NUCLEAR DATA AND MEASUREMENTS SERIES

ANL/NDM-121

# A Suggested Procedure for Resolving an Anomaly in Least-Squares Data Analysis Known as "Peelle's Pertinent Puzzle" and the General Implications for Nuclear Data Evaluation

by

Satoshi Chiba and Donald L. Smith

September 1991

**ARGONNE NATIONAL LABORATORY,**
**ARGONNE, ILLINOIS 60439, U.S.A.**

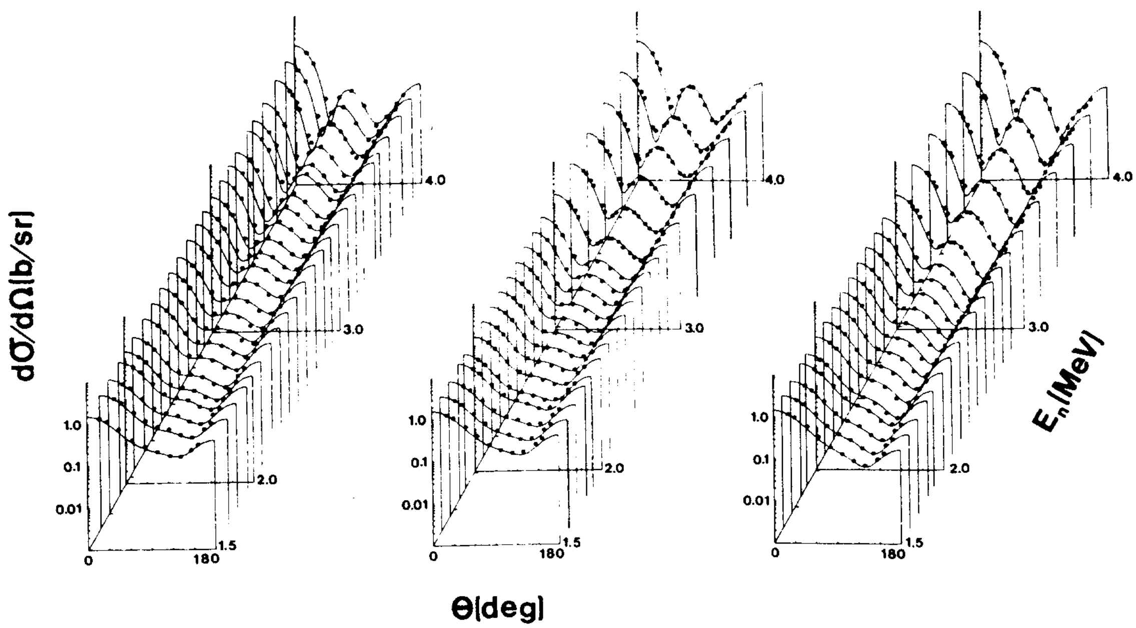
# NUCLEAR DATA AND MEASUREMENTS SERIES

ANL/NDM-121

**A SUGGESTED PROCEDURE FOR RESOLVING AN ANOMALY IN LEAST-SQUARES DATA ANALYSIS KNOWN AS "PEELLE'S PERTINENT PUZZLE" AND THE GENERAL IMPLICATIONS FOR NUCLEAR DATA EVALUATION**

Satoshi Chiba and Donald L. Smith

September 1991




ARGONNE NATIONAL LABORATORY, ARGONNE, ILLINOIS

Operated by THE UNIVERSITY OF CHICAGO

for the U. S. DEPARTMENT OF ENERGY

under Contract W-31-109-Eng-38

Reproduced from the best available copy.

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831
Prices available from (615) 576-8401, FTS 626-8401

Available to the public from the
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, VA 22161

ANL/NDM-121

# A SUGGESTED PROCEDURE FOR RESOLVING AN ANOMALY IN LEAST-SQUARES DATA ANALYSIS KNOWN AS "PEELLE'S PERTINENT PUZZLE" AND THE GENERAL IMPLICATIONS FOR NUCLEAR DATA EVALUATION*

by

Satoshi Chiba** and Donald L. Smith

Engineering Physics Division
Argonne National Laboratory
9700 South Cass Avenue
Argonne, Illinois 60439
U.S.A.

February 1990 (Original Manuscript Completed)

September 1991 (Revised Manuscript Completed)

NUCLEAR DATA EVALUATION. Peelle's Pertinent Puzzle. Absolute errors. Percent errors. Systematic errors. Random errors. Covariance matrices. A priori parameters. A posteriori parameters. Best-estimate parameters. Bayes' theorem. Data weighting in evaluations. Correlations. Least-squares estimators. Chi-square statistic. Enhanced errors. Modifications to the nuclear data evaluation code GMA.

---

* This work was supported by the U.S. Department of Energy, Energy Research Programs, under contract W-31-109-Eng-38.

** Permanent address: Department of Physics, Japan Atomic Energy Research Institute, Tokai-mura, Naka-gun, Ibaraki-ken 319-11, Japan.

## NUCLEAR DATA AND MEASUREMENTS SERIES

The Nuclear Data and Measurements Series presents results of studies in the field of microscopic nuclear data. The primary objective is the dissemination of information in the comprehensive form required for nuclear technology applications. This Series is devoted to: a) measured microscopic nuclear parameters, b) experimental techniques and facilities employed in measurements, c) the analysis, correlation and interpretation of nuclear data, and d) the evaluation of nuclear data. Contributions to this Series are reviewed to assure technical competence and, unless otherwise stated, the contents can be formally referenced. This Series does not supplant formal journal publication, but it does provide the more extensive information required for technological applications (e.g., tabulated numerical data) in a timely manner.

## TABLE OF CONTENTS

## INFORMATION ABOUT OTHER ISSUES OF THE ANL/NDM SERIES

A list of titles and authors for previous issues appears in each report of the Series. The list for reports ANL/NDM-1 through ANL/NDM-75 appears in ANL/NDM-76, and ANL/NDM-91 contains the list for reports ANL/NDM-76 through ANL/NDM-90. Below is the list for ANL/NDM-91 up to the current report. Requests for a complete list of titles or for copies of previous reports should be directed to:

Section Secretary
Applied Nuclear Physics Section
Engineering Physics Division
Building 314
Argonne National Laboratory
9700 South Cass Avenue
Argonne, Illinois 60439
U.S.A.

**ANL/NDM-91** A.B. Smith, P.T. Guenther and R.D. Lawson, *On the Energy Dependence of the Optical Model of Neutron Scattering from Niobium*, May 1985.

**ANL/NDM-92** Donald L. Smith, *Nuclear Data Uncertainties (Vol.-I): Basic Concepts of Probability*, December 1988.

**ANL/NDM-93** D.L. Smith, J.W. Meadows and M.M. Bretscher, *Integral Cross-section Measurements for* $^{7}Li(n,n't)^{4}He$, $^{27}Al(n,p)^{27}Mg$, $^{27}Al(n,\alpha)^{24}Na$, $^{58}Ni(n,p)^{58}Co$, *and* $^{60}Ni(n,p)^{60}Co$ *Relative to* $^{238}U$ *Neutron Fission in the Thick-target* $^{9}Be(d,n)^{10}B$ *Spectrum at* $E_d = 7$ *MeV*, October 1985.

**ANL/NDM-94** A.B. Smith, D.L. Smith, P. Rousset, R.D. Lawson and R.J. Howerton, *Evaluated Neutronic Data File for Yttrium*, January 1986.

**ANL/NDM-95** Donald L. Smith and James W. Meadows, *A Facility for High-intensity Neutron Irradiations Using Thick-target Sources at the Argonne Fast-neutron Generator*, May 1986.

**ANL/NDM-96** M. Sugimoto, A.B. Smith and P.T. Guenther, *Ratio of the Prompt-fission-neutron Spectrum of Plutonium-239 to that of Uranium-235*, September 1986.

**ANL/NDM-97** J.W. Meadows, *The Fission Cross Sections of* $^{230}Th$, $^{232}Th$, $^{233}U$, $^{234}U$, $^{236}U$, $^{238}U$, $^{237}Np$, $^{239}Pu$, *and* $^{242}Pu$ *Relative* $^{235}U$ *at 14.74-MeV Neutron Energy*, December 1986.

**ANL/NDM-98** J.V. Meadows, *The Fission Cross Section Ratios and Error Analysis for Ten Thorium, Uranium, Neptunium and Plutonium Isotopes at 14.74-MeV Neutron Energy*, March 1987.

**ANL/NDM-99** Donald L. Smith, *Some Comments on the Effects of Long-range Correlations in Covariance Matrices for Nuclear Data*, March 1987.

**ANL/NDM-100** A.B. Smith, P.T. Guenther and R.D. Lawson, *The Energy Dependence of the Optical-model Potential for Fast-neutron Scattering from Bismuth*, May 1987.

**ANL/NDM-101** A.B. Smith, P.T. Guenther, J.F. Whalen and R.D. Lawson, *Cobalt: Fast Neutrons and Physical Models*, July 1987.

**ANL/NDM-102** D.L Smith, *Investigation of the Influence of the Neutron Spectrum in Determinations of Integral Neutron Cross-section Ratios*, November 1987.

**ANL/NDM-103** A.B. Smith, P.T. Guenther and B. Micklich, *Spectrum of Neutrons Emitted From a Thick Beryllium Target Bombarded With 7-MeV Deuterons*, January 1988.

**ANL/NDM-104** L.P. Geraldo and D.L. Smith, *Some Thoughts on Positive Definiteness in the Consideration of Nuclear Data Covariance Matrices*, January 1988.

**ANL/NDM-105** A.B. Smith, D.L. Smith, P.T. Guenther, J.W. Meadows, R.D. Lawson, R.J. Howerton, and T. Djemil, *Neutronic Evaluated Nuclear-data File for Vanadium*, May 1988.

**ANL/NDM-106** A.B. Smith, P.T. Guenther, and R.D. Lawson, *Fast-neutron Elastic Scattering from Elemental Vanadium*, March 1988.

**ANL/NDM-107** P. Guenther, R. Lawson, J. Meadows, M. Sugimoto, A. Smith, D. Smith, and R. Howerton, *An Evaluated Neutronic Data File for Elemental Cobalt*, August 1988.

**ANL/NDM-108** M. Sugimoto, P.T. Guenther, J.E. Lynn, A.B. Smith, and J.F. Whalen, *Some Comments on the Interaction of Fast Neutrons with Beryllium*, November 1988.

**ANL/NDM-109** P.T. Guenther, R.D. Lawson, J.W. Meadows, A.B. Smith, D.L. Smith, and M. Sugimoto, *An Evaluated Neutronic Data File for Bismuth*, November 1989.

**ANL/NDM-110** D.L. Smith and L.P. Geraldo, *A Vector Model for Error Propagation*, March 1989.

**ANL/NDM-111** J.E. Lynn, *Fifty Years of Nuclear Fission*, June 1989.

**ANL/NDM-112** S. Chiba, P.T. Guenther, and A.B. Smith, *Some Remarks on the Neutron Elastic- and Inelastic-scattering Cross Sections of Palladium*, May 1989.

**ANL/NDM-113** J.E. Lynn, *Resonance Effects in Neutron Scattering Lengths*, June 1989.

**ANL/NDM-114** A.B. Smith, R.D. Lawson, and P.T. Guenther, *Ambiguities in the Elastic Scattering of 8-MeV Neutrons from Adjacent Nuclei*, October 1989.

**ANL/NDM-115** A.B. Smith, S. Chiba, D.L. Smith, J.W. Meadows, P.T. Guenther, R.D. Lawson, and R.J. Howerton, *Evaluated Neutronic File for Indium*, January 1990.

**ANL/NDM-116** S. Chiba, P.T. Guenther, R.D. Lawson, and A.B. Smith, *Neutron Scattering from Elemental Indium, the Optical Model, and the Bound-state Potential*, June 1990.

**ANL/NDM-117** Donald L. Smith and Luiz P. Geraldo, *An Evaluation of the Nb-93(n,n')Nb-93m Dosimeter Reaction for ENDF/B-VI*, November 1990.

**ANL/NDM-118** J.W. Meadows, *Characteristics of the Samples in the FNG Fission Deposit Collection*, December 1990.

**ANL/NDM-119** S. Chiba, P.T. Guenther, A.B. Smith, M. Sugimoto and R.D. Lawson, *Fast-neutron Interaction with Elemental Zirconium, and the Dispersive Optical Model*, June 1991.

**ANL/NDM-120** A.B. Smith, P.T. Guenther, J.F. Whalen, and S. Chiba, *Fast-neutron Total and Scattering Cross Sections of Ni-58 and Nuclear Models*, April 1991.

# A SUGGESTED PROCEDURE FOR RESOLVING AN ANOMALY IN LEAST-SQUARES DATA ANALYSIS KNOWN AS "PEELLE'S PERTINENT PUZZLE" AND THE GENERAL IMPLICATIONS FOR NUCLEAR DATA EVALUATION*

by

Satoshi Chiba** and Donald L. Smith

Engineering Physics Division
Argonne National Laboratory
9700 South Cass Avenue
Argonne, Illinois 60439
U.S.A.

## ABSTRACT

Modern nuclear-data evaluation methodology is based largely on statistical inference, with the least-squares technique being chosen most often to generate best estimates for physical quantities and their uncertainties. It has been observed that those least-squares evaluations which employ covariance matrices based on absolute errors that are derived directly from the reported experimental data often tend to produce results which appear to be too low. This anomaly has come to be known as "Peelle's Pertinent Puzzle" (PPP) because the validity of certain evaluations afflicted in this manner has been called into question by R.V. Peelle. The anomaly, as originally posed by Peelle through a specific example, is discussed briefly in this report, and a procedure for resolving it is suggested. The method involves employing data uncertainties which are derived from errors expressed in percent. These percent errors are used, in conjunction with reasonable a priori estimates for the quantities to be evaluated (rather than the individual experimental values), to derive the covariance matrices which are required for applications of the least-squares procedure. This approach appears to lead to more rational weighting of the experimental data and, thus, to more realistic evaluated results than are obtained when the errors are

(continued on the following page)

---

* This work supported by the U.S. Department of Energy, Energy Research Programs, under contract W-31-109-Eng-38.

** Permanent address: Department of Physics, Japan Atomic Energy Research Institute, Tokai-mura, Naka-gun, Ibaraki-ken 319-11, Japan.

## ABSTRACT

(continued from the preceding page)

based on the actual data. The procedure is very straightforward when only one parameter must be estimated. However, for those evaluation exercises involving more than one parameter, this technique demands that a priori estimates be provided at the outset for all of the parameters in question. Then, the least-squares method is applied iteratively to produce a sequence of sets of estimated values which are anticipated to converge~~nce~~ toward a particular set of parameters which one then designates as the "best" evaluated results from the exercise. It is found that convergence usually occurs very rapidly when the a priori estimates approximate the final solution reasonably well. In fact, the procedure is observed to be quite robust in that convergence is not too difficult to achieve even when the a priori estimates aren't very good. Some examples are given in this report to illustrate the problem and to demonstrate the approach suggested here for its resolution. Some general implications for the practice of nuclear data evaluation are also discussed.

# 1. SOME COMMENTS ON THE LEAST-SQUARES METHOD

Data evaluation is comprised of specific procedures by which the available information, usually experimental in nature but sometimes consisting of both experimental and theoretical components, is employed to generate recommended "best" values for those particular parameters being evaluated. In modern practice, such procedures are quite often based on statistical inference. The least-squares method is the statistical approach most commonly used in contemporary nuclear data evaluation for inferring best values for parameters from the available information [Smi88b]. Various computer codes based on this method have been developed for use in routine statistical data evaluation applications [e.g., Poe81 and Smi82b]. For this pragmatic reason, there is a strong motivation to continue to utilize the least-square technique in spite of some troublesome complications which are frequently encountered in dealing with real data.

Use of the least-squares method is quite well justified on the basis of fundamental mathematical theory. It is beneficial to elaborate on this point before we proceed with a discussion of the specific issue of this report. There are two distinct philosophical approaches to the science of statistics, although both share many technical features in common. Classical statisticians adhere to the belief that all statistics is based on probability theory and that, in turn, probability provides a formal mathematical model for dealing with those phenomena in nature which reflect the consequences of the relative frequency of occurrence of certain well-defined events. Therefore, classical statistics is deeply rooted in set theory. On the other hand, Bayesian statisticians have adopted a more liberal point of view. Even though the mathematical formalism is generally identical, the concept of relative frequency is expanded in the Bayesian framework to include consideration of relative rational degrees of plausibility for those hypotheses which are suggested to explain physical phenomena. Bayes' theorem [e.g., Smi88a] plays a vital role in both of these approaches, but its interpretation and applications are much more severely restricted in the classical tradition. For this reason, Bayesian statistics offers scientists more opportunities for the application of statistical inference, but with higher levels of associated risk. Fortunately, the least-squares method emerges as a viable approach in <u>both</u> classical and Bayesian statistics. We will now review the procedure briefly and explore its origins and justification from both of these statistical points of view.

The least-squares problem can be formulated in a number of different ways [e.g., Fro86, Smi81, Man81, Mar71 and Zeh70]. Without any loss of generality, we shall consider only the following one: Let

$y_{exp}$ be a vector of n experimental values, corresponding to a set of observables Y. Associated with $y_{exp}$ is a covariance matrix V which represents the data uncertainties. For the present, we will defer consideration of exactly how V should be determined. Given this information, we wish to deduce the best values for a set of m parameters S (the evaluated results). A functional relationship between the observables Y and the parameters S must be postulated, i.e., calculated values are given by $y_{calc} = Y(S)$, in order to apply the method. Many physical problems permit the assumption of a linear regression model, i.e., the existence of a linear relationship between the observables Y and the parameters S. One then specifies a matrix A (often called the design matrix) such that $y_{calc} = AS$. If such a relationship does not exist naturally, it can be forced through approximation by introducing a priori estimates for the parameters and then using Taylor series expansions to obtain linear relationships which involve small increments of both the observables and the parameters. This step can lead to some complications, but let us not be distracted here by this distinct and somewhat irrelevant issue. In classical statistics, application of the least-squares method amounts to satisfying the following condition:

$$Q = (y_{exp} - y_{calc})^{+} V^{-1} (y_{exp} - y_{calc}) = \text{minimum}, \tag{1}$$

or, equivalently, $\nabla Q = 0$, where the symbol " $\nabla$ " signifies the gradient operator with respect to the parameter set S and the symbol "+" denotes matrix transposition. Q is known as a quadratic form. If the quantities $y_{exp}$ and V are independent of S, and $y_{calc} = AS$, then Eq. (1) is satisfied by the expression

$$S = V_S A^{+} W y_{exp}. \tag{2}$$

Furthermore, the covariance matrix for the solution S is given by the formula

$$V_S = (A^{+} W A)^{-1}, \tag{3}$$

where $W = V^{-1}$. A detailed derivation of this result can be found in a report by Smith [Smi81]. By this approach we are able to obtain all that we care to know, namely, best-estimate parameters S and their uncertainties, as defined by $V_S$. However, if Y is not linearly related to S, or if V depends upon S, we are confronted with a problem because Eqs. (2) and (3) no longer provide a rigorous solution to the condition expressed by Eq. (1). In the present report it will be assumed that Y and S are linearly related, but not necessarily that V is independent of S. We will return to a discussion of this issue in Section 3.

If we take S, as given by Eq. (2), and substitute it into the expression for Q given in Eq. (1), then that value of Q which is obtained is usually referred to as $\chi^2$. It is interpreted as a specific value of the chi-square statistic which is commonly employed in chi-square tests for the statistical significance of hypotheses (i.e., interpretations of the data). Since the expected value of the chi-square statistic is $\langle\chi^2\rangle = f = (n-m)$, the consistency of the data $y_{exp}$, the appropriateness of the linear regression model (as embodied in the design matrix A), and the quality of the solution S can be tested quantitatively by comparing the value of $\chi^2$ thus obtained with f. If $\chi^2/f > 1$ the consistency is not good. We should then proceed as follows: i) Attempt to identify and correct discrepant data wherever that is feasible. ii) Selectively increase certain errors in the data, where good judgment indicates that it is warranted. Only rarely should data be rejected. iii) As a last resort, enhance all the elements of the solution covariance matrix $V_S$ by the factor $\chi^2/f$. If $\chi^2/f < 1$, no further action is required. In other words, the evaluated data are considered to be consistent. Strictly speaking, such a statistical interpretation of $\chi^2$ is proper only if the data $y_{exp}$ are normally distributed. We shall return to this point later in the present section.

It is important to realize that the procedure based on Eq. (1) is simply postulated in classical statistics. This least-squares method is but one of several such ad hoc procedures which have been suggested for use in statistical inference, i.e., for the generation of "estimators" that are used to derive best values of certain physical parameters from the available information [e.g., Mar71, Smi81 and Zeh70]. The relative merits of these various procedures are usually judged by two general criteria: i) how practical they are to use and ii) the extent to which they produce estimators that exhibit certain properties which have been deemed to be mathematically desirable. Therefore, let us consider what additional desirable features least-squares estimators do exhibit. One feels intuitively that good estimators should make optimal use of all the available experimental data, that is they should be <u>sufficient estimators</u>. In this regard, we note that the expressions for S and $V_S$ do appear to utilize all the relevant input data. However, it is not easy to prove that a statistic is sufficient [Zeh70]. Another desirable feature for estimators is consistency. <u>Consistent estimators</u> will produce numerical values which converge toward the "true value" for large data sets. This is the case for least-squares estimators. Then, there is the matter of bias. An estimator is an <u>unbiased estimator</u> if the expected value of that estimator equals the true value of the parameter being estimated. So long as we adhere to linear regression models, least-squares estimators are unbiased, quite independent of the underlying distributions [Zeh70]. The Gauss-Markov theorem, as amended by Aitken for correlated data [Ait57], states that, within the framework of a linear regression model, as described above, the estimator for S which results from an application of the least-squares condition will be of minimum variance. The validity of this theorem does not depend upon the data being normally distributed! However, the assumption that the data actually are normally distributed leads to an additional benefit, namely, that one is rigorously justified in using the chi-square statistic, as indicated above, for significance testing. So, from the point of view of classical statistics, the least-squares method is amply supported as a procedure for data evaluation because it satisfies several worthy criteria for generating good estimators.

What about the Bayesian point of view? The approach is much different from the classical one, but we will now show that the end result turns out to be essentially the same. The ideas are thoroughly discussed in papers by Froehner [e.g., Fro86] and Jaynes [e.g., Jay68, Jay73, Jay76, Jay78, Jay80], so the present description of this fairly complex issue is kept fairly brief.

In Bayesian parameter estimation, mean values and an associated covariance matrix for the set of parameters in question are calculated directly from a probability function. In principle, this function incorporates all the possible information about these parameters that can be known. Actually, we are required to determine only the first and second moments of this distribution to meet our needs. The derivation of an expression for this probability function formally starts with a consideration of Bayes' theorem. We can express Bayes' theorem in terms of the notation which has already been developed in this report. We tacitly assume that the parameters represented by S are continuous rather than discrete. The formula we obtain from Bayes' theorem is

$$p(S|y_{exp}) = L(y_{exp}|S)p(S) \;/\; [\int L(y_{exp}|S')p(S')dS']. \tag{4}$$

Here, $p(S|y_{exp})$ is the a posteriori probability density function for the parameters S. It represents our knowledge of S after acquisition of the data set $y_{exp}$. The function p(S) is the a priori probability function. It represents our knowledge of S before obtaining the data set $y_{exp}$. $L(y_{exp}|S)$ is called the likelihood. It represents the probability (or likelihood) that parameter set S would produce data set $y_{exp}$. The factor [...] in the denominator of Eq. (4) is just a normalization constant which insures that

$$\int p(S|y_{exp})dS = 1. \tag{5}$$

The range of integration in Eqs. (4) and (5) is over all values of the parameters (represented by either S' or S, as the case may be) that are allowed by the physical nature of the problem.

Let $S_i$ be a typical parameter from the set S. The Bayesian estimate for this parameter is given by the expression

$$< S_i > = \int S_i p(S|y_{exp})dS, \tag{6}$$

while the elements of the covariance matrix $V_S$ are given by

$$V_{Sij} = \int (< S_i > - S_i)(< S_j > - S_j)p(S|y_{exp})dS. \tag{7}$$

It appears that nothing could be simpler than applying this approach, if only we knew the exact form of $p(S|y_{exp})$!

Neglecting the normalization constant (it can always be deduced from numerical integration if required), we see that Eq. (4) takes the form

$$p(S|y_{exp}) \propto L(y_{exp}|S)p(S). \tag{8}$$

We will assume, for the sake of the present discussion, that the only information available to us is the data set $y_{exp}$ and its associated uncertainties. The function p(S) must then be "non-informative", i.e., $p(S) \propto 1$ over the range of integration for S. So, under these assumptions we have

$$p(S|y_{exp}) \propto L(y_{exp}|S). \tag{9}$$

Jaynes recognized that Shannon's principle of maximum entropy from information theory [e.g., Sha48 and SW49] could be used to provide an explicit expression for the likelihood function $L(y_{exp}|S)$. The correct form of L certainly depends on the exact nature of the information available for the estimation procedure. Here it is assumed that a collection of data $y_{exp}$ and its associated covariance matrix V are given, nothing more or nothing less. Under these conditions, it is found that

$$\begin{aligned} L(y_{exp},V|S) &\propto \exp[-(1/2)Q] \\ &= \exp[-(1/2)(y_{exp}-y_{calc})^{+}V^{-1}(y_{exp}-y_{calc})]. \end{aligned} \tag{10}$$

The quadratic form, $Q = (y_{exp}-y_{calc})^{+}V^{-1}(y_{exp}-y_{calc})$, which appears in the argument for the exponential function of Eq. (10), is positive so long as V is positive definite, which it always should be if it is properly selected to represent physical uncertainties [GS88]. Consequently, L is maximized when Q is a minimum. Note also that we have now included V explicitly as an argument of L to emphasize the point that the form of L, as given by Eq. (10), is valid <u>only</u> when the available information consists exclusively of $y_{exp}$ and V.

We could choose to proceed with the calculations indicated by Eqs. (6) and (7), using Eq. (10) along with an appropriate normalization constant to represent $p(S|y_{exp})$. When is this a practical approach? If Y and S are linearly related and V is independent of S, then the a posteriori probability distribution will be truly normal and the Bayesian solutions for S and $V_S$ can be determined easily. They are given by Eqs. (2) and (3), and are therefore entirely consistent with the least-squares results. Otherwise, the required calculations could be prodigious, even when handled numerically using a high-speed computer. This is particularly true if the data array $y_{exp}$ and parameter set S are both large. The Bayesian approach would be very unappealing in such situations, perhaps impossible. However, there is an approximation which often is made to alleviate the situation. It assumes that an acceptable solution for S is the one with the greatest likelihood, i.e., the one which maximizes L. This solution is called the most probable or maximum-likelihood solution, and it is conceptually quite different from the true Bayesian solution which must be obtained by evaluating expected values according to Eqs. (6) and (7). Referring to Eqs. (1) and (10), it is evident that the approach resulting from this maximum-likelihood approximation is entirely equivalent to the least-square method since it requires Q to be a minimum. This approximation will lead to results which can be quite different numerically from the Bayesian values when the a posteriori distribution is significantly different from a normal distribution, e.g., if it is strongly skewed. Our understanding of the particular anomaly which is described in Section 2 will ultimately evolve from a thorough appreciation of this important point.

Summarizing, we see that in classical statistics the least-squares method is simply a postulated procedure for generating parameter estimates which possess several desirable properties. However, it is also evident that this method can be derived more fundamentally from an application of Bayes' theorem and the principle of maximum entropy, provided that, when forced to do so, we are able to justify resorting to the maximum-likelihood approach as an approximation to the explicit calculation of expected values. In view of these various considerations, it is easy to see why evaluators are reluctant to abandon the least-squares method, in spite of some of the difficulties which it can pose in practical applications.

## 2. THE ANOMALY KNOWN AS "PEELLE'S PERTINENT PUZZLE"

R.W. Peelle, in an informally distributed memorandum, has suggested that the manner in which least-squares methodology is conventionally applied to data evaluation sometimes tends to produce rather strange results [Pee87]. He questioned whether they ought to be believed if they contradict intuition. This issue has come to be known as Peelle's Pertinent Puzzle (PPP). Here is the puzzle as expressed in Peelle's own words from the original memorandum:

*"Suppose we resolve to combine experimental values using weighted least-squares adjustment with the weight matrix taken to be the inverse of the full variance-covariance matrix of the input data. This approach is encouraged to achieve the minimum-variance result and to obtain an output variance-covariance matrix that properly reflects the input data base."*

*"We are required to obtain the weighted average of two experimental results for the same physical quantity. The first result is 1.5, and the second 1.0. The full covariance matrix of these data is believed to be the sum of three components. The first component is fully correlated with standard error <u>20%</u> of each respective value. The second and third components are independent of the first and of each other, and correspond to <u>10%</u> uncertainties in each experimental result."*

*"The result is a weighted average of about 0.87 ± 0.23, a value outside the range of the input values! Under what conditions is this the reasonable result that we sought to achieve by use of an advanced data reduction technique?"*

*"One's first reaction is to blame the non-intuitive result on the discrepant nature of the input data. This is not the reason, because the whole input data covariance matrix can be scaled up without changing its 'shape' until the data are consistent. The input data are indeed strange, but a similar if muted effect would occur if a less-odd example of this type were offered."*

*"My own first reaction to this puzzle was to have confidence in the method, and imagine experiments that would yield such a data covariance matrix and for which the odd-looking result is reasonable."*

*"However, what if I try to justify such a covariance matrix for an experiment in which two effectively independent 10% activation experiments are performed using the same foil for which the mass is known to 20%? For that case we feel strongly that the best answer is 1.25, with an uncertainty greater than 20% that depends on how any discrepancy is resolved in the activation part of the experiment. If that feeling is correct, evaluators need a very clear way to know that this combination problem should not be analyzed using the technique described in the first paragraphs above."*

*"My own effort to explain the weighted least-squares evaluation technique do not yet indicate the level of data reduction at which one must apply the full least-squares data combination to obtain unambiguous results."*

The issue that Peelle raised had not gone unnoticed by the community of nuclear data evaluators. It was certainly observed as early as 1978 by Perey [Per78]. Perey showed that in averaging two partially correlated data values one can expect the evaluated result to fall below either data value if the elements of the data covariance matrix V satisfy one or other of the relationships $V_{11} < V_{12}$ or $V_{22} < V_{12}$ (note that $V_{21} = V_{12}$). <u>This is indeed quite proper if one accepts the given absolute values of the covariance matrix elements as being correct</u>. The same point was later echoed by W. Mannhart [Man81]. This result follows from the estimator formula obtained by an application of the basic formalism, as given by Mannhart in Eq. (8), Appendix 2 of his report. This formula is reproduced here for convenience:

$$y = [(V_{22}-V_{12})y_1 + (V_{11}-V_{12})y_2]/[(V_{11}-V_{12}) + (V_{22}-V_{12})], \tag{11}$$

where, y is the evaluated value, $y_1$ and $y_2$ are the data values, and $V_{11}$, $V_{12} = V_{21}$ and $V_{22}$ are the elements of the data covariance matrix. If $V_{11} < V_{12}$ or $V_{22} < V_{12}$, then <u>negative weighting factors</u> will obviously result, with a depressing effect on the evaluated value y. If the data are uncorrelated, i.e., if $V_{12} = V_{21} = 0$, then one can easily show that Eq. (11) takes the form:

$$y = [(V_{11}^{-1})y_1 + (V_{22}^{-1})y_2]/(V_{11}^{-1} + V_{22}^{-1}). \tag{12}$$

Since $V_{11} > 0$ and $V_{22} > 0$, y will always fall between $y_1$ and $y_2$ for uncorrelated data. If $V_{22} > V_{11}$, then y will lie closer to $y_1$, i.e., $y_1$ will be more heavily weighted than $y_2$. The opposite will be true if $V_{22} < V_{11}$.

Our first step in examining PPP was to see if we could reproduce Peelle's result, 0.87 ± 0.23. In trying to do this we had to overcome a modest problem. But, first, let us list the formulas which we found to be useful in this analysis. They correspond, respectively, to

Eqs. (114), (183) and (184) from a report by Smith [Smi81]:

$$\chi^2 = (y_1-y)^2 W_{11} + 2(y_1-y)(y_2-y)W_{12} + (y_2-y)^2 W_{22}, \tag{13}$$

$$E_y^{-2} = (1-C_{12}^2)^{-1}(E_1^{-2} + E_2^{-2} - 2C_{12}E_1^{-1}E_2^{-1}), \tag{14}$$

$$y = [(E_1^{-2} - C_{12}E_1^{-1}E_2^{-1})y_1 + (E_2^{-2} - C_{12}E_1^{-1}E_2^{-1})y_2]/ \\ [(E_1^{-2} - C_{12}E_1^{-1}E_2^{-1}) + (E_2^{-2} - C_{12}E_1^{-1}E_2^{-1})], \tag{15}$$

where $E_1^2 = V_{11}$, $E_2^2 = V_{22}$, $V_{12} = V_{21} = C_{12}E_1E_2$. The elements of the correlation matrix C satisfy the properties $C_{11} = C_{22} = 1$ and $C_{12} = C_{21}$. Furthermore, $W_{11}$, $W_{12} = W_{21}$ and $W_{22}$ are defined as the elements of W, the inverse of the covariance matrix V (i.e., $W = V^{-1}$).

We proceed now to consider PPP. Our initial interpretation of this particular problem was that there exists a 20% fully-correlated error and two distinct 10% random error components for each of the two data points. Furthermore, we understood that the elements of the covariance matrix V were to be calculated by applying these percent errors to the measured values, i.e., that

$$V_{11} = E_1^2 = (0.1y_1)^2 + (0.1y_1)^2 + (0.2y_1)^2, \tag{16}$$

$$V_{12} = V_{21} = C_{12}E_1E_2 = (0.2y_1)(0.2y_2), \tag{17}$$

$$V_{22} = E_2^2 = (0.1y_2)^2 + (0.1y_2)^2 + (0.2y_2)^2, \tag{18}$$

where $y_1 = 1.5$ and $y_2 = 1.0$.

With these values, an analysis based on Eqs. (13)-(15) yielded the result: $y = 1.00$, $E_y = 0.24$ and $\chi^2 = 3.33$. We were very puzzled by the fact that this result is quite different from the one reported by Peelle. Subsequently, we were able to resolve this discrepancy by assuming that actually there was only one random error component of 10% for each data point, in addition to the 20% fully-correlated error, in the problem that Peelle had originally envisioned. Thus, Eqs. (16) - (18) ought to be replaced by the following expressions:

$$V_{11} = E_1^{\ 2} = (0.1y_1)^2 + (0.2y_1)^2, \tag{19}$$

$$V_{12} = V_{21} = C_{12}E_1E_2 = (0.2y_1)(0.2y_2), \tag{20}$$

$$V_{22} = E_2^{\ 2} = (0.1y_2)^2 + (0.2y_2)^2. \tag{21}$$

With these values, we obtained the result: $y = 0.88$, $E_y = 0.22$ and $\chi^2 = 5.88$. This is in good agreement with what Peelle reports. This minor ambiguity in the precise interpretation of Peelle's problem does not detract in any way from the underlying issue associated with PPP, so it should not be a cause for concern.

Peelle makes no mention of the chi-square statistic ($\chi^2$) in his memorandum; however, this important parameter ought to be considered in order to quantify our confidence in the results. Two data points ($n = 2$) averaged to produce a single result ($m = 1$) corresponds to one statistical degree of freedom ($f = n-m = 1$). The expected value of $\chi^2$ is f for consistent, normally distributed data (i.e., $< \chi^2 >/f = 1$). Since the values $\chi^2 = 3.33$ or $5.88$ (depending on the particular interpretation of the problem) are both significantly larger than unity, this implies that we should have a rather low degree of confidence in the evaluated results. Only by enhancing the input errors $E_1$ and $E_2$ by the factor $(\chi^2/f)^{1/2}$ or, alternatively, by increasing $E_y$ by the same factor, is it possible to achieve a satisfactory level of confidence in the evaluated result. Of course, as Peelle correctly points out, this will not alter the magnitude of the evaluated parameter.

We believe that Peelle indeed has raised a very fundamental issue which is of crucial importance to the practice of nuclear data evaluation. Since it is obviously undesirable to abandon the process of data evaluation by the least-squares method, we are of the opinion that the resolution of this problem must involve determination of a viable alternative method for deriving the elements of those covariance matrices which are required for least-squares parameter estimation. Our suggestion for achieving this objective is thoroughly discussed in Section 3.

## 3. A SUGGESTED PROCEDURE FOR DEALING WITH THE PUZZLE

We have observed that the difficulty discussed in Section 2 (for convenience we will continue to refer to it as "PPP", even in a more general context) can be alleviated if the required experimental data covariance matrix elements which are needed for least-squares analysis are computed from percent errors and best-estimate values for those parameters being evaluated, rather than directly from the experimental data as is usually done (e.g., in Peelle's example). Since these best-estimate values are not known in advance, in most instances it is necessary to provide reasonable a priori determinations of their magnitudes and then iterate the various solution steps until acceptable convergence is achieved. In this section we will discuss the reasoning which led us to make this suggestion.

As discussed in an earlier report by Smith [Smi82a], most experimental determinations of nuclear quantities (e.g., cross sections) involve extensive manipulations of raw data. Final experimental results are derived through a composition of the actual measured quantities and (usually) a number of computed corrections. Often these procedures can be formally represented by analytical models in which the experimental quantities $y_i$ are expressed as products of several primary measured or calculated factors. This is expressed by the formula

$$y_i = \prod_{\lambda=1}^{\Lambda} F_{i\lambda}. \tag{22}$$

The various factors $F_{i\lambda}$ will have associated uncertainties which may be correlated between the various data points (index "i"), but usually are <u>not</u> correlated between the various experimental attributes which enter into the product (index "$\lambda$"). These attributes tend to be independent because of their distinct physical natures (e.g., count rates, detector calibrations, radiation absorption or scattering parameters, etc.). Consequently, overall uncertainties for the quantities $y_i$ can be derived in terms of the uncertainties for the individual factors $F_{i\lambda}$ by means of error-propagation, according to the well-known formula

$$(E_{yi}/y_i)^2 = \sum_{\lambda=1}^{\Lambda} (E_{Fi\lambda}/F_{i\lambda})^2. \tag{23}$$

The essential point is that fractional errors (or equivalent percent errors) in the individual multiplicative components of Eq. (22) are propagated via addition in quadrature according to Eq. (23) to generate overall fractional errors for the derived results. <u>Therefore,</u>

it seems quite reasonable to consider fractional errors (or equivalent percent errors) instead of absolute errors in the data as a means for expressing the confidence we have in the experimental values with which we must deal in performing an evaluation. When comparing various data sets which purport to result from measurements of the same physical quantities, it is suggested that one should refer to percent errors and their correlations in deciding how to weight the given values for evaluation purposes.

If we choose to work with percent errors in the data, and also to preserve the least-squares evaluation method (with its requirement for an input data covariance matrix) then the obvious choice is to compute the matrix elements by multiplying a priori estimated values for the parameters by the corresponding fractional errors (equivalent to the percent errors). If, instead, we try to compute these quantities using percent errors and actual experimental values, we are likely to end up with inappropriate weighting factors for the data. Peelle has demonstrated that the consequences can be quite significant if there are strong correlations and inconsistencies in the data [Pee87]. One can illustrate rather simply the fallacy of utilizing experimental data to calculate weighting factors directly. Consider the following situation: Suppose that we make a series of independent observations of counts-per-minute (N) recorded by a particular radiation detector (under conditions of stochastic equilibrium). Let us suppose that the values we obtain are 83, 102, 99, 115, 107, ... etc. The $\sqrt{N}$ rule for statistical errors is a well-known result from Poisson statistics [e.g., Mar71 and Zeh70], so we might be tempted to assume that the corresponding errors in our observations are 9.11, 10.1, 9.95, 10.7, 10.3, ... etc. Application of the least-squares method would suggest that we then weight our data by the factors 0.0120, 0.00980, 0.0101, 0.00870, 0.00935, ... etc., respectively. These weightings are quite distinct, so the implication that this should be done is nonsense. In our quest to obtain a best estimate for N, we are justified in treating each of these values with equal confidence because we understand, intuitively, that the measured values (i.e., the data) are acquired through sampling from a common, universal probability distribution. Our chosen value for the error associated with these data represents an estimate of the standard deviation of this distribution. Let us assume that the best value is actually 100 counts-per-minute. The proper absolute error to assume for each value would then be 10 (corresponding to 10%). This result is consistent with our suggestion that the covariance matrix for the data set be computed using percent errors and the best-estimate value. Several independent measurements of the same physical quantity, each made with the same degree of confidence (as expressed by the percent error), can be viewed statistically as sampling from an event space in accordance with a well-defined probability function. The variances associated with the individual measurements should then be identical, even though the values actually obtained will vary. This statement is absolutely true if each of the quantities is considered to be governed by the same type of probability distribution (e.g., a normal distribution), and it should be approximately true even if each trial is governed by a different type of probability distribution.

It is unfortunate that this approach leads us inevitably to generate a covariance matrix V for the data which depends upon plausible a priori knowledge of the solution S. In fact, this is not the only complication we face. We shall see below that, in many applications of the least-squares method, $y_{exp}$ will also depend to some extent on S. To see how this can come about, we review the steps which are normally taken in performing an evaluation based entirely on experimental data:

Step 1: A comprehensive set of all the relevant data is gathered from the literature.

Step 2: Based on information provided in the documentation for each data set, on supplementary information available to the evaluator, and on the evaluator's own good judgment, some adjustments may be made to the original results (including the uncertainties) in order to insure that they indeed do correspond to what it is claimed that they represent, and to insure that appropriate data weighting factors are generated for the evaluation process.

Step 3: The evaluator then develops a linear regression model (i.e., a design matrix A) which relates the adjusted information to the parameters he has selected to evaluate. If the data base is large, he also may find it convenient to reduce the size of the arrays which must be manipulated, through the averaging of equivalent quantities, even though this step can lead to biased estimators [Poe81].

Step 4: Finally, the least-squares procedure is applied to obtain the evaluated results and their uncertainties.

Let us represent the experimental data emerging from Step 2 by the array $y_{data}$. For mathematical convenience, we choose to represent symbolically by the matrix operator F(S) all those procedures in Step 3 which are taken to reduce the number of raw data points, to shift them to equivalent values at grid energies, etc. This is where the dependence of $y_{exp}$ on S can enter. Thus, $y_{exp} = y_{exp}(S) = F(S)y_{data}$. The array $y_{exp}$ will have a dimension which is often smaller than that for $y_{data}$, but under no circumstances will have a larger one. The covariance matrix $V = V(S)$ with which we are concerned in the least-squares evaluation procedure corresponds to $y_{exp}$, not to $y_{data}$.

Since V and $y_{exp}$ (possibly) depend upon S, it would appear that, in our derivation of estimators for S and $V_S$, we should re-write the least-squares condition of Eq. (1) in the form

$$Q(S) = [y_{exp}(S) - y_{calc}]^{+}W(S)[y_{exp}(S) - y_{calc}] = \text{minimum}, \tag{24}$$

where $W(S) = V^{-1}(S)$. Although not indicated explicitly here, it is clear from the discussion in Section 1 that $y_{calc}$ also depends on S. It is most inconvenient that the relatively simple expressions for S and $V_S$ which follow from Eq. (1), and are reflected in Eqs. (2) and (3), are no longer strictly valid when we try to satisfy the condition of Eq. (24) by adjusting S. The essence of our proposition is that this problem be circumvented by replacing Eq. (24) by the condition

$$Q(S,S_0) = [y_{exp}(S_0) - AS]^{+}W(S_0)[y_{exp}(S_0) - AS] = \text{minimum}. \tag{25}$$

Eq. (25) is an approximation to Eq. (24) which is mathematically quite tractable and, hopefully, provides reasonable solutions to practical evaluation problems. $S_0$ in Eq. (25) represents a reasonable a priori estimate of S. We see that a solution (in terms of $S_0$) can be obtained from the expressions

$$S = G(S_0)y_{exp}(S_0) = G(S_0)F(S_0)y_{data}, \tag{26}$$

$$V_S(S_0) = [A^{+}W(S_0)A]^{-1}, \tag{27}$$

$$G(S_0) = V_S(S_0)A^{+}W(S_0). \tag{28}$$

Before we proceed any further, let us pause to examine what happens when we apply our suggested approach specifically to Peelle's example (see Section 2). We will denote the constant parameter which is required to determine the elements of the covariance matrix V by $S_0$. If we substitute $S_0$ for $y_1$ and $y_2$ in Eqs. (19) - (21) we obtain: $V_{11} = 0.05S_0^2$, $V_{12} = V_{21} = 0.04S_0^2$ and $V_{22} = 0.05S_0^2$. It is seen that the unknown factor $S_0^2$ conveniently cancels in the least-squares analysis for this problem, and we are led unambiguously to the solution $S = 1.25$, $V_S = 0.0784$ and $\chi^2/f = 4.00$. This is a reasonable result even though the confidence level is low due to the large value of $\chi^2/f$. This deficiency could be remedied easily by multiplying $V_S$ by the factor 4.00. The fortuitous cancellation of the unknown factor $S_0^2$

in this analysis is characteristic of what happens when only a single parameter is evaluated, and a linear model relates the data to that parameter. Otherwise, iteration is required and matters are more complicated.

The quantity $S_0$ which we have introduced in this formalism appears to be rather arbitrary. Earlier, we referred to it as simply a reasonable a priori estimate of S, knowing full well that usually we are not aware of just how to make an optimal choice for $S_0$ at the outset. Since the a posteriori solution for S which is derived from Eq. (26) is clearly dependent upon the ad hoc choice of an a priori (on $S_0$), we are inevitably led to pursue an iterative procedure in order to seek convergence toward the best possible final solution for S which the data (and the present approach) will allow. Each step of this iteration procedure incorporates the familiar least-squares methodology which is embodied in all of the computer codes that are currently in widespread use for data evaluation [e.g., Poe81 and Smi82b]. Of course, if the final solution to every problem has to be obtained through iteration, we must be very concerned about the whole matter of convergence. We will now introduce some notation to simplify the following discussion. If we let $H(S) = G(S)F(S)$ and, for convenience, $y = y_{data}$, Eq. (26) can be written in the simplified form

$$S = H(S_0)y = G(S_0)F(S_0)y. \tag{29}$$

The iteration procedure can then be represented symbolically by the following formulas

$$\begin{aligned} S_1 &= H(S_0)y, \\ S_2 &= H(S_1)y, \\ &\;\vdots \\ S_{i+1} &= H(S_i)y \qquad (i \geq 2), \end{aligned} \tag{30}$$

where $S_0$ is our initial estimate of the solution S. The iteration procedure is continued until $S_{i+1} \approx S_i$, that is until the difference between $S_{i+1}$ and $S_i$ is as small as we like. Then we can say that an acceptable degree of self consistency has been achieved.

A very important consideration is whether this iteration process will actually converge to a well-defined solution S? Furthermore, to be viable, this procedure should lead to an S which is independent of or, at worst, only weakly dependent on, the initial choice of a priori $S_0$! We shall now derive a condition for convergence. To do this we note that $S_0 = S + (S_0 - S) = S + dS_0$. Referring to Eq. (30), the first stage of iteration yields

$$
\begin{aligned}
S_1 &= H(S_0)y = [H(S + dS_0)]y \\
&= [H(S) + dH(S)]y \\
&\approx H(S)y + [(dH/dS)dS_0]y \\
&= S + [(dH/dS)y]dS_0 .
\end{aligned}
\tag{31}
$$

We have employed a first-order Taylor series expansion in order to obtain an approximate expression for the matrix operator $H(S + dS_0)$. Care must be taken to preserve the order for those matrices appearing in Eq. (31) because matrix multiplication is usually not commutative. If we assume that y has dimension (n,1) and both S and $dS_0$ have dimension (m,1), then the expressions above are consistent if and only if H and dH have dimension (m,n) and [(dH/dS)y] has dimension (m,m). The quantity [(dH/dS)y] must be interpreted in the following way: It represents a matrix comprised of elements which are actually sums of products of components from y and partial derivatives of the elements of H with respect to the components of S. For convenience, let T = [(dH/dS)y]. Then, the elements of T have the form

$$
T_{kl} = \sum_{\alpha=1}^{n} [\partial H_{k\alpha}/\partial S_l]\ y_\alpha \qquad (k,l = 1,m), \tag{32}
$$

where $S_l$ is the l-th component of S, $y_\alpha$ is the $\alpha$-th component of y and $H_{k\alpha}$ is the $(k,\alpha)$-th element of H. It is easy to show that the second stage of iteration yields

$$
\begin{aligned}
S_2 &\approx S + [(dH/dS)y][(dH/dS)y]dS_0 \\
&= S + [(dH/dS)y]^2 dS_0 = S + T^2 dS_0 .
\end{aligned}
\tag{33}
$$

Continued iteration, according to Eq. (30), leads to the result

$$
S_i \approx S + [(dH/dS)y]^i dS_0 = S + T^i dS_0 \qquad (i > 2). \tag{34}
$$

Note that $T^i$ has dimension (m,m) for all $i > 0$. Iteration is terminated when the evaluator is convinced that further iteration will not change the current evaluated results significantly. We note that the matrix $T = [(dH/dS)y]$ depends only on S and the data, not on $S_0$. It is also clear that the iteration procedure indicated above will converge (i.e., $S_i \rightarrow S$ as $i \rightarrow \infty$) so long as

$$\lim_{i \rightarrow \infty} T^i = [0], \tag{35}$$

where [0] symbolizes a "zero" matrix with dimension (m,m). Let $\tau$ be defined by

$$\tau = \max\{|T_{kl}|\} \qquad (k,l = 1,m). \tag{36}$$

Thus defined, $\tau$ represents the maximum magnitude of any of the elements of T. It can then be shown that a sufficient (though not necessary) requirement for the condition reflected in Eq. (35) to be satisfied is that

$$\lim_{i \rightarrow \infty} (m^{i-1}\tau^i) = 0. \tag{37}$$

Eq. (37) represents a much stronger constraint on the magnitude of $\tau$ than the simple inequality $0 < \tau < 1$. Under the conditions indicated above, convergence is achieved without regard to $dS_0 = (S_0 - S)$, i.e., to the initial choice of $S_0$. Of course, the assumption that the magnitude of $dS_0$ is usually fairly small prompted us to consider employing a first-order Taylor series expansion in the first place. Furthermore, we shall see that the magnitude of $dS_0$ does influence the rapidity of convergence.

The success of our method depends heavily on achieving convergence, as described above. We now offer some crude arguments which support our conjecture that rapid convergence will usually be achieved in practical situations. We recall that $H(S) = G(S)F(S)$. An evaluator is free to choose an energy-grid structure for his evaluation which minimizes the sensitivity of F to S considerably, i.e., one for which F is so weakly dependent upon S that we can treat it as being nearly a constant matrix. Then $H(S) \approx G(S)F$. The sensitivity of H to S will then derive from the sensitivity of G to S. It is clear from Eqs. (27) and (28) that the dependence of G on S can be traced to a dependence of V on S. It is also evident that G involves products of elements of $V^{-1}$ (and therefore of V, indirectly) through the factor $V_S = (A^+V^{-1}A)^{-1}$. So, we are led to suspect that there will be a tendency toward cancellation of those terms in G which might otherwise depend strongly on S and, thus, that the sensitivity of G to S will also tend to be quite modest. G is totally independent of S for those evaluation problems involving only a single parameter ($m = 1$). Then, as we have seen, iteration is not required to obtain the solution, although S should be derived first in order to obtain the final representation of the covariance matrix $V_S$ and $\chi^2$. Of course, this is the case in Peelle's example, as described in Section 2.

We will now examine the matter of minimum variance since that is an issue which is of interest for applications of the least-squares method within the framework of classical statistics (see Section 1). For a fixed value of $S_0$, the least-squares criterion insures minimum variance by virtue of the Gauss-Markov theorem (as amended by Aitken for correlated data), but this variance depends upon $S_0$. What happens to the quadratic form Q and to the variance as the iteration process progresses toward a solution? We first need to define what we mean by minimum variance for a multi-parameter problem. The generalized variance for the data set $y_{exp}$ is defined as det(V), i.e., the determinant of the covariance matrix V [Fis63]. Likewise, the generalized variance for the solution S is $\det(V_S)$. We would like to have both Q and $\det(V_S)$ decrease as a result of the iteration process. In practice, this can be tested fairly simply by invoking some well-known properties of the associated matrices. First, we note that V can be written in the form $V = U\ C\ U$, where C is the correlation matrix and U denotes the diagonal matrix with elements

$$U_{ij} = V_{ii}^{1/2}\ \delta_{ij}, \tag{38}$$

and $\delta_{ij}$ is the Kronecker delta function. It can be shown easily that

$$W = V^{-1} = U^{-1}\ C^{-1}\ U^{-1}. \tag{39}$$

The inverse matrix $Z = U^{-1}$ is easy to determine. Since U is diagonal, the elements of its inverse Z are given quite simply by the expression

$$Z_{ij} = (1/V_{ii}^{1/2})\ \delta_{ij}. \tag{40}$$

Thus, the inverse of V can be derived readily from the inverse of C. We will now prove that C does not vary during the iteration process and, therefore, only one matrix inversion operation actually needs to be carried out in the analysis, no matter how many iterative steps are required to obtain a good solution for S! This is an important result because the inversion of matrices can be an arduous task when the matrix dimensions are large. It can be shown, based on a discussion by Smith [Smi87b], that the elements of V can be written as

$$V_{ij} = S_i S_j \sum_{\lambda=1}^{\Lambda} B_{\lambda ij}(0.01P_{\lambda i})(0.01P_{\lambda j}), \tag{41}$$

where $P_{\lambda i}$ is the percent error in the element $S_i$ of S and $B_\lambda$ is a micro-correlation matrix (both associated with uncertainty attribute $\lambda$). Since the elements of C are given by

$$C_{ij} = V_{ij}/(V_{ii}V_{jj})^{1/2}, \tag{42}$$

it is clear that the dependence upon S cancels. It also happens that the correlation matrix $C_S$ which corresponds to the solution covariance matrix $V_S$ is also unchanging with successive iterations. Furthermore, since

$$\det(V_S) = \det(C_S)(\prod_{i=1}^{m} V_{Sii}), \tag{43}$$

we can follow the behavior of the generalized variance for S by simply computing the product of the diagonal elements of $V_S$ (squares of errors) at each stage of the iteration, noting that $\det(C_S)$ is constant. Actually, for this purpose it is more meaningful to compute the product of the enhanced variances, i.e., the quantities $(\chi^2/f)V_{Sii}$. The technique is discussed further in Example 1 of Section 4.

Let us turn now to a consideration of the practical aspects of iteration. Our experience in applying this method has involved utilization of a recently revised version of the GMA code package (consisting of DATGMA and GMA). The GMA code package was first developed by Poenitz [Poe81]. The transformation operation represented by F(S) is embodied in DATGMA while that for G(S) is carried out by GMA. We have found that it is desirable to pursue a double-iteration procedure, as shown schematically in Fig. 1. Outer iteration involves application of both DATGMA and GMA while inner iteration involves only GMA.

A few words are in order here concerning the matter of adjusting data to selected grid energies and then averaging them to reduce the size of the experimental array to be handled in the evaluation. Such a procedure can lead to some bias in the estimators, but this small price may be more than compensated by the advantage to be gained by reducing the sizes of matrices V which have to be inverted in applications of the least-squares method [Poe81]. Smith [Smi87a] has described a rigorous procedure for "collapsing" data. However, we find that the following approximation to this method often is quite satisfactory, and it is considerably easier to implement in computer programming. For this reason, we have utilized this approach in our revision of the GMA evaluation code package. The technique involves collecting all those data values from a particular experiment which end up being shifted to a particular energy grid, and then averaging them in accordance with weight factors based only on percent random errors. This procedure reduces random error while retaining the common systematic error. A simple example will suffice to demonstrate the utility of this approach. Suppose that two values, 100 and 91, are to be averaged to form a single result. The first value has 2% random error while the second has 3% random error. Each has a common systematic (100% correlated) error of 5%. The exact solution (based on an application of our method and consideration of the complete covariance matrix) is 97.23 ± 5.12 (5.3% total error). The present approximation leads to 97.23 with a 5.3% total error (1.7% attributed to random error and 5% to systematic error). These results are identical to the quoted accuracy.

The issue manifested in PPP also has implications in more complex Bayesian data adjustment applications [Smi82b]. In these problems it is generally assumed that the previously existing information is represented by an evaluated parameter set (called the a priori set) and its corresponding covariance matrix. In data adjustment it is desired to update this evaluation by the consideration of new data. The usual practice is to assume that the new information is independent of the old (uncorrelated to the a priori). The method of generalized least squares is commonly used to analyze these problems. In accordance with present considerations, the covariance matrix for the a priori may present a problem if it was produced by an evaluation in which the then-available data were improperly weighted (a distinct possibility in view of the discussions in this report). An obvious difficulty arises in deciding how to construct the covariance matrix for the new data sets. One choice would be to base it upon percent errors and the a priori parameter set. Iteration would be carried out, and successive steps in the process would lead to adjustments in this covariance matrix. However, it seems to us that this procedure would tend to jeopardize the assumed independence of new and prior information! The resolution of this issue is obviously a worthy topic for future investigations.

## 4. SOME EXAMPLES WHICH DEMONSTRATE THE PROCEDURE

This section is devoted to a discussion of two practical applications of the method we have described in this report. Both are examples involving real evaluations of nuclear data.

Example 1:

The first example we shall consider involves a differential cross section evaluation for the reaction $^{115}In(n,n')^{115m}In$. The details of the evaluation process have been reported by Smith et al. [Smi+90]. This evaluation is based entirely on experimental data. Our initial choice for the a priori $S_0$ (which we shall refer to as the "good" a priori) came from ENDF/B-V [BNL79]. A revised version of the GMA code package was employed in this analysis. It facilitated the application of our new method involving iteration, in which covariance matrices are calculated at each stage using percent errors and the current best-estimate values of those parameters to be estimated. Excellent convergence to the final solution S was obtained after the following sequence of steps: [DATGMA(1) + GMA(3) + DATGMA(1) + GMA(2)]. The values in (...) show the number of successive applications for the indicated code. To test the convergence properties further, we made a "wild" choice for the initial a priori $S_0$. The cross section was taken to be 1 mb at each grid energy (we shall refer to this as the "bad" a priori). Very good convergence to the same final solution S was ultimately obtained with the following sequence of iterative steps: [DATGMA(1) + GMA(3) + {DATGMA(1) + GMA(2)}(4)]. The notation {...}(4) indicates that the procedure represented by {...} was repeated four times. Progress in convergence toward the final solution, as measured by the value of chi-square ($\chi^2$), is traced in Fig. 2 for both choices of a priori $S_0$. The solution clearly converges more rapidly for the "good" a priori than for the "bad" a priori, but even for the latter the final solution S was obtained after surprisingly few iterative steps. We have not exhibited the corresponding progress that was made by the enhanced generalized variance, $(\chi^2/f)^m \det(V_S)$, as we proceeded through the iteration steps. Since the $\det(C_S)$ is constant, this could be traced by computing the product of the enhanced errors for the elements of the solution S. It was found that this quantity decreased rapidly during the first few iterative steps and subsequently stabilized as the final solution was approached. Thus, it appears that iteration leads not only to reduced $\chi^2$, but also to a significant decrease in the generalized variance.

The evaluation was also carried out using the original version of GMA, in which absolute covariance matrix elements are derived from percent errors and actual data values (old method). There were no iterations involved in this analysis. The results are compared in Figs. 3 - 5. It is clear that these two distinct approaches lead to quite different results. The evaluation produced by the new method yields a calculated (C) spectrum-average cross section for $^{252}Cf$ spontaneous-fission neutrons which is about 4% closer to the evaluated experimental (E) integral value than is obtained by using the ENDF/B-V differential cross section representation, i.e., C/E is closer to unity [Smi+90]. On the other hand, the old method yields a solution which produces an inferior C/E comparison, relative to the ENDF/B-V differential cross section representation! This outcome provided us with good motivation to embrace this new approach in preference to the alternatives.

Example 2:

Our second example involves an evaluation of 14-MeV cross sections. It is based on the work of Evain et al. [ESL85]. The reaction we have considered is $^{59}Co(n,p)^{59}Fe$, as discussed on pp. 115 - 120 of that report. We have drawn the specific information required for the present analysis from Table 2, p. 118, of the report. First, we analyzed the available information using the old method. Our resulting solution for the evaluated 14.7-MeV cross section is 56.6 ± 3.14 mb, with $\chi^2/f = 8.53$. The enhanced error is therefore 9.17 mb (16.2% error). This agrees quite well with the result reported by Evain et al. [ESL85]. We then re-analyzed the same information using our new approach. An a priori estimate $S_0$ of the solution was not required here to obtain S because only a single parameter is involved. However, the a posteriori solution is needed in order to evaluate the uncertainty and $\chi^2/f$. Our result from this analysis is 70.2 ± 3.67 mb, with $\chi^2/f = 6.16$. The enhanced error is therefore 9.10 mb (13.0% error). The difference in these two results is obviously quite large. However, the values do overlap within their uncertainties. Notice that the evaluated result from the new method is larger than the one obtained by the old method. We recall that Peelle's comments were prompted by his observation that evaluated results often appear to be lower than might be intuitively expected.

These two examples clearly demonstrate that the results from evaluations can be strongly influenced by the method used to compute the data covariance matrix. Naturally, this raises questions about the reliability of some of the results generated by earlier evaluation efforts, e.g., those produced by the work of Evain et al. [ESL85].

# 5. IMPLICATIONS FOR NUCLEAR DATA EVALUATION

Statistical data evaluation and data adjustment procedures have been used widely in nuclear data and nuclear technology applications for well over a decade. A number of codes have been employed for these purposes. Our knowledge of the specific procedures incorporated in these codes is derived mainly from our experience with the GMA package. As indicated above, those versions of the GMA-package codes which were originally used in this laboratory involved computation of data covariance information directly from the experimental data, a procedure which we now believe is very questionable as a result of the present investigation. It is evident from the preceding discussions that, in practice, this is not too serious an issue so long as the data errors are small, the scatter in these data is consistent with these errors, and there is no dominant systematic error component which is common to all of the data to be evaluated. This is likely to be the case for most evaluations of standards data and of certain other data, such as fission cross sections, where there exist abundant and relatively consistent data bases. However, even in such favorable instances, revealing questions need to be asked concerning the impact on earlier evaluations of specific procedures which were used to determine the requisite covariance matrices for the corresponding least-squares analyses.

The problems may be more severe in other contexts, e.g., where the data scatter considerably and there are significant discrepancies (as discussed in the examples of Section 4). Evaluators should examine the code packages which they now employ in their work to ascertain just how the required covariance matrices are determined. The emergence of this issue reinforces our long-held opinion that uninformed ("blind") applications of evaluation computer codes, especially those which are poorly documented and/or are attributable to authors other than the user(s), are guaranteed to lead to trouble!

Opinions tend to be divided within the nuclear data community concerning the desirability of iteration in performing evaluations. We find that iteration is essential for successful application of the method described in this report. This offers few conceptual difficulties (other than an increase of labor) for evaluations based entirely on data. However, as mentioned in Section 3, we foresee some conceptual problem areas associated with utilization of more complex procedures involving Bayesian priors in conjunction with the least-squares method (i.e., with generalized least-squares analysis).

## 6. CONCLUSIONS

We find that it is possible to resolve the issue raised in Peelle's Pertinent Puzzle (PPP) by deriving data covariance matrices used in least-squares estimation formulas from percent errors in the data and reasonable a priori estimated values for the parameters to be evaluated, rather than values acquired directly from the experimental data. So long as a single parameter is being evaluated, we can be sure that the specific value which is assumed for this a priori estimate is irrelevant, at least as it concerns determination of the solution itself. However, proper knowledge of this solution is ultimately needed in order to obtain a correct value for its uncertainty. The issue becomes more complicated in exercises involving the evaluation of more than one parameter. Here we find that reasonable a priori estimates for the parameter-array values are required, and iteration is necessary in order to arrive ultimately at an acceptable overall solution. A condition for convergence of this iteration procedure has been established, and it is found that adequate convergence can be anticipated with relatively few iterations for most applications of practical interest. The procedure appears to be quite forgiving (i.e., robust) insofar as the choice of initial estimates for the parameter values is concerned. The impact of PPP on nuclear data evaluations can be significant, especially for those cases involving discrepant data, large errors and significant correlations. Further study is needed to gain an understanding of how to properly approach generalized least-squares adjustment problems (i.e., those based on Bayesian methodology) within the framework of the concepts discussed in this report.

ACKNOWLEDGEMENTS

The authors are indebted to Dr. R.V. Peelle for calling "Peelle's Pertinent Puzzle" (PPP) to their attention, and to Drs. F. Froehner, H. Vonach and Z. Zhao for offering many helpful comments on this topic. This work has been supported by the U.S. Department of Energy, Energy Research Programs, under contract W-31-109-Eng-38. One of the authors (SC) is grateful to the Japan Atomic Energy Research Institute, Tokai, Japan, for sponsoring his stay at Argonne National Laboratory during that time interval when most of this work was carried out.

## REFERENCES


Ait57 A.C. Aitken, Statistical Mechanics, 8th Edition, Oliver and Boyd Press, London (1957).

BNL79 Evaluated Nuclear Data File -B, Version V (ENDF/B-V), National Nuclear Data Center, Brookhaven National Laboratory (1979).

ESL85 B.P. Evain, D.L. Smith and P. Lucchese, Report ANL/NDM-89, Argonne National Laboratory (1985).

Fis63 M. Fisz, Probability Theory and Mathematical Statistics, John Wiley and Sons, Inc., New York (1963).

Fro86 F. Froehner, Report KfK-4096, Kernforschungszentrum Karlsruhe, Karlsruhe, Federal Republic of Germany (1986).

GS88 L.P. Geraldo and D.L. Smith, Report ANL/NDM-104, Argonne National Laboratory (1988).

Jay68 E.T. Jaynes, "Prior Probabilities", IEEE Trans. Syst. Cybern. SSC-4, 227 (1968).

Jay73 E.T. Jaynes, "The Well-Posed Problem", Foundations in Physics 3, 477 (1973).

Jay76 E.T. Jaynes, "Confidence Intervals Versus Bayesian Intervals", Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science, eds. W.L. Harper and C.A. Hooker, Reidel Publishing Co., Dordrecht, Holland (1976).

Jay78 E.T. Jaynes, "Where Do We Stand on Maximum Entropy?" The Maximum Entropy Formalism, eds. R.D. Legine and M. Tribus, M.I.T. University Press, Cambridge, Massachusetts (1978).

Jay80 E.T. Jaynes, "Marginalization and Prior Probabilities", Bayesian Analysis in Econometrics and Statistics, ed. A. Zeller, North-Holland Publishing Company, Amsterdam, Holland (1980).

Man81 W. Mannhart, Report FMRB-84, ISSN 0341-6666, Physikalisch-Technische Bundesanstalt, Braunschweig, Federal Republic of Germany (1981).

Mar71 B.R. Martin, Statistics for Physicists, Academic Press, New York (1971).

Pee87 R.V. Peelle, "Peelle's Pertinent Puzzle", Informal memorandum dated October 13, 1987, Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A. (1987).

Per78 F.G. Perey, "Covariance Matrices of Experimental Data", Proceedings of the International Conference on Neutron Physics and Nuclear Data for Reactors and Other Applied Purposes, Harwell, U.K., OECD Nuclear Energy Agency, Paris, p. 104 (1978).

Poe81 GMA, "A Least-Squares Program for Nuclear Data Evaluation", V.P. Poenitz, Report BNL-NCS-51363, eds. B. Magurno and S. Pearlstein, Brookhaven National Laboratory, p. 249 (1981), as revised by M. Sugimoto (1987) and S. Chiba (1990).

Sha48 C.E. Shannon, Bell Syst. Tech. J. 27, 379 and 623 (1948).

Smi81 D.L. Smith, Report ANL/NDM-62, Argonne National Laboratory (1981).

Smi82a D.L. Smith, Report ANL/NDM-67, Argonne National Laboratory (1982).

Smi82b D.L. Smith, Report ANL/NDM-77, Argonne National Laboratory (1982).

Smi87a D.L. Smith, Nucl. Instr. and Methods in Phys. Research A257, 361 (1987).

Smi87b D.L. Smith, Nucl. Instr. and Methods in Phys. Research A257, 365 (1987).

Smi88a D.L. Smith, Report ANL/NDM-92, Argonne National Laboratory (1988).

Smi88b D.L. Smith, "Examination of Various Roles for Covariance Matrices in the Development, Evaluation, and Application of Nuclear Data", Nuclear Data for Science and Technology, ed. S. Igarasi, Saikon Publishing Company, Tokyo, p. 425 (1988).

Smi+90 A.B. Smith et al., Report ANL/NDM-115, Argonne National Laboratory (1990).

SW49 C.E. Shannon and W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, Urbana, Illinois (1949).

Zeh70 P.W. Zehna, Probability Distributions and Statistics, Allyn and Bacon, Inc., Boston (1970).

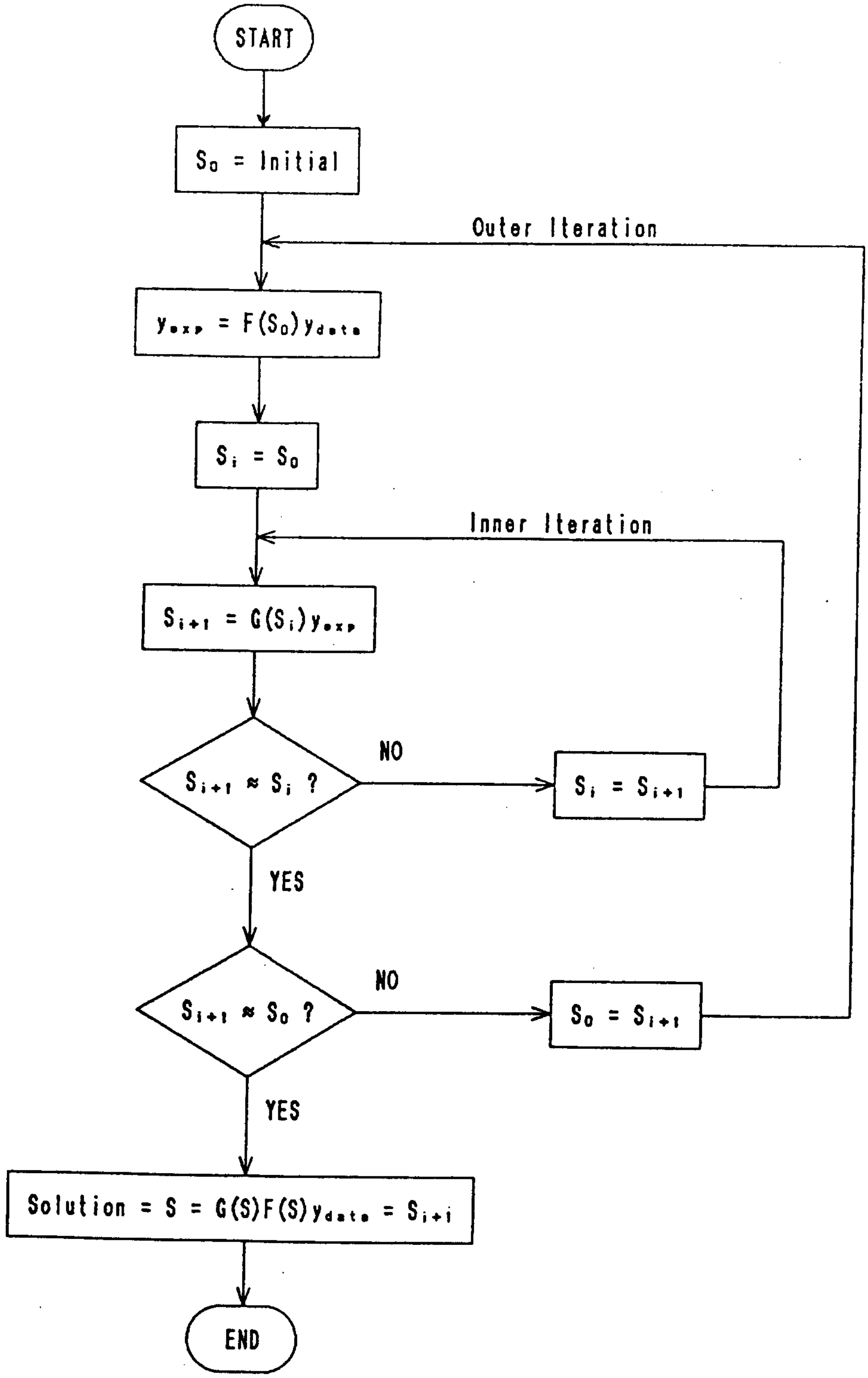



Figure 1: Flow chart for a typical evaluation procedure, as based on the present investigation. $S_0$ is the a priori parameter set to be used in data preparation (at the current outer iteration stage). $S_0$ is the a priori parameter set to be used in the least-squares analysis (at the current inner iteration stage), while $S_{i+1}$ is the a posteriori parameter set, which then serves as the a priori for the subsequent inner iteration step.

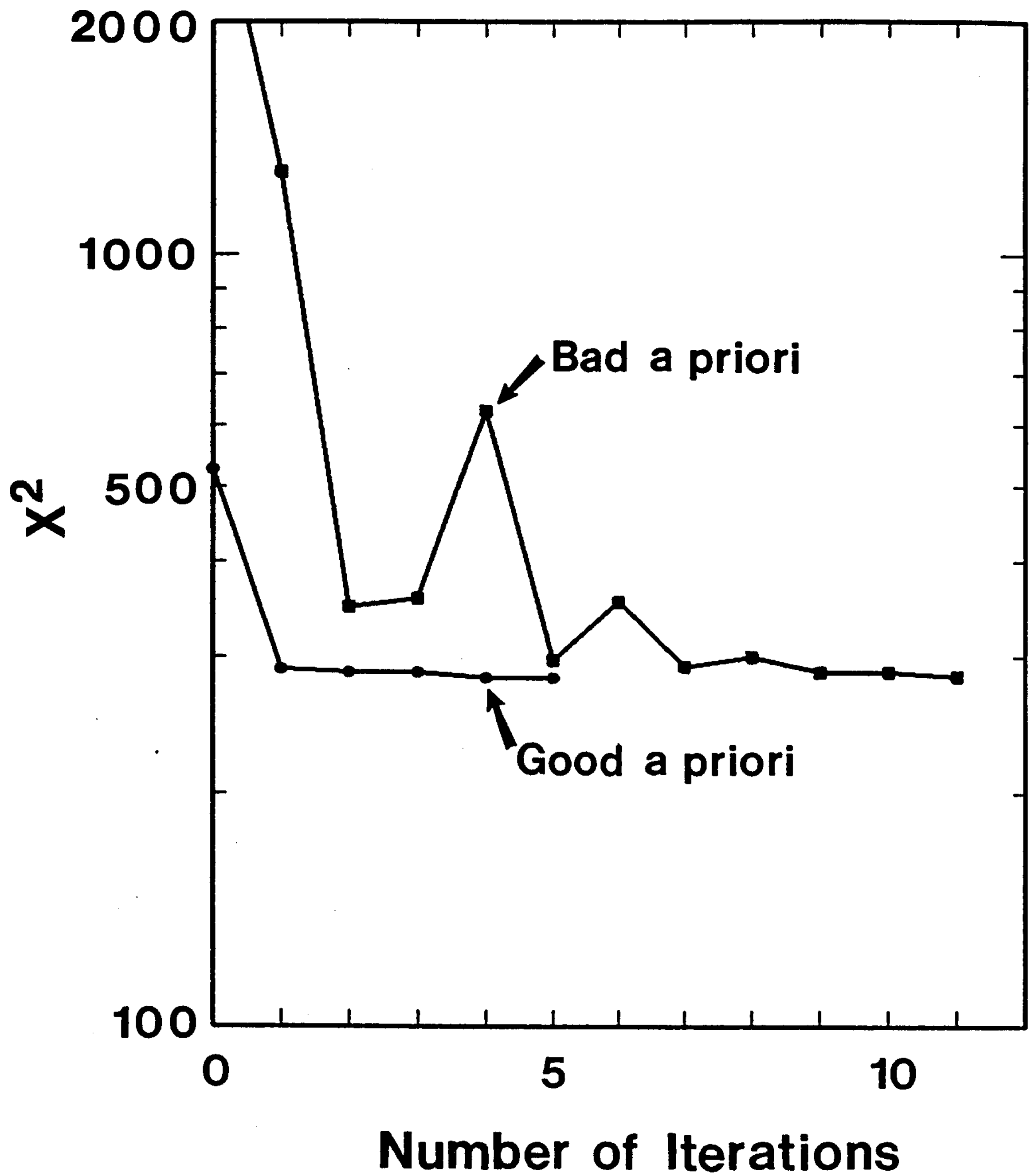



Figure 2: Convergence toward the final solution for Example 1 in Section 4 is traced through several iterative steps for two choices of an initial a priori, $S_0$, namely a "good" a priori and a "bad" a priori. The value of chi-square ($\chi^2$) is used to measure the progress at each step. Each of the "jumps" in $\chi^2$ (which occur at iteration Nos. 4, 6, 8, 10, respectively, for the "bad" a priori choice) corresponds to the start of a new "outer iteration" cycle, as indicated in Fig. 1.

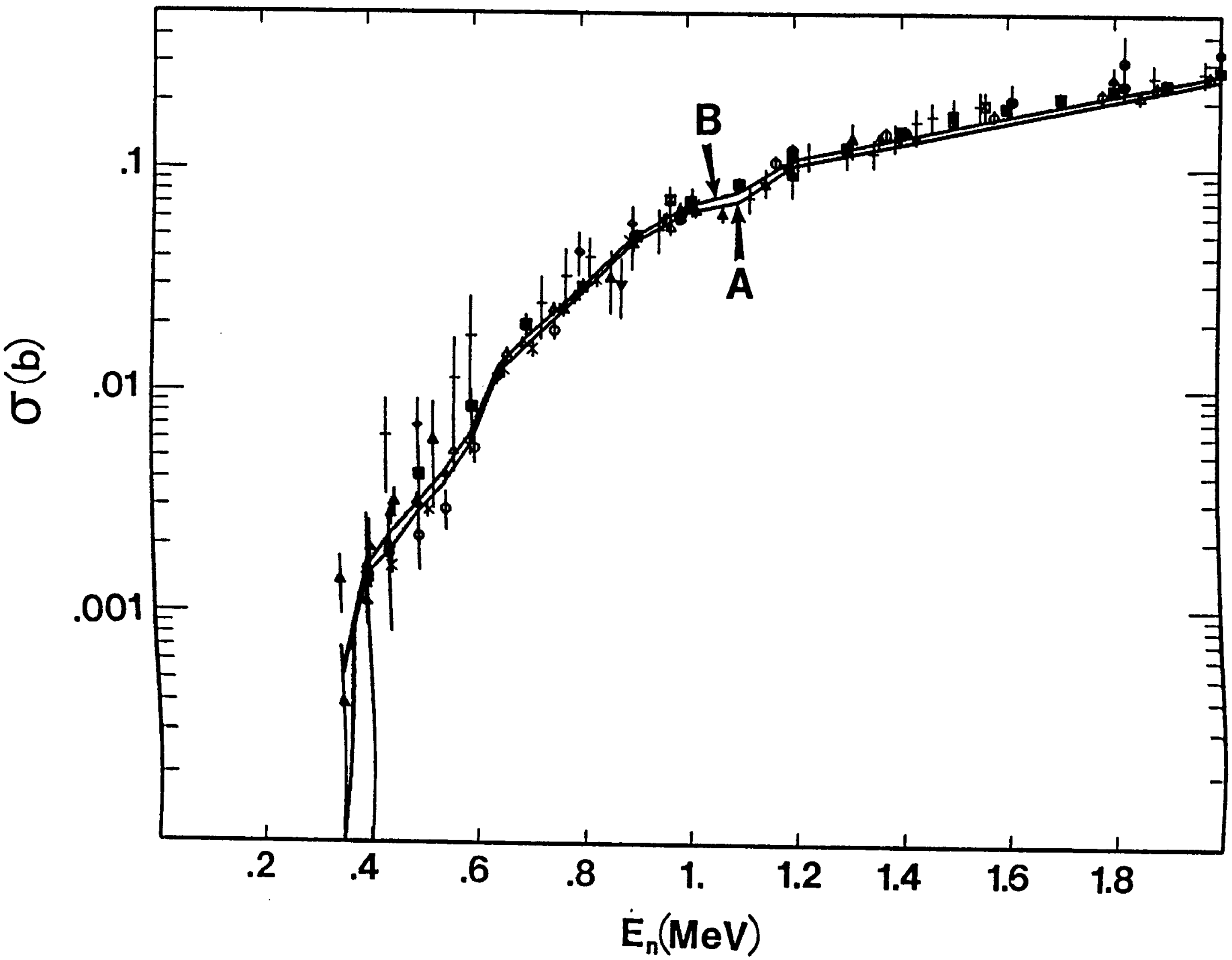



Figure 3 Comparison of evaluated results (threshold to 2 MeV) for the reaction $^{115}In(n,n')^{115m}$ following the application of two distinct least-squares evaluation procedures, as discussed in Section 4: A --- The data covariance matrix is derived from percent errors and actual measured values. B --- The data covariance matrix is derived from percent errors and best-estimate values for the parameters, as deduced from iteration.

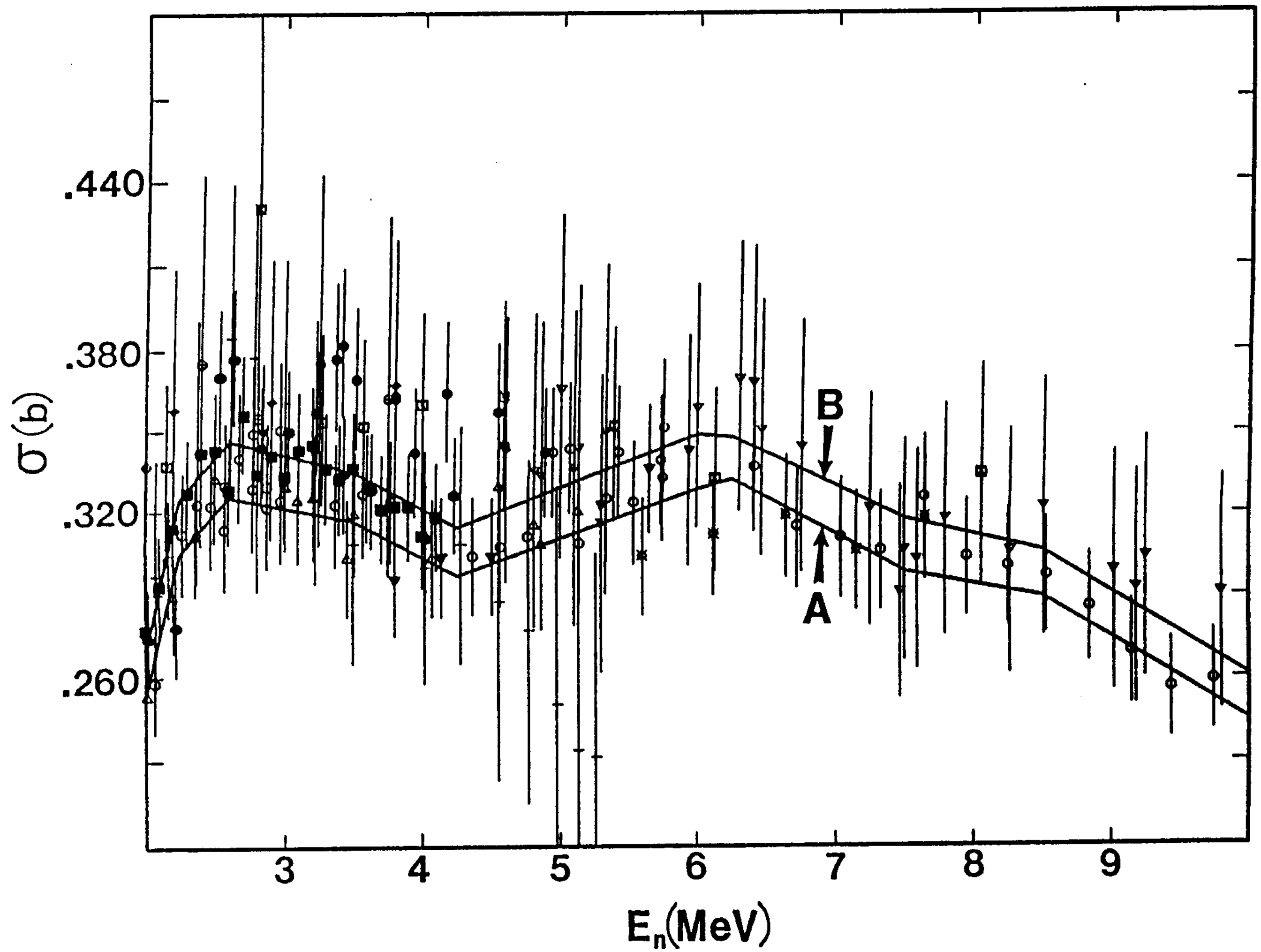



Figure 4: Comparison of evaluated results (2-10 MeV) for the reaction $^{115}$In(n,n')$^{115m}$ following the application of two distinct least-squares evaluation procedures, as discussed in Section 4: A --- The data covariance matrix is derived from percent errors and actual measured values. B --- The data covariance matrix is derived from percent errors and best-estimate values for the parameters, as deduced from iteration. It is clear that for this energy range the results from Method B provide a better representation of the experimental data than those from Method A.

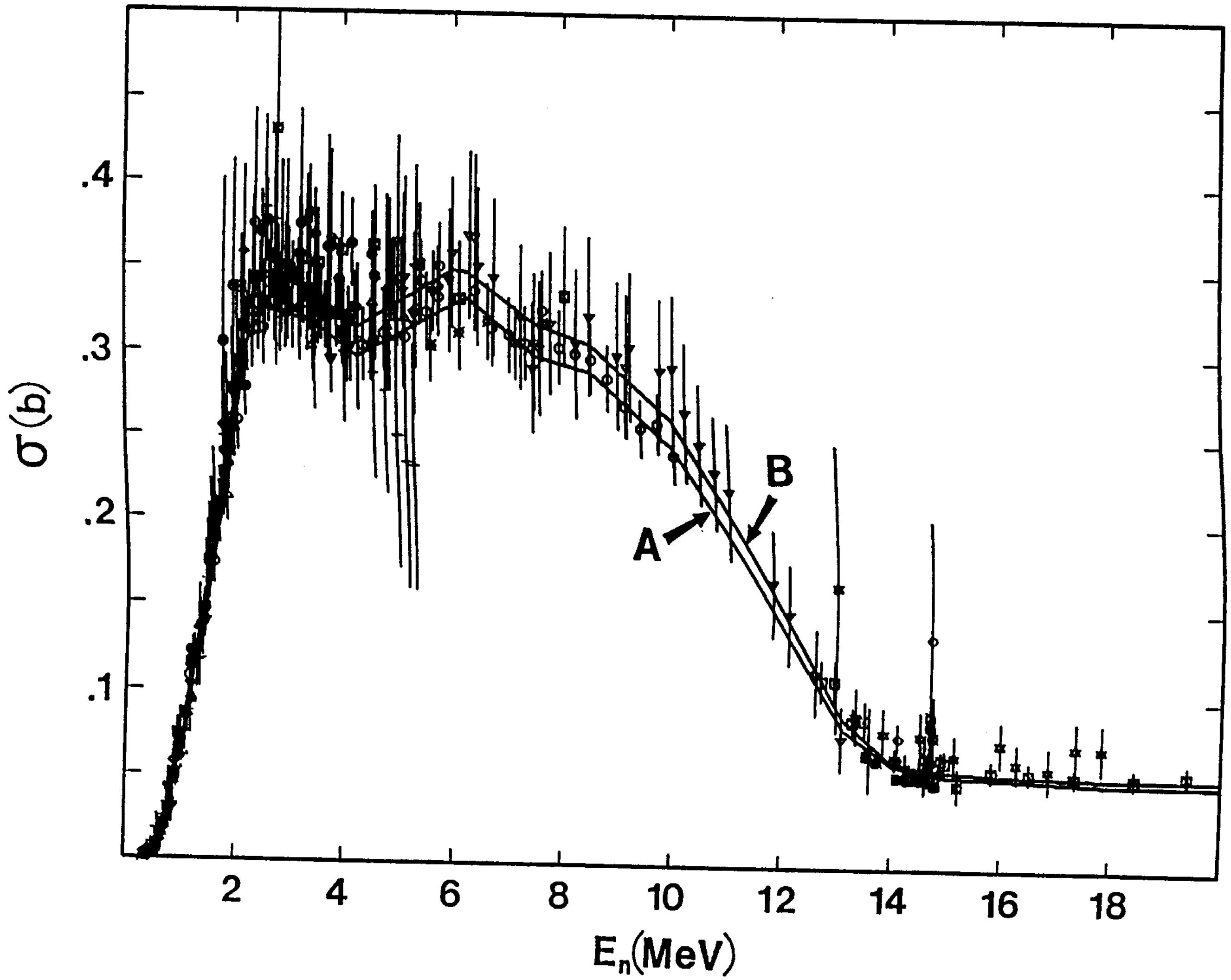



Figure 5: Comparison of evaluated results (threshold to 20 MeV) for the reaction $^{115}\mathrm{In}(n,n')^{115m}$ following the application of two distinct least-squares evaluation procedures, as discussed in Section 4: A --- The data covariance matrix is derived from percent errors and actual measured values. B --- The data covariance matrix is derived from percent errors and best-estimate values for the parameters, as deduced from iteration. The results from Method A fall below those of Method B over the entire energy range.