Solution of an "impossible" diffusion-inversion problem

J. Ross Macdonald
Department of Physics and Astronomy, University of North Carolina, Chapel Hill, North Carolina 27599-3255

(Received 19 October 1994; accepted 15 March 1995)

The recent conclusion of Craig and Thompson [Comp. Phys. **8**, 648 (1994)] that the inversion of spatial diffusion data to estimate accurately an initial delta-function temperature source distribution is an unstable, impossible problem, is demonstrated to be far too pessimistic. Instead of inverting such data with a method appropriate for continuous distributions as they did, inversion was carried out with one appropriate for a discrete distribution. The inversion involved simultaneous estimation of both line strengths and their positions. It yielded highly accurate estimates of the strength and position of the source, even with very large amounts of noise. Several inversions of noisy data are illustrated for the normalized time value used by Craig and Thompson and for very much longer times. Further, it is demonstrated that even for very noisy data it is possible to discriminate between the presence of one line or of several lines, and to conclude unambiguously whether a given source distribution is discrete or continuous. Finally, a closely related method appropriate for estimation of continuous source distributions was used to invert diffusion data calculated from a very narrow, continuous source distribution approximating a delta function. Even with noise present, such continuous source distributions can be reliably estimated, but the narrower the distribution the less well can it be distinguished from a single line, and the less important such discrimination becomes. © *1995 American Institute of Physics.*

INTRODUCTION

In a recent study concerning the difficulty of inverting Laplace transforms accurately, Craig and Thompson[1] highlighted the numerical difficulties of inversion by treating explicitly the recovery of a delta-function source distribution from noisy data arising from one-dimensional diffusion sampled discretely over a restricted range of distance, $x_{\min}$ to $x_{\max}$. They characterized this problem as "impossible," because they state that accurate estimation of the spatial temperature-distribution source function at $t=0$ from the spread-out, noisy diffusion data profile at some later time $t>0$ is "tantamount to reversing entropy." As I shall show, a good estimate is indeed possible for a delta-function source.

For an arbitrary source distribution, Allison[2] has cited references to related diffusion-type problems where one wishes to infer a past state of a system from its present state, called by him "travel in the past," or "the recovery of past events." Further, a recent inversion treatment of diffusion and convection of a groundwater contaminant[3] shows that even in the presence of noise it is possible to obtain an approximate, but useful, estimate of the time-dependent continuous-function source distribution. In the following, I shall follow Craig and Thompson by assuming that all quantities such as x and t are normalized and dimensionless.

Many inversion problems of physical interest involve the solution of the inhomogeneous Fredholm equation of the first kind for the unknown source function $f(y)$, where

$$u(x)=\int_{-\infty}^{\infty}K(x,y)f(y)dy; \tag{1}$$

$K(x,y)$ is the kernel or weight function; and $u(x)$ is the response function, usually represented by discrete data. Inversion to find $f(y)$, given the data and the form of $K(x,y)$, is well known to usually be an ill-posed problem for this integral equation, especially when $f(y)$ is a continuous function.[2,4] For such situations, the estimated source function is extremely sensitive to small changes in the data. By contrast, a well-posed computing problem is one where the solution exists and is unique, and small changes in the data result in only small changes in the computed result.[4]

For many situations, including diffusion, it is possible to express the kernel in convolution form. Then, Eq. (1) becomes, when parameterized for the one-dimensional diffusion problem mentioned above,

$$u(x,t)=\int_{-\infty}^{\infty}h_t(x-y)f(y)dy, \tag{2}$$

where the parameter t has been introduced to allow the specification of the time after $t=0$ at which a diffusion profile is measured. The general diffusion kernel may be written

$$h_t(z)\equiv\beta(Dt)^{-1/2}\exp(-z^2/4Dt),$$

where Craig and Thompson[1] have used $\beta=4\pi$, a value which will also be used here for easy comparison of results. But the conventional value, $\beta=(4\pi)^{-1/2}$, is consistent with the boundary condition $u(x,0)=f(x)$ because[5] in the limit $t\to0$, $h_t(z)\to\delta(z)$. In keeping with the approach of Ref. 1, the diffusion coefficient D is initially taken dimensionless and equal to unity.

Although both x and y are space variables here, for easy distinction I shall use the x variable only for the data and the y variable only for the source function. Because experimental data are discrete, let $x\to x_n$, with $0\leqslant n\leqslant N$. Since a distribution may be discrete, continuous, or possibly a combination of the two, it can be expressed as $f(y)=d(y)+c(y)$, where $d(y)$ includes all discrete lines and $c(y)$ all continuous distributions. For the present work, where only either discrete or continuous source distributions need be considered, take $f(y)=p(y)$, where $p(y)$ is either $d(y)$ or $c(y)$.

Consider now the estimation of M points of the $p(y)$ distribution, each located at y_m, where $1\leqslant m\leqslant M$. Then for the discrete case,

$$p(y)=d(y)\equiv\sum_{m=1}^{M}d_m\delta(y-y_m),$$

and d_m is the strength of the mth line. Here the d_m quantities do not represent points on a continuous curve since the distribution has been assumed to be discrete. The situation is different for a continuous source function, one where M discrete inversion estimates indeed approximate a continuous curve. Therefore, set $p(y_m)=c(y_m)\equiv c_m$. The integral of Eq. (2) immediately reduces to an exact sum when $f(y)$ is replaced by the above expression for $d(y)$, but it must be approximated by a numerical quadrature for the continuous case. Let w_m denote the required quadrature weights. It then follows that Eq. (2) may be written as

$$u(x_n,t)=\sum_{m=1}^{M}q_mh_t(x_n-y_m), \tag{3}$$

where $q_m=d_m$ for a discrete distribution and $q_m=w_mc_m$ for the continuous case. Then Eq. (3), which is necessarily approximate in the continuous case, readily allows $u(x_n,t)$ to be calculated when the pairs $\{q_m,y_m\}$ are known, and it further allows the $\{q_m,y_m\}$ quantities, which define the distribution, to be estimated when the $u(x_n,t)$ data values are known. It is important to emphasize that the present inversion approach thus involves a weighted summation over the kernel whether the source is discrete or continuous.

Data are not only discrete but usually contain errors, so one must generally estimate all significant $\{q_m,y_m\}$ values from noisy data, $u_{\text{meas}}(x_n,t)$. Here I shall present some inversion results for exact data and for noisy data of two types. To allow direct comparison with the results of Craig and Thompson,[1] I take $0\leqslant x_n\leqslant64$ and $y_1=32$. Exact discrete-source data were thus calculated from Eq. (3) using double-precision arithmetic with $M=1$ and $d_1=1$. For the present analysis, the two types of noise involved either proportional errors or additive errors, defined by

$$u_{\text{meas}}(x_n,t)=u(x_n,t)[1+\sigma_eP(0,I_n)], \tag{4}$$

and

$$u_{\text{meas}}(x_n,t)=u(x_n,t)+u(x_p,t)\sigma_eP(0,I_n), \tag{5}$$

respectively, where x_p is the value of x at the peak of the $u(x_n,t)$ response curve. Here $P(0,I_n)$ denotes a value

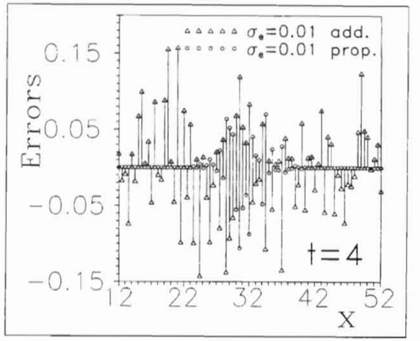

Figure 1. Additive and proportional errors for $t=4$, produced with seed 1 and with $\sigma_e=0.01$.

sampled randomly from a distribution having zero mean and unity standard deviation. Thus σ_e sets the standard deviation of the individual errors.

In the present analysis, all errors are drawn from a normal (Gaussian) distribution and are independent and uncorrelated. The relative standard deviation of the overall error distribution is just σ_e, but it is relative to each individual $u(x_n,t)$ value for the proportional errors case and relative to the maximum $u(x_p,t)$ value for additive errors. Thus, with the same choice of the seed used to obtain a pseudorandom number, the proportional error is the same as the additive error for a given value of σ_e only at $x=x_p$. Figure 1 compares such overall errors for $t=4$ and $\sigma_e=0.01$, which will be denoted loosely as a 1% error situation. For clarity, the proportional errors are shown with opposite sign to the additive ones, and we see that they decrease rapidly in magnitude as x_n moves away from $x_p=32$ since h_t decreases away from its maximum. Although experimental data are likely to contain a combination of independent proportional and additive errors with $(\sigma_e)_{\text{add}}<(\sigma_e)_{\text{prop}}$, we shall not investigate the effects of such composite errors here, but the fitting program used for the present inversions, LEVM, allows weighting with such a combination of error types.[6,7]

I. THE DV AND CV INVERSION METHODS

Equation (3) shows both how similar discrete-source and continuous-source inversions are, but it also makes clear the significant differences between them. One further common feature of great importance is the present treatment of the y_m position parameters. The most significant way in which the present inversion method differs from most previous ones is its inclusion of an arbitrary number M of source-function points whose positions are free to vary and are estimated as part of the fitting,[8] rather than being held fixed.[1,3] Thus, it is often appropriate to fit first with $M=1$, then 2, and continue incrementing until no further significant parameter estimates can be obtained, either because

0894-1866/95/9(5)/546/8/$6.00

and/or because errors in the data limit further estimation.

Now let us distinguish between the discrete and continuous inversion methods considered here. Let "DV" stands for "Discrete inversion with all source parameters free to Vary," and "CV" designate "Continuous inversion with all source parameters free to Vary." The DV and CV designations refer to the two kinds of analysis methods incorporated in Eq. (3). Because of the similarity between these approaches, it turns out that either one can be used to determine whether given data involve a discrete or a continuous source distribution. Thus, it is unnecessary to make any initial assumptions about the character of the distribution. In fact, inversion analysis of noise-free or moderately noisy data by either method allows an unambiguous determination to be made between any source-estimate points associated with a discrete distribution and any associated with a continuous distribution.[8] Once this determination has been made, however, DV inversion should be used for discrete-source situations and CV for continuous- or continuous-and-discrete-source conditions. Such discrimination is further discussed below.

DV inversion,[6–10] is simple in principle and involves just the weighted, nonlinear-least-squares fitting of all the $\{d_m, y_m\}$ parameters of Eq. (3), with the left-hand side replaced by $u_{\text{meas}}(x_n, t)$. CV inversion is similar but involves the weighted, nonlinear-least-squares estimation of the $\{c_m, y_m\}$ parameters.[8] A computer routine for DV inversion has been a part of the general impedance-spectroscopy complex-nonlinear-least-squares fitting program LEVM[6,7] for the past decade. This program was recently generalized to include the CV fitting procedure as well.[8] It involves extended trapezoidal quadrature with weights which account for nonequal spacing[11] of the y_m values and yields satisfactory results for the finite-length range $-\infty < y_{\min}$ to $y_{\max} < \infty$, especially when the c_m end points are adjusted to help correct for the effects of a truncated range.[8] It is worth remarking that y_m values obtained from the DV and CV inversions of the same continuous-source data generally differ somewhat. For the present inversions, LEVM was run with a stringent convergence criterion, and it itself involves some regularization[1,3] arising from the nonlinear-least-squares fitting procedure used.[8] A minor difference between the DV and CV methods is that DV d_m values should not be normalized, while CV c_m ones are generally normalized so that the area defined by the continuous-distribution curve is unity.

Because the DV and CV inversion approaches essentially differ only in their weighting and normalization, much of the present discussion of the DV method also applies to the CV one. When it is known *a priori* that all the distribution-strength parameters must be ≥ 0, the usual physical situation and the one appropriate for the present analysis, a constraint of this type is included in the inversion procedure, although with the present variable-y_m values it is not usually needed. But no other constraints are needed for inversion of either discrete or continuous data. In particular, no other *a priori* assumptions whatsoever about the form of the distribution need to be included in the inversion procedure. Whether the distribution is found to be discrete or continuous is entirely determined by the data fitting, not by an initial choice of the DV or CV approach.

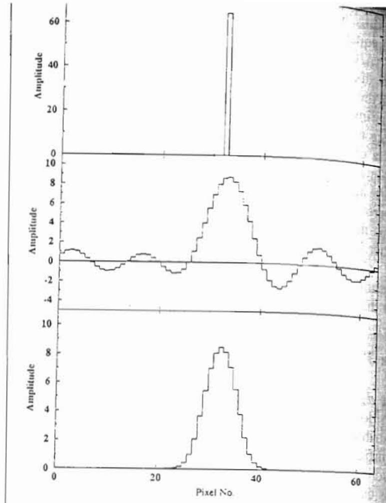

Figure 2. From top to bottom: representation of a delta-function source distribution centered at about 32; result of a second-order regularization inversion to find the source function using t=4 data with 1% proportional errors; and result of a maximum-entropy inversion of the same data, one which constrains the source estimates to be non-negative. The designation "Pixel No." refers to the present y variable. Reproduced from Ref. 1, Fig. 5.

One thus does not need to assume initially that the data arise from a discrete (or a continuous) distribution to find and estimate the proper distribution. But it should be noted that inversion of a discrete distribution is well-posed and that for a continuous one is generally ill-posed.

II. INVERSION RESULTS

A. Delta-function source distribution

1. Data with t=4

Although they characterized the recovery of a delta-function source distribution from noisy diffusion data as an impossible problem, Craig and Thompson[1] presented results of two different inversions for the choice $t=4$. Figure 2 reproduces their Fig. 5, which illustrated these results. Note the oscillations in the middle-panel inversion results. Craig and Thompson conclude that "the (maximum-entropy) recovery manages to exploit the bulk of the uncontaminated information in the data and can be plausibly interpreted as a minimum structure model of the source function." But even these results, whose peak occurs close to the correct value, are of low resolution and are so much broadened that one would be unlikely to think of comparing the result with a delta function unless prior information about the actual function were available.

% error	Proportional errors S_F	d_1	y_1	Additive errors S_F	d_1	y_1
0	1.26×10^{-13}	$1\pm1.4\times10^{-14}$	$32\pm9.4\times10^{-15}$	1.24×10^{-13}	$1\pm6.2\times10^{-15}$	$32\pm2.5\times10^{-14}$
1	0.0101	0.9998±0.0011	32.0002±0.0008	0.0101	1.0007±0.0032	32.0005±0.0128
30	0.3051	0.9945±0.034	31.989 ±0.023	0.3038	1.0214±0.094	32.011 ±0.0376
100	1.010	0.9893±0.112	31.923 ±0.063	1.013	1.071 ±0.320	32.022 ±1.194
200-1	2.029	0.9793±0.222	31.89 ±0.08	2.025	1.143 ±0.640	32.028 ±2.24
200-2	1.620	1.171 ±0.211	32.09 ±0.07	1.931	1.431 ±0.610	31.193 ±1.71
200-3	2.461	0.7640±0.210	31.97 ±0.07	1.899	1.062 ±0.599	33.009 ±2.26

It is reasonable to ask, Why were these unsatisfactory results obtained and are superior ones possible? The answer to the first part of the question is that data derived from an inherently discrete distribution were analyzed by methods developed for, and appropriate for, a continuous distribution. These methods start with a set of fixed points in y space, usually equally spaced, and calculate estimates of the source distribution at each of these points.[1,3] They have also been applied in the past to other discrete-distribution situations,[12,13] again leading to unsatisfactory results. As shown below, much superior results are indeed possible using DV inversion. In particular, such inversion with the y_m position parameters free to vary eliminates the oscillations usually characteristic of fixed y_m inversion methods[1,8,13] (see the middle panel of Fig. 2).

Some DV inversion results are shown in Table I for the $t=4$ case considered by Craig and Thompson. The x range used here was 12(0.5)52, and each calculated data value was truncated after 13 decimal digits. The results shown in the table were obtained using the LEVM program with function-proportional weighting[6,7] for the data with proportional errors and with unity weighting for those containing additive errors. The % error column lists values of $100\sigma_e$, and only the results of $M=1$ fits are shown. The ± values are estimated standard deviations of the parameter estimates. Finally, the quantity S_F is the relative standard deviation of the weighted data-fit residuals, and, as we see, it is an excellent estimator of σ_e when the latter is small.

For zero added errors, it is evident from Table I that the inversion method recovers the original parameters with even less estimated uncertainty than the error arising from the truncation of the data values. So far so good, but what happens when inversion is carried out with values of M of 2 or larger? For $M \geq 2$, one again finds essentially exact estimates for $\{d_1, y_1\}$ and estimates of d_m ($m>1$) of 10^{-8} or less, unequivocally demonstrating that the source distribution is discrete and involves only a single delta function.

In the present work, the same random-error seed has been used in forming all noise contributions, unless otherwise noted. It was used for the row marked 200-1 in the table, but different seeds were used for the noisy data employed for each of the last two rows. In Fig. 3 the exact data, the data with 200-1% additive errors, and the fit points are all compared. The uncertainties of the $\{d_1, y_1\}$ parameter estimates listed in the table are clearly much smaller for the proportional-error situation than for the additive-error one, just as one would expect. Three fits for 200% error are shown in the table to allow comparison of results for data with independent error sets, each with the same σ_e. These results show that even with very large errors in the data, DV allows one to obtain good estimates for the $\{d_1, y_1\}$ pair that defines a single delta function.

For $M \geq 2$ fits of noisy data constructed with a single delta-function source distribution, the situation is somewhat different. For proportional errors, one finds similar results to those mentioned above for the zero-added-error case, but for additive errors the standard deviations of all $m>1$ parameters are usually very large, even when the $\{d_1, y_1\}$ estimates themselves are reasonable. Again, such results allow one to conclude that even with large error present one can identify the presence of a single line, characterize it well, and rule out unambiguously the presence of two or more lines or of a continuous distribution. Further, earlier work on the electrical response of dielectric materials demonstrated how well two spectral response lines of much different strengths and nearly the same relaxation time can be resolved and quantified by DV inversion of their frequency response data.[8]

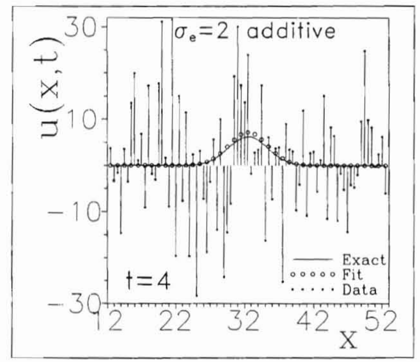

Figure 3. Data values: exact t=4 curve; inversion fit of data with $\sigma_e=2$, additive; and noisy data with $\sigma_e=2$, additive.

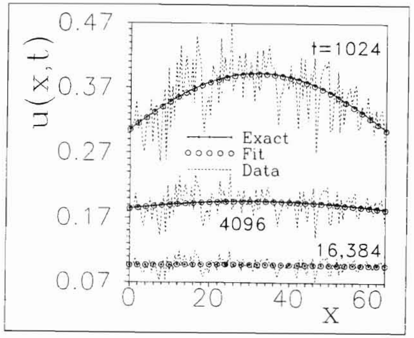

Figure 4. Data curves and points with σ_e=0.1 proportional errors for t=1024, 4096, and 16,384. For the noisy data, dotted lines connecting the discrete points have been included to guide the eye.

2. Retroactive inversion for large t

For $t=4$, the data curve is still relatively narrow, and we have seen that DV inversion of even very noisy $t>0$ data allows the $t=0$ source function to be very well estimated. But what happens when the measurement of the diffusion profile is made at much longer times? In this section, some inversion results that attempt to answer this question are presented. Take the x range as 0(0.5)64 and consider 10% proportional data errors analyzed with function-proportional weighting. Figure 4 shows the exact data, noisy data, and fit results for $t=1024$, 4096, and 16384. Since the exact data curves are nearly flat in this long-time region, there will not be much difference here between proportional and additive errors.

Inversions were made for data with $4 \leq t \leq 16384$ for each of the three seeds used for the 200%-error data in Table I. The results for y_1 and its estimated standard deviation are shown for $t=256$, 1024, 4096, and 16384 in Fig. 5. The response lines satisfy the equation

$$y_1 = (y_1)_{\text{exact}}(1+\alpha t) \qquad (6)$$

very closely. For seed 1, $\alpha \simeq -2.84 \times 10^{-6}$, while it is positive and larger for data calculated with the other two seeds. Although α is nearly proportional to σ_e for seed 3, it is about $+9.21 \times 10^{-8}$ for seed 1 and $\sigma_e=0.05$. The d_1 estimates were found to be independent of t for $t>4$ and were about 0.997, 1.012, and 1.008 for seeds 1, 2, and 3, respectively. As Fig. 5 shows, the errors in y_1 begin to be sufficiently large by $t=4096$ that differences between data constructed with different seeds begin to become significant. For this value of t, the ratio between the peak value of the exact data and the end values is only about 1.07, so the variation is less than σ_e itself. A Monte Carlo study would allow one to examine the bias in the y_1 estimates. For zero bias, the average α should approach zero as the number of independent replications is increased.

As t increases, the decreasing variation in the exact data begins to be swamped out by the noise when the data

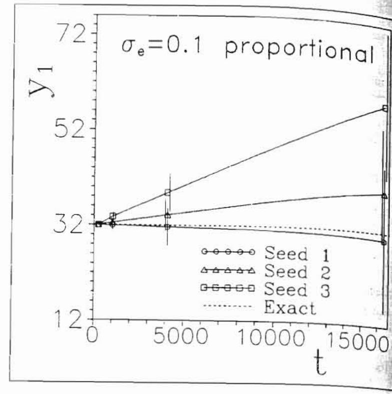

Figure 5. Dependence of inversion estimates of y_1, the line-source position estimate, on t for σ_e=0.1 proportional-noise data for three different random-error seeds. The data for seed 1 are presented in Fig. 4. In order to allow the vertical lines showing the ± one-standard-deviation range to be easily distinguished, those for seeds 2 and 3 have been displaced slightly to the left and right of their proper positions, respectively.

extend over a fixed range, as is the case here. But if one could extend the measured x range as t increased, this effect could be counteracted and good estimates of $\{d_1, y_1\}$ could then be obtained if σ_e remained constant. If this were possible, the present problem would not be impossible even for arbitrarily large t. For the nearly flat noisy data shown in Fig. 5, it is clear that an excellent estimate of the $t=0$ source distribution can be obtained from diffusion-profile measurements made long thereafter. Thus, the inversion problem is neither ill-posed nor particularly ill-conditioned. It is therefore indeed possible to infer some past events accurately long after their occurrence and so, in the sense of Craig and Thompson, to reverse the increase in entropy as diffusion progresses.

B. Approximate-delta-function continuous source distribution

1. Background

Although discrete source distributions made up of one or more ideal delta-function components are physically realizable to high approximation in spectroscopy,[8] where isolated spectral lines or relaxation times of negligible breadth appear, the assumption of a spatial temperature source of delta-function character, such as that considered for the above diffusion problem, is physically nonrealizable and can only be approximated in experimental situations. Therefore, it is important to investigate inversion for data constructed with an analytical approximation to a delta function for the source distribution, thus requiring us to

Table II. Comparison of exact and noisy inversion results for a continuous source distribution with t=4 and λ=4. Exact values are enclosed in parentheses.

	$\sigma_e=0$			$\sigma_e=0.001$		
M	S_F	c_m	y_m	S_F	c_m	y_m
1	0.0304	1.032 ± 0.0035	32.000±0.0022	0.0302	1.032 ± 0.0035	32.000±0.0022
2	2.5 × 10^{-4}	1.397 ± 0.0092 1.397 ± 0.0092	31.821±0.0011 32.179±0.0011	0.0011	1.387 ± 0.039 1.415 ± 0.039	31.820±0.0048 32.177±0.0047
3	1.3 × 10^{-6}	0.5328 ± 0.0075 (0.4947) 2.1774 ± 4.0 × 10^{-6} (2.2567) 0.5328 ± 0.0075 (0.4947)	31.692±2.4×10^{-5} 32.000±2.4×10^{-5} 32.308±2.4×10^{-5}	0.0010	7.5 × 10^{-5} ± 5.44 (2.3 × 10^{-23}) 0.5693 ± 0.8463 (1.5745) 1.1990 ± 0.1455 (1.1596)	30.181±0.051 31.850±7.6×10^{-4} 32.204±6.9×10^{-4}
3				0.0010	0.5794 ± 0.0134 2.1801 ± 0.0022 0.5738 ± 0.0146	31.7 32.002±1.6×10^{-4} 32.3

deal with a continuous distribution rather than a discrete one. A suitable approximation for finite λ is[5]

$$f(y) = \lim_{\lambda \to \infty} (\lambda/\sqrt{\pi}) \exp\{-[\lambda(y-y_0)]^2\}, \qquad (7)$$

which becomes exact in the large-λ limit. Here y_0 specifies the center of the distribution, again taken as 32. For finite λ, the diffusion data are then given by

$$u(x_n,t) = [4\lambda\sqrt{(\pi/Dt)}]\int_{-\infty}^{\infty} \exp\{-[(x_n-y)^2/4Dt]\} \times \exp\{-[\lambda^2(y-y_0)^2]\}dy, \qquad (8)$$

which is again a Gaussian.

The remaining results were obtained using the continuous-distribution inversion subroutine CV described earlier. Because this procedure involves the simultaneous determination of estimates of all the c_m and y_m values during the iterative, weighted least-squares inversion, one cannot use such techniques as Richardson extrapolation to improve integration accuracy. Since source distributions are usually normalized to unity area, as is that of Eq. (7), proper comparison between such a known distribution and an estimated one requires area normalization of the latter as well, using the same approximate integration procedure as that used during the inversion. The c_m values presented here have been so normalized. Note that since area normalization cannot be used for the choice $M=1$, the line in Table II for this value used DV inversion as above, while the others involve CV continuous-function inversion. Thus, if only a single discrete line were present, exact inversion should yield the value $d_m=1$ for this case.

2. CV inversion of error-free data

To illustrate the estimation of a very narrow continuous distribution, I take $\lambda=4$ in Eq. (7), yielding a source width at half height of about 0.4, from about $y=31.8$ to $y=32.2$. When the noise-free data calculated with this value from Eq. (8) are plotted with similar data calculated for an ideal delta-function source, the differences between the results are sufficiently small that neither a linear nor a semilog plot allows them to be adequately distinguished. Therefore, Fig. 6 shows the relative difference between them,

$$R(x_n,t) \equiv [u_{df}(x_n,t) - u_{adf}(x_n,t)]/u_{df}(x_n,t), \qquad (9)$$

where "df" identifies the delta-function data and "adf" the approximate-delta-function data. Results are shown for $12 \leq x_n \leq 52$. Note that the data values cover a range from about 2π at their peaks down to about 10^{-10} at their extremes. This relative difference curve shows that the adf data are very slightly flattened and wider than the df data.

Figure 7 and Table II show source-distribution estimates at $t=4$ for added proportional errors of zero and 0.0001. Let us first discuss the $\sigma_e=0$ results. Table II shows that as M increases, the fit relative standard deviation, S_F, decreases extremely rapidly, indicating that only a few c_m points are needed to fit the data very well. For the $M=7$ results shown in Fig. 7, S_F has decreased to the extraordi-

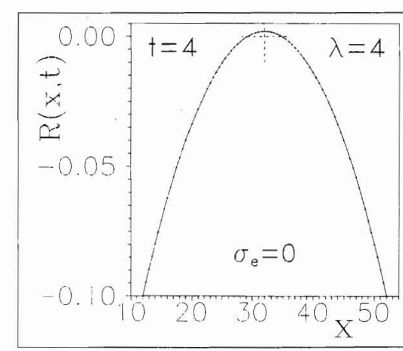

Figure 6. The relative difference between data sets calculated with exact and with approximate delta-function source distributions. No noise added.

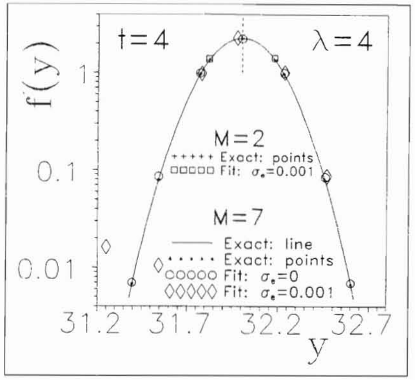

Figure 7. Comparison of the exact continuous source distribution and inversion estimates obtained with $M=2$ and 7, both with and without proportional errors present.

narily low value of 9×10^{-12}. A virtue of the present inversion procedure is that it provides estimates of the standard deviations of the estimated c_m parameters. For $M=7$, the y_m and c_m relative standard deviations are of the order of 10^{-6} and 10^{-4}, respectively.

The asterisk points on the $M=7$ curve were calculated from Eq. (7) using the y_m values obtained from the fitting, while the open circles are the fit results at these same positions. If the fit were perfect, the asterisks would lie at the center of the circles. Although the overall fit is clearly very good, it is evident that the second points from the bottom are slightly too high and deviate from their correct values by much more than their estimated uncertainties. The departure from a perfect estimate of the actual source distribution, even when the fit of the data involves an S_F which is approaching the accuracy of the data, is evidence of the ill-conditioning which must be expected for inversion of continuous distributions such as the present one. Nevertheless, it is remarkable that its effects are as small as they are here.

3. CV inversion of noisy data

The above results make it clear that with no added error, one can obtain accurate enough estimates of points on the distribution that it can be completely distinguished from other types of distributions and that its parameters λ and y_0 can themselves be determined with high accuracy. But in real life one usually does not have the luxury of dealing with data with vanishingly small errors. Data involving $\sigma_e=0.001$ proportional errors were used to obtain the $M=1$, 2, and 3 results of Table II and Fig. 7. First note that there is negligible difference between the $\sigma_e=0$ and 0.001 results for $M=1$. This is because the goodness of fit is limited by fitting a continuous distribution by a single-line discrete one for this choice of M. Although such fitting yields good estimates of the relevant single-line parameters, it leads to a value of S_F much larger than σ_e.

As the results in Table II indicate, when $\sigma_e>0$, S_F cannot be appreciably less than σ_e. Clearly, this restriction limits the maximum number of significant source-parameter values that can be estimated from the data by the present method. For $M=2$, the results in Table II and those plotted in Fig. 7 show that for $\sigma_e=0.001$ one can still obtain good estimates, but by $M=3$, the estimates have deteriorated badly. The quantities in parentheses in the middle $M=3$ section of Table II are exact values which should be compared with the fit values immediately above them. Since LEVM allows any parameters to be free or fixed, the last $M=3$ part of the table shows what happens when two y_m's are held fixed during the inversion. Much improved estimates are then obtained, with a standard deviation of the overall fitting errors of about 7%. These results, and those with $M=1$ and 2, are sufficient to again allow one to identify both the type of distribution and to estimate its parameters.

Such estimation may be accomplished with direct nonlinear least squares fitting, again using LEVM. Although there are too few $\{c_m, y_m\}$ values for the noisy $M=2$ and 3 results in Table II to allow good fitting statistics to be obtained, fits of the $M=2$ and bottom $M=3$ results to Eq. (7) nevertheless yielded for (λ, y_0) the estimates (3.95,32) and (3.84,32), respectively. An important virtue of the present CV and DV inversion methods is that one need not know the form of $f(y)$ or the most appropriate value of M when carrying out the inversion to determine these quantities. But if an inversion estimate of discrete $f(y)$ values suggests that a particular analytic form of the source distribution is likely, one can fit the noisy experimental data to an equation such as Eq. (8) and obtain estimates of the relevant parameters of both the kernel and the source function. For the present noisy data, one such fit using Eq. (8) yielded (3.9,32) and 0.9998 for (λ, y_0), and D, respectively, quite close to their exact values of 4, 32, and 1. Even if an inaccurate initial value of D were used for the inversion, determination of an adequate parameterized shape for the source function should be possible, and then direct fitting could be used to obtain good estimates of all the parameters, including D.

The basic problem addressed here is one of distinguishing in the presence of noise between a single delta-function line and a narrow continuous distribution. The larger λ, the narrower the distribution, and thus the smaller the zero-error value of S_F will be for a given small value of M such as 2. Therefore, as one might expect, for data with small but non-negligible errors the harder it will be to distinguish between a delta-function source distribution and a narrow approximate one. But, luckily, the narrower the continuous distribution, the better its approximation to a single line, and the less important it will be to distinguish between the two. As the source distribution becomes narrower and narrower, the better it will be able to be represented by a single spectral line, and the greater will be its robustness to added noise, as demonstrated for the above DV results. Conversely, for a given noise level, the wider the source distribution, the larger the number of statistically significant source parameters that can be estimated before the noise becomes limiting.

Incidentally, one would expect from the above results and discussion that no meaningful values of c_m and y_m could be obtained for $\sigma_e=0.001$ and $M\geq7$. And, in fact, this is true in some sense. But, suppose, for example, that one initially obtains $M=7$ fit-parameter estimates either for zero-error data (not a viable possibility for real data!) or from a fit of the noisy data with evenly spaced, fixed y_m values,[8,13] and then uses these estimates as initial values in the fitting of the noisy data with all parameters free and $\sigma_e=0.001$. One then usually obtains a fit with S_F near 0.001 and with several of the parameter estimates not very far from the original ones. Such estimates are shown by the points denoted by the open double triangle symbols in Fig. 7. We see that although the top four points are reasonably good, the bottom two are terrible, and the seventh, whose c_m estimate was about 3×10^{-6}, was far too small to fit on the figure.

The reader might still conclude that the top four points were close enough to allow distribution parameters to be estimated. But for the present $M=7$ fit the estimated relative standard deviations of the y_m parameters were of the order of 100 and those of the c_m values were of the order of 10^4. Thus, statistically speaking, one cannot distinguish any of them from zero. Nevertheless, some of the parameter values show only relatively small errors and are therefore still useful. Regularization inversion procedures for continuous distributions with noise, where the y_m values are fixed,[1,3] allow one to obtain a large number of points to define the source distribution, but it is unusual to find uncertainty estimates quoted for these points. The present results suggest that estimates of the standard deviations of all free-parameter estimates should be cited for all inversion results whenever such standard deviations are not negligible.

When the $\sigma_e=0.001$, $M=7$ data were also inverted using different, more random, starting values for the parameters and with, as usual, stringent convergence criteria, four of the c_m estimates were driven down to very small values, of the order of 10^{-5} or much less. The value of S_F on convergence was slightly smaller than that found above for $M=7$. Although the relative standard deviations of the three other c_m values were much smaller than those of the first fit, they were still appreciably greater than unity. These results strongly suggest that one should carry out a new fit with M no larger than three. When this was done, again with arbitrary starting values, the $M=3$ results listed in the middle of Table II were again found, bespeaking a true least-squares solution. Thus, fitting with an inordinate initial number of free parameters yields direct information on the largest value of M worth using even when only rough parameter standard deviation estimates are available. Further, fitting with different values of M allows one to distinguish between discrete source points, whose positions should remain nearly independent of M, and continuous-distribution points, whose positions would not. Such discriminatory power is one of the most important virtues of a variable y_m approach, as compared to one with fixed y_m values.

Finally, it is of interest to examine the effect of the number of data points on inversion results with the present methods. Suppose that we take an x range from 22 to 42, 41 points, instead of the 81 points used above for $t=4$, which extend from 12 to 52. The data then decrease from about 2π down to about 0.01. For the $M=7$, $\sigma_e=0$ inversion, S_F was about 3×10^{-13}, and all the parameter estimates were the same. But the $M=2$, $\sigma_e=0.001$ results were somewhat different. The estimated standard deviations of the y_m values were about ten times larger than those in Table II but are still essentially negligible. For the c_m's, however, the two standard deviation estimates were about 0.28 and 0.43, by no means negligible. But the actual errors of the two c_m values were only about 0.04, only slightly larger than those obtained with the 81-point fit. Thus, the estimates were appreciably better than indicated by their poorly estimated standard deviations. In this example, for noisy data a narrower data range leads to worse standard deviation estimates but to only a negligible increase in the actual errors of the parameter estimates.

ACKNOWLEDGMENT

I much appreciate the numerous suggestions for improvement of this work provided by Professor W. J. Thompson.

REFERENCES

1. J. D. Craig and A. M. Thompson, Comp. Phys. **8**, 648 (1994).
2. H. Allison, Math. Scientist **4**, 9 (1979).
3. T. H. Skaggs and Z. J. Kabala, Water Resour. Res. **30**, 71 (1994).
4. E. Isaacson and H. B. Keller, *Analysis of Numerical Methods* (Wiley, New York, 1966), pp. 21–22.
5. J. R. Macdonald and M. K. Brachman, Rev. Mod. Phys. **28**, 393 (1956), see p. 416.
6. J. R. Macdonald and L. D. Potter, Solid State Ion. **23**, 61 (1987).
7. J. R. Macdonald, Electrochim. Acta **35**, 1483 (1990).
8. J. R. Macdonald, J. Chem. Phys. **102**, 6241 (1995).
9. B. A. Boukamp and J. R. Macdonald, Solid State Ion. **74**, 85 (1994).
10. J. R. Macdonald, J. Electroanal. Chem. **378**, 17 (1994).
11. J. R. Macdonald, Electrochim. Acta **38**, 1883 (1993).
12. J. L. Salefran and Y. Dutoit, J. Chem. Phys. **74**, 3056 (1981).
13. F. D. Morgan and D. P. Lesmes, J. Chem. Phys. **100**, 671 (1994).