

# Determination of glucose in a synthetic biological matrix with decimated time-domain filtered near-infrared interferogram data

Ndumiso A. Cingo<sup>a</sup>, Gary W. Small<sup>a,\*</sup>, Mark A. Arnold<sup>b</sup>

<sup>a</sup> Center for Intelligent Chemical Instrumentation, Clippinger Laboratories, Department of Chemistry and Biochemistry, Ohio University, Athens, OH 45701, USA

<sup>b</sup> Department of Chemistry and Optical Science and Technology Center, 230 Iowa Advanced Technology Laboratories, University of Iowa, Iowa City, IA 52242, USA

Received 6 July 1999; received in revised form 18 November 1999; accepted 3 December 1999

## Abstract

Fourier transform near-infrared interferogram data are used to determine the physiological levels of glucose in a synthetic biological matrix consisting of varying levels of bovine serum albumin and triacetin in a pH 7.4 phosphate buffer. Finite impulse response (FIR) digital filters are applied to short segments of the interferogram data, and the resulting filtered interferogram intensities are used as independent variables in constructing multivariate calibration models based on quadratic partial least-squares regression analysis. To increase the performance characteristics of the filters, the collected interferograms are decimated to reduce their frequency bandwidth. By use of a combination of optical and digital filtering, the spectral bands in the region of 4000 to 5000  $\text{cm}^{-1}$  are aliased to the range of 0 to 1975  $\text{cm}^{-1}$ . This allows the attenuation in the stopbands of the filters to be increased from 7 to 50 dB, thereby increasing their selectivity. By use of FIR filters centered on the glucose C–H combination band at 4397  $\text{cm}^{-1}$  and the triacetin C–H combination band at 4446  $\text{cm}^{-1}$ , a calibration model for glucose is constructed that achieves a standard error of calibration of 0.484 mM and a standard error of prediction of 0.618 mM over the 1–20 mM concentration range. This prediction performance represents a 20% improvement over the performance of a model built with the full-bandwidth data. The improved model is also observed to outperform a similar calibration model based on the conventional analysis of absorbance spectra. © 2000 Elsevier Science B.V. All rights reserved.

**Keywords:** Fourier transform spectrometry; Interferogram; Near-infrared; Glucose; Digital filtering

## 1. Introduction

Fourier transform near-infrared (FT-near-IR) spectrometry has recently come under study as a viable tool for analyzing clinically relevant analytes

in biological matrixes [1–3]. Our laboratories have been involved in an ongoing effort to develop data analysis methodology for use with such near-IR spectrometric data [4–9]. The ultimate goal of one of these projects is to develop a procedure for the noninvasive determination of glucose in human blood. If available, a noninvasive blood glucose measurement would provide a significant benefit to

\* Corresponding author. Tel.: +1-740-593-1748; fax: +1-740-593-0148.

diabetic patients who currently perform self-monitoring with methods that require a blood sample.

Previous studies have demonstrated the potential of combining digital filtering techniques with multivariate calibration methods for selectively modeling glucose concentrations in a variety of matrixes [4–7]. Conventional spectral analysis methods require the use of a background measurement, which is incompatible with a noninvasive determination since no glucose-free measurement of the sample matrix is possible. To overcome this obstacle, the direct use of single-beam spectral data in building glucose calibration models has been investigated and successfully implemented [8]. This approach, however, still requires the collection of a full FT-near-IR interferogram in order to be implemented, and thus requires a conventional FT spectrometer. The cost, mechanical complexity, and lack of ruggedness of such an instrument precludes its use as a dedicated glucose monitor for home use by diabetics. This is unfortunate since the fast scan speed, high optical throughput, and excellent wave number accuracy and reproducibility offered by an FT spectrometer make it an attractive choice for use in an application in which a small glucose spectral signal must be numerically isolated from the strong and overlapping background absorbance of the biological matrix.

Recent work performed in our laboratories has introduced a promising alternative method to address both the need to eliminate the use of a background measurement and the desire to simplify the FT spectrometer [8–10]. This approach implements an analysis of glucose by directly using short segments of bandpass digitally filtered interferogram data to build multivariate calibration models. The motivation for restricting the analysis to a short interferogram segment is to reduce the optical path difference (retardation) that must be achieved in the interferometer. If the required retardation is short enough, alternative interferometer designs are made possible that allow the advantages of FT spectrometry to be realized while making the instrument more practical in terms of cost, complexity, and ruggedness [11].

To demonstrate the feasibility of this approach, one of the previous studies used multiple Fourier filtering on interferogram data as a spectral preprocessing tool to isolate glucose information, as well as information pertaining to two major matrix interfer-

ents, bovine serum albumin (BSA) and triacetin. The resulting filtered interferogram data were submitted to a partial least-squares (PLS) regression algorithm to construct calibration models for use in estimating glucose concentrations [12]. While this approach demonstrated the feasibility of directly using short interferogram segments to determine glucose in a complex matrix, its drawback lies in the fact that the filtering was implemented in the frequency domain, using Fourier filters. The Fourier filtering step, which was shown to be crucial in isolating analyte and interference information, requires processing of a full-length interferogram in order to be successfully implemented. This limitation undermines the ultimate goal of developing a simplified spectrometer that collects only short interferogram segments to be used in a noninvasive determination of glucose.

In the current study, we implement the preprocessing digital filtering step in the time-domain. Filtering in the time-domain is implemented using finite impulse response (FIR) filters [13]. In this work, FIR filters are designed and applied to limited bandwidth interferogram data, thus taking advantage of their improved attenuation characteristics over narrow bandwidths. This is achieved by an instrumental setup that allows the transmission of a limited band of radiation, and the subsequent aliasing of such radiation to lower, radiation-free frequencies. The overall effect of this approach is to improve the selectivity of this analysis method for the determination of glucose.

## 2. Experimental

### 2.1. Apparatus and reagents

The data employed in this work have been described previously [12]. The data collection will be summarized here. Interferograms were collected with a Digilab FTS-60A FT spectrometer (Bio-Rad, Cambridge, MA) configured for near-IR operation. The instrument was equipped with a 100 W tungsten-halogen lamp, a CaF<sub>2</sub> beam splitter, and a liquid nitrogen-cooled InSb detector. Samples were contained in a 2-mm path length Infrasil quartz transmission cell. A K-band optical interference filter

(Barr Associates, Westford, MA) was placed before the sample cell holder to limit the near-IR radiation passing through the cell and striking the detector to the 4000–5000  $\text{cm}^{-1}$  range. Sample temperatures were kept at 37–38°C through the use of a water-jacketed cell holder and refrigerated temperature bath.

All solutions were prepared with reagent grade water which was generated by passing house distilled water through a Milli-Q Plus water purification unit (Millipore, Bedford, MA). Reagent grade glucose, potassium phosphate salts, 5-fluorourasil, and sodium benzoate were purchased from common suppliers. Triacetin and BSA (Cohn Fraction V powder, product no. A4503) were purchased from Sigma (St. Louis, MO).

## 2.2. Procedures

### 2.2.1. Standard solutions

Stock solutions of glucose, BSA and triacetin with concentrations of 50 mM (9 g/l), 190 g/l and 35 g/l, respectively, were prepared by dissolving pure reagent in 0.1 M, pH 7.4 phosphate buffer to which 0.05% (w/w) 5-fluorourasil or sodium benzoate had been added as a preservative. A factorial design experiment was constructed with 10 glucose levels (1, 3, 5, 7, 9, 11, 13, 15, 17, and 19 mM), four BSA levels (50, 65, 80 and 95 g/l), and four triacetin levels (1.4, 2.1, 2.8, and 3.5 g/l). This produced a total of  $10 \times 4 \times 4 = 160$  samples, with 16 different combinations of BSA and triacetin for each glucose level. The glucose levels were chosen to span a clinically relevant concentration range in human blood, while the BSA and triacetin levels spanned the full range of concentrations of total protein and triglycerides, respectively, that are encountered in human blood samples.

### 2.2.2. Data collection

Interferogram data were collected in triplicate for each sample by acquiring three consecutive interferograms of the same sample. The data were collected as 256 co-added, single-sided interferogram scans of 16 384 (16K) points of each replicate. The interferograms were sampled at every zero-crossing of the He–Ne reference laser, yielding a maximum spectral frequency of 15801  $\text{cm}^{-1}$ . The 160 samples were measured in a random order with respect to the

glucose, triacetin, and BSA concentrations. This ensured that there were no correlations between time-dependent data artifacts and concentration. Using software resident on the Bio-Rad SPC-3200 computer controlling the spectrometer, these interferograms were also Fourier processed with triangular apodization and Mertz phase correction, yielding single-beam spectra with a point spacing of 1.93  $\text{cm}^{-1}$ . The collected interferograms and computed spectra were transferred to a Silicon Graphics 4D/460 computer operating under Irix (version 5.3, Silicon Graphics, Mountain View, CA). All computer software used in processing the spectral and interferogram data was implemented in FORTAN 77. Fourier transform and multiple linear regression computations were performed with subroutines from the IMSL software package (IMSL, Houston, TX). All data analysis computations were performed with a Silicon Graphics Indigo<sup>2</sup> Impact 10000 workstation operating under Irix (version 6.2). The equivalence of the Fourier processing steps performed on the Bio-Rad and Silicon Graphics computers was verified.

Background spectra of the 0.1 M phosphate buffer were collected before the first, and then after every third or fourth sample. Spectra in absorbance units were subsequently obtained by ratioing the single-beam spectra of the samples to the most recently collected single-beam spectrum of the buffer. The background measurements were only used in the construction of absorbance spectra since the interferogram-based analysis method requires no separate measurement of the background.

The decimated (slower sampled) interferogram data sets were derived from the full-size (16K) interferogram set, using the following procedure. The centerburst in the filtered 16K interferograms was located, and starting with the centerburst, every other point was retained while the rest were discarded. This procedure resulted in a new set of interferograms containing 8192 (8K) points. This procedure was also used to obtain interferograms of 4096 (4K) and 2048 (2K) points by retaining every fourth and eighth point, respectively.

### 2.2.3. Data partitioning

Each of the interferogram data sets obtained as previously outlined was randomly partitioned into

calibration and prediction sets, consisting of 75% (120 samples, 360 interferograms) and 25% (40 samples, 120 interferograms) of the full data set (160 samples, 480 interferograms), respectively. The calibration set was further divided into a calibration subset and a monitoring set, in a ratio of 0.8 (96 samples, 288 interferograms) to 0.2 (24 samples, 72 interferograms), respectively. The replicate interferograms corresponding to a given sample were carried together into the selected subsets. The same calibration, monitoring, and prediction sets were used in the analysis of the various interferogram and spectral data sets. The calibration subset was used to develop PLS regression models, and the monitoring subset was used to evaluate the performance of these models. This procedure was used to optimize model sizes, digital filtering parameters, and interferogram and spectral ranges for the various data sets. Once the optimal parameters were selected, the calibration and monitoring subsets were recombined to form the full calibration set, which was then used to build the final model with the previously selected parameters. Models built in this way were tested with the prediction set, and the results reported in this paper are derived from this last step. The data in the prediction set were not used in any way in the development of the models, and thus served as an independent validation set.

### 3. Results and discussion

The constituents of the sample matrix used here have dominating near-IR features that have the potential to interfere in a spectral or interferogram-based determination of glucose focussed on the 4000–5000  $\text{cm}^{-1}$  region. Previous work has established that water, protein, and triglycerides constitute the most significant spectral interferences to glucose in this spectral region [5,7]. This sample matrix thus serves as a convenient simulant to an actual biological matrix such as blood plasma or serum.

Fig. 1 shows near-IR absorbance spectra of stock solutions of (A) glucose (50 mM), (B) triacetin (35 g/l) and (C) BSA (190 g/l) in the 4200–4800  $\text{cm}^{-1}$  region. Three relatively broad bands are observed in the glucose spectrum: an O–H combination band centered at 4700  $\text{cm}^{-1}$  and two C–H combination

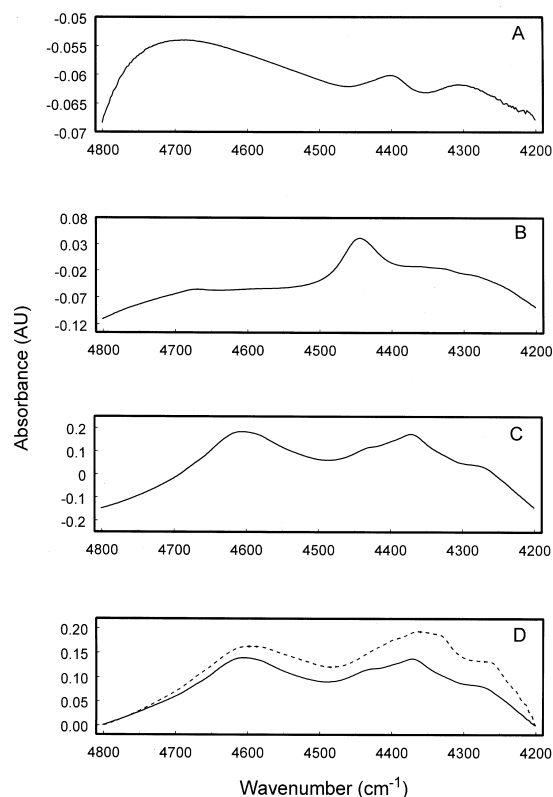


Fig. 1. Near-IR absorbance spectra of (A) 50 mM glucose, (B) 35 g/l triacetin, (C) 190 g/l BSA, and (D) (solid line) a mixture of 11 mM glucose, 2.1 g/l triacetin, and 80 g/l BSA in 0.1 M, pH 7.4 phosphate buffer. The dashed line in (D) is a spectrum of human serum in which the glucose, triglyceride, and total protein concentrations were 11 mM, 3.01 g/l, and 71 g/l, respectively. The point spacing is 1.93  $\text{cm}^{-1}$  in all spectra.

bands at 4400 and 4300  $\text{cm}^{-1}$ . The triacetin spectrum shows a large C–H combination band centered at 4450  $\text{cm}^{-1}$ , while the BSA spectrum has N–H and C–H combination bands centered at 4610 and 4360  $\text{cm}^{-1}$ , respectively, with the 4360  $\text{cm}^{-1}$  band exhibiting shoulders at 4400 and 4270  $\text{cm}^{-1}$ . The solid line in Fig. 1D is one of the collected mixture spectra in which the glucose, BSA, and triacetin concentrations were 11 mM, 80 g/l, and 2.1 g/l, respectively. The similarity of the mixture spectrum to the spectrum of the BSA stock solution confirms that the protein absorption bands are the dominant spectral features in the mixture spectrum. For comparison, the dashed line in Fig. 1D is an absorbance

spectrum of human serum [14]. The two spectra were translated to a minimum of zero absorbance to facilitate comparisons. The glucose, total protein, and triglyceride concentrations in the serum sample were 11 mM, 71 g/l, and 3.01 g/l, respectively. To compute the spectrum in absorbance units, the single-beam spectrum of the serum sample was ratioed to a background spectrum of phosphate buffer. Comparison of the spectra in Fig. 1D reveals significant similarity. The dominant features in the serum spectrum are also derived from the proteins present.

It has previously been established [4,5,12] in our laboratories that the  $4400\text{ cm}^{-1}$  band of glucose is the most useful band in the  $4000\text{--}5000\text{ cm}^{-1}$  region for implementing a quantitative analysis. The triacetin and BSA bands occur in close proximity to the glucose bands, and they also have relatively large intensities compared to the bands of glucose because of the higher concentration of BSA and triacetin in the biological matrix. Extraction of the glucose signatures from those of BSA, triacetin, and water thus represents a significant challenge.

### 3.1. Implementation of time-domain filtering

A procedure for implementing an interferogram-based analysis using an approximate linear relationship between analyte concentration and interferogram intensity derived from the Beer–Lambert law was recently developed in our laboratories [9]. The analysis procedure begins by bandpass filtering the collected interferogram data using filters whose frequency responses are centered on frequencies corresponding to an absorption band of the analyte of interest. Subsequently, short segments of the resulting filtered interferograms are used as input for building multivariate calibration models. It was also found in previous work that when there is more than one absorbing component varying in the sample matrix and the analyte bands are in close proximity to the interference bands, the assumption of an approximate linear relationship between analyte concentration and interferogram intensity is violated [15]. This problem of nonlinearity was addressed in the current work through the use of a quadratic PLS algorithm that was described by Wold et al. [16]. The

models used in the direct interferogram analysis have the form

$$c_i = b_0 + b_1 x_{i,1} + \dots + b_h x_{i,h} + b_{h+1} x_{i,1}^2 + \dots + b_{2h} x_{i,h}^2 \quad (1)$$

where  $c_i$  is the concentration of sample  $i$  predicted by the model, the  $b$  terms are regression coefficients, and the  $x$  values are PLS scores corresponding to the  $h$  latent variables computed by application of the PLS algorithm to the supplied filtered interferogram intensities in the chosen segment. Computation of the models with stepwise regression allowed deletion of terms in Eq. 1 that did not meet a significance level of 95% on the basis of partial  $F$ -values.

An interferogram normalization procedure that was described previously was also implemented here [9]. In this procedure, the intensity of each interferogram point is ratioed to the root sum of squares of intensities of all points in the interferogram. This procedure is implemented to minimize the effect of interferogram intensity variations that are caused by changes in interferometer alignment from day to day or imprecision in positioning the sample cell within the spectrometer. All results reported in this study for interferogram analysis were obtained from interferograms processed in this way.

Previous studies have implemented the filtering step described above by use of Fourier filters [9,10,12]. An equivalent filtering procedure can be implemented in the time-domain using FIR filters, which are time-domain approximations of the frequency response functions of the filters. The advantage obtained with the use of FIR filters is that the filtering procedure does not require Fourier processing of the full interferogram in order to be implemented. The analysis thus requires the collection of only a short interferogram segment.

FIR filters are generally characterized by the attenuation characteristics in their passband relative to their stopbands. A filter designed to transmit a certain range of frequencies while attenuating others will have little or no attenuation of signals in the passband, centered on the targeted frequencies, and high attenuation in the stopbands, which cover the remaining frequencies outside the passband.

A common problem encountered in the design of FIR filters is the effect of the width of the passband on the attenuation characteristics of the filter [12]. As

the passband width narrows, less attenuation can be achieved in the stopbands. In considering the ramifications of this design constraint, it is important to note that the passband width is defined relative to the total bandwidth of the data. For example, a filter with an absolute passband width of  $100\text{ cm}^{-1}$  would be a very narrow filter in a bandwidth of  $15801\text{ cm}^{-1}$ , and it would be difficult to achieve good attenuation performance in the filter stopbands. If the bandwidth of the data were only  $1975\text{ cm}^{-1}$ , however, the same passband width of  $100\text{ cm}^{-1}$  would represent a filter of moderate width, and it would be much easier to achieve good attenuation in the stopbands. Since spectral features pertaining to an analyte will have the same shape, and hence the same full-width-at-half-height (FWHH) regardless of whether the total bandwidth of the data is narrow or wide, this example illustrates that an improvement in filter performance can be obtained by reducing the data bandwidth.

The previous study [12] that implemented an interferogram analysis of glucose in a biological matrix used data that contained signals that covered a bandwidth of  $15801\text{ cm}^{-1}$ . The target range in this signal was the  $4000\text{--}5000\text{ cm}^{-1}$  region that contains

the C–H, N–H and O–H combination bands of glucose, triacetin, and BSA. The FWHH values of the absorption bands that have previously been found to be useful for quantifying glucose in this range are of the order of  $100\text{ cm}^{-1}$  or less. Analogous to the example cited above, this implies that if we are to transmit information pertaining to glucose selectively, we would have to design very narrow FIR filters since the desired  $100\text{ cm}^{-1}$  passband constitutes a fraction of only  $\sim 0.006$  of the  $15801\text{ cm}^{-1}$  bandwidth. The attenuation characteristics of such a filter may have limited success in isolating glucose information in the presence of the overwhelming BSA and triacetin interference.

Fig. 2 shows the attenuation characteristics of a 200-coefficient FIR filter centered at  $4397\text{ cm}^{-1}$ , corresponding to one of the glucose C–H combination bands, with a FWHH in the passband of  $98\text{ cm}^{-1}$ . This filter has an attenuation of approximately 7 dB in the stopbands relative to the passband. For reference purposes, the dB scale is logarithmic, and it requires an attenuation of 20 dB to accomplish a 10-fold attenuation in the signal. With this level of attenuation of frequencies in the stopbands, it may be difficult to transmit analyte frequencies selectively

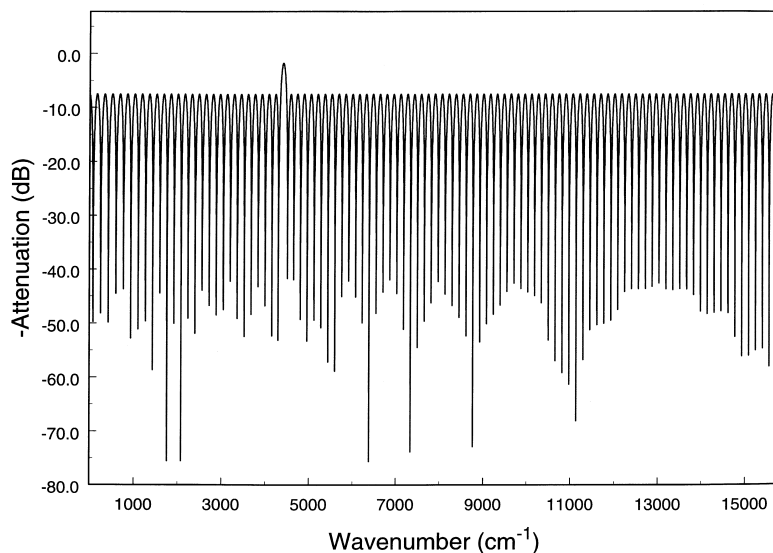


Fig. 2. Attenuation characteristics of a 200-coefficient FIR filter centered at  $4397\text{ cm}^{-1}$ , with a FWHH of  $98\text{ cm}^{-1}$ . This filter has an attenuation of approximately 7 dB in the stopbands relative to the passband, and is designed to operate over a total bandwidth of  $15801\text{ cm}^{-1}$ . Values plotted are the negatives of the attenuation in dB in order to orient the passband in the same direction as an absorbance spectrum.

when there is severe overlap between analyte and interference bands as is the case with glucose in a biological matrix.

### 3.2. Preliminary investigation of full-bandwidth data

To study the actual performance characteristics of FIR filters designed to operate on the 16K interferogram data, which spans a  $15801\text{ cm}^{-1}$  bandwidth, a grid search was performed in which the filter width, interferogram segment specification, and number of PLS factors used in the calibration model were optimized. Ten filters centered on the  $4397\text{ cm}^{-1}$  band of glucose were investigated. The filters had FWHH values ranging from  $44$  to  $142\text{ cm}^{-1}$ , increasing in approximate steps of  $11\text{ cm}^{-1}$ . The filter position of  $4397\text{ cm}^{-1}$  was chosen on the basis of an optimization performed in a previous study [12]. The frequency responses of these filters were analogous to that displayed in Fig. 2. Interferogram segment lengths that were studied were varied from 50 to 1000 points, in steps of 50. The locations of these segments were varied from interferogram point 1 to 1500 (relative to the centerburst, where the centerburst point is defined as point 1), in steps of 50, so that each segment size was realized at different locations throughout the entire 1 to 1500-point range. PLS model sizes as specified by the value of  $h$  in Eq. 1 were varied from 5 to 20. Models were constructed from the resulting filtered interferogram data and used to predict the glucose concentrations corresponding to the interferograms in the monitoring set. The performance of these models was evaluated using the standard error of prediction of the monitoring set (SEM)

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

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, and  $\hat{c}_{m,i}$  is the corresponding glucose concentration predicted by the model. This procedure was used to optimize model size, interferogram location, and filter widths. The mean of the standard error of calibration (SEC) and SEM corresponding to

each combination of model size, filter width, and interferogram starting and stopping point was used to rank the results. The value of SEC is computed similarly to that of SEM, with the denominator in Eq. 2 changed to reflect the loss of degrees of freedom corresponding to the number of terms in the model. The parameters that gave the best five results were selected, and using the larger calibration set, the five corresponding final calibration models were computed. These models were subsequently used to predict the concentrations corresponding to the interferograms in the prediction set.

Fig. 3A shows the mean standard error of prediction (SEP) obtained from the five models, and the best prediction results of the five. The calculation of SEP is analogous to that of SEM in Eq. 2. The model size, filter parameters, interferogram segment location, and calibration statistics corresponding to the best of the prediction results are listed in Table 1.

These results are best evaluated in comparison to those obtained from a conventional analysis of ab-

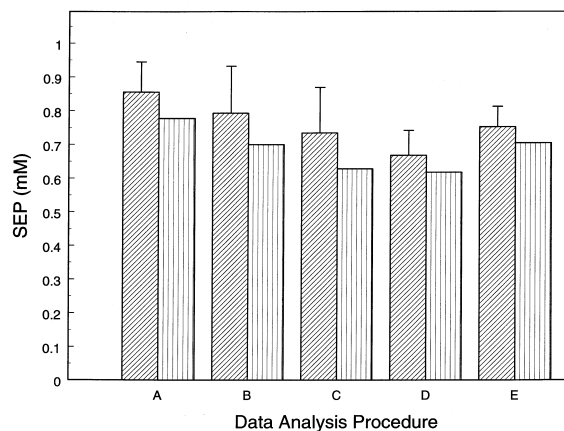


Fig. 3. Mean SEP values of the best five performing PLS models, with error bars denoting the upper 95% confidence limit associated with each mean value (left bar) and lowest SEP of the five models (right bar). These results were obtained with (A) 16K interferogram data filtered with a glucose filter ( $4397\text{ cm}^{-1}$ ), (B) absorbance spectral data, and 2K interferogram data filtered with (C) a glucose filter ( $446\text{ cm}^{-1}$  apparent frequency corresponding to an actual frequency of  $4397\text{ cm}^{-1}$ ), (D) a glucose ( $446\text{ cm}^{-1}$ ) and a triacetin filter ( $492\text{ cm}^{-1}$  apparent frequency corresponding to an actual frequency of  $4446\text{ cm}^{-1}$ ), and (E) a glucose ( $446\text{ cm}^{-1}$ ) and a BSA filter ( $654\text{ cm}^{-1}$  apparent frequency corresponding to an actual frequency of  $4606\text{ cm}^{-1}$ ).

Table 1  
Results from analyses with full-bandwidth data

Analysis type	$R^2$ (%)	SEC (mM)	SEP (mM)	Model size	Interferogram segment (points) or spectral range ( $\text{cm}^{-1}$ )
Interferogram	99.29	0.503	0.777	23 (3) <sup>a</sup>	301–1300 <sup>b</sup>
Absorbance	99.36	0.469	0.708	14	4319–4619

<sup>a</sup>The number of nonlinear terms in the model is indicated in parentheses.

<sup>b</sup>Relative to the interferogram centerburst.

sorbance spectra, in which the single-beam spectra of the samples were ratioed to background spectra of phosphate buffer. The absorbance spectral analysis results, performed in the 4000–5000  $\text{cm}^{-1}$  range, employed spectra with a point spacing of 1.93  $\text{cm}^{-1}$ . A grid search optimization procedure was used to obtain the optimal model size and spectral range in conjunction with a standard linear PLS regression analysis. Models that were explored ranged from 5 to 20 PLS factors in size, and the ranges used were 100 to 600  $\text{cm}^{-1}$ , varied in steps of 100  $\text{cm}^{-1}$ . These ranges were shifted across the 4000 to 5000  $\text{cm}^{-1}$  region in steps of 20  $\text{cm}^{-1}$ . The computed models were evaluated in a manner analogous to that described above.

Fig. 3B shows the mean SEP obtained from the best five models selected through the optimization procedure, along with the best of these prediction results. The parameters and statistics corresponding to this best SEP are also shown in Table 1. The results of the interferogram-based analysis are slightly worse than those obtained with an absorbance spectral analysis.

As indicated by the typical frequency response in Fig. 2, the results produced in the analysis of the interferogram data were obtained with the use of an FIR filter that has quite poor attenuation characteristics. It was hypothesized that a filter with better attenuation in the stopbands relative to the passbands would help in increasing the selectivity of the analysis for glucose information in the data, and thereby improve the overall predictive ability of the PLS regression models. However, to design filters with better performance characteristics requires that the total bandwidth over which the filter is applied be substantially reduced. One approach to overcoming this problem involves taking advantage of aliasing, a characteristic of FT spectrometry that can allow a

higher frequency target range to be observed at lower frequency in a narrower bandwidth [17]. Aliasing the target range to lower frequencies can thus permit the design of FIR filters with good attenuation characteristics in the stopbands relative to the passbands.

### 3.3. Data sampling rate and aliasing

In FT spectrometry, the interferogram corresponding to the spectrum of a sample is composed of signal intensities recorded at discrete intervals of optical retardation. The required sampling frequency is determined by the Nyquist theorem, which states that to record a time-varying sinusoidal waveform unambiguously, it must be sampled at least twice during each cycle. In a measurement that utilizes an FT spectrometer, this implies using a sampling rate that is at least twice the highest modulated spectral frequency. The lower spectral frequencies are automatically successfully sampled since the sampling frequency meets the Nyquist criterion for them.

It is a characteristic of FT spectrometry that unless precautions are taken to ensure that there are no frequencies contained in the signal which are higher than that dictated by the Nyquist sampling criterion, then those higher frequencies will be aliased to lower frequencies in the resulting computed spectrum, superimposed on the true spectral information found there [17]. One way to ensure that a limited range of frequencies is recorded is to employ an optical filter to exclude the unwanted frequencies. When an optical bandpass filter is used to allow only radiation of certain frequencies to be transmitted, then the sampling rate will determine where these frequencies appear in the computed spectrum. If a sampling rate that is below the Nyquist criterion for the transmitted radiation is used, then the resulting computed spec-

trum will show the transmitted radiation appearing at frequencies lower than its true frequencies.

In the data used in this study, a K-Band optical interference filter was employed to isolate the 4000–5000  $\text{cm}^{-1}$  spectral region. Fig. 4 shows a single-beam spectrum of this filter. The fact that frequencies below 4000  $\text{cm}^{-1}$  are suppressed ensures that if the signal is undersampled, it is aliased onto a region of the spectrum that contains essentially no light.

Fig. 5A–C show three versions of an absorbance spectrum of 50 mM glucose corresponding to different data sampling rates. The ranges shown all correspond to 4000–5000  $\text{cm}^{-1}$ . The spectrum in Fig. 5A is recorded at the highest sampling rate used in this study, one interferogram point sampled at every zero crossing of the He–Ne reference laser. The spectrum in Fig. 5B corresponds to a sampling rate that is 1/4 that of the original, and thus has a maximum frequency of 3950  $\text{cm}^{-1}$ . In Fig. 5C, the sampling rate is 1/8 that of the original, and the spectrum has a maximum frequency of 1975  $\text{cm}^{-1}$ . In Fig. 5A, corresponding to a maximum frequency of 15801  $\text{cm}^{-1}$ , the features of glucose in the 4200–4800  $\text{cm}^{-1}$  region appear at their proper frequencies since the maximum frequency is higher than this range. In

the slower sampled spectra, however, the features of glucose that normally occur in the 4200–4800  $\text{cm}^{-1}$  region now appear at 3100–3700  $\text{cm}^{-1}$  in Fig. 5B, and at 250–850  $\text{cm}^{-1}$  in Fig. 5C. When the actual frequencies alias once, the apparent frequencies can be computed as

$$\bar{\nu}_{\text{apparent}} = 2\bar{\nu}_{\text{Nyquist}} - \bar{\nu}_{\text{actual}} \quad (3)$$

where  $\bar{\nu}_{\text{apparent}}$  is the apparent wave number after aliasing,  $\bar{\nu}_{\text{actual}}$  is the actual wave number in the unaliased spectrum, and  $\bar{\nu}_{\text{Nyquist}}$  is the maximum wave number as defined by the Nyquist sampling rate (e.g., 3950  $\text{cm}^{-1}$  for the case depicted in Fig. 5B). In this case the glucose band at 4400  $\text{cm}^{-1}$  aliases to 3500  $\text{cm}^{-1}$ . When the data aliases a second time as in Fig. 5C, the calculation of the apparent frequency can be performed by substituting  $\bar{\nu}_{\text{apparent}}$  from the first aliasing into Eq. 3 as  $\bar{\nu}_{\text{actual}}$ . For the case in Fig. 5C,  $\bar{\nu}_{\text{Nyquist}}$  is now 1975  $\text{cm}^{-1}$  and  $\bar{\nu}_{\text{actual}}$  is 3500  $\text{cm}^{-1}$ . This procedure produces a new  $\bar{\nu}_{\text{apparent}}$  of 450  $\text{cm}^{-1}$ . Despite the changes in apparent frequencies, however, the general shapes and the absorbance levels are maintained in the undersampled spectra in relation to those observed in the true spectrum.

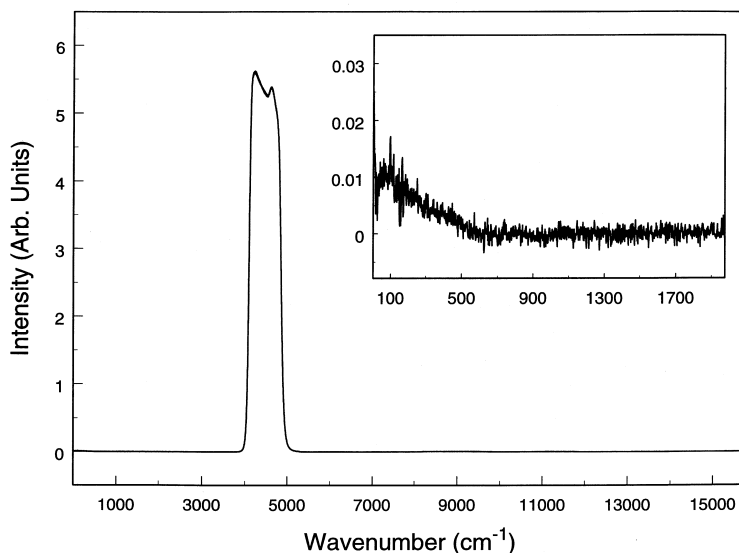


Fig. 4. Single-beam spectrum of a K-band interference filter in the 0–15801  $\text{cm}^{-1}$  bandwidth. Also shown as an insert is a single-beam spectrum of the interference filter in the 0–1975  $\text{cm}^{-1}$  bandwidth.

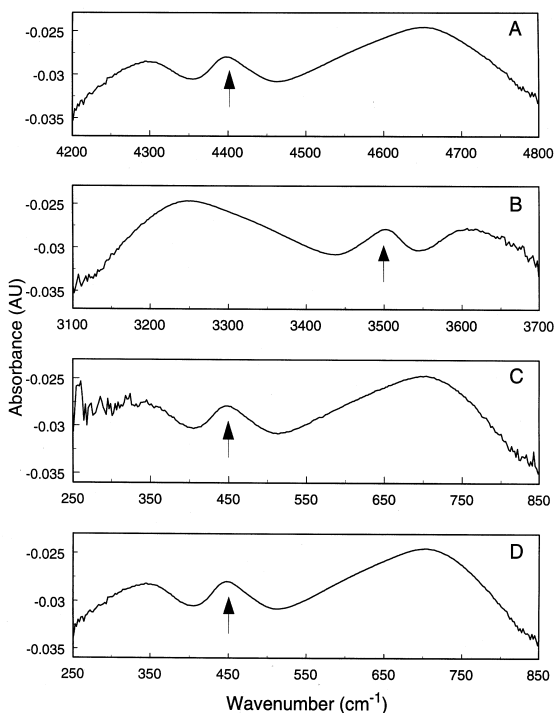


Fig. 5. Spectra of 50 mM glucose in phosphate buffer showing the range corresponding to the 4200–4800  $\text{cm}^{-1}$  region at three different sampling rates. The maximum digitized frequency in each of the spectra is (A) 15801  $\text{cm}^{-1}$ , (B) 3950  $\text{cm}^{-1}$ , and (C) 1975  $\text{cm}^{-1}$ . Plot (D) shows the same spectrum as that in plot (C) after application of the bandpass filter used in conjunction with the decimation procedure. The arrows denote the location of the glucose C–H combination band with a true position of 4400  $\text{cm}^{-1}$ .

A closer inspection of the slowest sampled spectrum reveals that the absorption bands at 0–700  $\text{cm}^{-1}$  appear to have a higher level of noise compared to the corresponding ranges in the faster sampled spectra. An analysis of the noise levels in the full data sets of spectra derived from interferograms of different sampling rates shows that the spectra produced from one level of aliasing have slightly higher noise than the original spectra, while the spectra produced from two levels of aliasing exhibit a significant noise increase.

Fig. 6 shows a plot of pooled root mean-square (RMS) noise levels in the different data sets in the 4300–4500  $\text{cm}^{-1}$  range. This region contains the

4400  $\text{cm}^{-1}$  band of glucose previously found to be useful for quantitative analysis. These noise levels were obtained by ratioing all combinations of the three replicate single-beam spectra of each sample to one another and converting to absorbance units, using a second order polynomial least-squares calculation to model the resulting noise spectra, and computing the RMS value of the deviations about this model in each noise spectrum. The overall noise level was calculated by pooling these RMS values over the 160 samples. It is clear in Fig. 6 that there is a slight increase in the noise level as the sampling rate is reduced to 1/4 the original rate (one level of aliasing), and a 50% increase when the sampling rate is further reduced to 1/8 of the original rate (two levels of aliasing).

The occurrence of this increased noise can be more clearly observed in the single-beam spectrum of the interference filter in the 0–1975  $\text{cm}^{-1}$  range, shown as an insert in Fig. 4. The 0–700  $\text{cm}^{-1}$  region has an elevated signal compared to the higher frequency region, with the normal noise features superimposed on it. Since the K-band interference filter effectively blocks light in this region and the InSb detector has essentially no response here, the signals observed arise from electronic noise and from detector and amplifier nonlinearities. This noise will adversely affect an analysis based on analyte signals aliased to this range, and thus reduce the expected advantage of the highly attenuating filters that can be designed to operate here.

To overcome this effect, a strategy was employed to digitally filter the noise that occurs outside the 4000–5000  $\text{cm}^{-1}$  range before undersampling the interferogram. This allows the aliased signal to fall on a region of the spectrum that contains relatively low noise, and also minimizes the occurrence of aliased noise that originates from the 5000–15801  $\text{cm}^{-1}$  region. Fig. 7 shows the attenuation characteristics of the 200-coefficient filter used for this purpose. For the bandwidth of 1975  $\text{cm}^{-1}$ , Fig. 5C and D show absorbance spectra of 50 mM glucose obtained without and with, respectively, the use of the filter in conjunction with the decimation of the interferogram. The level of noise in the 0–700  $\text{cm}^{-1}$  range has clearly been reduced in Fig. 5D. Fig. 6 shows the corresponding pooled RMS noise levels. It is clear that use of the filter effectively reduces the

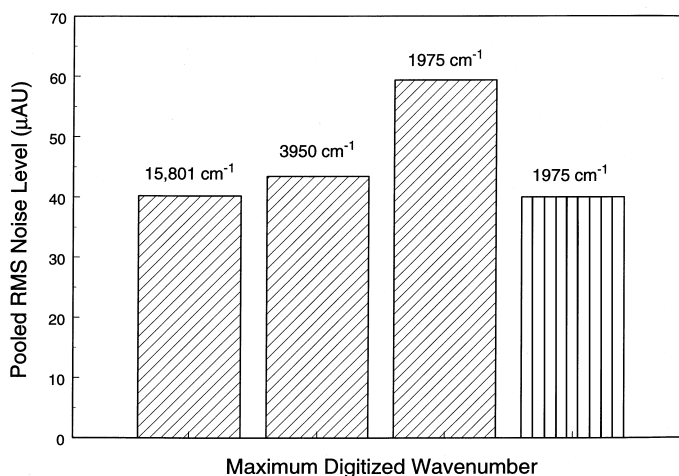


Fig. 6. Pooled RMS noise levels in the 4300–4500  $\text{cm}^{-1}$  range of the spectra corresponding to different sampling rates. Also shown is the noise in the data set with a maximum wave number of 1975  $\text{cm}^{-1}$  before and after (rightmost bar) application of the bandpass filter used in conjunction with the decimation of the data.

noise in the undersampled data to a level that is equivalent to that found in the original data. On the basis of these results, the decimated interferograms obtained with the use of the filter were employed in all subsequent work, and the 1975  $\text{cm}^{-1}$  bandwidth was used.

### 3.4. Single FIR filter

In the undersampled data with a maximum frequency of 1975  $\text{cm}^{-1}$ , the 4397  $\text{cm}^{-1}$  band is observed at 446  $\text{cm}^{-1}$ . In the study performed with the decimated interferogram data, the following glu-

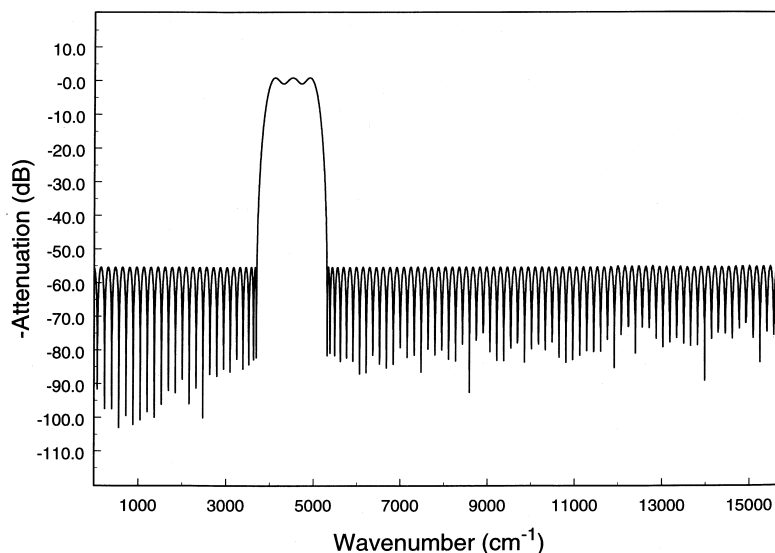


Fig. 7. Attenuation characteristics of a 200-coefficient FIR filter used prior to decimation of the original interferogram data. The passband is centered at 4500  $\text{cm}^{-1}$ , and has a FWHH of 1000  $\text{cm}^{-1}$ . The filter thus transmits the 4000–5000  $\text{cm}^{-1}$  range of frequencies. Values plotted are the negatives of the attenuation in dB in order to orient the passband in the same direction as an absorbance spectrum.

Table 2  
Results from analysis of decimated interferogram data

Analysis type	$R^2$ (%)	SEC (mM)	SEP (mM)	Model size	Interferogram segment
1 filter (glucose)	99.23	0.522	0.628	23 (3) <sup>a</sup>	31–180 <sup>b</sup>
2 filters (glucose, triacetin)	99.39	0.484	0.618	26 (7)	21–120 <sup>c</sup>
2 filters (glucose, BSA)	99.09	0.570	0.705	24 (4)	171–220 <sup>d</sup>

<sup>a</sup>The number of nonlinear terms in the model is indicated in parentheses.

<sup>b</sup>Relative to the interferogram centerburst.

<sup>c</sup>Segment for the triacetin filter.

<sup>d</sup>Segment for the BSA filter.

glucose filter parameters were explored: filter positions 441, 446 and 451  $\text{cm}^{-1}$  and filter widths 49.0, 57.5, 66.6, 75.1 and 83.6  $\text{cm}^{-1}$  at each filter position. The interferogram segment sizes studied ranged from 25 to 300 points in steps of 25, located between interferogram points 1 and 400, relative to the centerburst. These segments were shifted within the range of points 1 to 400 in steps of 10 until the entire range was covered. The filtered interferogram data were used to build quadratic PLS calibration models. The value of  $h$  in Eq. 1 was varied from 5 to 20. The resulting models obtained in the grid search optimization were evaluated as before by use of the

values of SEC and SEM. Table 2 shows the results obtained through this procedure.

As displayed in Fig. 3C, the results obtained with the decimated data show an improvement in the mean SEP and the lowest SEP obtained compared to those obtained using the full-bandwidth data. The optimal interferogram segment was 150 points long, located between points 31 and 180 (relative to the centerburst). The filter selected as optimal had a FWHH of 66.6  $\text{cm}^{-1}$ , and was centered at 446  $\text{cm}^{-1}$ . Fig. 8 shows the frequency response of this 200-coefficient filter. It is clear, when compared to the filter depicted in Fig. 2, that significantly better

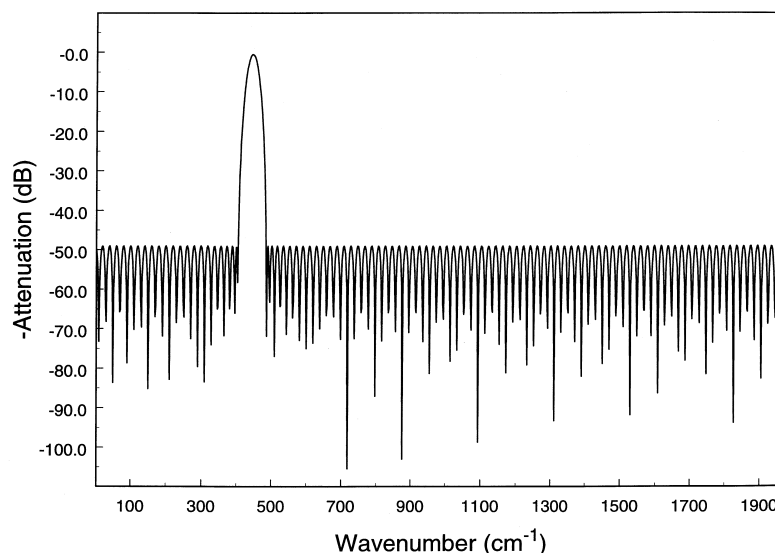


Fig. 8. Attenuation characteristics of a 200-coefficient FIR filter centered on the 4397  $\text{cm}^{-1}$  band of glucose band (apparent frequency 446  $\text{cm}^{-1}$ ), with a FWHH in the passband of 66.6  $\text{cm}^{-1}$ . This filter is designed to operate over the 1975  $\text{cm}^{-1}$  bandwidth. Values plotted are the negatives of the attenuation in dB in order to orient the passband in the same direction as an absorbance spectrum.

attenuation in the stopbands has been obtained by reducing the bandwidth of the data. The 50 dB of attenuation in the stopbands represents a 300-fold reduction in signals there relative to those in the passband. This increased attenuation enhances the selectivity of the analysis for glucose information.

### 3.5. Use of multiple FIR filters

The previous study that implemented an interferogram analysis based on multiple filters used various Fourier digital filters centered on glucose, triacetin and BSA absorption bands [12]. It was found that the use of multiple filters centered on the bands of the analyte and matrix interferences helped in enhancing the selectivity of the analysis for glucose information. The filters used were centered on the glucose band at  $4397\text{ cm}^{-1}$ , the triacetin band at  $4446\text{ cm}^{-1}$ , and the BSA band at  $4606\text{ cm}^{-1}$ . The corresponding positions in the  $1975\text{ cm}^{-1}$  bandwidth used in the current study are  $446$ ,  $492$  and  $654\text{ cm}^{-1}$  for the glucose, triacetin, and BSA filters, respectively. The strategy used to implement multiple filtering was to apply the FIR filters to as many identical copies of each raw interferogram as the number of filters being used, two in this case, and then optimize the segment

positions and sizes corresponding to each filter. The following filter combinations were used: (1) the glucose filter alone; (2) the glucose and the triacetin filters together; and (3) the glucose and BSA filters together. The interferogram segments arising from the use of each filter were then concatenated and used together in the PLS analysis.

Using the previously optimized glucose filter and interferogram parameters and keeping them constant, the grid search optimization procedure for the second segment was performed by varying the positions of the BSA and triacetin filters. These filters were applied to the respective copies of the interferograms, and the resulting filtered interferograms were used as the second segments in the analysis, concatenated to the first segment before being input into the quadratic PLS regression algorithm. The optimizations were performed as before, and only the first segment was kept constant throughout the optimization procedure, while the second segment was varied. Table 2 shows the results obtained by using filters centered on the BSA and triacetin bands to filter the second segments, while Fig. 3 shows the mean SEP values of the best five models and the corresponding lowest SEP.

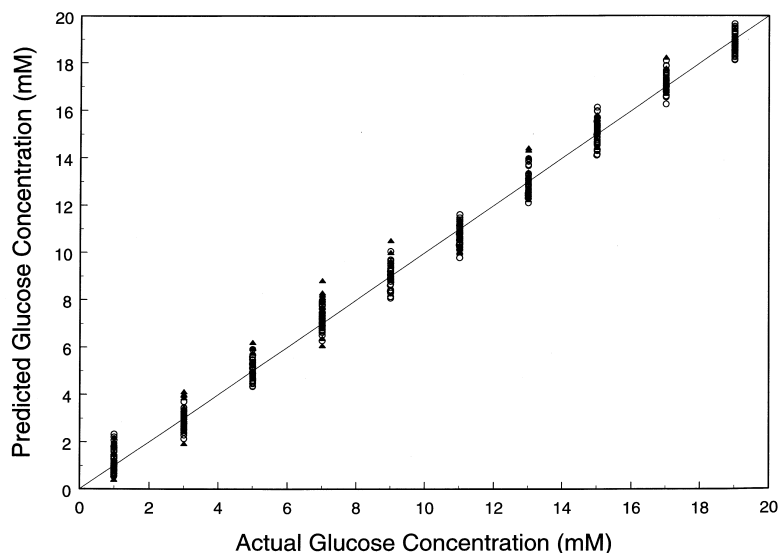


Fig. 9. Correlation plot of predicted vs. actual glucose concentrations produced with the best performing model based on decimated interferogram data and using glucose and triacetin filters. Open circles represent the calibration data, and the closed triangles indicate the prediction data.

The mean SEP obtained for the glucose and triacetin filter combination shows an improvement over the single glucose filter. A 20% improvement in SEP is noted relative to the results reported in Table 1 in which the full-bandwidth interferogram data were used. Improvement is also noted relative to the SEP obtained with the conventional analysis of absorbance spectra reported in Table 1.

The results of the glucose and BSA filter combination are not as good as those of the glucose/triacetin combination, but are comparable to those of the single glucose filter. This indicates that use of the BSA filter adds little or no new information to the PLS algorithm to help in selectively modeling glucose concentration information. The triacetin filter, on the other hand, does a good job of helping the PLS algorithm separate the glucose and triacetin concentration information. Since the triacetin filter is positioned on the triacetin band that is centered at  $492\text{ cm}^{-1}$  (actually  $4446\text{ cm}^{-1}$ ), this filter may be isolating some information from the  $446\text{ cm}^{-1}$  (actually  $4397\text{ cm}^{-1}$ ) glucose band, in addition to the triacetin band.

A correlation plot of predicted vs. actual glucose concentrations is shown in Fig. 9. The results plotted are produced by the best performing model using the glucose and triacetin filters with the decimated interferogram data. The calibration data are represented by the open circles while the prediction data are indicated by the solid triangles. The  $R^2$  value for this model was 99.39%. Both the prediction and calibration data are tightly distributed around the unity line throughout the concentration range, confirming that a strong correlation exists between the actual and predicted glucose concentrations. No outliers were detected when corresponding residual plots were examined, and there were no relationships noted between residuals and predicted glucose concentrations.

#### 4. Conclusions

The work presented here has further demonstrated the feasibility of implementing a determination of glucose in a synthetic biological matrix by direct analysis of filtered near-IR interferogram data. The matrix used in this work includes the most significant spectral interferences to the glucose determina-

tion. This lends confidence to the extension of the methodology to actual biological matrixes such as blood serum or plasma.

The best results obtained in the direct interferogram analysis were superior to those obtained in a conventional analysis of absorbance spectra. The advantage of this approach is that it is implemented without use of a background measurement, which would otherwise be impractical to obtain in a noninvasive spectral determination. The fact that only short interferogram segments are required for the analysis makes it compatible with alternative interferometer designs that may allow FT spectrometry to be used in a practical solution to the noninvasive glucose monitoring problem.

Comparison of the results presented in Tables 1 and 2 reveals that the calibration models require a relatively large number of terms and that substantially more terms are used in the interferogram-based models compared to the models computed with absorbance spectra. The large number of terms arises from the fact that the glucose signal is a very small component of the overall signal, and is spectrally overlapped with the signals arising from the other components of the sample matrix. The difference in the number of terms between the interferogram-based and spectral-based models arises partially from the use of the quadratic PLS method with the interferogram data since a larger pool of independent variables is submitted to the regression procedure in this case. Another factor is the removal of the water background signal in the spectral data through the use of the spectrum of phosphate buffer.

In comparing the models, interpretation of the PLS loading weights or spectral loadings is speculative at best, particularly for the interferogram-based method in which the loadings resemble waveforms. Relative to its performance with the absorbance spectra, the PLS calculation appears to be less efficient in decomposing the interferogram data into the underlying factors that allow the glucose concentrations to be modeled. The underlying reasons for this lower efficiency are not readily apparent. This issue remains under study.

The use of decimated interferogram data in this work has shown how selectivity for glucose can be significantly improved by taking advantage of the characteristics of FIR filters when applied over nar-

row bandwidths. The attenuation of the FIR filters is greatly improved by using narrow bandwidth data, which allows frequencies pertaining to the glucose molecule in a certain band to be transmitted selectively, while those of the background and matrix components are suppressed. The use of a filter in conjunction with the decimation step is crucial to prevent the spectral signals from being aliased onto noise originating from the detector and electronics.

Multiple FIR filters were also used in this study and showed that when information pertaining to the variation of some matrix components is submitted to the multivariate calibration algorithm, its ability to model glucose concentration changes can be enhanced.

### Acknowledgements

This research was supported by the National Institutes of Health under grant DK45126.

### References

- [1] I. Amato, *Science* 258 (1992) 892.
- [2] M.A. Arnold, in: G.R. Kost (Ed.), *Handbook of Clinical Laboratory Automation, Robotics, and Knowledge Optimization*, Wiley, New York, 1996, pp. 631–647.
- [3] H.M. Heise, in: D.M. Fraser (Ed.), *Biosensors in the Body: Continuous in vivo Monitoring*, Wiley, Chichester, UK, 1997, pp. 79–115.
- [4] M.A. Arnold, G.W. Small, *Anal. Chem.* 62 (1990) 1457.
- [5] L.A. Marquardt, M.A. Arnold, G.W. Small, *Anal. Chem.* 65 (1993) 3271.
- [6] G.W. Small, M.A. Arnold, L.A. Marquardt, *Anal. Chem.* 65 (1993) 3279.
- [7] S. Pan, H. Chung, M.A. Arnold, G.W. Small, *Anal. Chem.* 68 (1996) 1124.
- [8] G. Lu, X. Zhou, M.A. Arnold, G.W. Small, *Appl. Spectrosc.* 51 (1997) 1330.
- [9] M.J. Mattu, G.W. Small, *Anal. Chem.* 67 (1995) 2269.
- [10] M.J. Mattu, G.W. Small, *Appl. Spectrosc.* 51 (1997) 1369.
- [11] P. Hariharan, *Basics of Interferometry*, Academic Press, San Diego, CA, 1992, Chap. 3.
- [12] M.J. Mattu, G.W. Small, M.A. Arnold, *Anal. Chem.* 69 (1997) 4695.
- [13] A.V. Oppenheim, R.W. Schaffer, *Discrete-Time Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1988.
- [14] K.H. Hazen, M.A. Arnold, G.W. Small, *Anal. Chim. Acta* 371 (1998) 255.
- [15] N.A. Cingo, G.W. Small, in: J.A. de Haseth (Ed.), *Fourier Transform Spectroscopy: 11th International Conference*, American Institute of Physics, Woodbury, NY, 1998, pp. 231–234.
- [16] S. Wold, N. Kettaneh-Wold, B. Skagerberg, *Chemom. Intell. Lab. Syst.* 7 (1989) 53.
- [17] P.R. Griffiths, J.A. de Haseth, *Fourier Transform Infrared Spectroscopy*, Wiley, New York, 1986.