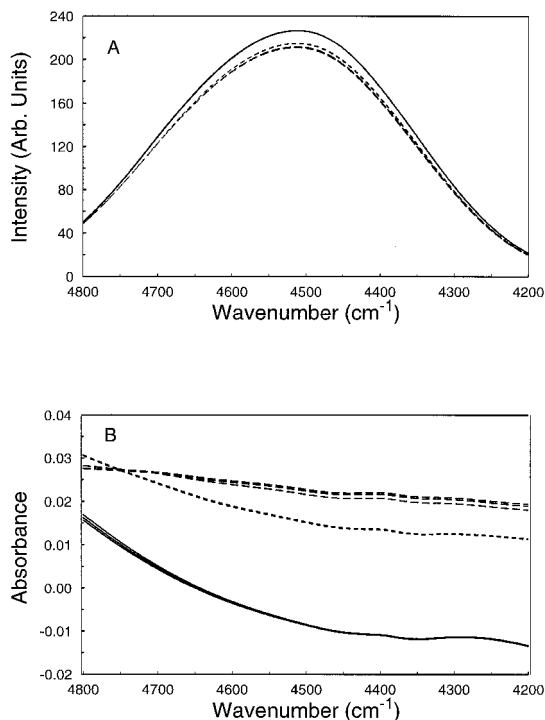


FIG. 2. Single-beam (A) and absorbance (B) spectra for a 17.98 mM glucose solution collected at 33 (solid line), 35 (short dashed line), and 37 (long dashed line) °C. Three replicate spectra are plotted corresponding to each temperature. Absorbance spectra were computed by use of a background spectrum of phosphate buffer collected at 37 °C.

trum is less sensitive to temperature variations in this spectral region.<sup>5</sup>

Figure 2 illustrates the effect of temperature on the glucose single-beam and absorbance spectra. Figure 2A plots nine single-beam spectra of a solution of 17.98 mM glucose. The spectra represent three groups of three replicates and correspond to data collected at 33 (solid line), 35 (short dashed line), and 37 (long dashed line) °C. The nine spectra are derived from the same data collection session and therefore are consistent in terms of the source energy reaching the sample. The single-beam intensity decreases with increasing temperature due to the shift of the water band centered at 3800  $\text{cm}^{-1}$  to higher frequency with increasing temperature. This increases the water absorbance in the 4300–4700- $\text{cm}^{-1}$  region, thereby decreasing the light intensity reaching the detector. An inspection of Fig. 2A also reveals that the effect is nonlinear. The change in intensity between the spectra collected at 33 and 35 °C is significantly greater than the corresponding change in the spectra collected at 35 and 37 °C.

Figure 2B illustrates the effect of temperature mismatches on computed absorbance spectra. The same three groups of three replicate spectra are plotted after conversion to absorbance. The line types are the same as in Fig. 2A. Each single-beam spectrum was ratioed to a spectrum of phosphate buffer collected at 37 °C. The result of the temperature mismatch between the glucose and buffer solutions is a change in the tilt and offset of the spectral baseline. The baseline change is most severe for the 33 °C glucose sample (solid line) due to the 4 °C mismatch with the buffer. The baseline artifacts introduced are larger in magnitude than the glucose absorption bands. The

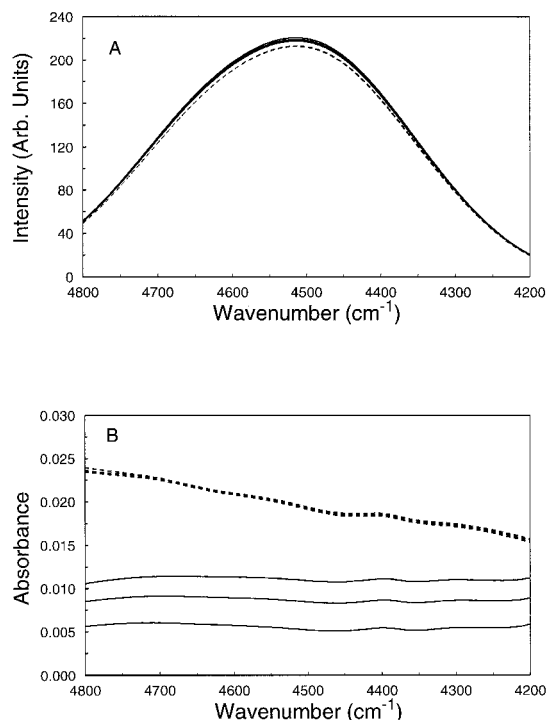


FIG. 3. Single-beam (A) and absorbance (B) spectra of a 12.08 mM glucose solution collected at 37 °C on each of two different days. The three replicate spectra collected on each day are indicated by the solid (day 1) and short dashed lines (day 2). Absorbance spectra were computed by use of background spectra of phosphate buffer collected at 37 °C on the same day as the sample spectra.

glucose bands observed in Fig. 1 can barely be distinguished in Fig. 2B due to the baseline artifacts. This point is particularly significant given that the 17.98-mM glucose spectra in Fig. 2 represent the upper end of the concentration range studied.

As described previously, some combinations of glucose concentration and temperature were run on two separate days to evaluate the day-to-day reproducibility of the data. Figure 3 is a plot analogous to Fig. 2 that compares two groups of three replicate single-beam and absorbance spectra of a 12.08-mM glucose solution measured at 37 °C. The replicate spectra from days 1 and 2 are plotted as solid and short dashed lines, respectively, in Figs. 3A and 3B. The computation of the absorbance spectra in Fig. 3B employed buffer spectra measured at 37 °C and collected on the same day as each glucose spectrum. Inspection of Figs. 3A and 3B reveals that, even under the same temperature conditions, there is significant day-to-day variability in the data.

**Assembly of Data Sets.** Except for the one constraint noted below, the interferograms and absorbance spectra were randomly partitioned into calibration and prediction sets. This procedure was performed by defining a “sample” as a combination of glucose concentration, temperature, and day of data collection. As described previously, there were 70 samples that met these criteria. The temperature effects observed in Fig. 2 justify the use of temperature as a variable in defining the independent samples for use in calibration and prediction. While significant day-to-day variability in the spectra was observed in Fig. 3, it was decided not to split samples with the same glu-

glucose concentration and temperature between the calibration and prediction sets. The 10 samples (i.e., the five samples from each of two days) that duplicated glucose concentrations and temperatures were all placed in the calibration set.

Of the 70 samples, 58 (83%) were assigned to the calibration set and the remaining 12 set aside as the prediction set. The calibration set was further randomly partitioned into a calibration subset and a monitoring set. Twelve samples (21% of the main calibration set) comprised the monitoring set. The replicate interferograms and spectra corresponding to each of the samples were carried along into the respective data sets. This step produced 135, 36, and 35 interferograms and spectra in the calibration subset, monitoring set, and prediction set, respectively. The interferogram and spectral data sets were identical in terms of which samples were assigned to the calibration, monitoring, and prediction sets.

The calibration subset was used for model building during the optimization of the various experimental variables involved, and the predictive capability of the computed models was assessed with the use of the monitoring set. Once the best combination of these variables had been established, the main calibration set (i.e., calibration subset + monitoring set) was used together with these optimal parameters to build the final model, which was evaluated with the use of the prediction set. This procedure allowed the predictive ability of the models to be assessed during the model optimization while still retaining some samples to serve as a final independent test set. In order to confirm the results generated through this "calibration-prediction" approach, a leave-one-out cross-validation was also performed by using the already determined optimal model parameters. In this procedure, the replicate interferograms or spectra of each sample were withheld in turn, a calibration model was constructed with the remaining data, and the concentrations of the withheld spectra or interferograms were predicted with the use of the computed model. By withholding the data from each sample in turn, an overall pooled estimate of the prediction errors was obtained.

**Interferogram-Based Analysis. Digital Filtering.** The interferogram-based quantitative analysis proposed consists of two steps that require the optimization of parameters. First is the application of digital filters directly to specific short segments of FT-IR interferograms collected with the analyte of interest in the optical path of the spectrometer. The digital filter will remove or suppress from the filtered interferogram the spectral information outside of the frequency region defined by the filter bandpass. Two procedures can be used for filtering an interferogram: Fourier filtering and time-domain digital filtering. In Fourier filtering, the fast Fourier transform (FFT) is applied to the raw interferogram to yield a single-beam spectrum with the analyte features superimposed on a large detector response background. The target analyte feature (e.g., the glucose band centered at  $4400\text{ cm}^{-1}$ ) is then isolated from the background by multiplication of the single-beam spectrum with a Gaussian-shaped function that is centered on the band of interest. The Gaussian function is termed the "frequency response function" of a digital filter. This multiplication truncates the background response to coincide with the frequency response

function. On inverse Fourier transforming this filtered spectrum, one obtains a filtered interferogram. The dotted line in Fig. 1 depicts an example frequency response function centered on the glucose absorption band at  $4400\text{ cm}^{-1}$ . The frequency response function is defined by its center position and full width at half-height (FWHH).

In the context of digital filtering, the Gaussian curve defines a passband of frequencies that the filter will pass. Those frequencies outside of the Gaussian curve define two stopbands of frequencies (i.e., one band on either side of the curve) that will be attenuated by the filter. The filter was applied to the single-beam spectrum by the multiplication of the spectrum and the frequency response function.

In the interferogram or time domain, the raw interferogram is filtered by convolving it with the time-domain representation of the frequency response function. As expressed by the convolution theorem of the Fourier transform, this time-domain function is the Fourier transform pair of the frequency response function. If the interferogram is considered to be a summation of cosine waveforms, the action of filtering the interferogram is simply to suppress the amplitude of those sinusoids whose frequencies lie outside the passband of the filter. The true convolution result is an infinite summation and therefore must be approximated in a practical implementation.<sup>6,7</sup> The approximation technique used in this work was designed in our laboratory specifically for use with interferogram data.<sup>8</sup>

The work described in this paper addresses the feasibility of coupling digital filtering and PLS regression for developing quantitative calibration models from the filtered interferogram data of glucose samples. Results from the use of both Fourier filters and time-domain filters will be compared.

*Motivation for Multivariate Calibration.* The second step in the interferogram-based analysis is the use of the filtered interferogram segment to construct a calibration model that allows analyte concentrations to be predicted. The feasibility of performing quantitative analysis of bandpass-filtered FT-IR interferograms has been demonstrated in our initial feasibility study.<sup>4</sup> In that work, the presence of a region in the filtered interferogram apparently related to the magnitude of the spectral absorption band was established and rationalized. Information directly related to the filtered analyte band was found to be located past a marker in the interferogram termed the "node point". The node arises as a result of the destructive interference between the positive-going Gaussian frequency response function and the negative-going absorption band superimposed on it. The location of the node, relative to the interferogram centerburst, was found to be dependent on the widths of both the analyte band and the frequency response function. Wider functions were observed to shift the node closer to the centerburst. To the right of the node, the interferogram magnitude increased with increasing magnitude of the analyte spectral band, while to the left of the node, an inverse relationship existed between interferogram and spectral band magnitudes.

By making several key assumptions, we derived the approximate mathematical relationship between the intensity of the filtered interferogram past the node and

concentration.<sup>4</sup> One of these assumptions was that only a single analyte band was isolated by the bandpass filter. In our initial study, this assumption held because the carbon disulfide solvent had negligible absorbance within the passbands of the filters used to isolate the targeted benzene or nitrobenzene bands. This factor led to the successful use of a univariate calibration model on the basis of the magnitude of a segment of the interferogram located past the node.

However, if other analyte bands or bands from other constituents in the sample matrix fall within the passband, their interferogram representations will add to the interferogram signature of the targeted analyte band. The filtered interferogram will then possess a more complex structure, and the simple univariate linear relationship between interferogram segment magnitude and concentration used in our previous work will no longer hold.

Figure 4 plots points 550–800 (relative to the centerburst) of the filtered interferograms corresponding to glucose solutions of concentration 17.98 (A), 7.83 (B), and 1.25 mM (C) collected at 37 °C. To obtain these interferograms, we used Fourier filtering with the frequency response function depicted by the dotted line in Fig. 1. The arrows in Fig. 4 point to the location of the node. As expected, the node position moves away from the centerburst as the glucose concentration is reduced. This response is due to the narrowing of the glucose spectral band as its intensity is reduced. The node point is no longer as distinct as observed previously in corresponding plots of filtered benzene or nitrobenzene interferograms.<sup>4</sup> This result is attributed to the contribution of the overlapping water absorption signature that has been passed by the filter. In this case, the filtering procedure cannot totally isolate the glucose signature. However, it is evident that the amplitude of the signal after the node does increase with glucose concentration.

**PLS Regression.** The principal goal of this work was to evaluate PLS regression<sup>9–11</sup> for its utility in building multivariate calibration models on the basis of bandpass-filtered interferograms, thereby helping to overcome the problem of overlapping signals in the interferogram segment. Given the intensities of the  $p$  points in each filtered interferogram segment and the corresponding known glucose concentrations for the  $n_c$  interferograms in the calibration set, the PLS algorithm decomposes the mean-centered interferogram data matrix into a set of  $h$  underlying factors that model the covariance between the interferogram and concentration information. The value of  $h$  represents a parameter to be optimized.

The computed factors are linear combinations of the input interferogram intensities and serve to extract information from the interferogram that is relevant in modeling the variation in glucose concentration. As illustrated in Fig. 4, the filtered interferogram takes the form of a damping cosine wave. Since there is an underlying functional relationship between the points on this waveform, there is significant correlation among the columns of the  $n_c \times p$  interferogram data matrix. Thus, there is some redundancy of information in the data matrix. For this reason, the optimal value of  $h$  is typically much less than  $p$ .

**Optimization Experiments.** In addition to finding the optimal value of  $h$ , other experimental parameters that need to be optimized are the position and width of the

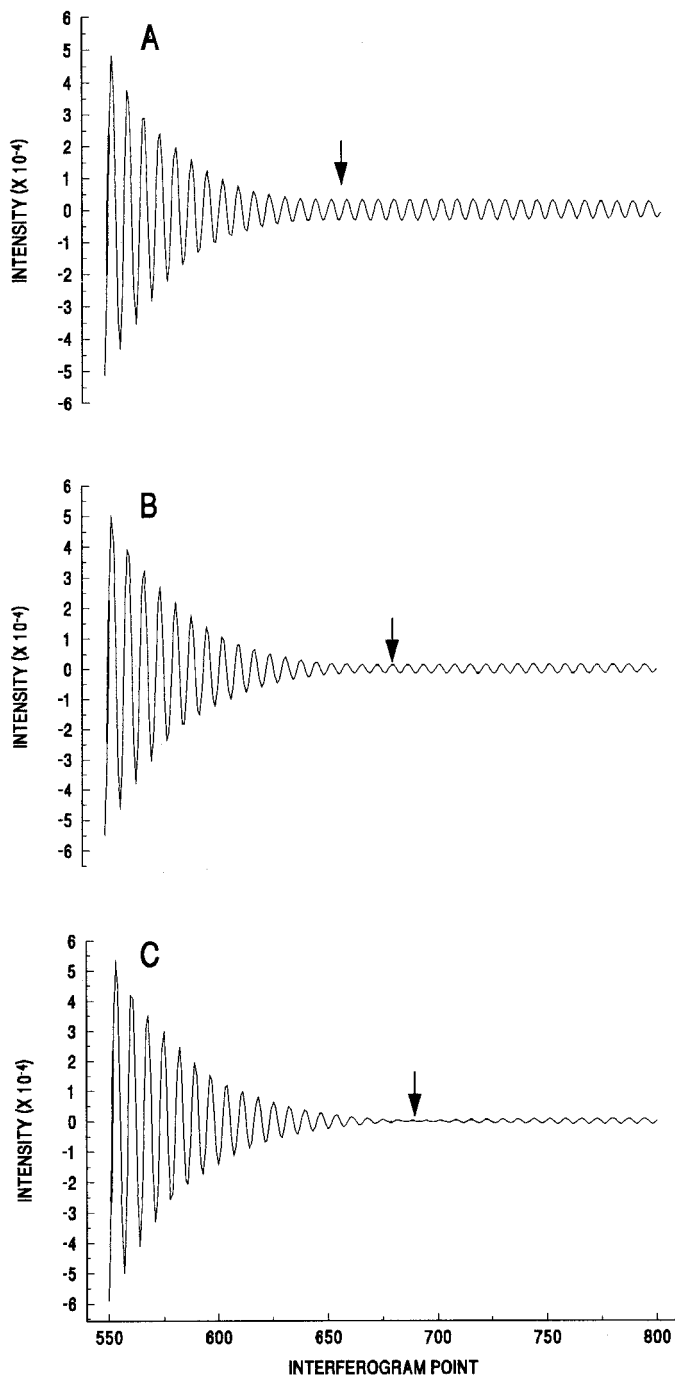


FIG. 4. Segments of filtered interferograms. Interferogram segments A–C correspond to aqueous glucose standards of concentration 17.98 (A), 7.83 (B), and 1.25 mM (C) collected at 37 °C, after filtering with a Gaussian frequency response function centered at 4397.1  $\text{cm}^{-1}$  and with an FWHH of 84.9  $\text{cm}^{-1}$ . The characteristic node points in the interferogram are denoted by the arrows. Note the movement of the node toward the centerburst with increasing concentration and the corresponding increase in interferogram amplitude after the node.

filter passband and the interferogram segment location and length. Since the sample matrix used here involves two absorbing species, it was instructive to adjust the filter position to locations away from the center of the analyte band. Experiments directed to the optimization of these parameters were first implemented by using Fourier filtering, then repeated with time-domain filtering.

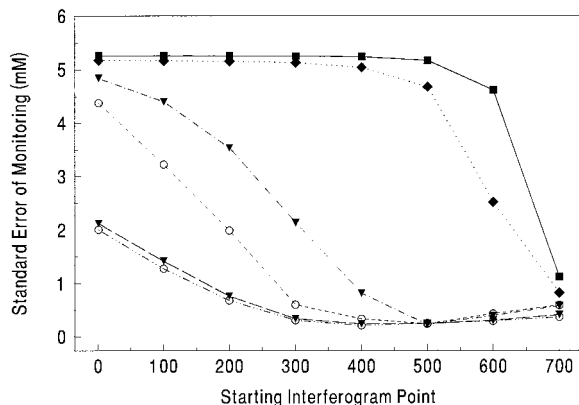


FIG. 5. Standard error of monitoring vs. interferogram starting point for calibration models generated with 300-point filtered interferogram segments. The Fourier filter used was centered at  $4406.7\text{ cm}^{-1}$  and had an FWHH of  $72.7\text{ cm}^{-1}$ . Results are presented for models generated with one (solid line), two (dotted line), three (dot-dash line), four (short dashed line), five (long dashed line), and six (dot-dot-dot-dashed line) PLS factors.

The optimization step took the form of a series of factorial experimental designs in which discrete levels of interferogram segment length, segment starting point, and number of PLS factors ( $h$ ) were specified, and all combinations of these levels were investigated for a series of filter positions and widths. Segment lengths of 50, 100, 150, 200, 250, 300, and 500 points were studied. The starting points of the interferogram segments tested were shifted in steps of 100 points, starting from point 1 and ending at point 1000, relative to the centerburst. In previous work, points beyond 1000 were observed to add little useful information.<sup>4</sup> The number of PLS factors used was varied from 1 to 6. The filters investigated were centered at the following positions: 4387.4, 4397.1, 4400.9, 4404.8, 4406.7, 4408.6, and  $4412.5\text{ cm}^{-1}$ . The FWHH values tested were 63.6, 68.1, 72.7, 77.2, 81.8, 86.3, and  $90.8\text{ cm}^{-1}$ . A bisection procedure was used to select the optimal values for filter position and width. Three pairs of position and width values were studied first, and the best results from the factorial designs were used to bracket the optimal range for each filter parameter. The bracketed ranges were then split in half, and the investigation was repeated. This process continued until the results converged.

During the optimization experiments, the current parameters being evaluated were applied to construct a model from the interferograms comprising the calibration subset. The glucose concentrations corresponding to interferograms in the monitoring set were then predicted by use of the computed model. The statistic used to assess the predictive ability was the standard error of monitoring (SEM), defined as

$$\text{SEM} = \sqrt{\frac{\sum_{i=1}^{n_m} [c_{m,i} - \hat{c}_{m,i}]^2}{n_m}}, \quad (1)$$

where  $n_m$  is the number of interferograms in the monitoring set,  $c_{m,i}$  is the actual glucose concentration associated with the  $i$ th interferogram in the monitoring set,

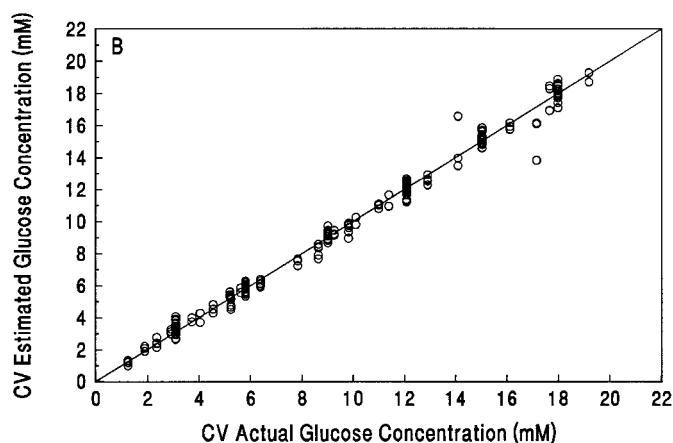
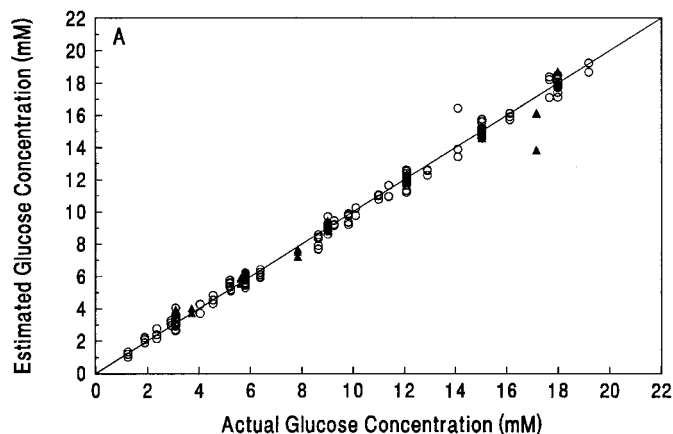


FIG. 6. Estimated vs. actual glucose concentrations for the optimal calibration model based on Fourier-filtered interferogram segments. (A) Samples in the calibration and prediction sets are indicated by open circles and closed triangles, respectively. (B) Cross-validated estimated concentrations are plotted against actual concentrations.

and  $\hat{c}_{m,i}$  is the corresponding glucose concentration predicted by the model.

Figure 5 displays a plot of SEM vs. interferogram starting point for segments that were 300 points in length. The optimal filter specifications consisting of a center position of  $4406.7\text{ cm}^{-1}$  and FWHH value of  $72.7\text{ cm}^{-1}$  were used. From this plot, it is clear that the optimal segment consists of points 501–800, coupled with a three-factor PLS calibration model.

Given these optimized parameters, the calibration subset and monitoring set were combined, and a new model was generated. This model was characterized by computing the standard error of calibration (SEC), and the model was subsequently applied to estimate the concentrations corresponding to the interferograms in the prediction set. The standard error of prediction (SEP) was used to assess the predictive ability of the model. The computations of SEC and SEP are analogous to that of SEM, with the exception that the degrees of freedom for the SEC calculation are adjusted for the model size.

Figure 6 shows two correlation plots corresponding to the best results generated from the two calibration approaches: (A) calibration-prediction mode and (B) leave-one-out cross-validation. The calibration samples (i.e., calibration subset + monitoring set) are indicated in Fig.

6A by open circles, while prediction set samples are indicated by closed triangles. The experimental variables used were a Fourier filter centered at  $4406.7\text{ cm}^{-1}$  and with an FWHH of  $72.7\text{ cm}^{-1}$ , a 300-point interferogram segment located between points 501 and 800, and a model based on three PLS factors. The calibration-prediction mode yielded a model with an SEC of  $0.3887\text{ mM}$  and SEP of  $0.6998\text{ mM}$ . Similar results were obtained with leave-one-out cross-validation: a pooled SEC of  $0.4517\text{ mM}$  and a pooled cross-validated SEP of  $0.4646\text{ mM}$ . Both plots show two outlier points corresponding to the third replicate of the glucose standards at concentrations of  $14.09$  and  $17.15\text{ mM}$ . The presence of one of these outliers in the prediction set is the principal cause of the SEP value exceeding the value of SEC.

**Time-Domain Filtering.** As noted above, the interferogram filtering process can also be achieved in the time-domain. This is the preferred ultimate implementation, as the FFT is not required, and the entire procedure can be based on a short interferogram segment.

As noted previously, the practical implementation of digital filters in the time domain requires an approximation to the convolution sum. Having established the position of the filter using the Fourier filtering approach, we varied the width of the passband, since the actual FWHH achieved by the time-domain filter approximation is typically wider than the target width specified in the filter design step. A further ramification of the time-domain filter approximation is that, as attempts are made to achieve a very narrow passband, the attenuation in the stopbands tends to degrade. Given the maximum spectral frequency of  $15798.57\text{ cm}^{-1}$ , the previously optimized passband FWHH of  $72.7\text{ cm}^{-1}$  represents less than 0.5% of the total spectral bandwidth. This is an extremely narrow filter specification. If the price of achieving a narrow passband is too much loss of attenuation in the stopbands, the filter will be effectively useless, as it will not do a good job of isolating the passband frequencies.

Twenty time-domain filters were generated, all centered at  $4406.7\text{ cm}^{-1}$ , but with varying FWHH. The interferogram segments studied included segment sizes of 100, 200, 250, 300, 400, and 500 points starting from points 301 to 1000. This is the region that was found to contain information that correlated with glucose concentration when analyzing the Fourier-filtered data. The size of the calibration models was varied from 1 to 6 PLS terms. The best results were obtained with the following combination of the experimental parameters: filter centered at  $4406.7\text{ cm}^{-1}$ , actual FWHH of  $145\text{ cm}^{-1}$ , a 250-point interferogram segment located between points 601 and 850, and a calibration model based on three PLS factors.

This model was evaluated with both the calibration-prediction and leave-one-out cross-validation procedures. In the calibration-prediction evaluation, an SEC value of  $0.3311\text{ mM}$  and an SEP value of  $0.6950\text{ mM}$  were obtained. The results of cross-validation were a pooled SEC value of  $0.4031\text{ mM}$  and a cross-validated SEP value of  $0.4258\text{ mM}$ . Correlation plots in both cases were virtually identical to those displayed in Fig. 6. The two outlier points mentioned previously were also evident.

Table I is a summary of the results from both filtering techniques and the two model evaluation procedures. As

**TABLE I. Comparison of Fourier and time-domain filtering techniques.**

| Filtering technique    | Interferogram segment <sup>a</sup> | PLS factors | SEC(mM) | SEP(mM) |
|------------------------|------------------------------------|-------------|---------|---------|
| (a) Fourier            |                                    |             |         |         |
| Calibration-prediction | 501–800                            | 3           | 0.3887  | 0.6998  |
| Cross-validation       | 501–800                            | 3           | 0.4517  | 0.4646  |
| (b) Time-domain        |                                    |             |         |         |
| Calibration-prediction | 601–850                            | 3           | 0.3311  | 0.6950  |
| Cross-validation       | 601–850                            | 3           | 0.4031  | 0.4258  |

<sup>a</sup> Relative to interferogram centerburst.

expected, the cross-validation approach gives slightly better prediction errors than the calibration-prediction mode because of the availability of more data for use in generating the calibration model. The cross-validation procedure uses all the samples in the construction of the model except the one set aside for prediction, and therefore the models tend to encode more data variation than the models computed from fewer samples in the calibration-prediction approach.

**Spectral-Based Analysis.** Given the success of the interferogram-based analysis described above, the question arises as to how the calibration and prediction results compare with those obtained in a conventional analysis of absorbance spectra. Previous work with a different subset of the spectral data resulted in the identification of four spectral ranges that appeared useful in constructing PLS-based calibration models.<sup>5</sup> These were  $4850\text{--}4220$ ,  $4811\text{--}4457$ ,  $4457\text{--}4354$ , and  $4354\text{--}4227\text{ cm}^{-1}$ . The first range includes the three glucose bands shown in Fig. 1, while the next three ranges address the specific glucose bands individually. As noted previously, the spectral analysis used the same assignment of samples to the calibration, monitoring, and prediction sets as was used in the analysis of the interferogram data. The calibration models were evaluated on the basis of the calibration-prediction approach.

Prior to submission to PLS regression, the absorbance spectra were Fourier filtered to remove noise and baseline variation. A previously designed grid-search protocol was used to optimize the bandpass position and width of a Gaussian-shaped filter.<sup>12</sup> The mean and standard deviation of the Gaussian frequency response function (i.e., the center position and width of the filter passband) were both varied from  $0.001$  to  $0.1 f$  in increments of  $0.002 f$ , giving a total of 2500 evaluations for each of the PLS factors studied. The bandpass position and width values are expressed in digital frequency units ( $f$ ), where  $0.5 f$  is the maximum harmonic frequency of the data. Calibration models based on 1–15 PLS terms were assessed for each of the four spectral ranges. For each model size (i.e., number of PLS factors), the filter parameters that produced the best SEC and SEM values were selected as optimal. Therefore, 15 optimal filters were realized for each of the four spectral ranges.

In order to select the number of significant PLS factors necessary to account for most of the variation in the data, an  $F$ -test was performed on the SEM values for each model as described by Haaland and Thomas.<sup>13</sup> The con-

**TABLE II. Results of spectral-based analysis.**

| Spectral range (cm <sup>-1</sup> ) | PLS factors | SEC (mM) | SEP (mM) |
|------------------------------------|-------------|----------|----------|
| 4354–4227                          | 2           | 0.4093   | 0.7976   |
| 4457–4354                          | 1           | 0.2561   | 0.6486   |
| 4811–4457                          | 1           | 0.4987   | 0.7837   |
| 4850–4220                          | 3           | 0.2396   | 0.6115   |

confidence level selected for this work was 99.9% (i.e.,  $\alpha = 0.001$ ).

As in the interferogram-based analysis, once the optimal parameters were established, the calibration subset and monitoring set were combined to form the main calibration set with which the final calibration model was built. The predictive capability of this model was then tested with the use of the prediction set. The best results for each of the four spectral ranges are summarized in Table II. For all four spectral ranges, the best filter was centered at  $0.023 f$  and had a width specification of  $0.003 f$ .

**Comparison of Spectral and Interferogram-Based Analyses.** Tables I and II can be used to compare the results from the spectral- and interferogram-based analyses. The best results from the spectral analysis are slightly better in terms of the SEC and SEP statistics. The same outlying observations are observed in both cases. While the best interferogram-based and spectral-based calibration models both required three PLS factors, a similarly performing spectral-based model required only one factor. A reduced number of factors was possible in the spectral analysis since it was based on the use of absorbance spectra in which the aqueous background was largely removed by ratioing to a buffer spectrum. The baseline artifacts induced by the temperature mismatches between the sample and background spectra did not adversely affect the analysis due to the use of the Fourier-filtering preprocessing step. The use of absorbance spectra to construct the calibration models provided additional selectivity to the analysis that was not available in the interferogram-based processing. This factor, coupled with the occurrence of the interferogram node point shifting with glucose concentration, serves to explain the additional model terms required in the interferogram-based analysis. It should be emphasized, however, that an advantage of the interferogram-based approach is

that a viable calibration model can be developed without the need for a background measurement.

## CONCLUSION

The feasibility of performing quantitative analysis based on short segments of bandpass filtered interferograms has been demonstrated. The combination of digital filtering and PLS regression was observed to provide sufficient spectral selectivity to allow viable calibration models to be generated for an application in which filtering alone could not prevent the presence of overlapping compound signatures in the interferogram segment.

The successful use of time-domain filtering allowed the analysis to be based entirely on a short segment of the interferogram, thereby making possible the collection and use of "low-resolution" interferograms. This factor, coupled with the removal of the requirement for a background spectral measurement, may provide significant benefits to the use of FT-IR spectroscopy in nonlaboratory monitoring applications.

## ACKNOWLEDGMENTS

This research was supported by the National Institutes of Health under grant 1-R01-DK45126-01A1. The Department of the Army is acknowledged for providing the Silicon Graphics 4D/460 computer system. Preliminary results of this research were presented at the Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, Chicago, Illinois (1994).

1. D. T. Sparks, R. Lam, and T. L. Isenhour, *Anal. Chem.* **54**, 1922 (1982).
2. S. L. Monfre and S. D. Brown, *Appl. Spectrosc.* **46**, 1699 (1992).
3. S. L. Monfre and S. D. Brown, *Appl. Spectrosc.* **46**, 1711 (1992).
4. M. J. Mattu and G. W. Small, *Anal. Chem.* **67**, 2269 (1995).
5. K. H. Hazen, M. A. Arnold, and G. W. Small, *Appl. Spectrosc.* **48**, 477 (1994).
6. A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing* (Prentice Hall, Englewood Cliffs, New Jersey, 1989).
7. J. H. McClellan and T. W. Parks, *IEEE Trans. Circuit Theory CT-20*, 697 (1973).
8. G. W. Small, A. C. Harms, R. T. Kroutil, J. T. Dittilo, and W. R. Loerop, *Anal. Chem.* **62**, 1768 (1990).
9. P. Geladi and B. R. Kowalski, *Anal. Chim. Acta* **185**, 1 (1986).
10. H. Martens and T. Næs, *Multivariate Calibration* (John Wiley and Sons, New York, 1989), pp. 116–166.
11. K. R. Beebe and B. R. Kowalski, *Anal. Chem.* **59**, 1007A (1987).
12. G. W. Small, M. A. Arnold, and L. A. Marquardt, *Anal. Chem.* **65**, 3279 (1993).
13. D. M. Haaland and E. V. Thomas, *Anal. Chem.* **60**, 1193 (1988).