THE RESOLUTION OF DISCREPANCIES AMONG NUCLEAR DATA

ROBERT PEELLE
Consultant, Nuclear Data Measurement and Evaluation Section
Oak Ridge National Laboratory, P. O. Box 2008
Oak Ridge, Tennessee, USA 37831-6356

ABSTRACT

Significant differences among input data occur in the evaluation of nuclear data because it is difficult to achieve experimental results with the accuracy required for some applications. Types of "discrepancies" are classified. The means are reviewed by which an evaluator may treat discrepancies in the process of evaluation. When all means fail that are based on how the discrepant data were obtained, the perplexed evaluator must sometimes combine discrepant data based just on the stated values and uncertainties; techniques for treating such challenges are compared. Some well-known data discrepancies are examined as examples.

1. Introduction

Particularly for neutron reactions, experimenters often must strain to achieve the level of accuracy required by nuclear data users. In most such cases theory provides some guidance, but model calculations yield insufficient absolute accuracy for those cross sections technology requires be known within a few percent. Moreover, in some cases there has been difficulty in reaching consensus on whether certain cross sections should be accurately represented in file formats and, if so, how best to accomplish the goal. An example is that only in the recent data files for the MeV region has the ENDF/B system included formats to represent continuum yields that are functions of both secondary angle and energy. Finally, differential data sets have been subjected to application-embedded integral tests of data and methods which themselves are difficult to perform and interpret. In this context all types of discrepancies are to be expected.

Typically, microscopic evaluated neutron data are used as input for complex transport calculations designed to estimate system parameters of practical interest. First order sensitivity theory is used to estimate the effects of cross section changes on these calculated parameters. The same sensitivity coefficients allow propagation of data uncertainties and covariances to give the uncertainties in, and correlations among, the calculated parameters. Differences between calculated and measured integral parameters for test systems can be analyzed using sensitivity coefficients only to the extent that evaluators have included covariances in the data files for the reactions of importance. Covariance data are also required for integral benchmark data.

Data discrepancies are resolved by thoughtful persons who find problems in experimental measurements, rather than by "canned" procedures applied to published
results. Good decisions are often reached by distillation of ephemeral clues. Therefore, this paper attempts to categorize, clarify, and comment on methods used to make progress in the presence of discrepancies, rather than to provide preferred recipes for action. It is hoped to remind evaluators of the present possibilities and encourage them to seek improved strategies.

2. Types of Nuclear Data Discrepancies

Dialog about "discrepancies" would be enhanced if workers more carefully distinguished among the situations that commonly arise.

"Truly discrepant" experimental values differ so much relative to the stated uncertainties that the nominal probability is $< 10^{-3}$ that such disagreement would occur by chance. (Beware non-normal probability density functions!) This is a case that pains evaluators as old records are reviewed, values and/or uncertainties are altered with the grudging permission of authors, and somewhat arbitrary techniques are sometimes employed in attempts to resolve the apparent problem.

"Important differences" among values, judged relative to application needs, occur when data are not truly discrepant but differ enough to concern users. Frequently, more accurate experiments need to be sought, for there is little point in debating which inadequate values to adopt. Careful analysis of uncertainties by experimenters allows this situation to be recognized.

"Discrepancies with integral experiments" are important differences between the results of integral experiments and the corresponding values computed from a differential cross section data base. When all data include covariance information and analysis biases and uncertainties are accounted for, true discrepancies sometimes remain. Since such comparisons usually require complicated computational models, the study of such discrepancies must go beyond the experimental values themselves.

"Discrepancies with theory" occur when experimental values for some quantities are systematically inconsistent with model calculations. Such discrepancies often occur in angle or energy distributions for a range of incident energy and a series of nuclides. Perhaps the theory being used can't account for an observed trend, or perhaps there are uncompensated systematic errors in the experiments in these regions. Such systematic discrepancies should not be treated as random uncertainties in assessing the uncertainty of a set of model parameters, but rather as an indication of the need to refine either the nuclear model or the experiments.

3. Evaluating Discrepant Data

Here we consider true discrepancies among comparable data elements, and wish to itemize the approaches used by evaluators to deal with the discrepancy to allow presentation of a combined value and its uncertainty. Assume that the experimenters have given equivalently defined numerical uncertainties as well as any pertinent covariance data.
A typical evaluator first updates values to correct apparent mistakes in analysis and to include modern values of any auxiliary information previously used in reducing the raw data. Data sets should be used according to the quantities measured, even if these are not the quantities reported by the authors. For example, values reported as cross sections in journals may have been measured as unnormalized ratios to some standard cross section. By the time these initial steps have been completed, the existence of an apparent discrepancy may be recognized. If the data are important, the evaluator should ask the experimenters to comment on the discrepancy and any weak points in the documentation. The evaluator should also consider initiating a request for a measurement with improved technique or most careful execution of a previously used technique. This can be done by direct contact with experimenters and/or by submitting the item to the appropriate nuclear data request list. In the interim, some means to combine the existing data will usually be required.

An evaluator may next look for legitimate reasons to increase uncertainties on specific data elements based on an author's failure to include numerical values for all acknowledged uncertainties, failure to recognize some inevitable sources of uncertainty, or documentation that does not justify unusually small reported uncertainties. (The "needed increase" is not a justification here.) A clue in this search may be large differences between uncertainties reported for different experiments performed using similar techniques; however, sometimes the conditions were indeed much more favorable for some of the experiments. The difficulties of this evaluation process are legendary.

If the data remain discrepant, the evaluator may consider the following steps:

(a) Examine the raw data for the experiments that give the discrepant results. If the features of interest are unclear in the raw data, there may be justification to reject a data set or to enlarge the uncertainties quoted for an experiment. Examples occur in the evaluation of resonance parameters.

(b) Increase the uncertainties on older data as once discussed by Poenitz, preferably in some systematic way that doesn't depend on recognition of a discrepancy. Equipment and techniques were generally superior in 1985 relative to those in 1965, and experimenters had the advantage of more experience. Note the weakness, however, in general use of weighting according to newness; nuclear data facilities and scientists may well be less able to produce precise measurements in 1995 than in 1985.

(c) Assign a minimum uncertainty for any measurement of a given type, based in part on the known capabilities of the experimental technique. In the limit, each experiment would be given equal weight, but it is probably self-defeating to set the minimum uncertainty too large. For multipoint data sets this idea must be elaborated to consider correlation patterns.

(d) Consider trends apparent in other data from the same author and experimental equipment, perhaps for nuclides for which a greater wealth of relevant data is available, that could justify rejection of a data set or assignment of an

**DISCLAIMER**

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.
expanded uncertainty. Sometimes data nearly impossible to explain by theory can be rejected.

If discrepancies persist after these techniques have been used or rejected, one may perform a weighted average and accept the result. A few apparent discrepancies are indeed expected. Often, the uncertainty of the quoted weighted average is increased by scaling the variances and covariances of all input values by a common factor to make the chi-square just equal to the number of degrees of freedom. This means that each input is given a relative weight in accord with its author's estimate, and the scatter among the experimental results determines the uncertainty of the average. The combined result itself remains unchanged, even though it is likely that the (unknown) proper correction of the input data would change the data correlation pattern and the average value.

If reliance on the above approaches seems inadequate, the evaluator may consider those in the next section.

4. Numerical Techniques For Addressing Discrepancies

The techniques considered below are more controversial because they do not refer to the experiments that produced the discrepant values. They can be attractive and perhaps necessary if discrepancies remain after detailed evaluation of the underlying experiments or if there is no opportunity for such expert evaluation. Use provides the data evaluator a transparently impersonal "objective" approach and thereby some protection against authors of errant data. However, primary reliance on such techniques will reduce the likelihood that the evaluator can make an intellectual contribution by pointing to corrections needed in the presented data. More important, experimenters would have less incentive to document their work carefully and thereby might be less likely to find their own previously unsuspected errors. The paragraphs below refer to some recent papers on this popular topic.

Froehner exhibited a Bayesian formalism in terms of unrecognized errors on each data element. The approach has the merit that it is derived in a clear-cut way from simple assumptions. Knowledge concerning the relative "reliability" of the experiments (with respect to unrecognized errors) can be included. The evaluated results from this formulation move smoothly from the limit for data consistency to the limit for extreme data inconsistency. The former yields the usual weighted average with the original uncertainties. The latter leads (for equal reliability) to an unweighted average with uncertainty tied to the value of the parameter $\tau$ for the width of the assumed density function of the unrecognized errors. The text of ref. 11 does not prescribe a specific choice of this parameter $\tau$. Stepping outside the spirit of Froehner's work, one could adjust $\tau$ to achieve consistency. For the case of equal reliability, the updated uncertainties on the input data would be formed by combining the original uncertainties with $\tau$ in quadrature. The uncertainty in the combined result is taken from the extended uncertainties just as in a weighted average. Note
that "outlier" values are not automatically taken to be suspect. Froehner has considerably extended this approach to deal with the scale of the $\tau$ values.\textsuperscript{12}

Statistical criteria are used by Yaborov\textsuperscript{13} to identify which of a series of data elements are "mismatched" in that they are believed to have uncertainties larger than postulated. In the proposed scheme, the uncertainties of these mismatched data are then scaled up. Note that in this approach and the others below the uncertainty will never be scaled up on a value close to the average, even though in fact its uncertainty may have been badly underestimated.

The paper of James, Mills, and Weaver\textsuperscript{14} considers each experiment result in terms of its "normalized residual," the difference between it and the weighted mean of the remaining measurements, divided by the standard deviation of that difference. They suggest reducing the weight of the data element that has the largest normalized residual if it is greater than a selected limiting value. The weight is reduced by the amount just needed for the modified residual to have the chosen limiting value. The defined normalized residual has nice properties for this type of analysis. The present author is concerned that this method could lead to inappropriate deweighting of outliers unless the user has previously identified and corrected all unreasonably small uncertainties.

Various techniques are discussed and compared by Rajput and MacMahon,\textsuperscript{15} and a process termed the "Rajeval technique" is described. First, results are rejected as outliers that have residuals greater than a selected critical value. (Here, the residuals are calculated like the normalized residuals of the previous paragraph except that the unweighted mean is used.) Then the weighted residuals for the remaining values are examined to identify those deemed inconsistent according to a criterion that depends on the number of measurements. The uncertainties on these values are increased until the criterion for consistency is met.

While all these approaches are thoughtful efforts that have performed encouragingly in specific applications, some seem to require rather arbitrary selection of critical values. The present author suggests that a potential user try any prospective methods against a few test cases with anomalies of the types expected in the data to be examined, just to be sure odd cases will be handled appropriately. For example, selection of a technique might depend on whether one expects biases in some of the values rather than badly underestimated standard deviations. Overall, it seems safer to utilize a technique that has a clear if idealized theoretical foundation.

5. Examples of Resolution of Nuclear Data Discrepancies

A few examples of nuclear data discrepancies of the last 20 years are reviewed below. Talks later in this session cover other examples. A discrepancy file is maintained by international nuclear data organizations.\textsuperscript{16}
5.1 Fission Neutron Multiplicity of $^{252}\text{Cf}$.

The average number of fast neutrons emitted per fission $\bar{\nu}$ is the single parameter to which reactor physics calculations are most sensitive.\(^{17}\) Fortunately it can be measured well. The value for spontaneous fission of $^{252}\text{Cf}$ has long been used as a reference standard. By 1970, in addition to ongoing experiments, there were two careful experimental measurements based on counting of prompt neutrons absorbed in liquid scintillator, and others equally careful based on time-uncorrelated measurements of emitted neutrons absorbed in manganese baths or analogous instruments. Depending on the corrections applied, the results were very weakly inconsistent with the quoted uncertainties. However, the scintillator tank results appearing to be larger by 1%, a few times the $1/4\%$ uncertainty quoted on the weighted mean of about 3.73 neutrons/fission.\(^{18}\)

Teams of expert experimenters combed experiment documentation for obscure errors, and authors received repeated requests for additional experimental details. Numerous re-evaluations of the data observed in particular experiments were performed and several were adopted. H. Kouts in the keynote address to the 1975 nuclear data conference labeled the situation a scandal.\(^{19}\) It became apparent that improved data were required to support a needed evaluated uncertainty in the neighborhood of 0.2%, while the measurements existing then had quoted uncertainties of greater than 0.4%. Considering the importance of the data, experts were not willing to rely too much on reduction of uncertainty through averaging. In this light, it appeared that additional measurements were needed using techniques that would take advantage of all the insight gained over the years. Spencer and Gwin performed such a painstaking experiment and achieved a claimed uncertainty of about 0.2%.\(^{20}\) For now the controversy has abated even though their value is 1% higher than the average of the others.

Looking back at the situation, the data set after corrections in 1972 showed important differences among the data, but only borderline discrepancy. When an improved experiment could be completed and documented to avoid many of the concerns about earlier experiments, it appeared that the need had been met. Moreover, the whole set of thermal neutron constants had become internally consistent with a $^{252}\text{Cf}$ $\bar{\nu}$ value of 3.768 ± .005.\(^{21}\)

5.2 Fission Cross Section of $^{235}\text{U}$.

Because of the direct importance of fast neutron fission in $^{235}\text{U}$ and its use as a standard cross section, work was strongly underway in the 1960s toward achieving 1% uncertainty in this cross section from 0.1 MeV to at least 15 MeV. Great efforts were made to improve techniques. An apparently excellent measurement of this era was performed by White\(^{22}\) at several energies up to 14 MeV. His documentation was so detailed and thoughtful that other workers assumed that his results would endure
even though his values were about 5% below previously favored ones. Viewed from a more recent perspective, he worked against major disadvantages including lack of a pulsed source to allow flight-time discrimination against backgrounds from fission at lower energy. Except at 5.4 MeV, his original values, quoted to 2-3%, are all larger than those in ENDF/B-VI by 0.7 to 3 times his uncertainty estimates.

The process of upgrading data continues, though only extremely precise and somewhat diverse experiments are generally encouraged. Since the ENDF/B-VI evaluation, new broad-range data were obtained including energies above 15 MeV\textsuperscript{23} where the data base had been relatively thin, and notable differences were recorded. Another evidence that this and other standards data are not yet mature is that complete covariance data are not yet released for ENDF/B-VI, and the values of the reported evaluated uncertainties are questioned. The uncertainties given by the data combination procedure used for the evaluation do suggest that the 1% accuracy goal has been reached from 0.2 to 15 MeV for 70 average cross section values covering that interval.

This example illustrates that while skillfully obtained and documented data can be affected by systematic error, persistent effort by a variety of approaches can be rewarded by gradually convergent results. The evaluator is left with a puzzle how to fairly treat the results from the decade prior to the work of White.

5.3 Energy Balance in Evaluations.

Upon the release of ENDF/B-V, it was discovered that when evaluations were used to compute the kinetic energy released locally in a material (KERMA) there were energy balance discrepancies.\textsuperscript{24} That is, after taking into account the Q values of reactions, the incident and outgoing energies were not equal. In the MeV region, a major source of this discrepancy was that measured reaction gamma-ray emission spectra were often placed directly in the file, while the neutron energy distributions for reactions feeding those emissions were evaluated based on other data. There were bound to be distressing differences between the two sources of information even if those differences were within known uncertainties.

All (excluding fission) or most of the neutron cross sections and secondary distributions in these regions are obtained from nuclear model calculations that are benchmarked against the available data. When these models include de-excitation gamma-ray cascade treatments they produce photon yields vs. incident energy that are included in the benchmarking against experimental observations.\textsuperscript{25} However, much of the experimental photon data were obtained using elemental samples, while the evaluations are best performed isotope by isotope.

In many of the current evaluations,\textsuperscript{26} energy imbalance in the evaluation has been avoided by using model codes to give the ENDF/B gamma-ray data required for applications. (In others, various adjustments have been applied to force energy balance.) Energy distributions from neither secondary neutron nor photon data are included directly in the file, but both types of data are examined during the process
of setting up the model calculations to achieve a good compromise parameter set. The presence of any discrepancy between them may be deduced from the separate comparisons with model calculations. In that sense, the evaluations have been made internally consistent, but serious inconsistencies between neutron and photon data may remain.

6. Concluding Remarks

Numerical techniques such as weighted least-squares data combination have most validity with consistent or nearly consistent data; in face of discrepancy the use of technical judgement is necessary. For important data, overcoming either real discrepancies or important differences is likely to be slow and thereby expensive work.

Evaluators need to be bold though diplomatic, and should not fear to exercise and document scientific judgement related to updating values and uncertainties in the work of others. (Note that there is substantial disagreement on this point!)

Use of theory as an evaluation tool is virtually required as incident energies rise above the threshold for emission of second and third particles, and numerous reaction channels open that may not have been observed in any experiment. Evaluation theorists need to remain aware of the level of experimental verification available, and experimenters need to seek and perform critical measurements to help theorists control systematic modeling error.

7. Acknowledgement

The author acknowledges the many workers who generated nearly all the ideas compiled above. Helpful reviews were received from D. C. Larson and L. W. Weston. Only in a few cases could sources be identified. Research sponsored by the Office of Energy Research, Division of Nuclear Physics, U. S. Department of Energy, under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.

8. References

3. ibid. P. Collins, p. 159ff.


