SOME GUIDELINES FOR THE EVALUATION OF NUCLEAR DATA

Donald L. Smith
Technology Development Division
Argonne National Laboratory
Argonne, Illinois 60439

March 20, 1996

Introduction

Modern data evaluation methodology draws upon basic principles from statistics. It differs from earlier *ad hoc* approaches which are completely subjective (e.g., eye guides to data) or are objective in a limited sense (e.g., combinations of reported data by a simple least-squares procedure without regard to correlations in the data errors or a careful scrutiny of the data included in the evaluation). In addition to utilizing more rigorous mathematical procedures, modern evaluation methodology involves taking great care to insure that the data which are being evaluated are equivalent to what has been assumed in the evaluation model and that the values are consistent with respect to the use of standards and other fundamental physical parameters. This short memorandum cannot substitute for more comprehensive treatments of the subject such as can be found in the listed references. The intent here is to provide an overview of the topic and to impress upon the reader that the evaluation of data of any sort is not a straightforward enterprise. Certainly evaluations cannot be carried out automatically with computer codes without considerable intervention on the part of the evaluator.

There are two types of information (data). One is objective data based on experimental measurements. The other is subjective data which, in the case of basic nuclear quantities, often emerge from nuclear model calculations. It is rare that there is sufficient experimental information upon which to base a comprehensive evaluation. Usually it is necessary to merge the complementary processes of measurement and modeling in order to generate such an evaluation. Furthermore, various nuclear quantities are not independent. For example, an evaluated file for a particular isotope or element, as it appears in ENDF/B or any other national or international file, consists of many interrelated components (e.g., partial cross sections) corresponding to various reaction channels. Partial cross sections must add up to the total cross section. Unitarity of the S-matrix appearing in theoretical calculations generally insures that this will be the case when these quantities are derived from nuclear models. However, this will not happen for experimentally derived quantities. Completely different experiments and techniques are involved in measuring individual partial cross sections (or combinations thereof), often leading to a rather messy state of affairs for the evaluator to sort out in carrying out an evaluation. What is measured is rarely equivalent to what one seeks to obtain. The relationship between what is measured (or calculated) and what is sought must be specified in order to carry out a proper evaluation. The experimenter or model calculator ought to be aware of this, but frequently this is not the case so it is left to the evaluator to bridge the gap in understanding. Ideally an evaluator ought to be well versed in all aspects of nuclear model calculations, nuclear data measurements and the analysis of measured data so that all the features of the raw materials which must be employed in his evaluation are well understood. Realistically this happens rarely, so comprehensive evaluations such as those appearing in ENDF are often the

result of collaborations involving individuals with various complementary skills. This used to be feasible in-house at many of the individual laboratories in the U.S. doing nuclear data research, since in earlier times the resources available were far more extensive than they are now. Due to staff reductions, retirements, laboratory closings, etc., it is far less common now to find under one roof all the necessary skills needed to perform a comprehensive evaluation properly. Today inter-laboratory collaboration is essential. Adequate funding is also required to support the personnel involved in this labor-intensive activity.

Modern Theory of Data Evaluation

In rather abstract terms, the process of data evaluation reduces to the following: Given a data set $\mathcal{D}$ (which may include both objective and subjective information), determine what is the most likely (best) set of values for the evaluated quantities represented by a vector $\mathbf{p} = (p_1,p_2, \ldots , p_k, \ldots , p_K)$. The methodology described here is based on the application of three fundamental principles: i) Bayes' Theorem, ii) the Principle of Maximum Entropy, iii) the Generalized Least-squares Method. These principles are somewhat interrelated, as discussed in the references below. The following formalism is a fully probabilistic one in the sense that it offers a prescription for generating a probability distribution function $p(\mathbf{p})$ that embodies all the information available concerning the parameters $\mathbf{p}$.

Bayes' Theorem and the Principle of Maximum Entropy

In the present context, Bayes' Theorem assumes the form

$$p(\mathbf{p}|\mathcal{D}) = C\mathcal{L}(\mathcal{D}|\mathbf{p})p_a(\mathbf{p}), \tag{1}$$

where $p_a(\mathbf{p})$ is the *a priori* probability distribution that describes the knowledge of $\mathbf{p}$ before any new information is acquired, $\mathcal{D}$ represents the newly obtained information, $\mathcal{L}(\mathcal{D}|\mathbf{p})$ is the likelihood that the parameters $\mathbf{p}$ could have led to the data set $\mathcal{D}$, $p(\mathbf{p}|\mathcal{D})$ is the *a posteriori* probability distribution for $\mathbf{p}$ (after the new information became available) and C is a positive constant which insures that the *a posteriori* distribution is normalized, i.e., that the requirement $\int p(\mathbf{p}|\mathcal{D})d\mathbf{p} = 1$ is satisfied when integration is carried out over the entire space of physically reasonable parameters $\mathbf{p}$.

Suppose that the experiments and/or calculations which generated the data set $\mathcal{D}$ involve a collection of J physical quantities denoted collectively as $\mathbf{y} = (y_1,y_2, \ldots ,y_j, \ldots ,y_J)$. The generation of data entails uncertainties, therefore let $\mathbf{V}_y$ represent the covariance matrix (error matrix) for these data. Thus, $\mathcal{D}$ is represented by the values $\{\mathbf{y},\mathbf{V}_y\}$. It is assumed that given parameter set $\mathbf{p}$ it is possible to calculate a set of J quantities $\mathbf{f}(\mathbf{p}) = [f_1(\mathbf{p}),f_2(\mathbf{p}), \ldots ,f_j(\mathbf{p}), \ldots ,f_J(\mathbf{p})]$ which are equivalent to the data values $\mathbf{y}$ (one-to-one). The Principle of Maximum Entropy enables a relatively simple expression for $\mathcal{L}(\mathcal{D}|\mathbf{p})$ to be written down directly, namely,

$$\mathcal{L}(\mathcal{D}|\mathbf{p}) \propto \exp\{(-1/2)[\mathbf{y}-\mathbf{f}(\mathbf{p})]^{+}\mathbf{V}_y^{-1}[\mathbf{y}-\mathbf{f}(\mathbf{p})]\}, \tag{2}$$

where "$^{+}$" signifies matrix transposition and "$^{-1}$" signifies matrix inversion. $\mathbf{V}_y$ is required to be positive definite. If the *a priori* knowledge includes a parameter set $\mathbf{p}_a$ and corresponding positive definite covariance matrix $\mathbf{V}_a$, then the Principles of Maximum Entropy states that

$$p_a(\mathbf{p}) \propto \exp\{(-1/2)[\mathbf{p}-\mathbf{p}_a]^+\mathbf{V}_a^{-1}[\mathbf{p}-\mathbf{p}_a]\}, \quad (3)$$

which is a multivariate normal distribution.

The Generalized Least-squares Method

The Generalized Least-squares Method (GLSM) follows from imposing a maximum-likelihood condition on Eq. (1), namely, that the GLSM solution for **p** is the one for which the *a posteriori* probability distribution achieves its maximum value. Because of the nature of the exponential function, combining Eqs. (1)-(3) leads to the requirement

$$[\mathbf{y}-\mathbf{f}(\mathbf{p})]^+\mathbf{V}_y^{-1}[\mathbf{y}-\mathbf{f}(\mathbf{p})] + [\mathbf{p}-\mathbf{p}_a]^+\mathbf{V}_a^{-1}[\mathbf{p}-\mathbf{p}_a] = \text{minimum}, \quad (4)$$

provided that the new and prior knowledge are essentially independent (a point which an evaluator must always keep in mind when collecting input data and prior information for a GLSM evaluation). If the relationship between **p** and **f**(**p**) is non-linear, in general it can be quite difficult to find a solution which satisfies Eq. (4). There are ways to do this based on numerical integration involving the probability distribution but this will not be discussed here. However, if the model is linear, i.e., if **f**(**p**) = **Ap**, then the *a posteriori* probability distribution $p(\mathbf{p}|\mathbf{y})$ is a multivariate normal distribution. The matrix **A** is often referred to as the design (or sensitivity) matrix. A complete description of the relationship between the acquired data and the parameters to be derived from the evaluation process is contained in **A**. Even if the relationship between **p** and **f**(**p**) is non-linear it may still be possible to linearize the problem via the approximate relationship

$$\mathbf{f}(\mathbf{p}) - \mathbf{f}(\mathbf{p}_a) \approx \mathbf{A}(\mathbf{p}-\mathbf{p}_a), \quad (5)$$

where the elements of matrix **A** are given by the expression $a_{jk} = [\partial f_j/\partial p_k]$ evaluated at $\mathbf{p} = \mathbf{p}_a$. The approximation in Eq. (5) is valid as long as the solution **p** does not differ too much from the prior estimate $\mathbf{p}_a$. In practice, most evaluations rely on being able to use this approximation, and therefore experienced evaluators try to set up an evaluation process so that this condition is reasonably well satisfied.

For the linear (or "linearized") model, the solution to Eq. (4) is contained in the following four equations which form the basis of data evaluation by GLSM:

$$\mathbf{p} = \mathbf{p}_a + \mathbf{V}_a\mathbf{A}^+(\mathbf{Q}+\mathbf{V}_y)^{-1}[\mathbf{y}-\mathbf{f}(\mathbf{p}_a)], \quad (6)$$

$$\mathbf{Q} = \mathbf{A}\mathbf{V}_a\mathbf{A}^+, \quad (7)$$

$$\mathbf{V}_p = \mathbf{V}_a - \mathbf{V}_a\mathbf{A}^+(\mathbf{Q}+\mathbf{V}_y)^{-1}\mathbf{A}\mathbf{V}_a, \quad (8)$$

$$(\chi^2)_{min} = [\mathbf{y}-\mathbf{f}(\mathbf{p}_a)]^+(\mathbf{Q}+\mathbf{V}_y)^{-1}[\mathbf{y}-\mathbf{f}(\mathbf{p}_a)]. \quad (9)$$

Two features of this solution are worth pointing out here. First, the solution yields not only a parameter vector **p** (the evaluation itself) but also a corresponding covariance matrix $\mathbf{V}_p$ representing the uncertainties in the evaluated quantities. Second, there is a statistical test provided *gratis* in the form of the quantity $(\chi^2)_{min}$. This quantity obeys a chi-squared distribution with J degrees of freedom. A comparison with standard tables of the

chi-square distribution enables the evaluator to determine whether the input data and/or evaluation model are consistent. If inconsistencies are found then the input information and evaluation model must be examined to try to discover the source of the problem.

Practical Considerations in Data Evaluation

The procedure sketched out above is deceptively simple. In order to emphasize this point it is worthwhile examining each of the quantities appearing in Eqs. (6)-(9).

p (parameters to be evaluated):

It may seem obvious what it is that one wishes to evaluate but this is not always the case. For example, if an evaluated reaction cross section is desired vs. incident energy, this is really a continuous function. How should it be represented? One approach is to give "point" cross sections, namely, a set of energies and corresponding cross section values such that one can reconstruct the desired curve through interpolation. Another approach is to give group cross sections, namely, interval-average cross sections for well-defined energy intervals. If the desired quantity is a derived value, e.g., a Maxwellian-spectrum-average capture neutron capture cross section then this needs to be well defined before the evaluation process begins.

$\mathbf{p}_a$ and $\mathbf{V}_a$ (prior parameter values and their uncertainties):

In the GLSM method it is necessary to start from somewhere, even if it is only a guess. The prior parameters $\mathbf{p}_a$ might be values from an earlier evaluation (in which case the new evaluation should include only information not reflected in the earlier evaluation) or they may result from model calculations which are to be "adjusted" by the inclusion of new experimental data via the GLSM method (data merging). The associated covariance matrix $\mathbf{V}_a$ needs to be generated in a consistent way (e.g., it must be positive definite). This is not easy to do if the prior values are merely estimates, or if they are based on calculations using models that are very sensitive to fundamental nuclear interaction constants and that are not well validated to begin with. It seems rather intimidating to be forced to provide something as input to the codes which implement GLSM in the face of such sketchy knowledge. However, it should be comforting to know that assumed prior parameters with large errors generally carry very little weight in the GLSM process, and the solution tends to be heavily dominated by the new information if that is both extensive and relatively accurate. Still, this happy state of affairs can be thwarted if the correlations existing in the covariance matrices $\mathbf{V}_a$ and $\mathbf{V}_y$ are too strong, posing yet another potential pitfall for the wary evaluator!

y and $\mathbf{V}_y$ (new data and their uncertainties):

The most important thing to know here is what the data actually represent. Are the energies well established? What was the neutron spectrum in which they were measured? What standards were used? Are the various data collected from the literature truly independent or are there common sources of uncertainty? These and many other questions force the evaluator to examine the data and their documentation very carefully, and it is often necessary to adjust these data for changes in standards, to transform to new energy grid points, etc. This process of adjusting data prior to their evaluation is the most time consuming part of evaluation work, and often it is the most arbitrary one since poor documentation of published data is a notorious problem. Only when the input data are properly prepared can one hope to get reasonable results

from an evaluation, regardless of the procedure used. This simply cannot be done by a "machine approach" without the aid of human scrutiny.

f(**p**) or **A** (the model which relates the data to the evaluated parameters):

This is a test of the evaluator's skill. The elements of matrix **A** can be generated easily enough from the selected model, either analytically or via numerical procedures. What is taxing is knowing just how a given piece of data relates to the parameters to be evaluated when the data in question are either undocumented or relatively poorly documented. Often it is necessary to reject certain data points because crucial information is lacking. For example, if a cross-section published in 1957 indicates an energy "14 MeV" could this mean 13.9 MeV or 14.1 MeV? If the physical quantity is known to vary rapidly with energy this is a crucial matter. Often the evaluator has to look at the original paper and, from a description of the experimental setup, try to answer the question. Early works frequently fail to indicate which standards were used or to give actual values for these standards when they are mentioned. Based on the date of the work and clues in the documentation it may be possible for an evaluator to estimate what was used in the original data analysis with reasonable reliability. Frequently that is not possible. If care is not taken to relate what was measured to what is sought then the evaluation process reduces to an exercise not unlike that of comparing apples and oranges.

$(\chi^2)_{min}$ parameter (test for confidence in the GLSM evaluation):

If all the data are reasonably consistent with the assumed uncertainties, and if the evaluation model is consistent with the input data, then $(\chi^2)_{min} \approx J$ (number of degrees of freedom) should result from the analysis embodied in Eqs. (6)-(9). If $(\chi^2)_{min} >> J$, then there are inconsistencies which need to be resolved by the evaluator. This may entail looking at all the data sets to see if they are discrepant or if the assumed errors are too small. It may also entail looking at the evaluation model which relates the data and parameters to see if it is somehow faulty. In any case, the evaluator must do something! An evaluation with a low degree of confidence (large chi-square value) is simply unacceptable.

Finally, it should be mentioned that computational round-off errors associated with the adjustment of data or with the GLSM evaluation process (which often involves the inversion of large matrices) can lead to inferior evaluated results. Evaluators need to insure that their analyses are carried out using adequate numerical precision.

Summary

Data evaluation, like making good wine or cheese, involves not only good quality ingredients but also depends critically on the "art of the evaluator". Combing the literature and experimental data files for the raw materials needed in evaluations has been likened to archaeology. A good evaluator must be a very patient individual. Modern data evaluation concepts, as embodied in GLSM, provide an unbiased approach to the merging of all types of data which become known to an evaluator, once it has been assembled, examined critically and put into a unified format for analysis. There are various codes that can do the actual GLSM calculations, depending upon the nature of the data (e.g., SAMMY, GLUCS, GMA, GLSMOD, UNFOLD, BAYES, etc.). The particular software which is used is generally of less importance than understanding the nature of the data employed and verifying its fidelity.

References

This list gives a good starting point for a more extensive study of the literature on this subject.

1. Donald L. Smith, "Covariance Matrices and Applications to the Field of Nuclear Data", Report ANL/NDM-62, Argonne National Laboratory (1981).

   [A basic primer on the concepts of modern data evaluation. Formulas are derived and some simple examples are worked through in detail].

2. Donald L. Smith, "Non-evaluation Applications for Covariance Matrices", Report ANL/NDM-62, Argonne National Laboratory (1981).

   [Extends the material presented in Report ANL/NDM-62 by giving numerous simple examples from everyday nuclear applications including detector calibrations, etc. Note that the methods used in analyzing experimental nuclear data are identical to those employed in nuclear data evaluation].

3. Donald L. Smith, Probability, Statistics, and Data Uncertainties in Nuclear Science and Technology, American Nuclear Society, LaGrange Park, Illinois (1991).

   [A comprehensive treatment of basic probability theory and the development of modern nuclear data analysis and evaluation methodology. Numerous simple examples are given and a detail search of the literature on nuclear data evaluation methodology up to 1990 is documented in the reference list. Available in hardback (269 pages) from the American Nuclear Society Press, LaGrange Park, Illinois 60525, USA. Price is $25 with a 10% discount available for ANS members (credit card orders accepted).]

4. Nuclear Data Evaluation Methodology, ed. Charles L. Dunford, World Scientific Press, Singapore (1993).

   [An extensive collection of papers on nuclear data evaluation reflecting the status of development up to 1992. Contributions to a conference on data evaluation held at Brookhaven National Laboratory.]

5. Donald L. Smith, "A Least-squares Computational 'Tool Kit'", Report ANL/NDM-128, Argonne National Laboratory (1993).

   [A handy reference on the basic principles of data evaluation by the least-squares method (both simple and generalized). Examples are given and computer codes that are useful for data analysis and evaluation are described.]

6. A. Pavlik, M.M.H. Miah, B. Strohmaier and H. Vonach, "Update of the Evaluation of the Cross Section of the Neutron Dosimetry Reaction ^{103}Rh(n,n')^{103m}Rh", Report INDC(AUS)-015, IAEA Nuclear Data Section, International Atomic Energy Agency, Vienna (1995).

   [Well-documented description of a recent evaluation effort at the University of Vienna. The procedures associated with collecting data from the literature and reviewing and adjusting it in preparation for evaluation by the least-squares method are discussed very well in this report.]