Journal of Physics: Conference Series **663** (2015) 012009 doi:10.1088/1742-6596/663/1/012009

FABADA Goes MANTID to Answer an Old Question: How Many Lines Are There?

D Monserrat[1], A Vispa[1], L C Pardo[1], R Tolchenov[2,3], S Mukhopadhyay[4] and F Fernandez-Alonso[2,5]

[1] Grup de Caracterització de Materials, Departament de Física i Enginyeria Nuclear, ETSEIB, Universitat Politècnica de Catalunya, Diagonal 647, E-08028 Barcelona, Catalonia, Spain
[2] ISIS Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, United Kingdom
[3] Tessella Ltd., 26 The Quadrant, Abingdon Science Park, Abingdon OX14 3YS, United Kingdom
[4] ISIS Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, United Kingdom. Department of Materials, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom
[5] Department of Physics and Astronomy, University College London, Gower Street, London, WC1E 6BT, United Kingdom.

E-mail: `luis.carlos.pardo@upc.edu`

Abstract. Data fitting and model selection lie at the very heart of the scientific method, and these tasks are often tackled via the use of popular (and typically unquestioned) optimization algorithms of limited validity or flexibility, *e.g.*, least-squares methods. To circumvent these limitations, this work introduces the implementation within the MANTID framework of robust inference methods based on the Bayesian fitting algorithm FABADA. The power of this new MANTID tool is illustrated by revisiting an overarching and recurring question in the interpretation of quasielastic neutron scattering measurements, that is, how many spectral components may (or may not) be justified by the available experimental data.

1. Introduction

Data fitting and model selection underpin much of what is done with scientific data, to verify the adequacy of specific mathematical models and to obtain parameters that best describe the underlying physical phenomena. The classical way of performing data fitting involves the minimization of a so-called *cost function* (CF) such as the celebrated χ^2. For a given physical model, popular algorithms such as Levenberg-Marquardt (LM) seek to minimize the CF in order to find the parameter values (and associated uncertainties) that 'best describe' the data [1]. Some of the limitations associated with these classical approaches have been previously discussed by some of the authors [2, 3]. In brief, LM algorithms are successful when the χ^2 is well behaved, *i.e.,* if no local minima are encountered in the search for the global minimum. Moreover, LM is based on the assumption that the minimum is quadratic and that there is no correlation between fit parameters beyond linear couplings. With these restrictions in mind, model selection is then implemented on the basis of the value of the χ^2 divided by a term that penalizes the addition of extra model parameters, leading to a concomitant reduction in the overall number of degrees of freedom. Such is the rationale behind the concept of a reduced-χ^2 (χ^2_{red}) CF, commonly defined

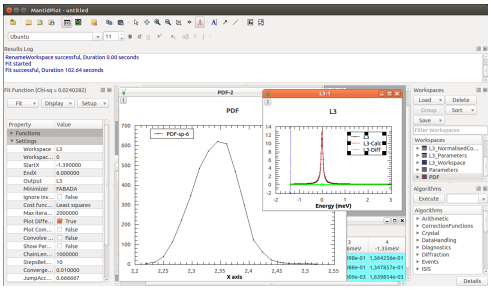

Figure 1. (Color online) Snapshot of MANTID program where the minimizer FABADA has been selected.

as $\chi^2_{red} = \chi^2 \cdot (N - m - 1)^{-1}$, where N is the number of data points and m is the number of fit parameters. Model selection based on χ^2_{red} is faced with three main drawbacks. In addition to the primary assumptions of the LM method discussed above, parameter correlation leads to an effective reduction in the number of real parameters. In the language of statistical analysis, this added complication implies that the CF does not conform to the expected χ^2 Probability Density Function (PDF) [4]. Last but not least, reliance on a single number (the χ^2) is unlikely to represent a general way of performing robust model selection.

Bayesian Inference Methods (BIMs) provide a suitable framework to address the above limitations. First and foremost, BIMs do not rely on specific assumptions about the underlying χ^2 landscape or on putative (and typically unknown) correlations between model parameters. In addition, jumps across local minima are possible [2] and the robustness of the results can always be tested by increasing the (fictive) temperature associated with artificial increments in data uncertainties [5]. Also, the interpretation and visualization of BIM results differ from the more widespread *frequentist* point of view in a number of fundamental ways. In BIMs, PDFs are obtained for each model parameter, as well as for the χ^2. Likewise, model-selection protocols can be implemented in a systematic manner via direct inspection of the χ^2 PDF. This approach can immediately establish the mono- or multi-modal character of this function, and the position of its most-probable value provides a suitable figure of merit for quantitative model selection. Moreover, analysis of the PDFs makes it possible to *assess the robustness* of any criteria we might have chosen to implement model selection. For example, if χ^2 PDFs corresponding to two different models display some overlap in χ^2 space, this result immediately implies that there exists a combination of parameters for one model which, given the available experimental data, is necessarily as likely as the corresponding parameter set for the second (alternative) model.

In the present work, we introduce the first implementation of BIMs within the MANTID framework [6, 7]. As a software package, MANTID is primarily aimed at the visualization and analysis of neutron and muon data using modern software-engineering paradigms, as recently illustrated in Refs. [8, 9, 10]. The added value of this new MANTID tool is illustrated by considering its use in QuasiElastic Neutron Scattering (QENS) data analysis, superseding previous BIMs [11, 12, 13]. The implementation has been done in the release 3.4 and has been added as a new minimizer in the main menu of MANTID as it can be seen in figure 1.

2. A Closer Look Under the Hood: The FABADA Algorithm

Before describing the implementation of BIMs in MANTID, we provide a summary of the underlying Fitting Algorithm for Bayesian Analysis of DAta, hereafter FABADA [3]. Bayes' theorem relates the conditional probabilities between two events [4] and therefore provides the starting point to calculate the likelihood that a given hypothesis is supported by the experimental

data. This task may be accomplished by taking proper account of our prior state of knowledge concerning the starting hypothesis as well as the likelihood that the experimental data are described by the proposed hypothesis. In what follows, we assume no previous knowledge about the hypothesis, a situation that amounts to assigning a constant probability to the prior.

Typically, neutron-scattering experiments rely on the acquisition of a given number of neutron counts per channel. When count levels are high, the underlying Poisson distribution describing experimental uncertainties becomes Gaussian, and the probability that a given model is supported by the experimental data becomes proportional to $\exp\left(-\frac{\chi^2\{P_i\}}{2}\right)$. In this expression χ^2 is defined as usual, that is, as the quadratic sum of the deviation between data and model weighted by uncertainties in the former (for more details, see Refs. [2, 3]). As expected, the value of χ^2 depends on the specific values of the model parameters. This condition can be made explicit by writing $\chi^2\{P_i\}$ where P_i is used to denote the set of model parameters. FABADA, or any other Markov Chain Monte Carlo (MCMC) method, changes one of the parameters in P_i in a random manner, and accepts the change if the CF (χ^2 in our case, or the energy in Monte Carlo simulations) diminishes. Since there are errors associated with the data (or thermal agitation in the case of a Monte Carlo simulation), there is also a non-zero probability to accept changes in P_i that nonetheless lead to an increase of the CF. In the case of data fitting, the probability of accepting an increase in the CF can be written as

$$\frac{P[H(P_i^{new})|D]}{P[H(P_i^{old}]|D)} = exp\left(-\frac{\chi^2_{new}-\chi^2_{old}}{2}\right) \tag{1}$$

where $H(P_i^{new})$ denotes the hypothesis (model) described by the parameter set P_i^{new}, D represents the experimental data, and $P[H|D]$ denotes the conditional probability that hypothesis H holds given D. This additional allowance for otherwise "illegal" moves that increase the CF enables jumps over small barriers in the χ^2 landscape. Higher barriers can also be avoided by multiplying data uncertainties by a fictive temperature, as detailed in Refs. [2, 5].

The main limitation of the above algorithm is related to the extent one needs to change parameter values so that the optimization process becomes equally sensitive to all of them. To circumvent this difficulty, FABADA uses an adaptive MCMC scheme that avoids unwanted bias to specific parameters. This method is best understood by considering a simple example. Let us assume that a given parameter is changed by an amount Δ_i following the prescription $P_i^{new} = P_i^{old} + \rho\Delta_i$ where ρ is a random number. To make all parameters relevant during the fitting procedure, FABADA changes the value of the step Δ_i in an adaptive manner based on the acceptance ratio. Parameters characterized by a low acceptance ratio would then be given large values of Δ_i so that changes in χ^2 become large. This condition leads to large values of $\chi^2_{new} - \chi^2_{old}$ in Eq. (1) and low acceptance probabilities. The opposite happens to parameters with a high acceptance ratio. Convergence implies that all parameters must be accepted with equal probability, a condition that generally requires different values of Δ_i. The specific values of Δ_i are also related to the curvature of the (multidimensional) χ^2 PDF along the direction of a given parameter and its associated error. A small value of Δ_i implies a strong curvature and a small error.

3. Bayesian Model Selection: How Many Lines Are There?

Falsifiability in science [14] usually operates via model selection. For this reason, understanding the conceptual underpinnings of this process is an important exercise deserving careful thought and consideration. One of the most widely used model-selection criteria is predicated upon the notion that a given model is better than others because it provides a "better fit" to the available data. This condition is only true when the number of parameters across models is the *same* and all parameters remain *uncorrelated*, as correlation diminishes their effective number. In the

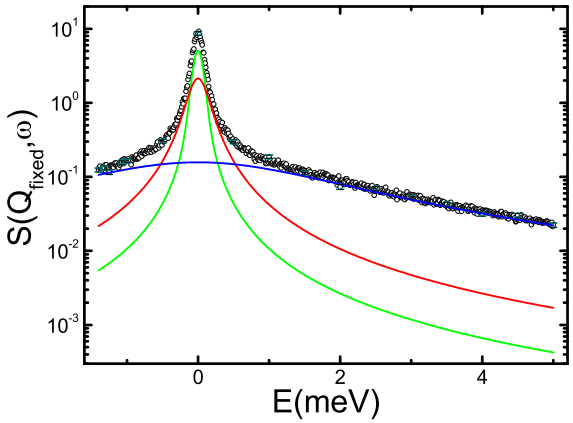

Figure 2. (Color online) Synthetic QENS spectrum to test model selection (black circles). Note the Log-Lin representation of these data. Green, red, and blue lines have been used to depict three L modes with spectral HWHMs 0.02, 0.2, and 1 meV, respectively.

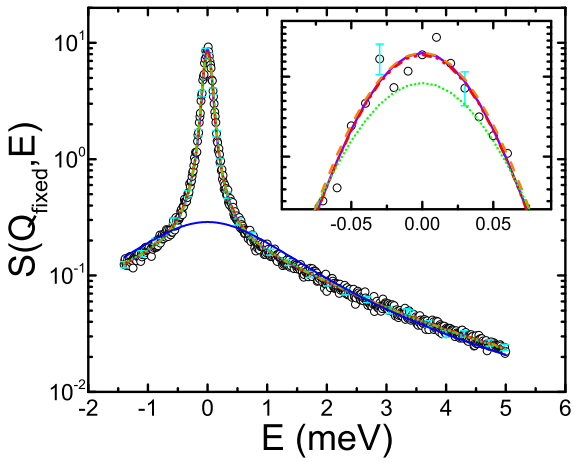

Figure 3. (Color online) Fits to the synthetic QENS data (black circles) using an increasing number of L modes: L1 (solid blue line), L2 (dotted green), L3 (dashed-dot red), L4 (dashed orange), and L5 (solid purple). Note the logarithmic (linear) scales on the main figure (inset).

(metaphorical) words of von Neumann: 'With four parameters I can fit an elephant, and with five I can make him wiggle his trunk' [15]. A model with more parameters usually gives a better fit to the data, and, thus, we must be careful to assess whether the data *and their associated errors* support the new model.

In the context of neutron data analysis, we provide below an alternative answer to an old question posed by Sivia and Carlile in their seminal study *How Many Lines Are There?* [16]. The advantage of the method presented in this work is its flexibility since any parametric model, analytical or not, can be used. We briefly recall that QENS data analysis is oftentimes complicated by the fact that a number of (*a priori* unknown) dynamical processes are all centered around the elastic response of the material. In the limit of very different timescales associated with these processes, visual inspection of the data can be a convenient starting point to construct a suitable prior. The situation becomes less clear when spectral widths become comparable to each other, the extreme case being the occurrence of accidental degeneracies which necessarily lead to a reduction of the dimensionality of our available parameter space. To address this question in a systematic manner, we have generated synthetic QENS data as the linear addition of three equally weighted Lorentzian (L) modes convolved by a Gaussian resolution function with a Half-Width-Half-Maximum (HWHM) of 0.005 meV. The three L modes have been generated with spectral HWHMs of 0.02, 0.2, and 1 meV. To emulate realistic experimental conditions, a Gaussian-distributed relative error of 6% has been added to the QENS spectrum after convolution. These data are shown in Fig. 2.

All data fits have been performed using the current implementation of FABADA in MANTID [7]. To explore model selection with this new MANTID tool, Fig. 3 shows best-fit results for a total of five different models described by an increasing number of L modes, hereafter denoted as Ln where n is the number of L modes. Visual inspection of these results makes it clear that L1 is not capable of describing the data satisfactorily. Beyond this case, a cursory 'χ-by-eye' analysis of these data is no longer possible. A much closer look at specific regions of the spectral response might still allow us to discover that L3 is slightly better than

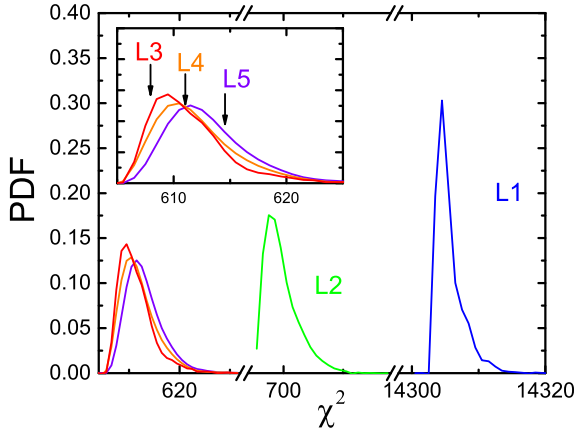

Figure 4. (Color online) χ^2 PDFs for the L1-L5 models shown in Fig. 3. The inset shows an enlargement around L3-L5, showing that L3 is favored over the other two. This figure uses the same color coding as in Fig. 3.

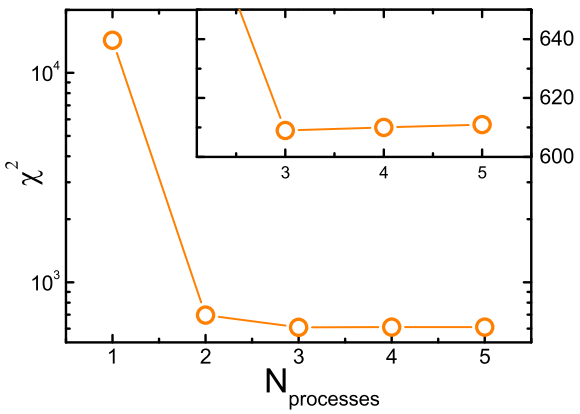

Figure 5. Most-probable values of the χ^2 PDF for models L1-L5. Note the Log(Lin) scale in the main panel (inset).

L2 (*cf.* inset in Fig. 3). For L3, L4, and L5, the quality of the fits remains virtually unchanged. Thus, in this particular case, even a simple model-selection protocol using a χ^2_{red} criterion would favor L3 over more complex descriptions of the experimental data.

To put these considerations on firmer ground, Fig. 4 shows the χ^2 PDFs associated with each Ln model. As it became evident from a visual inspection of Fig. 4, the most probable value of the χ^2 PDF for L1 peaks at $\chi^2 \sim 14305$, more than one order of magnitude above other models. For L3, L4, and L5, we see that these three PDFs share the same onset corresponding to the lowest χ^2 (best fit), as expected from the results shown in Fig. 3. However, the *maximum* of the PDF shifts to higher χ^2 values in going from L3 to L5. This trend is a consequence of an increase in the number of parameters, leading to an overall broadening and concomitant upward shift of the corresponding χ^2 PDF. Thus, *the addition of more parameters beyond L3 is clearly not justified by the experimental data*. From a physical viewpoint, the most interesting situation relates to the differences between L2 and L3. Fig. 4 provides a quantitative answer to this question. As shown in this figure, the PDF for L2 peaks at $\chi^2 \sim 700$ and does not overlap with the corresponding one for L3. In probabilistic terms, the immediate consequence of this result is that *no combination of L2 parameters can provide a better fit than L3*. Therefore, *L3 is favored over L2*, a result that we have established in a robust and quantitative manner.

Although direct inspection of χ^2 PDFs represents a safer way of performing model selection, it might also be desirable to have a simple measure of fit quality. The results presented above make it clear that the most-probable value of the χ^2 PDF could be used to this end. Fig. 4 also brings to the fore a common misconception in model selection, as the minimum χ^2 values associated with a particular model (the choice in popular minimization algorithms) is characterized by a negligibly small probability. To close this discussion, Fig. 5 shows these most-probable values as a function of the number of L modes. As one would expect on the basis of our analysis, its value decreases until it reaches a minimum for L3, followed by a slight increase thereafter due to the broadening of the underlying PDF (see inset in Fig. 4). The attentive reader might also have realized that this procedure is tantamount to the implementation of *Ockham's Razor on our QENS data using models L1-L5*.

4. Conclusions & Outlook

Using the Bayesian fitting algorithm FABADA, we have introduced the implementation of BIMs in the MANTID framework. This new tool enables robust data fitting and model selection well beyond more popular and widespread data-analysis methodologies. To illustrate its use, we have revisited an old (yet still topical) question relating to the number of spectral modes present in QENS data. From this exercise, we highlight the merits associated with the use of the *entire* PDF of the CF instead of relying on single-valued criteria such as the minimum χ^2. This approach allows us to ascertain in a quantitative manner whether any combination of parameters for a given model could provide a better description of the experimental data relative to other models. Although it is true that in this simple case a classical model selection would have performed quite good, we would like to point out that the presented method is able to deal with some cases where classical fitting algorithms would not be able to provide a right answer.

We would like to emphasize that this method does not rely on any assumptions relating to the shape of the underlying χ^2 landscape or possible correlations between fit parameters. For the QENS data explored herein, we have only assumed that the data arises from a counting experiment with sufficiently high count rates such that the underlying Poisson distribution becomes Gaussian. Relaxing this assumption has already been implemented with success [3] and could find further use in the analysis of neutron experiments suffering from inherently low signal levels. Moreover our approach is able to include prior knowledge to the fitting, for example coming from previous determinations as shown in the work of Henao et al. [5]. Classical methods based on χ^2 are not able to deal with non-Gaussian distributed parameters, small number of counts or including previous information.

This work was supported by the Spanish MINECO (grants No. FIS2011-24439 and FIS2012-39443-C02-01) and by the Government of Catalonia (grants No. 2009SGR-1003 and 2009SGR-0581). F.F.-A. gratefully acknowledges financial support from the UK Science and Technology Facilities Council.

References

[1] Acton F S 1990 *Numerical Methods that (Usually) Work* (The Mathematical Association of America)
[2] Pardo L C, Rovira-Esteva M, Busch S, Moulin J F and Tamarit J Ll 2011 *Phys. Rev.* E **84** 046711
[3] Pardo L C, Rovira-Esteva M, Busch S, Ruiz-Martín M D and Tamarit J Ll 2011 *J. Phys.: Conf. Ser.* **325** 012006
[4] Sivia D S and Skilling J 2006 *Data Analysis: A Bayesian Tutorial* (Oxford University, New York)
[5] Henao A, Rovira-Esteva M, Vispa A, Tamarit J Ll, Guardia E and Pardo L C 2013 *J. Phys.: Condens. Matter* **25** 454217
[6] Arnold O *et al* 2014 *Nucl. Instrum. Meth.* A **764** 156
[7] The latest version of the MANTID software package can be downloaded free of charge from dx.doi.org/10.5286/SOFTWARE/MANTID
[8] Jackson S 2014 *Rutherford Appleton Laboratory Technical Report RAL-TR-2014-010* (Didcot, UK). Url: epubs.stfc.ac.uk/work/12240844
[9] Mukhopadhyay S 2014 *Rutherford Appleton Laboratory Technical Report RAL-TR-2014-005* (Didcot, UK). Url: epubs.stfc.ac.uk/work/12135206
[10] Jackson S, Krzystyniak M, Seel A G, Gigg M, Richards S E and Fernandez-Alonso F 2014 *J. Phys.: Conf. Ser.* **571** 012009
[11] Howells W S, Garcia Sakai V, Demmel F, Telling M T F and Fernandez-Alonso F 2010 *Rutherford Appleton Laboratory Technical Report RAL-TR-2010-006* (Didcot, UK). Url: epubs.stfc.ac.uk/work/52040
[12] Fernandez-Alonso F, Bermejo F J, McLain S E, Turner J F C, Molaison J J and Herwig K W 2007 *Phys. Rev. Lett.* **98** 077801
[13] Fernandez-Alonso F, McLain S E, Taylor J W, Bermejo F J, Bustinduy I, Ruiz-Martin M D and Turner J F C 2007 *J. Chem. Phys.* **126** 234509
[14] Popper K R 2002 *Conjectures and Refutations: The Growth of Scientific Knowledge* (Routledge Classics)
[15] Dyson F 2004 *Nature* **427** 297
[16] Sivia D S and Carlile C J 1992 *J. Chem. Phys.* **96** 170