Improved Baseline Recognition and Modeling of FT NMR Spectra

Sergey Golotvin* and Antony Williams†

*Advanced Chemistry Development, Russian Office, 6 Bakuleva Street, Moscow 117513, Russia; and †Advanced Chemistry Development,
90 Adelaide Street, Suite 702, Toronto, Ontario M5H 3V9, Canada

E-mail: golotvin@acdlab.garnet.ru, tony@acdlabs.com

A new approach to baseline correction is presented. A smoothed spectrum is used for both baseline area recognition and modeling. To complete the model an interpolation technique is employed over the signals area; then the model is subtracted from the spectrum giving a flat baseline. The method has been shown to give excellent results even for spectra with large baseline distortions due to different origins. © 2000 Academic Press

Key Words: **NMR baseline recognition and modeling; NMR baseline correction algorithm; NMR data postprocessing.**

Nowadays a lot of information (peaks positions, integrals volumes) is obtained from NMR spectra automatically in the support of NMR laboratories in walk-up environments. This requires high-quality spectra prior to automated analysis. Among the sources of error in any quantitative measurements in NMR spectra are distorted baselines. Baseline distortions in NMR spectra can arise due to the "dead time" problem of pulsed NMR (*1*), nonlinearity of the filter-phase response (*2*), the discrete nature of the Fourier transform (*2, 3*) instrumental instabilities, and other miscellaneous reasons (*2, 4*). Although some baseline problems can be avoided by adjusting acquisition parameters during the run-time of the NMR experiment (*2*) and exploiting digital filtering and oversampling (*5*), postprocessing data treatment offers a more general way of correcting baseline distortions. Most popular approaches include reconstruction of the first points of the fid (*1, 6*) and approximation of the baseline in the frequency domain using Fourier series (*7*), polynomials (*8–10*) and functions of special form (*11*). On the other hand, a majority of NMR desktop software allows the user to manually set the points belonging to the baseline and interpolate between them using analytical functions to completely model the baseline. While the results can often be good enough the method requires manual intervention and cannot be used for batch processing. On the other hand, the quality of automated procedures is rarely sufficient when the baseline has severe distortions. The failures are generally due to both inadequate types of analytical functions used for modeling and poor recognition of the baseline.

We have recently developed a new method for baseline correction which is governed by two easily adjustable parameters. The essence of the method is attributed to two parts, baseline recognition and baseline modeling.

Step 1: Baseline recognition. The reliability of the automated baseline recognition is the decisive step in our approach. Since the ultimate definition of "baseline" is "not containing peaks," it is natural to employ for its recognition the same tool that is used for peak identification. To decide whether the ith point belongs to the baseline, it is placed in the center of a rectangle with a width of N spectral points, which is the first parameter of our procedure. Among these N points the minimal and maximal values are found. If their difference does not exceed the noise standard deviation multiplied by a definite factor n (the second parameter of our technique), the ith point is considered to belong to baseline:

$$(y_i^{max} - y_i^{min}) < n\sigma_{noise}. \qquad [1]$$

In Eq. [1], y_i^{max} and y_i^{min} stand for the maximum and minimum points within the rectangular box with a center of the ith point; σ_{noise} is the noise standard deviation. To estimate the latter quantity the spectrum is broken down into 32 regions. For each ith region the corresponding noise standard deviation σ_i is calculated. The minimal σ_i is taken for the final σ_{noise} estimation.

It is essential to avoid spikes in the baseline area (single points belonging to the peak). For this purpose we require not only that the ith point satisfy Eq. [1] but also that one or two of its neighbors does. We found that a noise multiplier n varies very little from spectrum to spectrum and can be set to a constant value, say from 2 to 4. Variations of the width of the rectangular function are discussed below.

Step 2: Baseline modeling. It is natural for visual inspection to have the baseline be a flat straight line with no constant offset. Our idea is to use the smoothed spectrum to model the baseline over a spectrum area free of signals. The final baseline is then constructed from fragments of the smoothed spectrum connected by straight lines. We have used an averaging of the

neighboring data points of the spectrum, an operation equivalent to convolution with a rectangular function. Thus, the smoothed spectrum is calculated according to

$$S(i) = \sum_{k=-M}^{k=M} y(i + k)/(2M + 1). \qquad [2]$$

The number $2M + 1$ of points is the width of the convolution window used for the smoothed spectrum calculation. This parameter was found to affect the final results very little, so M was set to a constant value of 20 points for routine 1D spectra. The convolution window for points at the spectrum edges is not wrapped, instead a number of points smaller than $2M + 1$ is used. Though we have used the rectangular function to average data, the choice of the convolution function can be arbitrary, as well as the method of smoothing. Using a smoothed spectrum for baseline modeling has the advantage that it makes no assumption about the shape or functional form of baseline distortions, making the algorithm of general use.

The same advantage is inherent in the "model-free" approach to baseline correction suggested by M. S. Friedrichs (*12*). Considering NMR peaks as just local extrema, the baseline is defined by the criterion that the number of local extrema above the line equals the number of local extrema below the line. The baseline is constructed by finding the median of the extrema within a window centered about each point in a spectrum. Our approach is similar in the respect that it also uses a localized region for calculating the baseline. The "model-free" method has the particular requirement that the density of signals is small relative to the noise extrema (otherwise the baseline is biased upward for positive peaks and in the resultant spectrum the heights and integrals of the signals are reduced). Our method does not suffer such limitations and can be applied to simple 1D spectra, not only to multidimensional spectra which have low signal densities. The drawback of our piecewise model is that small discontinuities can form. The distortions, however, are comparable with the noise standard deviation and can be neglected.

Finally the baseline model is subtracted from the spectra. This is a straightforward process, and we do not consider it to be a separate step. Our method is very simple and can be applied to a very wide range of baseline distortions.

Algorithm steps are demonstrated in Fig. 1. The huge water hump observed in the top spectrum resulted from poor water suppression. It introduced a distortion of an unknown form, which can hardly be fitted by analytical functions. The spectrum in the middle of Fig. 1 presents the baseline model of fragments of the smoothed spectrum and regions interpolated by straight lines. It is obvious from the figure that the interpolation did not distort the peak heights. The spectrum has a digital resolution of approximately 6 points/Hz and was corrected with the parameters $N = 61$, $n = 3$.
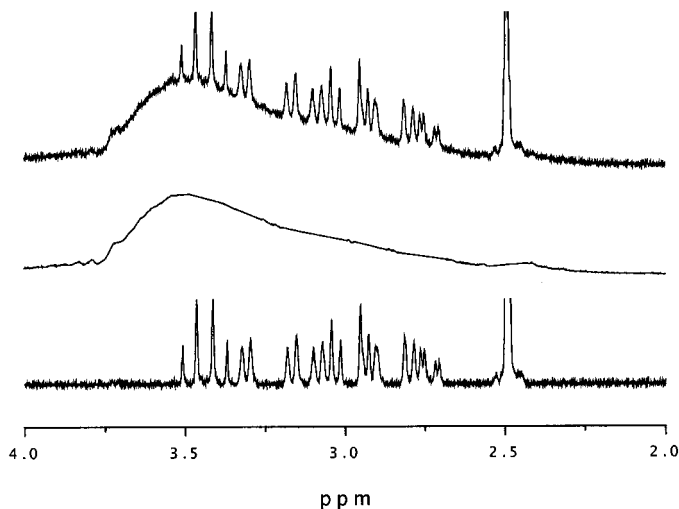

FIG. 1. Original ^{1}H 400 MHz spectrum with a water hump (top); the baseline model consisting of fragments of the smoothed spectrum and regions interpolated by straight lines (middle); the spectrum after baseline subtraction (bottom).

An increase in the width of the window used for baseline recognition will increase the contribution of distant points, and the local baseline structure is not reflected correctly. Here, N is approximately inversely proportional to the width of the detected baseline distortions. Figure 2 shows the experimental ^{1}H 300 MHz spectrum of a hexitol derivative and two corrected spectra obtained with different widths of the window used for baseline recognition. In the middle spectrum, with a narrower window, the broad signal at 4.1 ppm was considered as a baseline distortion. In the bottom spectrum, with an increased window width, the signal of interest was not detected as part of the baseline and was therefore preserved in the resultant spectrum.

^{19}F spectra are notorious for filter-induced baseline distortions due to the large spectral window. At the top of Fig. 3, a typical 100 kHz wide spectrum of an unknown mixture in DMSO is shown. The bad baseline results from inappropriate adjustment of the receiver off delay. The application of our method allows us to obtain an almost perfectly flat baseline, making possible unbiased quantitative analysis. It is worth noting also that all small peaks in the spectrum are preserved after the baseline correction. Performance of our approach in this case is comparable with that of BCF, a rather sophisticated method of baseline correction addressing explicitly the "dead time" problem (*1*).

Out-of-phase residual solvent signals present another source of baseline distortions. Generally these resonances are incompletely suppressed during the acquisition time and are filtered out by corresponding processing steps. The routines digitally filter the time-domain data (*13, 14*) but severely attenuate signals near the solvent. To avoid this problem, a dedicated algorithm IFLAT (*11*) was developed to treat residual solvent signals. It calculates a linear combination of base functions that
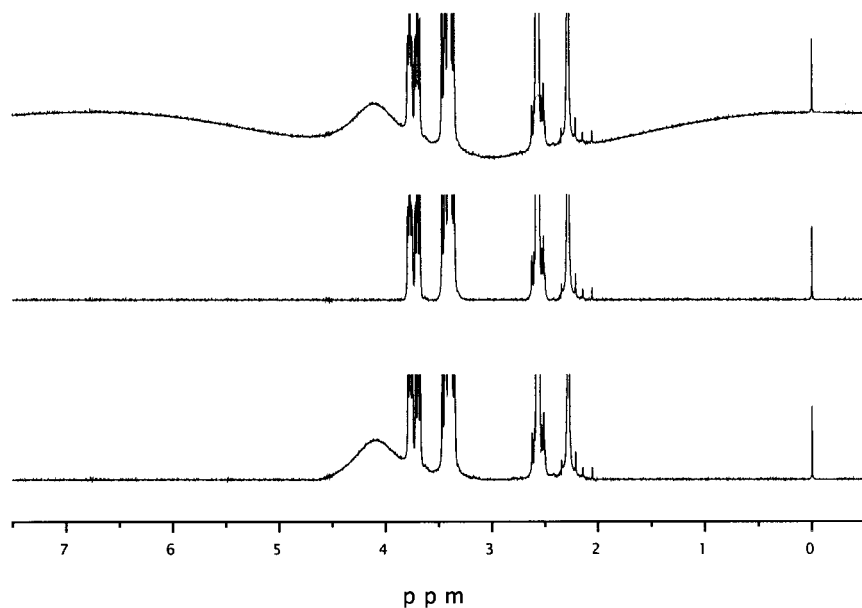
FIG. 2. The experimental ^{1}H 300 MHz spectrum of a hexitol derivative in DMSO-d_6 spectrum (top), corrected spectrum with $N = 31$, $n = 3$ (middle), and corrected spectrum with $N = 61$, $n = 3$ (bottom).

best represents the baseline distortion. Since these functions include the absorptive and dispersive components of a signal centered at the solvent resonance frequency, the method can be regarded as a type of solvent-suppression technique itself. This complicates using this routine for the correction of distortions of unknown shapes. Our approach to the problem of removing residual solvent signals consists in applying a suitable solvent suppression technique followed by a general-purpose baseline
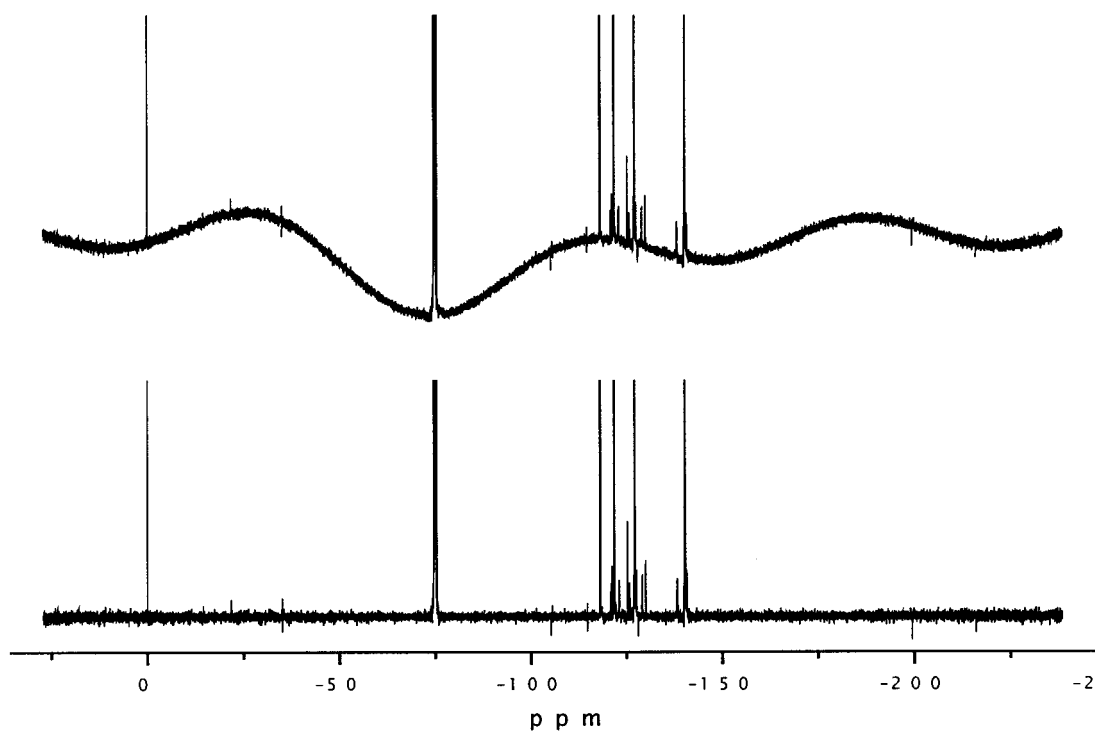

FIG. 3. Original ^{19}F 376 MHz spectrum of an unknown mixture in DMSO (top) and baseline corrected (bottom). Baseline correction parameters were $N = 31$, $n = 3$.
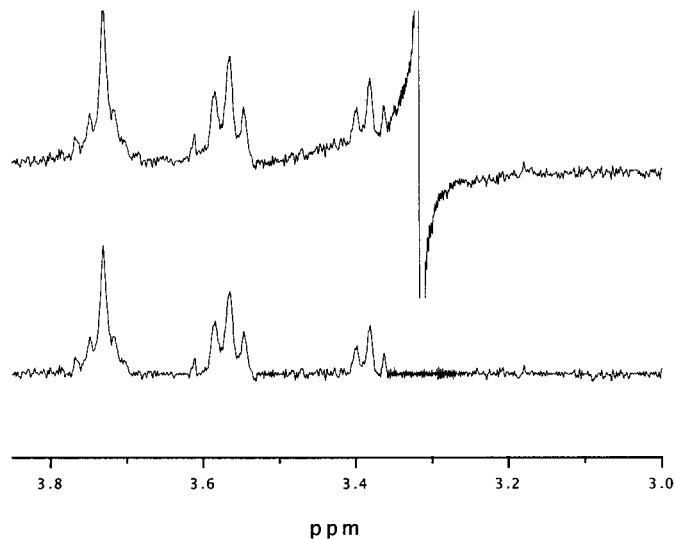
FIG. 4. The region in the vicinity of residual water signal in a flow-through probe ¹H 400 MHz spectrum. Original spectrum (top) and corrected spectrum (after removal of the residual water signal by processing in the frequency domain (*15*) followed by baseline correction by our method) (bottom).

correction routine. If time-domain filtering does attenuate signals in the vicinity of the solvent resonance, then frequency-domain processing (*15*) should be used. Figure 4 illustrates our approach. At the top of Fig. 4 the region in the vicinity of the residual water signal of a typical flow-through probe ¹H 400 MHz spectrum is shown. A triplet signal at 3.38 ppm is observed close to the residual water resonance (the difference does not exceed 30 Hz). After frequency-domain solvent suppression (*15*) followed by baseline correction, severe distortions are removed, and the signals of interest can be integrated.

We have applied our automatic baseline correction procedure to many 1D NMR spectra and as yet have encountered no major problems. Experience with several 2D spectra is also encouraging. Due to adequate baseline modeling and recogni-

tion, as well as the intrinsic simplicity of our technique, we have shown it to be superior and therefore of inherent value for desktop postprocessing.

REFERENCES

1. A. Heuer and U. Haeberlen, A new method for suppressing baseline distortions in FT NMR, *J. Magn. Reson.* **85,** 79–94 (1989).

2. C. Tang, An analysis of baseline distortion and offset in NMR spectra, *J. Magn. Reson. A* **109,** 232–240 (1994).

3. R. Freeman, "A Handbook of Nuclear Magnetic Resonance," Longman Scientific & Technical, Essex (1988).

4. D. Marion and A. Bax, Baseline distortion in real Fourier transform NMR spectra, *J. Magn. Reson.* **79,** 352–356 (1988).

5. A. Belguise, Numerical magnetic resonance, *Analusis* **23**(1), M57–M62 (1995).

6. D. Marion and A. Bax, Baseline correction of 2D FT NMR spectra using a simple linear prediction extrapolation of the time-domain data, *J. Magn. Reson.* **83,** 205–211 (1988).

7. D. S. Stephenson and G. Binsch, Automated analysis of high-resolution NMR spectra. II. Illustrative applications of the computer program DAVINS, *J. Magn. Reson.* **37,** 409–430 (1980).

8. G. A. Pearson, A general baseline-recognition and baseline-flattening algorithm, *J. Magn. Reson.* **27,** 265–272 (1977).

9. W. Dietrich, C. H. Rudel, and M. Neumann, Fast and precise automatic baseline correction of one- and two-dimensional NMR spectra, *J. Magn. Reson.* **91,** 1–11 (1991).

10. D. E. Brown, Fully automated baseline correction of 1D and 2D NMR spectra using Bernstein polynomials, *J. Magn. Reson. A* **114,** 268–270 (1995).

11. C. Bartels, P. Guntert, and K. Wuthrich, IFLAT—A new automatic baseline-correction method for multidimensional NMR spectra with strong solvent signals, *J. Magn. Reson. A* **117,** 330–333 (1995).

12. M. E. Friedrichs, A model-free algorithm for the removal of baseline artifacts, *J. Biomol. NMR* **5,** 147–153 (1995).

13. D. Marion, M. Ikura, and A. Bax, Improved solvent suppression in one- and two-dimensional NMR spectra by convolution of time domain data, *J. Magn. Reson.* **84,** 425–430 (1989).

14. K. J. Kross, Improved digital filtering technique for solvent suppression, *J. Magn. Reson. A* **101,** 220–224 (1993).

15. P. Tsang, P. E. Wright, and M. Rance, Signal suppression in the frequency domain to remove undesirable resonances with dispersive lineshape, *J. Magn. Reson.* **88,** 210–215 (1990).