Physica B 182 (1992) 341–348
North-Holland

PHYSICA B

Bayesian analysis of quasielastic neutron scattering data

D.S. Sivia, C.J. Carlile and W.S. Howells
Rutherford Appleton Laboratory, Chilton, Didcot, Oxon, UK

S. König
Physik-Department E22, Technische Universität München, D-8046 Garching, Germany

We consider the analysis of quasielastic neutron scattering data from a Bayesian point-of-view. This enables us to use probability theory to assess how many quasielastic components there is most evidence for in the data, as well as providing an optimal estimate of their parameters. We review the theory briefly, describe an efficient algorithm for its implementation and illustrate its use with both simulated and real data.

1. Introduction

The analysis of quasielastic neutron scattering (QENS) data, in common with all data analysis, is an exercise in scientific inference: given a finite set of noisy measurements and a knowledge of the experimental setup, we wish to learn about some object of interest. In the case of QENS we are interested in the rotational and diffusive motions of atoms and molecules.

If we have a physical model for the system then the data analysis problem usually reduces to a least-squares calculation for obtaining the best estimate of its parameters. In the absence of such a model, QENS data (assumed to be) collected at constant momentum transfer are often analysed in terms of the sum of a few Lorentzians:

$$d(\varepsilon) = \left[A_0\delta(\varepsilon) + \sum_{j=1}^{N} A_j \frac{w_j}{\pi(\varepsilon^2 + w_j^2)} \right] \otimes r(\varepsilon) + b(\varepsilon) + \sigma(\varepsilon), \quad (1)$$

where d are the data, as a function of energy transfer ε, r is the instrumental resolution function, b is the background signal and σ is the noise. The inferred values of the amplitudes and widths $\{A_j, w_j\}$ indicate the time-scale and abundance of the molecular motions; their variation with momentum transfer suggests their geometrical nature.

In section 2 we outline the theory for the Bayesian analysis of the problem formulated in eq. (1); this enables us to assess how many quasielastic components there is most evidence for in the data, as well as providing an optimal estimate of their parameters. An algorithm for the practical implementation of the theoretical results is described in section 3 and its use is illustrated in section 4. After a poignant anecdote in section 5, we conclude with section 6.

2. The method

In 1946, Cox [1] showed that any method of inference which satisfies simple rules for logical and consistent reasoning must be equivalent to the use of ordinary probability theory, as originally developed by Bayes [2] and Laplace [3]. The use of probability theory in this way is now called the Bayesian approach, to distinguish it from its earlier rejection as a tool for scientific

Correspondence to: D.S. Sivia, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, UK.

inference by the schools of conventional statistics. For a good review of the Bayesian approach and its historical development, the reader is referred to the paper by Loredo [4]; it also contains references to many excellent papers by Jaynes (e.g. refs. [5–7]). Let us outline the Bayesian approach for the analysis of the "sum of Lorentzians" problem in eq. (1).

2.1. Parameter estimation

First consider the case when the number of Lorentzian components N is known; the problem then reduces to one of estimating $2N + 1$ parameters from the data $(A_0; \{A_j, w_j\}, j = 1, 2, \ldots, N)$. Our inference about the amplitudes and widths of the Lorentzians is summarised by the conditional probability distribution function (PDF)

$$\text{prob}(A_0, \{A_j, w_j\} \mid N, d) ,$$

where "|" means "given". The best estimate of the parameters is given by the maximum of this PDF; their reliability is given by its width or spread.

In order to compute $\text{prob}(A_0, \{A_j, w_j\} \mid N, d)$ we need to use Bayes' theorem; this relates the PDF we require to one which we can calculate and to another which encodes our prior knowledge:

$$\begin{aligned} &\text{prob}(A_0, \{A_j, w_j\} \mid N, d) \\ &\quad \propto \text{prob}(d \mid A_0, \{A_j, w_j\}, N) \\ &\qquad \times \text{prob}(A_0, \{A_j, w_j\} \mid N) . \end{aligned} \tag{2}$$

The term $\text{prob}(A_0, \{A_j, w_j\} \mid N)$ on the far right is called the *prior* PDF and represents our state of knowledge (or ignorance) about the parameters before we have analysed the data. Our prior state of knowledge is modified by the so-called *likelihood function*,

$$\text{prob}(d \mid A_0, \{A_j, w_j\}, N) ,$$

which tells us how likely it is that we would have obtained our particular data set if we were given a set of (trial) parameter values. The product of the prior PDF and the likelihood function yields the *posterior* PDF we require.

If we make the assumption that the data are independent (so that one measurement does not affect another) and subject to additive Gaussian noise, then the likelihood function reduces to the familiar form $\exp(-\frac{1}{2}\chi^2)$, where χ^2 is the usual sum-of-squared-residuals misfit statistic. If we also make the simplifying assumption that our prior state of ignorance is best represented by the assignment

$$\text{prob}(A_0, \{A_j, w_j\} \mid N) = \text{constant} ,$$

then the maximum of the posterior PDF is given by the maximum likelihood solution. In other words, our best estimate of the amplitudes and widths of the Lorentzians is given by that set of parameters which minimises χ^2. This, then, provides us with the justification for using the method of least-squares.

2.2. Model selection

Let us move on to consider the case when the number of Lorentzian components is not known; the problem now becomes one of model selection. For example, is there one quasielastic component or two? The procedure for addressing this problem sketched out below is essentially the same as that found in ch. 5 of the book by Jeffreys [8]; details can also be found in Gull [9], Bretthorst [10] and Sivia and Carlile [11].

Basically, we need to compute the posterior PDF for the number of Lorentzians: $\text{prob}(N \mid d)$. We begin by using Bayes' theorem:

$$\text{prob}(N \mid d) \propto \text{prob}(d \mid N) \times \text{prob}(N) . \tag{3}$$

Assuming that we have no *a priori* reason to prefer a particular number of quasielastic components, we can take the prior PDF $\text{prob}(N)$ to be uniform and absorb it into the proportionality constant. To proceed further, we express $\text{prob}(d \mid N)$ as a *marginal* distribution over the joint PDF for the data and the amplitude and width parameters:

$$\text{prob}(N \mid d) \propto \int\int\cdots\int \text{prob}(d, A_0, \{A_j, w_j\} \mid N)\, dA_0\, d^N A_j\, d^N w_j . \quad (4)$$

Finally, we expand the joint PDF in terms of a conditional (likelihood) PDF and a prior PDF:

$$\text{prob}(N \mid d) \propto \int\int\cdots\int \text{prob}(d \mid A_0, \{A_j, w_j\}, N) \times \text{prob}(A_0, \{A_j, w_j\} \mid N)\, dA_0\, d^N A_j\, d^N w_j . \quad (5)$$

This is in fact just the normalisation constant for Bayes' theorem in eq. (2). There we ignored this term because it was irrelevant for the parameter estimation problem. For model selection, however, it turns out to be the crucial factor; it is often referred to as the *evidence*.

If we make some simplifying assumptions, similar to those needed to justify the use of least-squares, then the multiple integral of eq. (5) can be approximated by (e.g. see ref. [11])

$$\text{prob}(N \mid d) \propto \frac{N!}{[w_{\max} A_{\max}]^N} \times \frac{(4\pi)^N \, e^{-\chi^2_{\min}/2}}{\sqrt{\text{Det}(\nabla\nabla\chi^2)}} , \quad (6)$$

where $\chi^2_{\min}$ is the best-fit value of χ^2, assuming N Lorentzians, and $\text{Det}(\nabla\nabla\chi^2)$ is the determinant of the *Hessian* matrix (evaluated at $\chi^2_{\min}$). $A_{\max}$ and $w_{\max}$ are the estimates of the maximum amplitude and width of any Lorentzian. Before describing a simple algorithm for the evaluation of $\text{prob}(N \mid d)$ given by eq. (6), let us mention how probability theory allows us to deal with systematic uncertainties.

2.3. Dealing with systematic uncertainties

In the preceding sections, we have approximated the likelihood function as $\exp(-\frac{1}{2}\chi^2)$. In order to compute χ^2, we must be able to calculate an ideal data-set given trial parameter-values for the amplitudes and widths. Equation (1) tells us that, in order to do this, we need to know the resolution function and the background signal. Usually we try to obtain a fairly good estimate of both by collecting data using a standard elastic scatterer and an empty cell, but can we do anything if we are not quite so fortunate?

The answer is yes, at least in principle: we just integrate out the *nuisance* parameters, characterising the systematic uncertainties, from the joint PDF (*marginalisation*). Suppose that we had an unknown flat background B; eq. (4), for example, would then become

$$\text{prob}(N|d) \propto \int\cdots\int \text{prob}(d, A_0, \{A_j, W_j\}, B|N)\, dA_0\, d^N A_j\, d^N w_j\, dB . \quad (7)$$

The analysis would follow along the same lines as before, leading to eq. (6), but with B now contributing to both $\chi^2_{\min}$ and the Hessian matrix.

Although probability theory allows us to deal with systematic uncertainties, we should not take this to mean that there is no need or value in trying to obtain a good estimate of the resolution function and background. Even when we can use the theoretical apparatus in practice, systematic uncertainties still lead to a (significant) reduction in the reliability of the inferred Lorentzian parameters.

3. An algorithm

In order to make use of the analysis of section 2, we need an algorithm for its practical implementation. A *Monte Carlo* procedure would be the most robust, both in terms of finding the optimal parameters and for evaluating the multiple integral of eq. (5); unfortunately, it also tends to be very intensive in computer time. Instead, we outline an algorithm which uses discrete 1-dimensional searches and linearised multidimensional optimisation, in combination with the approximation of eq. (6). We have used this simple algorithm with considerable success and found it to be fairly efficient.

First let us state our modelling assumptions. We assume that the data are described, as in eq.

(1), by a δ-function to represent elastic scattering plus a small number of Lorentzians to represent quasielastic scattering. The resolution function is assumed to be known and stored on finely digitised grid (suitable for computer manipulation). We take the background to be linear, being described by two (nuisance) parameters. To compensate for any experimental misalignment, we introduce a (nuisance) parameter for an ε-axis off-set.

The parameters $A_{\max}$ and $w_{\max}$, needed for eq. (6), are estimated from a cursory look at the data. $A_{\max}$ is given by the integrated intensity in the data, since no amplitude can exceed this value (assuming that the resolution function has been normalised to unity). $w_{\max}$ is given by the energy-range of the data, since Lorentzians of widths much greater than this cannot be distinguished from a flat background.

We start by refining the parameters of the linear background, the ε-axis off-set and the amplitude of a δ-function. This is done using a Newton–Raphson algorithm (e.g. see ref. [12]), which requires the computation of the first and second derivatives of χ^2 with respect to these parameters. Having found $\chi^2_{\min}$ and the determinant of $\nabla\nabla\chi^2$, we can use eq. (6) to calculate the posterior probability for their being no quasielastic components: prob($N = 0 \mid d$).

Next, we conduct an explicit 1-dimensional search for the optimal width of a possible first Lorentzian component. If we keep the ε-axis off-set parameter fixed, then the optimisation of the amplitudes of the δ-function and the first Lorentzian, for a given width, is just a simple linear problem. Having come close to the optimal solution, we refine all the parameters in the one-Lorentzian model simultaneously (using Newton–Raphson). We can then calculate the probability for there being just one quasielastic component: prob($N = 1 \mid d$).

We continue in this manner, adding one more Lorentzian at a time until a maximum in the posterior PDF for the number of quasielastic components is evident. For each additional component, the optimisation of the parameters is split into two parts. In the first stage there is an explicit search for the best width of the new Lorentzian: all other non-linear parameters are held fixed during this period, so that only the amplitudes and linear background are refined. In the second stage, all the parameters of the N-Lorentzian model are refined simultaneously.

Before we go on to illustrate the use of this algorithm, we should make a couple of additional remarks. It is helpful to use range-parameters such as $A_{\max}$ and $w_{\max}$ to scale the parameters to be optimised. That is to say, if we work in dimensionless units like $A_i/A_{\max}$ the optimal parameters will all be of a similar order. Working in these dimensionless units we can improve the stability of the matrix calculations, such as the inversion of the Hessian matrix to obtain the variance of the inferred parameters, by adding (a few times) the identity matrix to the Hessian matrix. Adding the identity matrix does not change the eigenvectors of the Hessian matrix and so does not alter the correlations between the inferred parameters. It does, however, put a lower bound on the eigenvalues, thereby encoding our expectation that the uncertainty of any inferred parameter should not exceed the order of the range-parameter (e.g. $\sigma(A_i) \leq A_{\max}$). In terms of the Newton–Raphson optimisation, such a "beefing-up of the diagonals" puts an upper bound on the amount by which any parameter is allowed to change during an iteration; such conservative control was found to be necessary in order to stop the Newton–Raphson refinement from "blowing up".

4. Examples

We begin our illustration of the use of the theory and algorithms described in section 2 and 3 with the aid of data generated in a computer simulation. These test data are shown in fig. 1(a) and were generated according to the model of eq. (1); the resolution function is given in fig. 1(b). For how many quasielastic components is there most evidence in these data, and what are their inferred parameters and reliabilities?

Carrying out an analysis of these data as described above we find most evidence for two Lorentzians. The logarithm (to base 10) of the

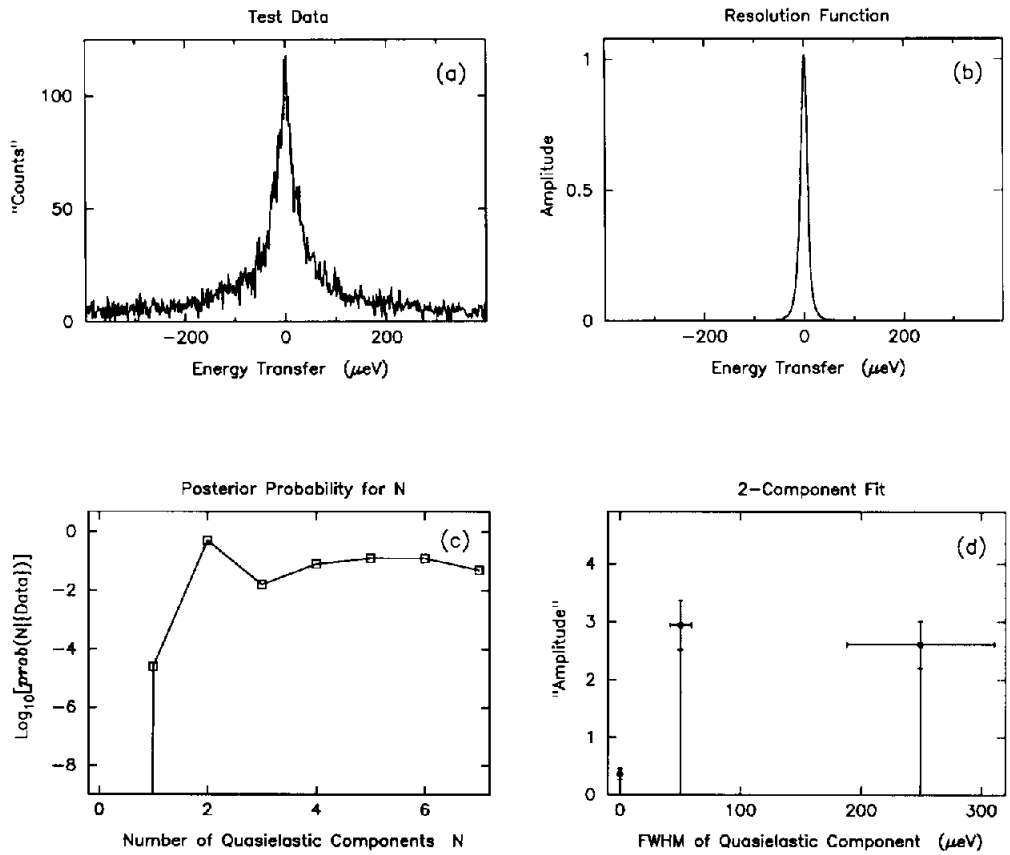

Fig. 1. An illustration of the Bayesian analysis using data simulated by a computer. (a) The test data. (b) The resolution function. (c) The logarithm (to base 10) of the posterior probability for the number of quasielastic components. (d) The inferred amplitude and width parameters and their estimated $1-\sigma$ error-bars for the optimal number of Lorentzian components.

posterior PDF for the number of quasielastic components, shown in fig. 1(c), indicates that the data support a 2-Lorentzian model over a 1-Lorentzian model by a factor of 10 000 to 1. Although the 2-Lorentzian model is favoured over greater numbers by a factor of about 10, we cannot rule out the possibility of there being more components at the 95% confidence level. The best estimate of the amplitude and width parameters of the two Lorentzians and the elastic line, along with their $1-\sigma$ error bars, is shown in fig. 1(d). The underlying spectrum which was used to generate the data of fig. 1(a) did indeed have two quasielastic components, with widths 50 and 250 μeV; the ratio of the amplitudes used for the elastic line and for the Lorentzians was 1:10:10.

The data of fig. 1 serve as a useful demonstration against the naive use of the χ^2 goodness-of-fit statistic as the sole criterion for model selection. Figure 2(b) shows the normalised residuals for the fit to the data in fig. 1(a) produced by the 1-Lorentzian model of fig. 2(a). Given that there were 453 data-points, the corresponding χ^2 value of 462 would be considered very respectable; the residuals are also free of any undue systematic variations. On the basis of the χ^2 statistic, therefore, we would consider it as brave (or even foolhardy) to assume the existence of more than one quasielastic component. The full probabilistic calculation, on the other hand, finds quite strong evidence for (at least) a second component in the data! The point is that while χ^2 is an important ingredient in answering

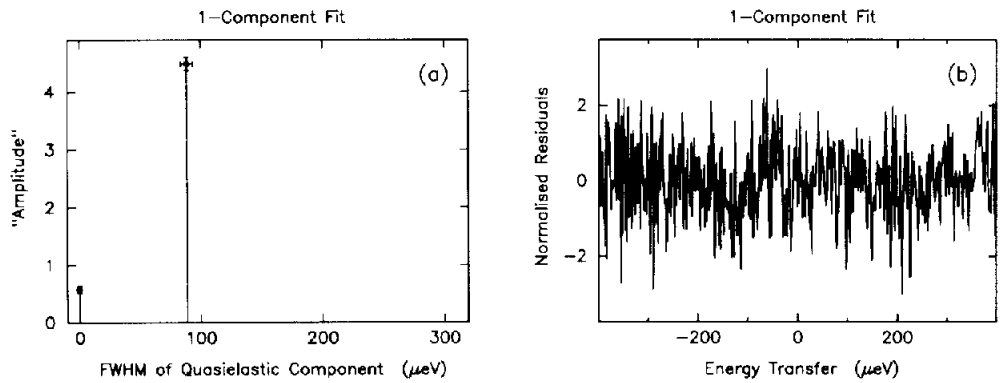

Fig. 2. (a) The best-fit parameters given by the data in fig. 1(a) assuming the existence of only one Lorentzian component. (b) The normalised residuals from the corresponding fit to the data.

the model selection problem, the analysis of section 2.2 tells us that there are other factors which must also be taken into account.

We now illustrate the model selection analysis procedure with data collected on IRIS at the pulsed neutron scattering facility ISIS; they are part of a study of König et al. [13] on the dynamics of water molecules between lecithin bilayers. In the first example, shown in fig. 3(a), we find most evidence for there being no quasielastic components in the data. Although the possibility of there being one Lorentzian cannot be ruled out at the 95% confidence level for any given value of the momentum transfer, the evidence for there only being an elastic line becomes quite strong if we make the assumption that the number of components is the same for all scattering angles. Since the data were the difference of measurements from low-hydration samples of H_2O and D_2O, taken at 0°C, the analysis confirms our expectation that we are looking mainly at bound water. Figure 3(b) shows the result obtained from corresponding data taken at 65°C; there is most evidence for one quasielastic component. As the lecithin bilayer exhibits a transition from a gel to a liquid crystalline phase at 55°C, the motional degrees of freedom of the lecithin, and thus bound water, increase and broadening is to be expected. Finally, fig. 3(c) shows the result from a highly-hydrated H_2O sample, taken in the liquid crystal phase; there is very strong evidence for (at least) two quasielastic components. This is also consistent with our physical intuition since we are now looking at the motions of both free and bound water as well as those of lecithin.

5. Interlude: what *not* to compute

Before finishing with the conclusions, we take the liberty of recanting a humorous, but poignant, anecdote from Forman Acton's 1970 book "Numerical Methods That Work" [14]. Half-way through his wonderful book, he has a section entitled "Interlude: what *not* to compute". In this, he considers several case studies; one of them comes under the heading of "exponential fitting".

He starts by saying that, despite the invention of the printing press, computational wisdom seems to be continually rediscovered rather than being learned from past experience. As an example, he uses the analysis of radioactive decay. In the first case, the relative abundances of two known substances A and B, with decay rates a and b, are to be estimated from the data:

$$y(t) = \mathrm{A}\,\mathrm{e}^{-at} + \mathrm{B}\,\mathrm{e}^{-bt}, \tag{8}$$

where y is proportional to the number of counts at time t. The analysis reduces to a simple least-squares fit, generally requiring only a pocket calculator.

Unfortunately there is a companion problem that looks only slightly more complicated, until

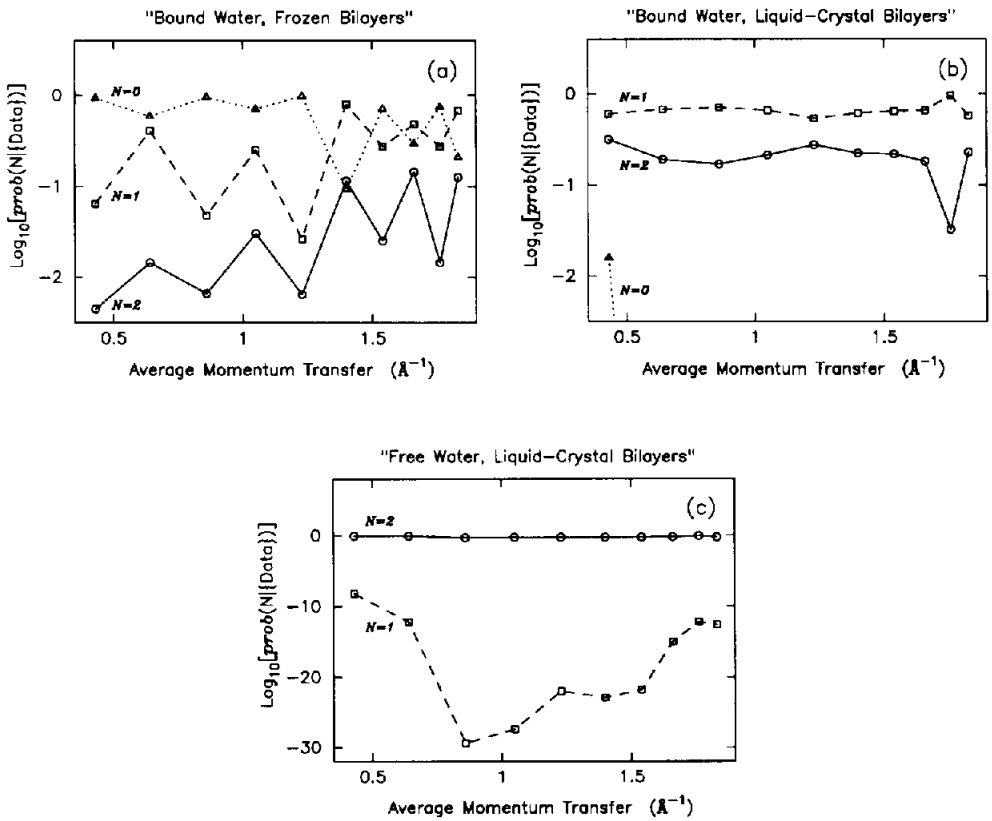

Fig. 3. Three examples showing the use of the probabilistic analysis for model selection with data taken by König et al. [13] on IRIS, at the pulsed neutron facility ISIS, using samples of H_2O/D_2O-hydrated lecithin bilayers as described in the text. The posterior probabilities for the number of quasielastic components are plotted on a logarithmic axis, to base 10.

you try it! In this case, the substances are not known; therefore, all four parameters (a, b, A and B) must now be estimated from the data. Acton states that the answer to this problem lies in the chemical, rather than the computer, laboratory because "... it is well known that an exponential equation of this type in which all four parameters are to be fitted is *extremely* ill conditioned. That is, there are many combinations of (a, b, A, B) that will fit most exact data quite well indeed (will you believe four significant figures?) and when experimental noise is thrown into the pot, the entire operation becomes hopeless." He concludes: "But those with Faith in Science do not always read the Book – and must be spanked or counselled."

This humorous anecdote becomes rather less amusing, however, when we realise that the analysis of QENS data is essentially equivalent to exponential fitting of the latter kind. Taking the Fourier transform of eq. (1), we find that

$$D(k) = \left[A_0 + \sum_{j=1}^{N} A_j \, e^{-2\pi |k \cdot w_j|} \right] \times R(k) + B(k) , \tag{9}$$

where $D(k)$, $R(k)$ and $B(k)$ are the kth Fourier components of the data, the resolution function and the background signal. The analysis described in section 2 automatically cautions us about this difficulty since, as can be seen in fig. 1(c), the posterior PDF for the number of Lorentzians becomes very flat after $N = 2$. This flatness contrasts with the significant falloffs found

in other applications of this type of analysis [9–11] so that probability theory is warning us that QENS data are relatively impoverished with respect to distinguishing between two and many quasielastic components.

6. Conclusions

We have briefly reviewed the Bayesian approach for the analysis of QENS data. In a more general context, the use of probability theory in this way provides a unified and logical method for data analysis; QENS is no exception. It allows us to: (a) quantitatively assess the number of components for which there is most evidence; (b) make optimal estimates of the quasielastic parameters and their reliabilities; (c) deal with known systematic uncertainties.

The analysis of QENS data is similar to the very ill-posed Laplace transform problem. This means that it is extremely difficult to infer reliably the parameters of more than a couple of quasielastic components. Such wisdom is, of course, already prevalent folklore. But then, to Laplace, "Probability theory is nothing but common sense reduced to calculation".

References

[1] R.T. Cox, Am. J. Phys. 14 (1946) 1.
[2] T. Bayes, Philos. Trans. Roy. Soc. London 53 (1763) 370.
[3] P.S. Laplace, Theorie Analytique des Probabilités (Courcier, Paris, 1812).
[4] T.J. Loredo, in: Maximum Entropy and Bayesian Methods (Dartmouth 1989), ed. P.F. Fougère (Kluwer, Amsterdam, 1990).
[5] E.T. Jaynes, Papers on Probability, Statistics and Statistical Physics, ed. R.D. Rosenkrantz (Riedel, Dordrecht, 1983).
[6] E.T. Jaynes, in: Maximum Entropy and Bayesian Methods in Applied Statistics, ed. J. H. Justice (Cambridge University Press, 1986).
[7] E.T. Jaynes, Probability Theory – the Logic of Science (in preparation).
[8] H. Jeffreys, Theory of Probability (Oxford University Press, 1939).
[9] S.F. Gull, in: Maximum Entropy and Bayesian Methods in Science and Engineering, Vol. 1, eds. G.J. Erickson and C.R. Smith (Kluwer, Amsterdam, 1988).
[10] G.L. Bretthorst, Bayesian Spectrum Analysis and Parameter Estimation (Springer-Verlag, New York, 1988).
[11] D.S. Sivia and C.J. Carlile, J. Chem. Phys. 96 (1992) 170.
[12] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, Numerical Recipes (Cambridge University Press, 1986).
[13] S. König et al., in preparation.
[14] F.S. Acton, Numerical Methods That Work (Harper & Row, New York, 1970).