Workshop summary on data treatment and techniques

R. N. Silver
Theoretical Division/Los Alamos Neutron Scattering Center
Los Alamos National Laboratory
Los Alamos, NM 87545
USA

M. W. Johnson
Rutherford-Appleton Laboratory
Chilton, Didcot, Oxon, OX11 0QX
UNITED KINGDOM

The goal of neutron scattering experiments is to measure the neutron scattering law, which is proportional to the correlation functions of the condensed matter systems of scientific interest. The data produced in neutron scattering experiments is equal to a convolution of the scattering law with a spectrometer resolution function, with the addition of Poisson noise due to the finite counting statistics and of background due to other physical processes taking place in the spectrometer. The data may be incomplete, may be measured at only a discrete set of points, and may have systematic errors. The problem of infering the scattering law from such data is central to the extraction of information from the neutron scattering technique.

The simplest, and most popular, approach is to assume that the raw data provides a first approximation to the scattering law. However, the raw data may poorly represent the scattering law because of distortions due to the measurement process. Especially for the time-of-flight techniques used in pulsed neutron sources, the raw data may be presented in a space of instrument variables (e.g. scattering angle, time-of-flight channel) which is different from the space of physical variables (e.g. momentum transfer, energy transfer) of interest. The ability to accurately display data in physical variables may be critical to real-time decisions about the conduct of experiments.

To date, most neutron scattering experiments have been analyzed by fitting the data with models using a minimum of parameters which, at least implicitly, assumes specific physical processes underlying the neutron scattering law. Such parameter estimation procedures depend on the accuracy with which one knows the resolution functions and backgrounds of the spectrometer. The paper by Bywater, Williams, and Carpenter in this workshop addresses the measurement of the pulse shapes of moderators from pulsed neutron sources which dominates the instrument resolution functions. Parameter estimation also depends on the sensitivity of the data to the parameters one wishes to determine, and on the physical validity of the fitting model. Such procedures can be well-controlled in many cases, e.g. the Rietveld profile refinement method for the analysis of powder diffraction data for neutron crystallography. They can also be poorly-controlled especially in cases where the prior physical knowledge is incorrect, e.g. fitting Gaussian peaks to data which are in fact Lorentzian broadened.

To go beyond these two traditional procedures for analyzing neutron scattering data, one can attempt to infer the scattering law directly from the data. Such statistical inference problems are inherently ill-conditioned because there may be an infinity of scattering laws all of which fit the data according to a chi-squared criterion. The problem is to use the data to make the best choice of scattering law, termed the ***image***, including whatever prior information about the scattering law one has such as sum rules, positivity, physical properties, etc.

Such image processing problems are not unique to neutron scattering. Sophisticated data analysis methods have been developed to handle similar problems in other fields of research such as radio astronomy, magnetic resonance imaging, computed x-ray tomography, etc. The most successful of these are the maximum entropy and Bayesian methods. Bayes' theorem provides a systematic approach to statistical inference. It states that the probability of the image after an experiment (the *Posterior*) is the product of the probability of the image before the experiment (the *Prior*) times the modification of the image probability by the data (the *Likelihood*). In the maximum entropy (*MaxEnt*) method the Prior is the exponential of the Shannon/ Jaynes entropy of the image relative to a starting model. The MaxEnt image reconstruction is obtained by maximizing the Posterior probability. The most important properties of MaxEnt are that it enforces the positivity of the scattering law, and it puts structure in the image only if it is warranted by the data. Moreover, it permits the incorporation of other forms of prior information such as physical knowledge and, therefore, it provides an iterative approach to image reconstruction. Because of the success of the maximum entropy method in other fields of research, and the obvious need for it in neutron scattering research, maximum entropy has recently been applied to time-of-flight experiments at the LANSCE and ISIS pulsed neutron sources. An excellent introduction with specific applications to pulsed neutron sources is presented by D. S. Sivia in these proceedings. This paper demonstrates that enormous improvements in image quality are obtainable with the maximum entropy method compared with the popular, "the raw data approximates the scattering law", philosophy.

Several of the workshop papers are concerned with more technical aspects of the application of maximum entropy method to pulsed neutron sources. The paper by Johnson and Litster provides empirical experience on the application of the maximum entropy method to the deconvolution of the typical spectra found in neutron scattering research. The paper by Soper addresses the prior knowledge which must be incorporated in the maximum entropy method in order to reliably extract the radial distribution function of amorphous materials and liquids from the incomplete scattering data obtained in neutron diffraction experiments. The paper by Silver, Sivia, and Pynn uses the maximum entropy technique to assess the relative ability of various instrument resolution functions to convey information about the scattering law.

More broadly, we emphasize that the application of modern methods of data analysis to neutron scattering may lead to a sea-change in how we analyze and display data, and even revolutionize the criteria for optimizing neutron scattering instrumentation and neutron sources. The appropriate philosophy is stated in the paper by Hanson: "...the data collection system includes both the spectrometer design and the subsequent data analysis... Optimization of the quality of the final data should also include the effects of the data processing that may be required for the proper interpretation of the data." Hanson goes on to provide a specific example of data analysis algorithm optimization for the case of the Algebraic Reconstruction Technique. The paper by Sivia, Silver and Pynn demonstrates by simulations that the maximum entropy data analysis procedure leads to a different optimization of instrument resolution functions than the popular, "the raw data approximates the scattering law," philosophy.

Neutron scattering is an inherently signal limited and expensive technique, and therefore it is imperative to optimize spectrometers, sources and experiments. The use of modern data analysis methods can lead to improvements of an order-of-magnitude or more in the information which can be extracted from neutron scattering data. Optimization of spectrometers and neutron sources based on modern data analysis procedures can potentially lead to further orders-of-magnitude gains. An information theory of spectrometer design should be a high priority in neutron scattering research. Such *software* approaches to advancing the state-of-the-art in neutron scattering research can be far more cost-effective than the conventional *hardware* approaches (e.g., increasing proton currents from accelerators, boosted targets, etc.) which have dominated prior meetings of the International Collaboration on Advanced Neutron Sources.

We believe that modern data analysis methods, such as maximum entropy image processing, will become a dominant theme in the future development of the neutron scattering technique.