

# Automatic Baseline Correction by Wavelet Transform for Quantitative Open-Path Fourier Transform Infrared Spectroscopy

LIMIN SHAO<sup>†</sup> AND PETER R. GRIFFITHS\*

Department of Chemistry, University of Idaho,  
Moscow, ID 83844-2343

A technique for automatically correcting the baseline of open-path Fourier transform infrared spectra is described in which the spectra are decomposed into high-frequency (details) and low-frequency (discrete approximations) information using a wavelet transform. After an appropriate number of iterations,  $n$ , the discrete approximation simulates the baseline of the spectrum. By setting the  $n$ th approximation to zero and reversing this process, the reconstructed signal contains only the high-frequency components in the original signal on a baseline that is approximately flat and at zero absorbance. When small molecules such as  $\text{NH}_3$  and  $\text{CH}_4$  are quantified by partial least-squares (PLS) regression, this process decreases the number of eigenvectors required for the analysis by two or three, increasing the ruggedness of the prediction. The baseline that is calculated is approximately the same for spectra measured at resolutions of 1, 2, 4, 8, and  $16\text{ cm}^{-1}$ . Surprisingly, the baseline that is calculated is not totally smooth, with artifacts that are caused by the analyte. The amplitude of the artifacts is directly proportional to the concentration of the analyte, so the accuracy of PLS regression is not degraded.

## Introduction

For quantitative atmospheric analysis by open-path Fourier transform infrared (OP/FT-IR) spectrometry, two multivariate algorithms, classical least-squares (CLS), and partial least-squares (PLS) regression, have been applied. CLS was the first to be applied to OP/FT-IR spectrometry and is still the more popular technique (1, 2) even though it rarely gives more accurate results than PLS (3). In practice, the accuracy of both techniques is improved if the spectral baseline is set accurately at zero absorbance units. The effect of a baseline offset or Beer's law nonlinearity on the results of PLS regression may be taken into account by adding more factors (eigenvectors) to the calibration matrix; however, it is generally recognized that the ruggedness of PLS regression is improved when the number of factors is minimized.

Many methods of manual and automated baseline correction have been developed (4), some of which require the operator to select appropriate "baseline points" while others are implemented automatically. Most of the baseline correction algorithms that have been applied to OP/FT-IR

spectra involve the operator manually choosing several wavenumbers at which no atmospheric specie (e.g.,  $\text{H}_2\text{O}$  or  $\text{CO}_2$ ) or analyte is believed to absorb. A curve such as a polynomial, spline, or sigmoid is then fit to these points and subtracted from the measured spectrum. Methods such as the noise median method (5), baseline standard deviation (6), threshold-based classification (7), and signal removal (8) fall into this category. If the baseline points are selected incorrectly, however, a significant quantitative error can result. The lower the concentration of the analyte, the more important it is to estimate the baseline correctly, but the more difficult it is to do so in practice.

The performance of most automated baseline correction methods is governed largely by two factors: the effect of noise and the capability of the algorithm to distinguish a broad absorption band, such as would be found in the spectrum of a fairly large molecule with unresolved rotational fine structure, from relatively rapid changes in the baseline caused by narrow spectral lines. Because trained spectroscopists are usually able to make this distinction, many manual baseline correction methods yield satisfactory results. However, any algorithm that requires significant human intervention is slow and subject to operator bias. OP/FT-IR measurements are capable of time resolution of less than a minute; if a manual baseline correction step is required for each spectrum, the time resolution will be dramatically degraded. It has been found that the accuracy of quantitative measurements that involve automatic baseline correction is often worse than that of manual modes in which human interactions are involved (9). Thus the development of automatic baseline correction methods is driven largely by the task of processing the enormous amount of data generated during continuous monitoring without loss of significant spectroscopic information. If OP/FT-IR spectrometers are to be operated unattended for many days, the development of an accurate automated baseline correction algorithm that is rugged enough to withstand the effect of changes of optical alignment and humidity levels is of critical importance. In this paper, we describe a rugged approach to baseline correction that is well suited to automated instrumental operation.

Two factors with respect to baseline correction should be considered. One is that in most cases, baseline correction is an intermediate step of the qualitative or quantitative analysis. Thus it is the *effect* of the baseline offset, rather than the baseline itself, that must be eliminated. Second, baseline correction is more necessary in quantitative than in qualitative analysis, since spectral feature lines for the identification are still recognizable with the baseline embedded. Based on these two considerations, accurate, automated baseline correction is needed for quantitative OP/FT-IR measurements. In this report, we demonstrate that wavelet transforms (WTs) can be employed for this purpose for OP/FT-IR spectra because of their high automation and robustness.

## Theory

**Basic Theory of Wavelet Transforms.** During the past decade, wavelet transforms have been successfully applied to signal processing in chemistry for such operations as data compression, denoising, and baseline correction (10). Wavelets are functions derived from a basis,  $\psi(t)$ , through dilation and translation:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), \quad (a, b, \in R, a \neq 0) \quad (1)$$

\* Corresponding author phone: 208-885-5807; fax: 208-885-6173; e-mail: pgriff@uidaho.edu.

<sup>†</sup> Current Address: Department of Chemistry, University of Science and Technology of China, Hefei, Anhui, 230026, People's Republic of China.

where  $a$  and  $b$  are parameters to control the dilation and the translation, respectively. By analogy to the Fourier transform, the wavelet transform and the inverse transform of function  $f(t)$  can be defined as this:

$$W_f(a,b) = \int_{-\infty}^{+\infty} f(t)\psi_{a,b}^*(t)dt \quad (2)$$

$$f(t) = \frac{1}{C} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_f(a,b)\psi_{a,b}\left(\frac{t-b}{a}\right) \frac{dadb}{a^2} \quad (3)$$

where

$$C = \int_{-\infty}^{+\infty} |\hat{\psi}(\omega)|^2 |\omega|^{-1} d\omega$$

Several discrete wavelet transform (DWT) algorithms have been developed (10), the most widely used of which is multiresolution signal decomposition (MRSD) designed by Mallat (11). MRSD, also called the Mallat pyramidal algorithm, has the following form:

$$A_{2^j}^d f(n) = \sum_{k=0}^{N-1} h(k-2n) A_{2^{j+1}}^d f(k) \quad (-J \leq j \leq -1) \quad (4)$$

$$D_{2^j}^d f(n) = \sum_{k=0}^{N-1} g(k-2n) A_{2^{j+1}}^d f(k) \quad (-J \leq j \leq -1) \quad (5)$$

where  $A_{2^j}^d f$  and  $D_{2^j}^d f$  denote the *discrete approximation* and *discrete detail* at resolution  $2^j$ , that contain low-frequency and high-frequency information, respectively, in the original signal.  $J$ , a positive integer, is the preset scale factor;  $n$  is the sequence of the data points in the approximation or the detail at resolution  $2^j$ ;  $N$  is the length of the approximation at resolution  $2^{j+1}$ ; and  $h$  and  $g$  are discrete filters. From these discrete approximations and details, the original signal can be reconstructed by the following procedure.

$$A_{2^{j+1}}^d f(n) = \sum_{k=0}^{N-1} h(n-2k) A_{2^j}^d f(k) + \sum_{k=0}^{N-1} g(n-2k) D_{2^j}^d f(k) \quad (6)$$

Equations 4–6 show that MRSD can be viewed as a recursive operation. At a certain resolution, both the discrete approximation and the detail have the same number of data points, but half that of the previous discrete approximation or detail. After each decomposition, the number of data points of the discrete approximation or detail is halved, whereas the reconstruction doubles the number of data points. Thus when the original spectrum is reconstructed, its length is also restored. The original spectrum should have enough data points that the desired number of decompositions can be performed. If not, the spectrum should be interpolated.

**Baseline Correction Method Based on the Wavelet Transform.** According to the theory of MRSD, after  $n$  decomposition operations, the original signal is decomposed into one discrete approximation (the  $n$ th one) and  $n$  details. The spectral information is not lost, but is rather redistributed into high- and low-frequency components. The  $n$ th discrete approximation contains low-frequency component(s), and the  $n$  discrete details contain the high-frequency components of the original signal.

The original signal can be completely reconstructed from the discrete approximation and details. However, if the  $n$ th approximation is set to zero, the low-frequency information is lost and the reconstructed signal contains only the high-frequency components in the original signal. In OP/FT-IR spectroscopy, the baseline fluctuates slowly, but the spectral information, especially of small molecules with resolvable rotational fine structure, varies at higher frequency. We may

take advantage of this kind of frequency distribution to discriminate between the baseline and the spectral information using a wavelet transform.

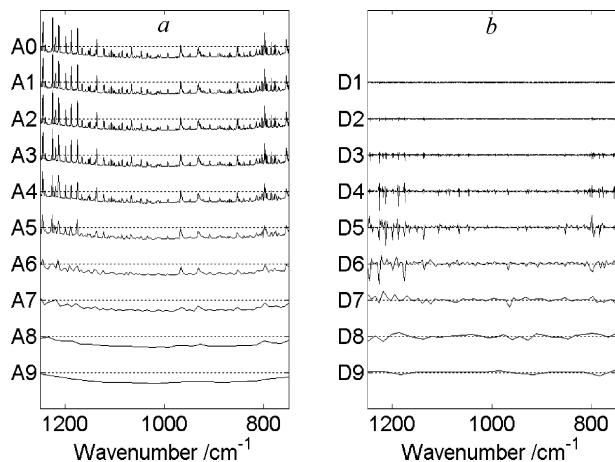
Baseline correction comprises two steps. The original spectrum is first decomposed  $n$  times, with  $n$  being set large enough that no relatively sharp lines are seen in the  $n$ th approximation, i.e., there is no high-frequency information remaining in this approximation. Second, the approximation that features the low-frequency content of the spectrum (which mainly contains the baseline information) is set to zero, and the reconstruction is carried out. The reconstructed signal is the baseline-corrected spectrum.

## Experimental Measurements and Data Analysis

OP/FT-IR measurements were carried out in June and July 2004, and January, March, and June 2005 on fields and a dairy farm in southern Idaho in a cooperative project for atmospheric monitoring with the Northwest Irrigation and Soil Research Laboratory (NWSL) of the United States Department of Agriculture (USDA). The OP/FT-IR spectrometer was manufactured by MDA (Atlanta, GA), and incorporated a Bomem Michelson 100 interferometer, telescope, a cube-corner array retroreflector, and a Sterling-engine-cooled mercury cadmium telluride detector. All OP/FT-IR spectra were computed from interferograms measured with a maximum optical path difference of 1 cm (nominal resolution of  $1 \text{ cm}^{-1}$ ), and computed with a zero-filling factor of 8 and medium Norton-Beer (MNB) apodization.

Quantitative determinations of ammonia and methane were performed by PLS regression of the OP/FT-IR spectra. The PLS calibration data set is prepared with simulated OP/FT-IR spectra synthesized using the following procedure. A total of 54 interferograms were measured at several different path-lengths, temperatures, and relative and absolute humidities over a pristine field upwind of the cattle pens, so that the range of background conditions likely to be encountered during the subsequent analytical measurements was represented. From these interferograms, spectra were calculated at  $1 \text{ cm}^{-1}$  resolution using the MNB apodization function. Transmittance spectra were calculated from the ratio of long-path single-beam spectra and a short-path single-beam background spectrum measured with the retroreflector located within 2 m of the telescope. These spectra were then converted to a linear absorbance format to yield a series of background spectra for the calibration data set. A unapodized high-resolution ( $\Delta\nu = 0.125 \text{ cm}^{-1}$ ) reference spectrum of ammonia in the absorbance ( $-\log T$ ) format obtained from Dr. Pamela Chu of the National Institute of Standards and Technology was then multiplied by 54 different factors so that each one corresponded to a known different path-integrated concentration. The resolution of the reference spectrum was high enough that resolution errors were minimized (12, 13). Each of the 54 spectra was then converted to transmittance and its Fourier transform was computed. The Fourier domain array was then truncated to an optical path difference of 1 cm (corresponding to a nominal resolution of  $1 \text{ cm}^{-1}$ ) and multiplied by a MNB apodization function before conversion back to a linear absorbance format. The 54 spectra of  $\text{NH}_3$  with known path-integrated concentrations were then added to the background spectra in a random manner, and the calibration set for  $\text{NH}_3$  is obtained. For the methane calibration, the global average concentration of  $\text{CH}_4$  was used instead.

For the ammonia calibration, spectra in the region from 1250 to  $750 \text{ cm}^{-1}$  were regressed against the path-integrated concentration of the analyte using the leave-one-out approach with between 1 and 20 factors. The predicted residual error sum squared (PRESS) was then plotted against the number of factors,  $N$ , used for the calculation. The optimal number of factors required to represent the calibration



**FIGURE 1.** Discrete approximations (a) and details (b) from the decomposition of the experimental OP/FT-IR spectrum by wavelet transform. A0 is the original spectrum.

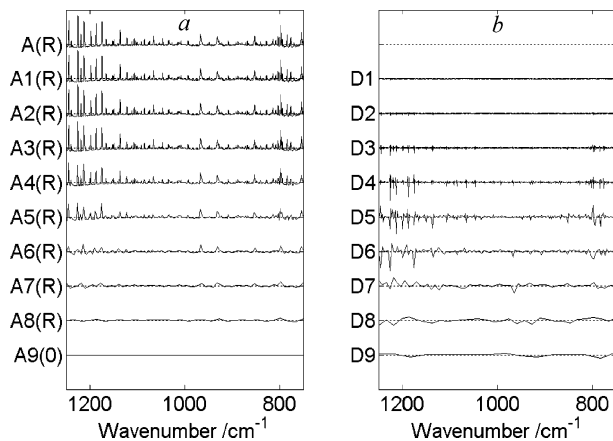
spectra accurately is given by the value of  $N$  giving the lowest PRESS. A similar procedure was used for methane, but in this case, the region from 3200 to 2850  $\text{cm}^{-1}$  was used.

For the wavelet transform, the Daubechies wavelet was used (14), and the length of the discrete filters ( $h$  and  $g$  in eqs 4 and 5) was set at 4. All manipulation of spectra and data processing were performed using MATLAB 7.0.1 (The MathWorks Inc., Natick MA) on the Windows 2000 operating system.

## Results and Discussion

**Procedure and Parameters of Baseline Correction by Wavelet Transform.** Figure 1 shows the results of one OP/FT-IR spectrum after nine decompositions. The spectra on the left side of the figure show the discrete approximations after each decomposition, and the spectra on the right show the details; A0 is the original spectrum with 8192 data points; the  $i$ th discrete approximation or detail has  $8192/2^i$  data points. The fact that the changes in the first two or three spectra are very small and that spectra D1, D2, and D3 are almost zero is expected because the original spectrum was calculated with a zero-filling factor of 8; thus there are seven interpolated data points for every independent datum. It can be seen that all discrete approximations from A1 to A9 show similar slow fluctuations that are caused by changes in the baseline of the spectrum, whereas this information is not seen in the discrete detail spectra. In Figure 1(a) it can also be seen that sharp spectral lines caused by rotational transitions in the spectrum of water vapor and ammonia are gradually separated from the approximations, and completely disappear after the ninth decomposition. The ninth approximation (A9) is seen to only contain the baseline information. The result of setting this spectrum to zero and reconstructing the spectrum is shown in Figure 2. The reconstructed approximation, A(R) in Figure 2(a), is the baseline-corrected spectrum obtained by this method. The comparison between A(R) in Figure 2(a) and A0 in Figure 1 (the original spectrum) clearly indicates that the spectral baseline has been largely corrected.

The results of this procedure show that the number of decompositions,  $n$ , is a very important parameter. Only after enough decompositions had been effected could the baseline be separated entirely from other high-frequency information. In the above instance, if the original spectrum were decomposed six instead of nine times, and A6 is set zero, then the high-frequency spectral information in A6 would be lost as well as the baseline, which would lead to an error in quantification. In our investigation, the number of decom-



**FIGURE 2.** Reconstructed discrete approximations (a) and discrete details (b). A9(0) means the zeroed ninth approximation.

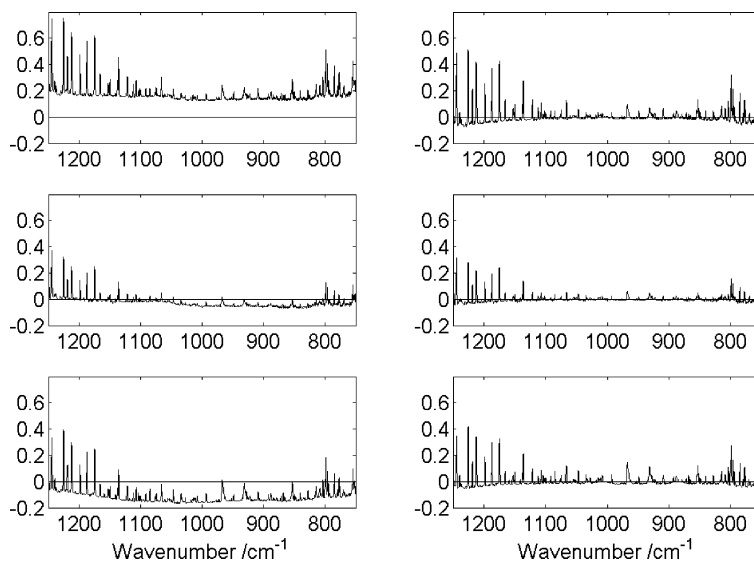
positions required was determined by visual inspection to find when sharp spectral features were completely absent.

To optimize the number of decompositions, three absorbance spectra at resolution of 1  $\text{cm}^{-1}$  with zero-filling factors of 4, 8, and 16 were prepared from the same interferogram, so the number of data points were 4096, 8192, and 16 384, respectively. The necessary decomposition times for baseline correction by the WT were found to be 8, 9, and 11 respectively, following the same criterion described above. Thus when the spectral resolution and the wavenumber region remain unchanged, an inverse correlation exists between the number of decompositions required and the spacing of the spectral data points. In the zero-filling process, zeros are added to the end of the measured interferogram, which results in added interpolated data points to the spectrum after the Fourier transform. It is important to apply zero filling before baseline correction using wavelet transforms since the target spectrum must have enough points to ensure a series of decompositions. Thus all spectra in this investigation were computed with a zero-filling factor of 8. The wavelet transform appears to be able to differentiate the interpolated points from the independent points. In our investigation, application of the wavelet transform was found to yield a few discrete details with values being close to, but not quite equal to, zero, as seen in Figures 1.D1 and 1.D2. WT theory indicates that these near-zero values are classified to be of high frequency. Based on this evidence, we believe the zeros that were added to the interferogram make a small contribution to the near-zero-valued discrete details and, therefore, increase the decomposition times necessary for the baseline correction.

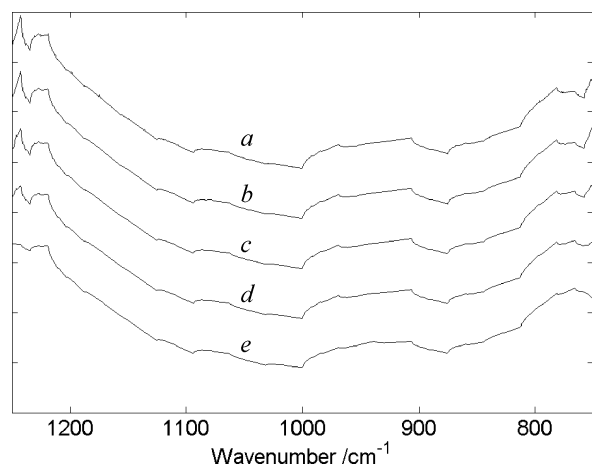
Although the number of decompositions, the type of wavelet, and the discrete filter length will be expected to affect the accuracy to which the baseline may be corrected, the differing effects of these parameters can be largely compensated in a PLS regression by applying the same process to the calibration, validation, and prediction spectra.

**Performance of the Baseline Correction by Wavelet Transform.** To evaluate the performance of the wavelet transform on baseline correction, three OP/FT-IR spectra were computed from interferograms measured at a resolution of 1  $\text{cm}^{-1}$  under various conditions (time of day and year, temperature, path-length, etc.) and the result of applying the procedure outlined above to these spectra is shown in Figure 3. Different baseline patterns of these spectra can be seen from the first column of this figure. Baseline correction by application of the wavelet transform was applied to these spectra with same parameters given above, and the results are shown in the second column of Figure 3. All the baselines were corrected satisfactorily, although the original spectra





**FIGURE 3.** OP/FT-IR spectra (the first column) measured under different conditions and the results (the second column) after baseline correction by wavelet transform.



**FIGURE 4.** Baselines obtained by wavelet transform at resolutions of (a) 1  $\text{cm}^{-1}$ ; (b) 2  $\text{cm}^{-1}$ ; (c) 4  $\text{cm}^{-1}$ ; (d), 8  $\text{cm}^{-1}$ ; (e) 16  $\text{cm}^{-1}$ . Curves are offset for clarity.

differ significantly and the baseline-corrected intensity of all the sharp spectral features is unchanged. However, at each end of the baseline-corrected spectra, the baseline is negative. The cause of this phenomenon will be discussed later in this paper.

The effect of applying this procedure to spectra measured at different resolutions was also investigated since the baseline should be identical at all resolutions. Five spectra were computed from the same interferogram at resolutions of 1, 2, 4, 8, and 16  $\text{cm}^{-1}$ . Zero-filling factors were set at 8, 16, 32, 64, and 128 so that all the spectra have the same number of points, and can be processed by the wavelet transform without changing any parameter. As shown in Figure 4, all the baselines show a high similarity even when the resolution is as low as 16  $\text{cm}^{-1}$ , when all the rotation lines of  $\text{NH}_3$  or  $\text{H}_2\text{O}$  are overlapped and no point on such a spectrum could provide effective baseline information for the correction. Most baseline correction methods based on the baseline estimation fail when spectra are measured at such a low-resolution that no region of the spectrum can be considered to represent only the baseline.

**Cause of the Distortions.** The use of the wavelet transform for baseline correction exhibits the joint merits of objectivity and automation. However, when the baselines between 1000 and 900  $\text{cm}^{-1}$  in Figure 4 are examined, it can be seen that

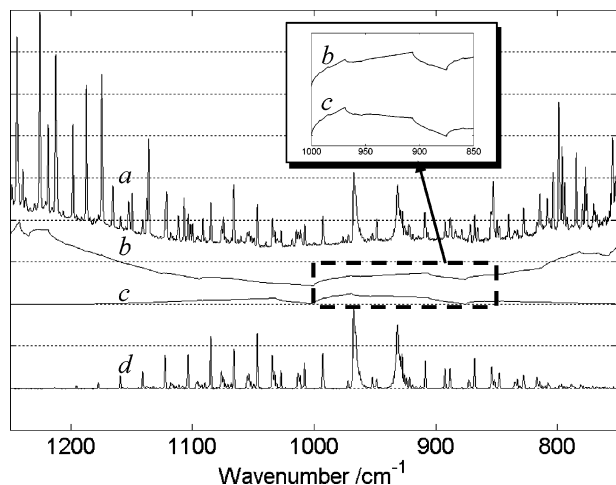
the baseline obtained is not a smooth curve. To explore the origin of these unexpected baseline discontinuities, the same procedure was applied to an OP/FT-IR spectrum with a curved baseline (Figure 5a) and to the reference spectrum of  $\text{NH}_3$  in which no baseline error was present (Figure 5d). As shown by Figure 5b and c, a nonzero baseline is calculated for both spectra. By comparing the two “baselines” in the inset of Figure 5, it can be seen that the distortion is created by the spectrum of  $\text{NH}_3$ . Therefore, the baseline obtained by this procedure is a mixture of the true baseline and the spectral distortions introduced by narrow spectral lines.

As discussed above, the baseline is corrected by setting the last approximation to zero before reconstructing the spectrum. Thus the spectrum can never be completely restored. By comparing the two “baselines” in the inset of Figure 5, the distortion between 900 and 1000  $\text{cm}^{-1}$  can be attributed to the wavelet transform of the sharp ammonia lines. The strong lines in the spectrum of water vapor (1250–1150 and 820–760  $\text{cm}^{-1}$ ) also cause distortions in the baseline, as can be seen in the baseline-corrected spectra in Figure 3. We therefore believe that analogous distortions are introduced whenever the spectral features are sharp. However, the cause of this distortion is not known at this time.

Apparently these distortions do not allow a perfect baseline to be calculated. In quantitative analysis, however, there are two concerns. The more obvious concern is whether the distortions can be minimized. The second is whether the amplitude of the distortion caused by the bands of the analyte ( $\text{NH}_3$  in this case) is linearly proportional to the analyte concentration. If it is, Beer’s law may be expected to be obeyed and multivariate statistical algorithms such as PLS may be expected to predict the concentration of the absorbing species accurately.

Our results suggest that baseline error curves have two characteristics, one of which is component-specific, i.e., a certain distortion is generated from the spectrum of each analyte, and the other is that the distortion introduced by a particular analyte does not change whether it is in a mixture or a pure sample since the wavelet transform is a linear operation. Provided all spectra used in the quantification are processed by the wavelet transform with the same parameters as the OP/FT-IR spectrum, the analyte-dependent error curve should not affect the quantitative accuracy.

**Effect of Distortions on Quantification.** To investigate the effect of the wavelet transform on quantification, two spectra of  $\text{NH}_3$  were prepared with known path-integrated

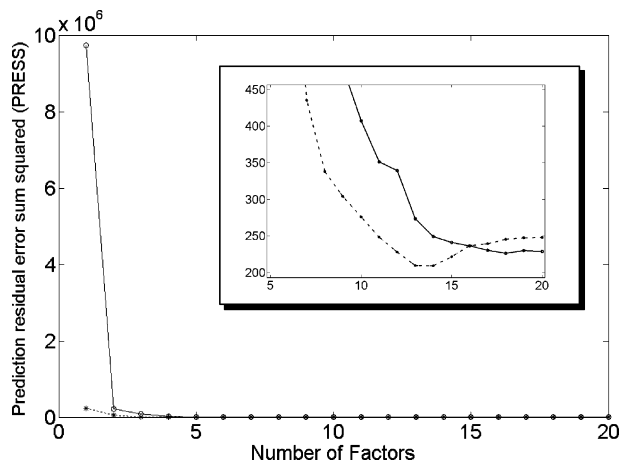


**FIGURE 5.** An OP/FT-IR spectrum (a) and a standard reference spectrum of  $\text{NH}_3$  (d). Curves b and c are baselines obtained by the baseline correction method based on wavelet transform from spectra a and d. Part of these “baselines” is expanded for clarity in the inset.

concentrations; one serves as the analyte to be quantified, and the other is the reference. To investigate nonlinear effects on quantification, a high-resolution ( $\Delta\nu = 0.125 \text{ cm}^{-1}$ ) absorbance spectrum of air-broadened  $\text{NH}_3$  with a path-integrated concentration of 1.00 ppm-m was multiplied by 1 and 10, respectively. The Fourier transforms of these spectra were then calculated and each array was truncated to an optical path difference of 1 cm (i.e., deresolved to a resolution of  $1 \text{ cm}^{-1}$ ) and apodized with the MNB function. Finally inverse Fourier transforms were performed to yield two absorbance spectra at a resolution of  $1 \text{ cm}^{-1}$  (the same as our field spectra) with their nominal path-integrated concentrations being 1 ppm-m and 10 ppm-m. When the 1.0 ppm-m spectrum was scaled and subtracted from the 10 ppm-m spectrum, the least-squares best fit was found when the scaling factor was 9.969, i.e., the path-integrated concentration of the analyte in the second spectrum was calculated to be 9.969 ppm-m. After baseline correction with the wavelet transform, the error was reduced to 0.02%.

The baseline discontinuities were initially believed to be caused by the fact that even the sharp lines in the vibration-rotation spectrum of small molecules such as  $\text{NH}_3$  contain low-frequency components. Rotational lines in the air-broadened vibrational spectrum of a small molecule have a Lorentzian shape, with the broadening coefficient usually being between 0.1 and 0.2  $\text{cm}^{-1}/\text{atm}$ . Even though the slope of the line near the center is very sharp, Lorentzian lines have an extended tail over which the slope is far smaller, i.e., the information in the wings of Lorentzian lines is largely of low frequency while that near the center is mainly high frequency. Thus baseline correction using the wavelet transform, or for that matter any algorithm that removes the low-frequency components of a spectrum, should always introduce an error when the spectrum contains Lorentzian features. Since the tail of Gaussian lines is less pronounced than that of Lorentzians, any baseline correction method that is based on distinguishing between the high and low-frequency components should be more successful in applications where the line shapes are Gaussians.

We originally believed that this argument explained why the baselines in the spectra shown in Figure 3 are never corrected accurately in spectral regions where there are several strong water vapor lines. Unfortunately, the argument does not hold up in practice. When either Lorentzian or Gaussian lines were synthesized and subjected to the baseline correction method outlined above, the same type of artifacts



**FIGURE 6.** Plot of PRESS values against number of factors used in a PLS cross validation for ammonia without (solid line) and with (broken line) baseline correction.

were still seen in the baseline. Changing the length of the array or applying a more symmetrical wavelet (e.g., the symmlet) did not affect this result. In any event, even though some distortion is introduced by the application of the wavelet distortion, our results showed that the quantitative accuracy is significantly improved. We have shown that this conclusion holds true even when the resolution is as low as  $8 \text{ cm}^{-1}$ .

**Application to PLS.** In building a PLS model the first step is to determine the appropriate number of factors. Generally speaking, the smaller the number of factors for a given analysis, the more robust is the method. For OP/FT-IR analysis by PLS regression, it is generally found that ten or more factors are required to account for such variables as resolution errors, the effect of detector nonlinearity, the effect of temperature on vibration-rotation spectra of the analyte molecules and atmospheric  $\text{H}_2\text{O}$  and  $\text{CO}_2$ , and baseline and frequency shifts caused by misalignment of the telescope, retroreflector, and interferometer. If the effect of baseline changes could be compensated, e.g., by the application of wavelet transforms, the number of factors required to describe OP/FT-IR spectra should be able to be reduced and the robustness of the process increased.

A calibration set comprising 54 spectra was prepared in the way described in Experimental measurements and data analysis. These spectra were then subjected to cross-validation using the “leave-one-out” procedure to find the optimum number of factors required for PLS regression for ammonia in the spectral region from 1250 to  $750 \text{ cm}^{-1}$ . The plot of the PRESS value vs the number of factors,  $N$ , used in the calibration when the baseline of the spectra had not been corrected is shown as the solid line in Figure 6. The corresponding data obtained after the baseline had been corrected are shown as the broken line. Clearly, the effect of baseline correction is to dramatically reduce the PRESS value even when only 1 or 2 factors were used. The inset figure in Figure 6 shows that a clear minimum in the plot is seen for baseline corrected spectra (when  $N = 13$ ), whereas five more factors were required for the minimum to be reached if baseline correction had not been performed. Even so, the percentage of the variance explained by the calibration was larger for spectra that had been baseline corrected, as shown in Table 1. There was no statistical difference between the root-mean-square error of cross validation (RMSECV) for the concentration of ammonia calculated using 18 factors from spectra that had not been baseline corrected and using 13 factors from spectra that had been baseline corrected using the wavelet transform. Because the number of factors required had been reduced, the robustness of the calibration was presumably improved.

**TABLE 1. Parameters of PLS Models for Ammonia and Methane without and with (a/b) Baseline Correction by Wavelet Transform**

molecule	factors	RMSECV	explained variance (%)
NH <sub>3</sub>	18/13	2.04/1.97	96.1/99.6
CH <sub>4</sub>	9/7	6.09/6.31	98.4/98.2

Similar results were obtained in the case of methane as shown in Table 1. In this case, baseline correction reduced the number of factors required to give a minimum of the PRESS plot from 9 to 7. The explained variance is again reduced by baseline correction, and the RMSECV values calculated by both approaches using the optimum number of factors are statistically indistinguishable. Thus again, the stability of the regression model is again increased by reducing the number of factors required.

### Acknowledgments

This work was funded under Cooperative Agreement Cooperative Agreement 58-5368-3-269 with the USDA/NWISL, Agricultural Research Service, Northwest Irrigation and Soils Research Laboratory, Kimberly, Idaho.

### Literature Cited

- (1) *Compendium of Methods for the Determination of Toxic Organic Compounds in Ambient Air*, 2nd Ed. EPA/625/R-96/010b; Center for Environmental Research Information, Office of Research and Development, U.S. Environmental Protection Agency: Cincinnati, OH, 1997.
- (2) Russwurm, G. M.; Childers, J. W. Open-path Fourier transform infrared spectroscopy. In *Handbook of Vibrational Spectroscopy*; Chalmers, J. M., Griffiths, P. R., Eds.; John Wiley and Sons: Chichester, UK, 2002.

- (3) Hart, B. K.; Griffiths, P. R. Effect of resolution on quantification in open-path Fourier transform infrared spectrometry under conditions of low detector noise: I. classical least squares regression; II partial least squares regression. *Environ. Sci. Technol.* **2000**, *34*, 1337–1345; 1346–1351.
- (4) Schulze, G.; Jirasek, A.; Yu, M. M. L.; Lim, A.; Turner, R. F. B.; Blades, M. W. Investigation of selected baseline removal techniques as candidates for automated implementation. *Appl. Spectrosc.* **2005**, *59*, 545–575.
- (5) Friedrichs, M. S. A model-free algorithm for the removal of baseline artifacts. *J. Biomol. NMR.* **1995**, *5*, 147–153.
- (6) Pearson, G. A. A general baseline-recognition and baseline-flattening algorithm. *J. Magn. Reson.* **1977**, *27*, 265–272.
- (7) Dietrich, W.; Rudel, C. H.; Neumann, M. Fast and precise automatic baseline correction of one- and two-dimensional NMR spectra. *J. Magn. Reson.* **1991**, *91*, 1–11.
- (8) Lieber, C. A.; Mahadevan-Jansen, A. Automated method for subtraction of fluorescence from biological raman spectra. *Appl. Spectrosc.* **2003**, *57*, 1363–1367.
- (9) Hachey, M. R.; Boruta, M. *Group Baseline Correction for Multivariate Analysis*; Pittcon: Orlando, FL, 2005.
- (10) Shao, X.; Leung, A. K. M.; Chau, F. T. Wavelet: a new trend in chemistry. *Acc. Chem. Res.* **2003**, *36*, 276–283.
- (11) Mallat, S. G. A theory for multiresolution signal decomposition: the wavelets representation. *IEEE Trans. Pattern Anal. Machine Learn.* **1989**, *11*, 674–693.
- (12) Zhu, C.; Griffiths, P. R. Extending the range of Beer's law in FT-IR spectrometry. Part I: theoretical study of Norton-Beer apodization functions. *Appl. Spectrosc.* **1998**, *52*, 1403–1408.
- (13) Shao, L.; Griffiths, P. R.; Chu, P. M.; Vetter, T. W. Quantitative vapor-phase infrared spectrometry of ammonia. *Appl. Spectrosc.* **2006**, *60*, 254–260.
- (14) Daubechies, D. *Ten Lectures on Wavelets*; SIAM Press: Philadelphia, PA, 1992.

Received for review September 13, 2006. Revised manuscript received June 21, 2007. Accepted June 22, 2007.

ES062188D