PERGAMON

Journal of Quantitative Spectroscopy & Radiative Transfer 68 (2001) 179–193

Journal of Quantitative Spectroscopy & Radiative Transfer

www.elsevier.com/locate/jqsrt

Baseline subtraction using robust local regression estimation

Andreas F. Ruckstuhl[a,*], Matthew P. Jacobson[b], Robert W. Field[b], James A. Dodd[c]

[a]*Departement P, Zurcher Hochschule Winterthur (Zurich University of Applied Sciences), Postfach 805, 8401 Winterthur, Switzerland*

[b]*Department of Chemistry and G.R. Harrison Spectroscopy Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

[c]*Space Vehicles Directorate, Air Force Research Laboratory, Hanscom Air Force Base, MA 01731-3010, USA*

Received 6 September 1999

Abstract

A technique entitled robust baseline estimation is introduced, which uses techniques of robust local regression to estimate baselines in spectra that consist of sharp features superimposed upon a continuous, slowly varying baseline. The technique is applied to synthetic spectra, to evaluate its capabilities, and to laser-induced fluorescence spectra of OH (produced from the reaction of ozone with hydrogen atoms). The latter example is a particularly challenging case for baseline estimation because the experimental noise varies as a function of frequency.

Keywords: Baseline removal; LOWESS; Robust fitting; Smoothing

1. Introduction

The robust baseline estimation (RBE) technique that is introduced in this paper is a technique for baseline removal; that is, in spectra that consist of sharp features superimposed upon a continuous, slowly varying baseline, it is designed to permit the separation of the two components, spectrum and baseline. There exist, of course, numerous techniques for baseline removal from spectra, because the problem is ubiquitous within spectroscopy. In many cases, baselines are simply removed "by eye". There is nothing inherently wrong with such an approach, but we believe that, even for a "mundane" task such as baseline removal, it is beneficial to use numerical tools that minimize the need for judgement calls and permit reproduction of the results by others. The ability

* Corresponding author. Tel.: +41-52-267-7812; fax: +41-52-267-7822.
E-mail address: andreas.ruckstuhl@zhwin.ch (A.F. Ruckstuhl).

0022-4073/00/$ - see front matter
PII: S0022-4073(00)00021-2

to extract quantitative peak intensity and lineshape information from spectra can in many cases be limited by the ability to remove an underlying continuous baseline, and in such cases numerical techniques of baseline estimation are clearly necessary.

To our knowledge, there exists no review article on techniques for estimating baselines in spectra and other signals. The available literature on the topic is scattered across many fields of research, including analytical chemistry/chemometrics [1–3], nuclear physics [4], X-ray spectroscopy [5–7], NMR spectroscopy [8–11], and even medical research [12]. In addition, many of the baseline estimation techniques discussed in the literature are tailored for specific data sets. A recent article by Phillips and Hamilton [1], however, does provide a useful overview of several general approaches to estimating baselines in spectra. One common approach involves automated peak rejection algorithms [9,11], which in principle identify those regions of a spectrum that can be accounted for solely by the baseline function; the baseline function in regions containing peaks can then be estimated by interpolation. Other approaches to estimating include those that are based upon digital filtering [6,7]; in the simplest application, a "high-pass" filter can be used to suppress a relatively slowly varying baseline, although often at a cost of distorting the remaining "sharp" components of the spectrum. Other techniques for baseline estimation are based upon statistical methods, including maximum entropy methods [1,5] and principal component analysis [13].

The RBE technique uses methods of robust local regression to estimate the baseline component of a spectrum. Among the attributes of the RBE technique are that it

1. takes into account the measurement error in the data in a natural way;
2. can be applied to a wide variety of baseline subtraction problems; and
3. requires minimal human intervention.

No specific claims are made that the RBE technique is better, faster, or more accurate than other techniques for baseline removal, but RBE does appear to provide a useful complement to the other techniques with which we are familiar.

The details of the RBE technique will be discussed in detail in Sections 3 and 4 below. First, however, we wish to make a few general comments on the use of statistical methods for baseline estimation. It is important to realize from the outset that the question "what is the baseline in my spectrum" is, in the absence of further information, an ill-defined question — in general, there exist an infinity of possible models for the baseline that are consistent with the data. A more well-defined question is, given a model family for the baseline function, which baseline function out of the model family best fits the data? However, the presence of peaks superimposed upon the baseline makes answering even such a restrictively posed question difficult. In essence, only *some* of the data is useful for determining the baseline function; data points on top of peaks are generally useless for determining the baseline. (A possible exception would occur if the line shapes and/or relative intensities of the peaks were known a priori. In this paper, we will make no assumptions about the lineshapes of the peaks contained in the spectra, other than that they must be sufficiently narrow that the baseline can be interpolated across the width of the peak.) If it is possible to estimate the baseline function at all, then the baseline must be assumed, at the very least, to be smooth and to vary slowly, so that the baseline can be safely interpolated across the width of a peak. Congested spectra (i.e., those in which a large fraction of resolution elements lie on peaks) present special challenges for baseline estimation; if a baseline can be determined at all for such a spectrum, then it is usually necessary to invoke stringent assumptions concerning the smoothness of the baseline.

Given these caveats, we believe that techniques of robust estimation may be useful for estimating the baselines present in many spectra. Robust or resistant estimation, in the simplest sense, refers to techniques of estimation that are less sensitive to outliers (extreme observations) than the conventional least-squares approach. It is not our intention to review techniques of robust estimation in general, since there exists a vast literature on the subject; for an introduction to the subject, see, for example, Chapters 1 and 8 of the text by Hampel et al. [14]. In estimating the baseline in a spectrum, the points in the spectrum that lie on top of peaks can be considered outliers, and thus one can imagine using a robust estimator to determine a baseline function by more-or-less ignoring those points that lie on peaks.

Several nontrivial issues must be faced before this simple idea can be implemented. The first is the choice of a functional form for the baseline to fit to the spectrum. In rare cases, a reasonable functional form may be known a priori, but a more general strategy is to estimate the baseline *locally* (i.e., over a sufficiently small section of the spectrum) by a low order polynomial (usually a line). Thus, the strategy for baseline estimation that we adopt is robust local regression estimation. In Section 2, we briefly review statistical concepts that are relevant to robust local regression estimation. The RBE technique itself is introduced in Section 3, and applied to synthetic data sets to illustrate its properties. Finally, in Section 4, RBE is applied to real experimental data, which are laser-induced fluorescence spectra of the OH radical. These spectra represent a particularly challenging application for baseline estimation because the measurement error varies discontinuously across the spectrum.

2. Robust local regression estimation

In this section, the concepts of local estimation and robust estimation are introduced, and differentiated from the usual least-squares approach to fitting data. We focus in particular on the LOcally WEighted Scatter plot Smoother (LOWESS) introduced by Cleveland [15], to which the RBE technique is closely related.

For a given data set (x_i, Y_i), $i = 1, \ldots, n$, where Y_i is the response and x_i is the predictor variable, one may fit a regression curve g through the data to establish the relationship between the response and predictor. Discrepancies between the regression curve and the data are often treated as noise. In other words, the data are regarded as deviations from the model

$$Y_i = g(x_i) + E_i \quad (i = 1, \ldots, n), \tag{1}$$

where E_i is an unknown error. Often, one can assume that the errors E_i are independent and identically distributed, with mean zero and variance σ^2. We will refer to σ as a scale parameter, because it is the natural unit with which to measure the size of deviations from the model.

If the form of the regression curve is known, up to unknown parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^{\mathrm{T}}$ (superscript T means transposed), as for example in simple linear regression modelling $[g(x; \boldsymbol{\theta}) = \theta_1 + \theta_2 x]$, then one of the most successful estimation techniques is the least-squares approach:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} [y_i - g(x_i; \boldsymbol{\theta})]^2. \tag{2}$$

In many situations, however, it is not possible to choose in advance an appropriate functional form for the the regression curve g in Eq. (1). If we can assume that g is smooth, then an alternate method for estimating the regression curve g is to apply linear regression modelling *locally*. That is, the regression curve $g(x)$ can be approximated as linear in a sufficiently small neighborhood around any given point x_o. One can simply apply the least-squares technique to a fraction of the data around x_o, or alternatively, one can incorporate a weight scheme into the local least-squares problem that decreases the influence of data points in proportion to their distance from x_o. That is,

$$\hat{\boldsymbol{\theta}}(x_o) = \arg\min_{\theta} \sum_{i=1}^{n} K\left(\frac{x_i - x_o}{h}\right)\{y_i - [\theta_0 + \theta_1(x_i - x_o)]\}^2, \tag{3}$$

where $K[(x_i - x_o)/h]$ is a weight function (also called a kernel weight) for the point i. The function K is a unimodal symmetric nonnegative function that is (nearly) zero outside the neighborhood of x_o, which is defined by $x_o \pm h$. Note, in particular, that the estimated parameters $\hat{\boldsymbol{\theta}} = (\hat{\theta}_0, \hat{\theta}_1)^T$ depend on x_o. Thus, $\hat{\theta}_0(x_o)$ is an estimate of $g(x)$ at x_o and is more appropriately named $\hat{g}(x_o)$.

The precise choice of the weight function K is not very important for the performance of the resulting estimators, both theoretically and empirically (see, e.g., Ref. [16, pp. 45, 72, 73], or Ref. [17, Section 3.2.6]). Thus, the functional form for K is chosen based upon other issues such as ease of computation. LOWESS uses the tricube kernel

$$K(u) = [\max(1 - |u|^3, 0)]^3, \tag{4}$$

which descends smoothly to zero and is zero outside the neighborhood defined by h (as opposed to a Gaussian function, for example, which has infinitely long "tails" where the function is close to, but not precisely, zero).

A more critical issue for local estimation is how wide the local neighborhood should be (i.e., the value of h). If we choose a very small h, the approximation error will be small. However, since the number of data points in the local neighborhood is also small, the variance of the estimated values $\hat{g}(x_o)$ will be large. On the other hand, the bandwidth h should not be made so large that the local linear approximation is no longer valid, which would create a large approximation error. Put in different words, when the bandwidth h is very small, the resultant estimate essentially interpolates the data points, whereas when h is very large, the solution is identical to the global linear regression solution. The choice of bandwidth for baseline estimation problems is discussed in Section 3.

The local linear regression estimate in Eq. (3), being based upon least-squares estimation, is sensitive to extreme observations (outliers) in the response variable in much the same way that ordinary least-squares estimates (Eq. (2)) are. When gross outliers may be present in a data set, an estimation procedure is needed that is more resistant to the extreme observations. It is important, however, to recognize that it is more difficult to identify unambiguously a point as an outlier when performing a local regression than it is when fitting a parametric (global) model. That is, the inherent flexibility of the local linear estimation approach makes it difficult to distinguish between systematic outliers that occur in patches and intrinsic features of the data such as jumps or peaks in the regression curve g.

Given these caveats, it is nonetheless possible to specify a robust local estimation procedure. One possible approach is to use Eq. (3) (i.e., local least-squares) to estimate $g(x)$ initially, but then use the residuals of this fit to assign "robustness weights" $w_r(x_i)$ to each point (x_i, Y_i), such that points with large residuals receive small robustness weights and vice versa. The regression curve $g(x)$ is then

refined by performing a weighted least-squares fit, according to

$$\hat{\boldsymbol{\theta}}(x_o) = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} w_r(x_i)\, K\left(\frac{x_i - x_o}{h}\right)\{y_i - [\theta_0 + \theta_1(x_i - x_o)]\}^2. \tag{5}$$

This fit can then be repeated iteratively to convergence, with the robustness weights $w_r(x_i)$ always being determined from the previous iteration of the fit. A more detailed discussion of robustly weighted least-squares local regression, and in particular, the choice of the functional form for w_r, can be found in the appendix.

In the LOWESS procedure, the robustness weights w_r are determined using Tukey's bisquare weights

$$w_r(x_i) = \{\max[1 - (r_i/b)^2, 0]\}^2, \tag{6}$$

where $r_i = [y_i - \hat{g}(x_i)]/\sigma$. The parameter b determines the "robustness" of the procedure (i.e., how strongly the fit is influenced by outliers), and its choice involves a trade-off between losing too much information at the "good" data points and controlling the influence of "bad" data points. Cleveland [15] chose a tuning constant of $b = 4.05$; we use a slightly different value for the RBE procedure, the choice of which is discussed in Section 3. Note that the use of the robustness weights in Eq. (6) for the robust local regression procedure (Eq. (5)) can result in multiple local minima. In the appendix, it is demonstrated that other choices of robustness weights result in a unique solution. However, for purposes of baseline estimation, it is critical to choose robustness weights, such as those in Eq. (6), which disregard distant outliers totally, and these necessarily lead to the possibility of multiple minima in the fit. In practice, however, our implementation of the RBE technique (Section 3) appears to converge to a reasonable solution for most baseline estimation problems of interest.

In order to implement the LOWESS procedure, one final parameter that needs to be specified is the scale parameter σ (i.e., the experimental noise). In certain cases, σ can be estimated a priori. When this is not the case, however, it can be estimated using the (standardized) median of absolute values (MAV) of the (unstandardized) residuals

$$\hat{\sigma}_{\mathrm{MAV}} = \mathrm{median}(|y_i - \hat{g}(x_i)|)/0.6745. \tag{7}$$

Note that the denominator is required in order to make the estimate of the scale parameter consistent with the standard deviation of a Gaussian distribution.

To summarize, the LOWESS algorithm, to which the RBE is closely related, proceeds as follows:

1. For each point i, compute $\hat{g}(x_i)$ by using the local regression estimator (Eq. (3)) with the kernel weights defined by Eq. (4).
2. Use Eq. (7) to estimate the scale parameter, and calculate the robustness weights $w_r(x_i)$ by applying Eq. (6).
3. For each point i compute a new fitted value $\hat{g}(x_i)$ by using the robust local regression estimator (Eq. (5)) with kernel weights defined by Eq. (4).
4. Repeat steps 2 and 3 until convergence is achieved, which generally only requires 3–5 iterations. The final fitted values yield the estimated curve $\hat{g}(x_i)$.

For further implementation details of LOWESS, see Ref. [15].

LOWESS is a well-known specific implementation of a smoothing technique based on the local (polynomial) regression technique. Smoothing techniques have become an important tool in data analysis and subsequently there is a vast literature. Two of the more recent books are those of Simonoff [16] and Fan and Gijbels [17]. The latter is a book-length treatment of local polynomial estimation, whereas the former gives an introduction to many aspects of smoothing. Both books discuss the LOWESS procedure and explain robust smoothing briefly. For a more general account of robust estimation see, e.g., the book by Rousseeuw and Leroy [18] or the references given in the appendix. A challenging application for robust estimation in molecular spectroscopy is described in Ref. [19].

3. Baseline subtraction

We consider a spectrum to be defined by

$$Y(x_i) = g(x_i) + m(x_i) + E_i, \tag{8}$$

where $Y(x_i)$ is the observed signal, $g(x_i)$ is the baseline drift, $m(x_i)$ is the desired (baseline-free) spectroscopic signal at x_i, and E_i represents the measurement errors, which we again assume to be independent and Gaussian-distributed (mean 0, variance σ^2). Separating the three components in Eq. (8) is an ill-posed problem without additional information. It is possible to estimate the baseline drift $g(x_i)$ at x_i if the spectral signal $m(x_i)$ in the neighborhood of x_i is minimal by using, for example, the LOWESS procedure that is discussed in Section 2. Otherwise, it is impossible in general to estimate $g(x_i)$. However, if we can assume that the baseline is sufficiently smooth, then we can interpolate the baseline at the positions x_i at which non-negligible spectral signals $m(x_i)$ exist (i.e., interpolate the baseline across peaks in the spectrum). Here we will only consider linear interpolation across peaks, which is reasonable if the baseline varies slowly relative to the peak widths.

The RBE technique is a robust local regression procedure that is closely related to LOWESS, which was described in Section 2. The basic idea of the RBE procedure is to regard spectral points i as outliers if $|m(x_i)| \gg \sigma$ (this condition is only fulfilled for points that lie on peaks in the spectrum). In most spectra, all of the peaks point in the same direction. Without loss of generality, we assume that $m(x_i) \geq 0$, and thus we can consider the baseline to be asymmetrically "contaminated" by the peaks in the spectrum. In such a case, robust estimators that use weight functions w_r that are symmetric with respect to the residuals r_i will generate biased estimates. To counteract the bias, we can use asymmetric robustness weights such as

$$w_r(x_i) = \begin{cases} 1 & \text{if } r_i < 0, \\ [\max\{1 - (r_i/b)^2, 0\}]^2 & \text{otherwise,} \end{cases} \tag{9}$$

where $r_i = [y_i - \hat{g}(x_i)]/\sigma$. Although such an estimator can reduce the bias created by the asymmetric contamination of the baseline by the peaks, it should be kept in mind that it can also over-correct in some cases as well as introduce bias when there is no spectral signal m in the neighborhood. One could imagine using a robust estimator with asymmetric robustness weights in neighborhoods with significant spectral signal, and a symmetric robust estimator otherwise.

However, the baseline estimate in regions with no spectral signal is usually of little consequence to the analysis of the data. In addition, in numerical tests we have found that the bias created by the asymmetric robust weight functions is minor, even when there are no peaks. For most problems of interest, the bias inherent to the data (i.e., unidirectional peaks) is a greater concern. Finally, it should be noted that the use of asymmetric robust weights helps to ensure that the fit converges to an acceptable solution.

To utilize the RBE procedure, we must select an appropriate bandwidth h to define the local neighborhood, and tuning constant b to specify the robustness of the procedure. We choose as the tuning constant $b = 3$, which corresponds to choosing a more robust estimator than Cleveland's original proposal of $b = 4.05$ for the LOWESS procedure. This choice is not arbitrary but rather resulted from optimizing the value of b against the two synthetic examples that are presented in this section (i.e., what value of b gives a baseline estimate that is closest to the true baseline?). In both synthetic examples, $b = 3.0$ was nearly optimal, and because the two synthetic examples are quite different, we expect this value to be appropriate for many baseline subtraction problems, although it should be kept in mind that other values may be appropriate for specific problems. From a qualitative standpoint, a highly robust estimator is needed for baseline estimation problems so that small peaks (and the "wings" of large peaks) do not bias the estimate (the points on top of large peaks are not a concern because they receive nearly zero weight functions according to Eq. (6)). See the appendix for further discussion of the choice of robustness weights and the trade-off between efficiency and robustness.

With respect to the bandwidth h, from a practical standpoint, most spectra consist of evenly spaced points x_i, so that the bandwidth can be conveniently defined by the number of points d in the local neighborhood that ranges from $x_o - h$ to $x_o + h$. Whether defined by d or by h, bandwidth selection is critical to the success of robust local estimation. A number of suggestions have been advanced (see, e.g., Ref. [16, Section 5.3], or Ref. [17, Chapter 4]) for automatically determining an appropriate bandwidth from the data, but these approaches would lead to reasonable bandwidths for estimating the "total" signal $g(x_i) + m(x_i)$ (baseline plus peaks), which is not our goal.

A more general consideration is the following: if, in a local neighborhood of x_o consisting of d data points, at least $d/2$ of them are seriously affected by the spectral signal m, then the robust local regression estimator is more likely to estimate $g(x_o) + m(x_o)$ than $g(x_o)$. To avoid such a failure, we can require that d be large enough such that, at the very least, less than half of the points in the local neighborhood for any x have significant spectral signal m. The smallest possible value of d we refer to as $d_\circ$; in a spectrum where all of the peaks are well-resolved, $d_\circ$ would be roughly twice the linewidth of the widest peak. As a general recommendation, a value of d of 2.5 times the linewidth of the widest peak is reasonable and should not oversmooth the data, as long as the baseline varies slowly relative to the peak widths.

As an additional safeguard against the local neighborhood being too small, we also omit those values x_i for which $w_r(x_i) = 0$ (i.e., those with substantial signal content m) when determining the local neighborhood. That is, since the outliers appear in patches (peaks), the bandwidth h can be enlarged in such regions to make sure that there is sufficient unperturbed information to determine the baseline. In a technical sense, the bandwidth $h_\circ$ at x_o is defined by the the dth smallest value among $|x_i - x_o|$, for those values of i for which $w_r(x_i) \neq 0$. Note that this definition implies that the bandwidth varies as a function of spectral position.

To illustrate the properties of the RBE procedure, we consider two synthetic spectra, where the components $g(x_i)$ and $m(x_i)$, and the standard deviation of the error σ (scale parameter), are known. In the case of the synthetic spectrum in Fig. 1 we chose a value of $d = 500$ for the bandwidth, and as a first test, we assumed that the scale parameter was known ($\sigma = 1.0$), and did not attempt to determine it from the data itself. Both the top and bottom panels of Fig. 1 show the solution obtained by RBE, which is quite satisfactory given that the error in the baseline estimate is within σ of the true baseline over most of the spectrum.

The largest discrepancies occur at the end of the spectrum and at the edges of the peaks, where the RBE procedure has some difficulty discriminating between "good" and "bad" data points. Relatively poor performance of the RBE near the first or last point of a spectrum is likely unavoidable, because the local neighborhood is by necessity asymmetrically positioned with respect to these points. The baseline estimate at the ends of the spectrum is generally irrelevant, however. A greater concern is the bias of the baseline estimate in the middle of the peaks. This bias can be observed to correlate with the second derivative of the true baseline g; the bias is positive when $g'' > 0$ and negative when $g'' < 0$, a result that is consistent with theoretical considerations of the bias of the local regression procedure (see Refs. [16, Section 5.2.2] or Fan and Gijbels [17, Section 3.2.1]). In any case, this bias is always less than 1σ for this example (less than the measurement error), which would be quite acceptable for most applications.

We have also estimated the baseline for the example in Fig. 1 assuming that the standard deviation σ of the measurement error is not known. However, the scale estimator in Eq. (7), which was the original proposal used by LOWESS, however, yields a value of 1.48 for the final RBE solution in Fig. 1 (i.e., nearly 50% larger than the known value). Clearly, the scale estimator in Eq. (7) is influenced to an unacceptable degree by the spectral signal m. A simple refinement is to calculate the scale estimator using only those observations for which the residuals $r_i = Y(x_i) - \hat{g}(x_i)$ are smaller than $b\hat{\sigma}_{\mathrm{MAV}}$ (that is, as with the determination of the bandwidth, we exclude those points for which $w_{\mathrm{r}}(x_i) = 0$). We call this estimator $\hat{\hat{\sigma}}_{\mathrm{MAV}}$, and for the synthetic example of Fig. 1, we obtain a value of 1.04 for the scale parameter using this refined estimator, which is quite close to the true value of 1.00.

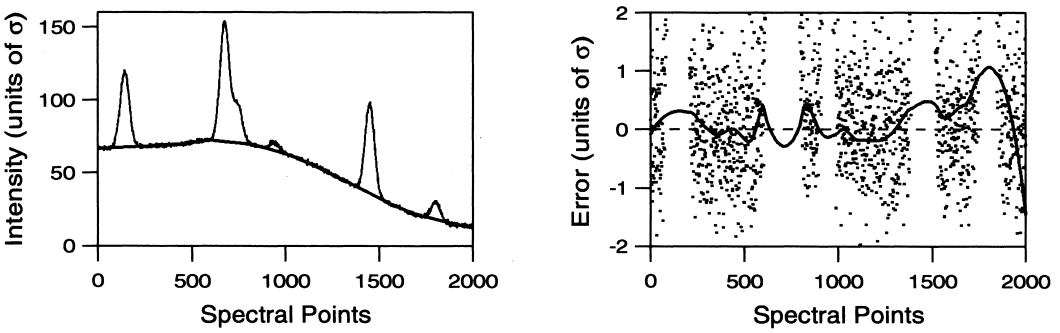

Fig. 1. Simple synthetic spectrum used to illustrate the properties of the RBE technique. The dashed line represents the true baseline and the solid line is the fitted baseline, using the tuning parameters $b = 3$ and $d = 500$. In the right panel, the true baseline has been subtracted from the spectrum and the fitted result, so that the true baseline is represented by the horizontal line at zero. The y-axis ("intensity") also has a restricted range so that the minor discrepancies between the true and fitted baselines can be examined.

The real experimental spectra discussed in Section 4 are more complicated than the synthetic example shown in Fig. 1. The second synthetic example that we consider in this section is designed to mimic the real examples, including the approximate signal to noise, resolution, and peak density. Most importantly, it mimics the property of the experimental spectra that the standard deviation of the measurement error is not constant, but rather is given by

$$\mathrm{var}[Y(x_i)] = \sigma^2/w_s(x_i), \tag{10}$$

where $w_s(x_i) > 0$ specifies how the measurement error varies across the spectrum. The variation of the scale parameter with respect to x is known in the real experimental data, and thus we assume it to be known here as well. To incorporate this additional information in the RBE procedure we include $w_s(x_i)$ as an additional weight component in the robust locally weighted regression estimator:

$$\hat{\vec{\theta}}(x_o) = \underset{\vec{\theta}}{\mathrm{argmin}} \sum_{i=1}^{n} w_s(x_i) w_r(x_i) K\left(\frac{x_i - x_o}{h}\right)\{y_i - [\theta_0 + \theta_1(x_i - x_o)]\}^2. \tag{11}$$

The robustness weights $w_r(x_i)$ must of course also be adjusted to account for the nonconstant σ. Instead of using $r_i = [y_i - \hat{g}(x_i)]/\hat{\sigma}$ in Eq. (6), we now use $r_i = \sqrt{w_s(x_i)}\,[y_i - \hat{g}(x_i)]/\hat{\sigma}$. Similar changes must be made for estimating the scale parameter σ, so that the estimator in Eq. (7) is replaced by

$$\hat{\sigma}_{\mathrm{wMAV}} = \mathrm{median}\{\sqrt{w_s(x_i)}\,|y_i - \hat{g}(x_i)|\}/0.6745. \tag{12}$$

Fig. 2 depicts the second synthetic spectrum, along with the true baseline (dashed line) and the solutions of the RBE procedure with $b = 3$ and $d = 125$ (solid line). The error in the baseline estimate is well within the measurement error $\{\mathrm{var}[Y(x_i)]\}^{1/2}$ across the entire spectrum. The estimated value for σ is 0.132, which is only 6% higher than the true value of 0.125.

4. Real examples

In this section we apply the RBE technique to real experimental spectra that present a substantial challenge for baseline estimation. The spectra in Figs. 3 and 4 were obtained using pulsed laser-induced fluorescence (LIF) detection of gas-phase hydroxyl (OH) vibration–rotation populations created by the reaction of hydrogen atoms with ozone ($H + O_3 \rightarrow OH + O_2$) [20]. The OH product is excited near 308 nm via optically allowed $A \leftarrow X$ transitions, and the resultant resonant fluorescence is detected using a photomultiplier tube fitted with a broad bandpass optical filter. The observed lines can all be identified with transitions that originate from rovibrational levels with $v = 0$ or 1.

The information that is desired to be extracted from these spectra is the OH quantum state population distribution, which is directly related to the intensities of the observed transitions. However, the baseline drift that can be clearly observed in Figs. 3 and 4, if uncorrected, would compromise the accuracy of the derived OH populations, particularly for the smaller peaks, for which the baseline variation could lead to factor-of-two errors or worse. Attempts to model the baseline drift based on, e.g., modest changes in the experimental pressure or temperature were

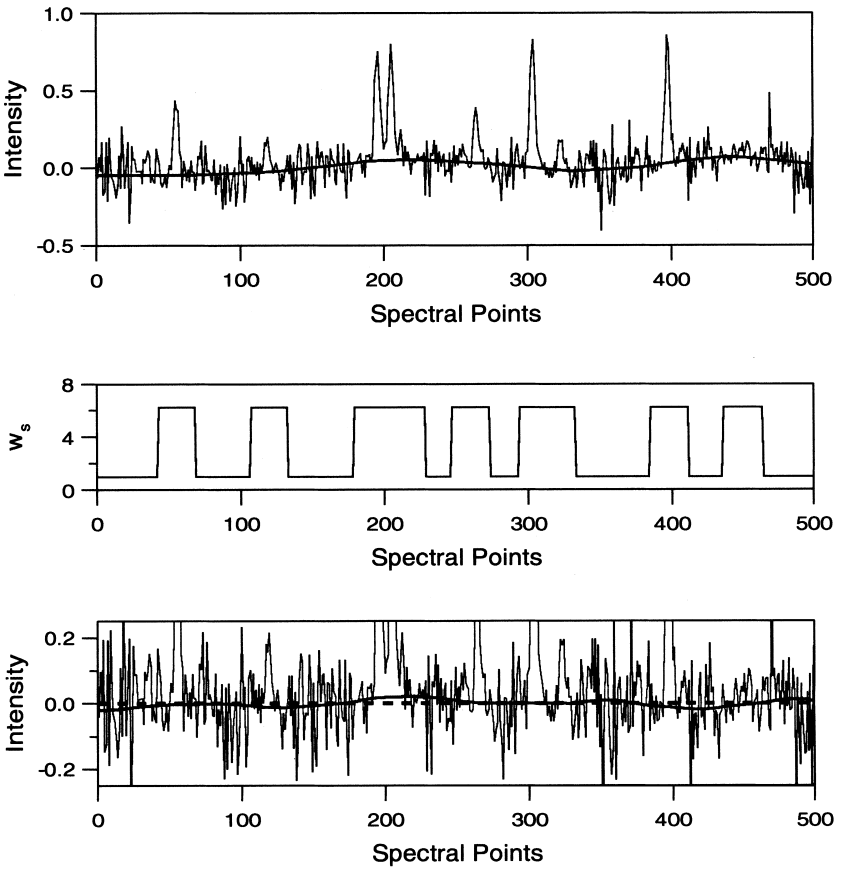

Fig. 2. Synthetic spectrum that is designed to mimic several key properties of the real experimental spectra in Section IV, including a known variation in the measurement error across the spectrum, which is represented in the second panel. The baseline estimate obtained by RBE, with $b = 3$ and $d = 125$ is depicted as a solid line in the top and bottom panels, and the true baseline is represented by a dashed line. In the bottom panel, the true baseline has been subtracted from both the fitted baseline and the spectrum, and the y-axis has been expanded.

unsuccessful, and thus the RBE technique was applied to numerically estimate and remove the baseline.

The solid line in Fig. 3 is the baseline estimate for the spectrum obtained by RBE, using the parameters $b = 3$ and $d = 750$. In contrast to the synthetic spectra in Section 3, the true baseline is of course unknown, and it is not possible to judge the accuracy of the fitted baseline. However, the solution obtained by RBE certainly appears to be reasonable. The estimated standard deviation of the measurement error (scale parameter) is $\hat{\sigma} = 0.0276$.

The spectrum in Fig. 4 differs from the spectrum in Fig. 3 in that the measurement error in Fig. 4 varies as a function of frequency due to the way in which the data was recorded. The spectral data acquisition process is quite time-consuming, and much of the time spent scanning occurs in spectral regions where there is no signal. To increase the efficiency of data collection, the scan rate can be increased in regions of no signal and decreased near OH transitions (particularly weak

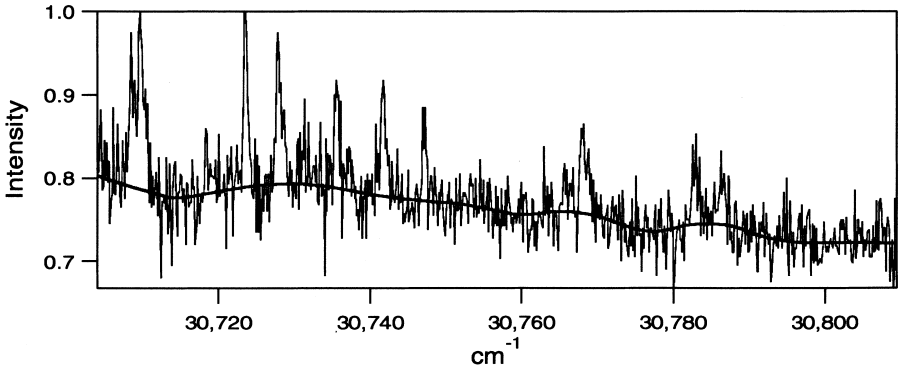

Fig. 3. Experimental laser-induced fluorescence spectrum of OH formed by reaction of hydrogen atoms with ozone. The baseline determined by RBE, with $b = 3$ and $d = 125$, is represented as a solid line.

ones), to improve the signal to noise. For the spectrum in Fig. 4, 16–64 laser shots were averaged for each data point near weak OH transitions; otherwise, only 8 shots were averaged.

As discussed in Section 3, we can include this extra information in the RBE procedure by incorporating the additional weights $w_s(x_i)$. These weights account for the variation in the scale parameter across the spectrum, but the absolute variance σ^2 must still be determined from the data. We arbitrarily choose to define σ^2 as the variance in the experimental spectrum for those data points (the majority) for which 8 laser shots were averaged. This choice corresponds to defining

$$w_s(x_i) = s_i/8 \quad (i = 1, \ldots, n), \tag{13}$$

where s_i is the number of laser shots averaged for each point.

The baseline estimate obtained by RBE is depicted as a solid line in Fig. 4, using $b = 3$ and $d = 200$. Once again, the solution appears to be reasonable, but no quantitative assessment of its accuracy is possible. Although the baseline drift is less severe in this example than in Fig. 3, accurate baseline estimation is nonetheless critical for accurate quantitative analysis. At the scan rate of 8 laser shots per data point, the estimated scale parameter is $\hat{\sigma} = 0.350$.

5. Summary

Robust local regression provides a simple but powerful approach for estimating baseline functions in spectra. The only requirement of the RBE technique is that the baseline must be smooth and vary slowly with respect to the peak widths. This procedure has two crucial parameters that must be specified by the user: the bandwidth d, and the tuning constant b for the robustness weights. We have provided recommendations for values of these parameters that we believe will be appropriate for many baseline subtraction problems ($b = 3$, and d is chosen to be at least 2.5 times the linewidth of the widest peak). It should be kept in mind, however, that these recommendations are not unique, and other choices could be useful for specific baseline estimation problems.

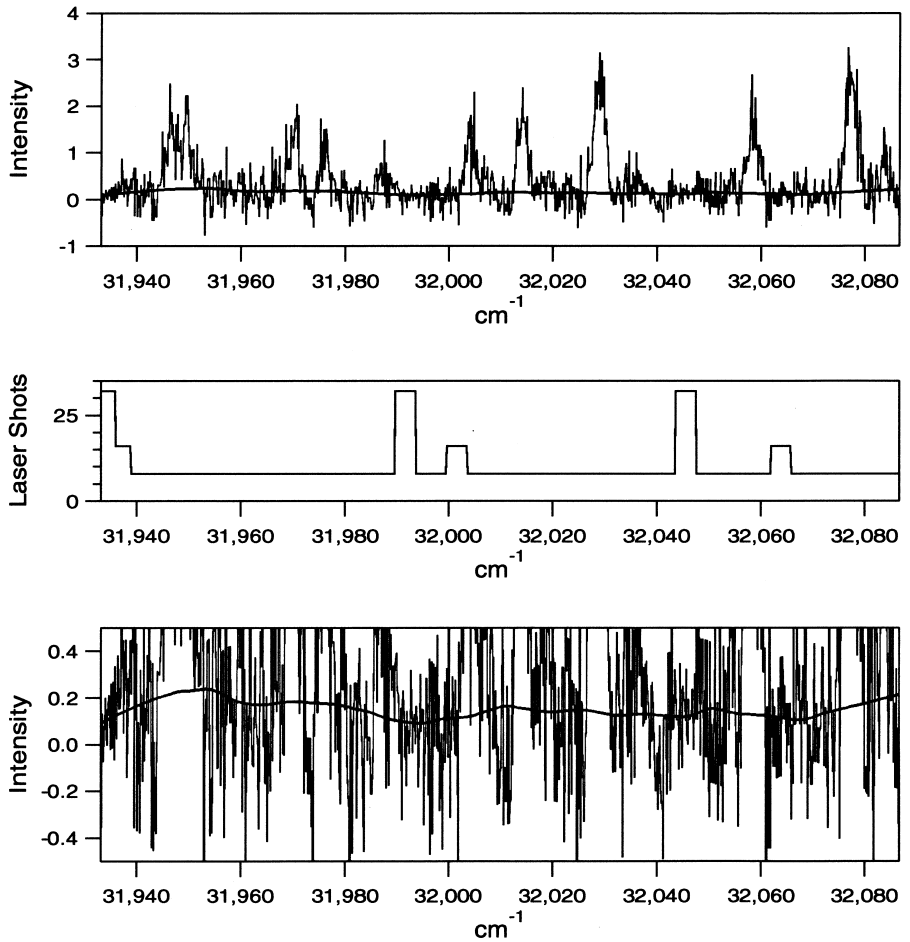

Fig. 4. Experimental laser-induced fluorescence spectrum of OH formed by reaction of hydrogen atoms with ozone. This spectrum differs from the one in Fig. 3 in that the measurement error varies across the spectrum due to a deliberately nonconstant scan rate, which is represented in the middle panel (number of laser shots averaged per point in the spectrum). The baseline determined by RBE, with $b = 3$ and $d = 200$ is represented as a solid line in the top and bottom (magnified view) panels.

Acknowledgements

This research was completed while A.F.R. was at the Australian National University, Canberra, Australia. M.P.J. gratefully acknowledges support from the Department of the Army under a National Defence Science and Engineering Graduate Fellowship, and from the Fannie and John Hertz Foundation. The work of M.P.J. and R.W.F. was supported by the Air Force Office of Sponsored Research, under Grant No. F49620-97-0040. J.A.D. gratefully acknowledges support from the National Science Foundation under Grant No. ATM-9714996, and from the Air Force Office of Scientific Research under Project 2303ES, Task 92VS04COR.

Appendix. Deriving robustness weights

In this appendix, we provide a brief discussion of how to choose appropriate robustness weights for a specific estimation problem. The basis for robust estimation is that, when the quadratic function in least squares estimation (see Eq. (3)) is replaced by a function $\rho(r)$ that goes to infinity at a slower rate as $|r|$ approaches infinity, the fit will show a more stable behavior when extreme observations are present. The robust optimization problem is defined by

$$\hat{\boldsymbol{\theta}}(x_o) = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} K\left(\frac{x_i - x_o}{h}\right) \rho\left\{\frac{y_i - [\theta_0 + \theta_1(x_i - x_o)]}{\sigma}\right\}. \tag{A.1}$$

The effect of extreme observations on a given estimator can be described by the so-called influence function (IF): if $\hat{g}(x_o)_{+[x,y]}$ denotes the estimator of g at x_o based on the sample augmented by a pair of data $[x, y]$, then

$$n[\hat{g}(x_o)_{+[x,y]} - \hat{g}(x_o)] \approx \mathrm{IF}(x, y) = \psi\left\{\frac{y - [\theta_0 + \theta_1(x - x_o)]}{\sigma}\right\} [M_0 + M_1(x - x_o)] K\left(\frac{x - x_o}{h}\right) \tag{A.2}$$

in which

$$\psi(r) = \frac{\partial}{\partial r}\rho(r), \tag{A.3}$$

n is the sample size, and M_0 and M_1 are suitable constants.

To bound the influence of extreme values on the estimator $\hat{g}(x_o)$, we must put a bound on the influence function. The influence function is a product of two factors, namely the *influence of residuals*

$$\psi\left\{\frac{y - [\theta_0 + \theta_1(x - x_o)]}{\sigma}\right\} \tag{A.4}$$

and the *influence of position*

$$[M_0 + M_1(x - x_o)] K\left(\frac{x - x_o}{h}\right). \tag{A.5}$$

The influence of position is bounded because the kernel weight function K is chosen to go to zero much faster than $|x - x_o|$ goes to infinity. The influence of residuals, on the other hand, may be bounded or unbounded depending on the choice of the function $\rho(r)$. The least-squares estimator has the ψ-function $\psi(r) = r$ [see Fig. 5(a)] and consequently, the influence of residuals is unbounded. To bound the (total) influence $\mathrm{IF}(x, y)$ it is sufficient to put a bound on ψ as, for example, in Huber's famous proposal

$$\psi(r) = \min[\max(r, -c), c] \tag{A.6}$$

[Fig. 5(b)], where c is a tuning constant, the choice of which is somewhat arbitrary. A standard procedure is to calibrate the tuning constant at the Gaussian distribution such that the efficiency loss with respect to the optimal estimator assuming Gaussian error is not too large. The value of

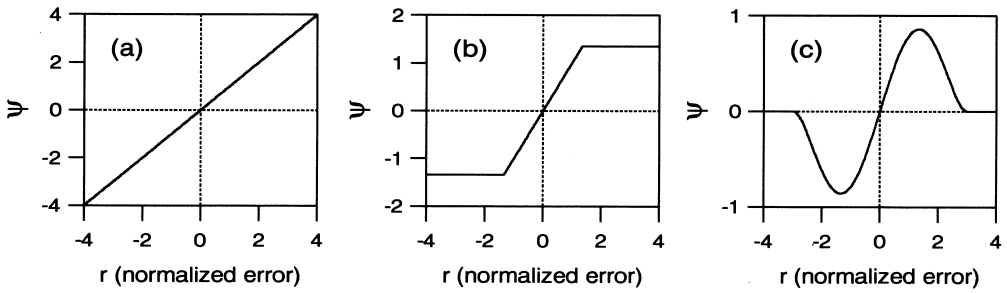

Fig. 5. ψ-functions of ordinary least squares (a), Huber's M-estimator with tuning constant 1.345 (b), and Tukey's bisquare M-estimator with tuning constant 3.0 (c). Since the ψ-function is proportional to the influence function, the influence of (distant) outliers in the response on the estimator can be inferred from these graphs.

$c = 1.345$ leads to an asymptotic efficiency of 0.95 relative to the most optimal procedure at the standard Gaussian distribution.

If there are very severe outliers, one often wants to disregard them altogether as is the case in baseline subtraction where the outliers are peaks that perturb the baseline signal. Using Huber's proposal, severe outliers still have some influence on the estimator $\hat{g}(x_o)$. We can achieve better performance when severe outliers are present by using a *redescending* ψ-function such as Tukey's bisquare ψ-function:

$$\psi(r) = r\{\max[1 - (r/b)^2, 0]\}^2. \tag{A.7}$$

Fig. 5(c) shows that distant outliers have no effect at all on this estimator. When calculating this estimator one must, however, proceed with caution since there might be local minima. Using the same argument as with Huber's proposal a standard choice of the tuning constant b is 4.685. When introducing the LOWESS procedure, Cleveland opted for a more robust estimator ($b = 4.05$) at the expense of some loss in efficiency with respect to the most optimal estimator. We have chosen an even more robust estimator for baseline estimation ($b = 3.0$), because in typical spectra a relatively large fraction of the spectral points lie on peaks, and are thus outliers from the standpoint of estimating the baseline.

The relationship between the influence function and the ψ-function allows us to design suitable robustness procedures for specific problems. Since Eq. (A.1) cannot be solved explicitly, an iterative procedure must be applied at any x_o. A simpler and more efficient approach is to determine robustness weights $w_r(x_i)$ globally and solve a weighted least-squares problem locally with weights

$$w_r(x_i)K\left(\frac{x_i - x_o}{h}\right), \tag{A.8}$$

where the robustness weights $w_r(x_i)$ are computed from

$$w_r(x_i) = \psi(r_i)/r_i, \tag{A.9}$$

with $r_i = [y_i - \hat{g}(x_i)]/\sigma$. The estimates $\hat{g}(x_i)$ are determined from the previous iteration. Hence we can solve the least-squares problem at x_o explicitly but repeat the smoothing a few times, adjusting the robustness weights at each iteration.

More detailed discussions of robustness are given, e.g., in Refs. [14,21,22] or [18]. Outliers are not the only source of deviations from a parametric model. See Chapters 1 and 8 of Hampel et al. [14] for a more detailed discussion.

References

[1] Phillips AJ, Hamilton PA. Improved detection limits in Fourier transform spectroscopy from a maximum entropy approach to baseline estimation. Anal Chem 1996;68:4020–5.
[2] McNulty DA, MacFie HJH. The effect of different baseline estimators on the limit of quantification in chromatography. J Chemom 1997;11:1–11.
[3] Xiao H, Levine SP. Application of computerized differentiation technique to remote-sensing Fourier transform infrared spectrometry for analysis of toxic vapors. Anal Chem 1993;65:2262–9.
[4] Burgess DD. A comparison of methods for baseline estimation in gamma-ray spectrometry. Nucl Instr Meth Phys Res 1984;221:593–9.
[5] Padayachee J, Prozesky V, von der Linden W, Nkwinika MS, Dose V. Bayesian PIXE background subtraction. Nucl Instr Meth Phys Res 1999;150:129–35.
[6] Kneen MA, Annegard HJ. Algorithm for fitting XRF, SEM and PIXE X-ray spectra backgrounds. Nucl Instr Meth Phys Res B 1996;109/110:209–13.
[7] Maxwell JA, Campbell JL, Teesdale WJ. The Guelph PIXE software package. Nucl Instr Meth Phys Res B 1989;43:218–30.
[8] Friedrichs MS. A model-free algorithm for the removal of baseline artifacts. J Biomol NMR 1995;5:147–53.
[9] Dietrich W, Rüdel CH, Neumann M. Fast and precise automatic baseline correction of one- and two-dimensional NMR spectra. J Magn Reson 1991;91:1–11.
[10] Marion D, Bax A. Baseline correction of 2D FT NMR spectra using a simple linear prediction extrapolation of the time-domain data. J Magn Reson 1989;83:205–11.
[11] Rouh A, Delsuc M-A, Bertrand G, Lallemand J-Y. The use of classification in baseline correction of FT NMR spectra. J Magn Reson Ser A 1993;102:357–9.
[12] Froning JN, Olson MD, Froelicher VF. Problems and limitations of ECG baseline estimation and removal using a cubic spline technique during exercise ECG testing: Recommendations for proper implementation. J Electrocardiol 1988;S149–57.
[13] Balcerowska G, Siuda R. Inelastic background subtraction from a set of angle-dependent XPS spectra using PCA and polynomial approximation. Vacuum 1999;54:195–9.
[14] Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA. Robust statistics: the approach based on influence functions. New York: Wiley, 1986.
[15] Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 1979;74:829.
[16] Simonoff JS. Smoothing methods in statistics. New York: Springer, 1996.
[17] Fan J, Gijbels I. Local polynomial modelling and its applications. Monographs on statistics and applied probability, vol. 66. London: Chapman and Hall, 1996.
[18] Rousseeuw PJ, Leroy AM. Robust regression and outlier detection. New York: Wiley, 1987.
[19] Ruckstuhl AF, Stahel WA, Dressler K. Robust estimation of term values in high-resolution molecular spectroscopy: application to the $e^3\Sigma_u^+ \rightarrow a^3\Sigma_g^+$ spectrum of T_2. J Mol Spectrosc 1993;160:434.
[20] Dodd JA, Lockwood RB, Hwang E, Miller S, Lipson SJ. Formation of OH ($v = 0, 1$) by the reaction of fast H with O_3. J Phys Chem 1999;103:7834.
[21] Hoaglin DC, Mosteller F, Tukey JW, editors. Understanding robust and exploratory data analysis. New York: Wiley, 1983.
[22] Huber PJ. Robust statistics. New York: Wiley, 1981.