

PRACTICAL ALTERNATIVES TO EXPLICITLY GENERATING AND INVERTING DATA COVARIANCE MATRICES

Nancy M. Larson
Oak Ridge National Laboratory
Building 6011, MS 6370
P O Box 2008
Oak Ridge, TN 37831-6370 USA
LarsonNM@ornl.gov

ABSTRACT

When fitting data to theory using generalized least squares or similar procedures (as, for example, in R-matrix analyses of cross section data), it is commonly accepted that the data covariance matrix should be explicitly generated and utilized, as a means of incorporating uncertainties inherent in the data-reduction process into the analysis process. In practice, however, when very large data sets are being analyzed, the data covariance matrix is seldom included because it is perceived as being difficult to generate, and too large to store and invert. In this paper, alternative procedures are described. One approach which is being used successfully in the SAMMY R-matrix code involves use of an implicit data covariance matrix; in this procedure, the mathematical formula for the data covariance matrix is manipulated to provide the inverse of the data covariance matrix without ever explicitly generating or inverting that matrix. Other procedures (such as fitting to raw data) are also discussed.

Key Words: covariance matrix, analysis, evaluation, least squares, Bayes

1. INTRODUCTION

All data evaluation schemes require the use of some type of fitting procedure, to obtain those values of the parameters which give the best fit to experimental data. The fitting procedure used in the SAMMY R-matrix code[1,2] is Bayes' method, sometimes called generalized least squares. (The perhaps-more-familiar least squares method can be considered a special case of Bayes' method.) In Section 2 of this paper, notation and equations for Bayes' method are given, and one possible derivation is described.

In either Bayes' method or least squares, the equations relating the prior (initial) values of the parameters to the posterior values (those values which provide a better fit to the data) require the inverse of the data covariance matrix. Because the size of that matrix may be prohibitively large, generating, storing, and inverting it can be costly and error-prone even with today's modern computer systems.

Fortunately, viable alternatives exist. One attractive, but often impractical, alternative is to fit directly to the measured data prior to any corrections to the data (in which case the data covariance matrix is intrinsically diagonal). A more practical method involves the implicit use of the data covariance matrix without the explicit generation, storage, or inversion of that matrix; see Section 3 for details. Another practical alternative is to include some, but not necessarily all, of the data-reduction processes within the analysis. Examples using each method are given in Section 4.

2. BAYES' EQUATIONS

The equations used for finding the best-fitting parameter values can be derived from Bayes' Theorem [3] and the following basic assumptions: (1) Both the data and the parameter values are chosen from normal (Gaussian) distributions. (2) The calculated value is a linear function of the parameters. Clearly neither assumption is exactly correct: (1) The raw data obey Poisson statistics; parameters obey a variety of statistics. (2) The model used to calculate the observable (e.g., the R-matrix theory used to calculate the cross section) is rarely linear with respect to the parameters (e.g., the cross section is not a linear function of the neutron width). Nevertheless, experience has shown that it is often possible to use the equations derived from these only-approximately-true assumptions to determine values of the parameters which give a good fit to the data.

A derivation of Bayes' Equations is given in the SAMMY manual [1], and will not be repeated here; instead we quote the results. Let P represent the parameter values and M the associated covariance matrix, D represent the experimental data to be fitted and V the associated covariance matrix, T represent the theoretical value calculated at the initial parameter values P , and G represent the partial derivatives of the theory with respect to the parameters (G is sometimes called the sensitivity matrix). Primes indicated the updated posterior values. In matrix form, Bayes' Equations can be written as

$$\begin{aligned} P' - P &= M' Y \quad \text{and} \quad M' = (M^{-1} + W)^{-1} \\ Y &= G^t V^{-1} (D - T) \quad \text{and} \quad W = G^t V^{-1} G \quad , \end{aligned} \quad (1)$$

where the second line of equations define Y and W .

To derive the least squares equations from Bayes' Equations requires one additional assumption, that the prior parameter covariance matrix M be diagonal and infinite (so that M^{-1} is zero).

In practice one must generally use a slightly more complicated form of these equations, to include an iterative procedure which partially compensates for non-linearities. Explicitly, the iterative equations take the form

$$\begin{aligned} P^{(n+1)} - P &= M^{(n+1)} Y^{(n)} \\ M^{(n+1)} &= (M^{-1} + W^{(n)})^{-1} \\ Y^{(n)} &= G^{(n)t} V^{-1} (D - T^{(n)} - G^{(n)} (P - P^{(n)})) \\ W^{(n)} &= G^{(n)t} V^{-1} G^{(n)} \quad , \end{aligned} \quad (2)$$

in which the superscript n denotes evaluation at the n^{th} set of parameter values. For simplicity's sake, in the remainder of this paper only the linear form, Eq. (1), will be discussed, but the reader should bear in mind that results are generally applicable when using Eq. (2).

3. GENERATING THE DATA COVARIANCE MATRIX

In practical applications, the data covariance matrix V is diagonal for those situations in which each data point is a separate and independent measurement. For example, in neutron time-of-flight measurements, the “raw” data are counts (as measured by the detector) per time channel (which is related to the energy of the neutron). Each “count” constitutes an independent measurement, and each obeys Poisson statistics.

For the data to be directly useful, they are converted from counts per time channel to cross section as a function of energy. This conversion, or data-reduction, process involves such operations as subtraction of background counts and multiplication by a normalization. It is the data-reduction process which leads to non-zero off-diagonal data covariance matrix elements.

To illustrate the derivation of the data covariance matrix, we first define some notation: let r_i represent the raw datum for channel i , and d_i represent the reduced datum for energy E_i . For a simple example in which only a constant normalization a and a constant background b are required to convert from measured counts to cross section, the data-reduction equation would involve two parameters,

$$d_i = a r_i - b \quad . \quad (3)$$

To generate the covariance matrix associated with the reduced data d_i , we begin by taking small increments of the data-reduction equation. For the simple example of Eq. (3), this gives

$$\delta d_i = \delta a r_i + a \delta r_i - \delta b \quad . \quad (4)$$

Multiplying by δd_j and taking expectation values gives the covariance between points i and j ,

$$\begin{aligned} V_{ij} &= \langle \delta d_i \delta d_j \rangle = \langle (\delta a r_i + a \delta r_i - \delta b) (\delta a r_j + a \delta r_j - \delta b) \rangle \\ &= \langle (\delta a)^2 \rangle r_i r_j + a^2 \langle \delta r_i \delta r_j \rangle + \langle (\delta b)^2 \rangle \\ &= \Delta^2 a r_i r_j + a^2 \Delta^2 r_i \delta_{ij} + \Delta^2 b \quad , \end{aligned} \quad (5)$$

in which all the cross terms have been dropped because each measurement (of a , b , r_i , and r_j) is assumed to be independent of the others. Explicitly, if there were only four data points, the lower half of the (symmetric) covariance matrix for the reduced data would be

$$\left(\begin{array}{cccc} (a^2 \Delta^2 r_1 + r_1^2 \Delta^2 a + \Delta^2 b) & & & \\ (r_1 r_2 \Delta^2 a + \Delta^2 b) & (a^2 \Delta^2 r_2 + r_2^2 \Delta^2 a + \Delta^2 b) & & \\ (r_1 r_3 \Delta^2 a + \Delta^2 b) & (r_2 r_3 \Delta^2 a + \Delta^2 b) & (a^2 \Delta^2 r_3 + r_3^2 \Delta^2 a + \Delta^2 b) & \\ (r_1 r_4 \Delta^2 a + \Delta^2 b) & (r_2 r_4 \Delta^2 a + \Delta^2 b) & (r_3 r_4 \Delta^2 a + \Delta^2 b) & (a^2 \Delta^2 r_4 + r_4^2 \Delta^2 a + \Delta^2 b) \end{array} \right) \quad . \quad (6)$$

For the general case, the data-reduction equation analogous to Eq. (3) would take the form

$$d_i = D(r_i, q_1, q_2, q_3, \dots) \quad (7)$$

where the variables q are data-reduction parameters. [There might in some cases also be dependence on other raw data points, a possibility which complicates the analysis only slightly and which will be ignored in this report.] The covariance matrix then can be readily derived as follows:

$$\begin{aligned} \delta d_i &= \frac{\partial D_i}{\partial r_i} \delta r_i + \sum_k \frac{\partial D_i}{\partial q_k} \delta q_k \\ V_{ij} = \langle \delta d_i \delta d_j \rangle &= \delta_{ij} \frac{\partial D_i}{\partial r_i} \Delta^2 r_i \frac{\partial D_i}{\partial r_i} + \sum_k \sum_l \frac{\partial D_i}{\partial q_k} \langle \delta q_k \delta q_l \rangle \frac{\partial D_j}{\partial q_l} \end{aligned} \quad (8)$$

Here the quantity $\langle \delta q_k \delta q_l \rangle$ represents the covariance matrix for the data-reduction parameters; we shall let Q represent this covariance matrix. Often Q will be diagonal, if each data-reduction parameter is measured separately. In general Q may have non-zero off-diagonal elements, as, for example, when a functional form is fitted to a measured background. In any case the size of this matrix is very small (\sim tens) compared to the number of data points (\sim hundreds of thousands in some cases).

3.1. Implicit Data Covariance Matrix

As seen above, it is *possible* to generate the full off-diagonal data covariance matrix for any measurement. Nevertheless, it is not *necessary* to do so. From Eq. (8), the data covariance matrix can be seen to consist of two pieces, the first of which is diagonal and the second of which is separable. We define matrices v and X as

$$v_{ij} = \delta_{ij} \frac{\partial D_i}{\partial r_i} \Delta^2 r_i \frac{\partial D_i}{\partial r_i} \quad , \quad \text{and} \quad X_{ik} = \frac{\partial D_i}{\partial q_k} \quad , \quad (9)$$

so that v is the diagonal term of the covariance matrix and X represents the partial derivative of the reduced data with respect to the data-reduction parameter. Then Eq. (8) can be written in matrix form as

$$V_{ij} = v_{ij} \delta_{ij} + \sum_k \sum_l X_{ik} Q_{kl} X_{jk} \quad , \quad (10)$$



where the boxes are used to indicated the relative size of the matrices; solid outlines indicate full off-diagonal matrices, and dashed outlines indicate diagonal matrices.

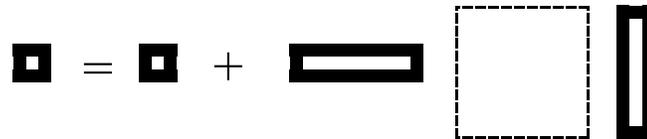
The inverse of V is can be formally derived as

$$\begin{aligned}
 V^{-1} &= (v + XQX^t)^{-1} = v^{-1} - v^{-1}X(Q^{-1} + X^t v^{-1}X)^{-1}X^t v^{-1} \\
 &= v^{-1} - v^{-1}XZ^{-1}X^t v^{-1}
 \end{aligned}
 \tag{11}$$



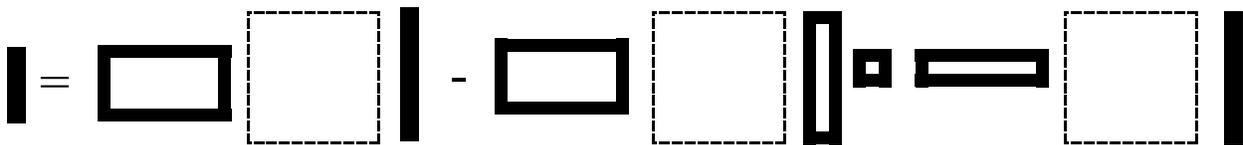
where Z is defined by

$$Z = Q^{-1} + X^t v^{-1}X
 \tag{12}$$



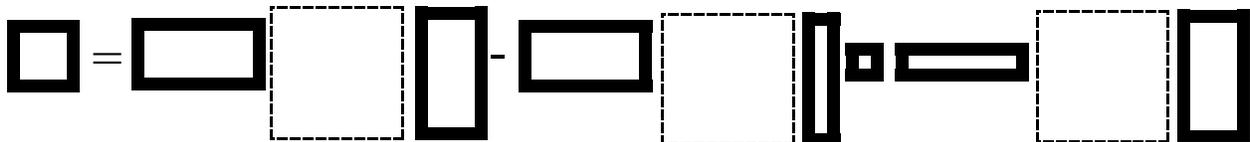
From Eq. (11) and (12), it is clear that one could find the inverse of V without explicitly generating V or inverting this large matrix. In addition, from Bayes' Equations (Eq. (1)), it is clear that there is no need to explicitly generate or store even the inverse of V . Instead, only the quantities Y and W are required:

$$\begin{aligned}
 Y &= G^t V^{-1} (D-T) \\
 &= G^t v^{-1} (D-T) - G^t v^{-1} X Z^{-1} X^t v^{-1} (D-T)
 \end{aligned}
 \tag{13}$$



and

$$\begin{aligned}
 W &= G^t V^{-1} G \\
 &= G^t v^{-1} G - G^t v^{-1} X Z^{-1} X^t v^{-1} G
 \end{aligned}
 \tag{14}$$



The implicit data covariance matrix option has been used successfully in the SAMMY R-matrix code for several years [4,5]. In the next section a simple example illustrates how this can be used in practice.

4. EXAMPLE

Let us assume that a neutron time-of-flight measurement has been made of three resonances for a fission cross section in the resolved-resonance region. The “raw data” consist of counts as a function of energy. [These data were, in fact, generated artificially, starting from a subset of a ^{241}Am fission cross section measurement, adding a constant background and normalization to simulate raw data.]

The data reduction process converts the counts to cross section via such operations as subtraction of background counts and/or multiplication by a normalization whose value is related to the incident neutron flux. In our example, we assume that the experimentalist has (erroneously) determined that the normalization a and background b have the values $a = 0.00105 \pm 0.00007$ and $b = 1.05 \pm 1.05$; the reduced data d_i are related to the raw data as in Eq. (3), and the associated data covariance matrix is as shown in Eq. (6).

The question of interest is the following: Is it possible to obtain an accurate representation of the “true” cross section by analyzing data which have been improperly reduced? The answer is yes, as shall be illustrated below.

4.1. Explicit Data Covariance Method

When the improperly-reduced data are analyzed by the R-matrix code SAMMY, using the full data covariance matrix, Bayes’ Equations provide an updated set of parameter values and associated covariance matrix. The cross sections calculated from the initial and from the final parameter values are shown in Fig. 1. Here the crosses represent the (reduced) data, the dashed curve represents the cross section evaluated from obviously-incorrect prior parameter values, and the solid curve is the cross section evaluated from the posterior parameters. While it is not immediately apparent from this figure that the posterior fit is accurate, we can readily convince ourselves that it is by noting the obvious background problem and adjusting the fitting curve upward, as shown in Fig. 2. The dotted curve in this figure is the unadjusted fit (identical to the solid curve of Fig. 1); the solid curve is the adjusted fit, which clearly does agree with the data.

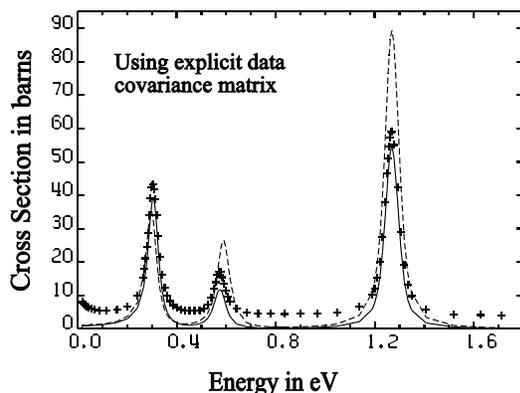


Figure 1. Fitting the reduced data using the explicit data covariance matrix. Crosses represent the incorrectly-reduced data, the dashed curve shows the calculation using prior parameter values, and the solid curve gives the calculation using posterior parameter values.

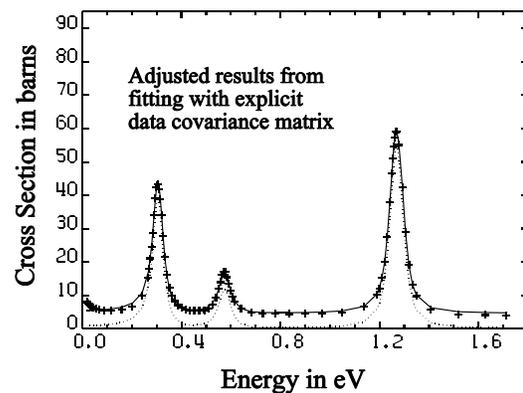


Figure 2. Adjusting the theoretical fit to include the proper background. The dotted curve represents the unadjusted fit (identical to the solid curve in Fig. 1). The solid curve is adjusted for the proper background, giving good agreement with the incorrectly-reduced data.

Resonance parameters generated by this method (fitting to reduced data using the explicit covariance matrix) and by other methods are shown in Table I. For our example, each resonance was described by five R-matrix parameters: energy E_λ and widths $\Gamma_{\lambda\gamma}$, $\Gamma_{\lambda n}$, $\Gamma_{\lambda f1}$, and $\Gamma_{\lambda f2}$. However, only those combinations which are relevant to fission cross sections for isolated resonances are included in the table; these are E_λ , $\Gamma_\lambda = \Gamma_{\lambda\gamma} + \Gamma_{\lambda n} + |\Gamma_{\lambda f1}| + |\Gamma_{\lambda f2}|$, and $\Gamma_{\lambda f} = |\Gamma_{\lambda f1}| + |\Gamma_{\lambda f2}|$.

4.2. Implicit Data Covariance Method

Virtually identical results are obtained using the implicit data covariance matrix, as described in Section 3.1; see Fig. 3 and Table I.

Table I. Comparison of parameter values.
 Values for χ^2 are normalized to the “raw data” value.

	prior values	explicit data covariance matrix	implicit data covariance matrix	raw data	hybrid method	diagonal part of covariance matrix	statistical uncertainties only
E_1	.3000 ± .0294	.3073 ± .0001	.3073 ± .0001	.3071 ± .0001	.3073 ± .0001	.3081 ± .0001	.3086 ± .0001
Γ_1	47.39 ± 4.70	43.60 ± 0.20	43.57 ± 0.20	43.41 ± 0.20	43.57 ± 0.20	62.09 ± 0.26	67.85 ± 0.22
Γ_{1f}	.3200 ± .0233	.3237 ± .0198	.3234 ± .0198	.3277 ± .0204	.3234 ± .0198	.4116 ± .0219	.4782 ± .0226
E_2	.5900 ± .0321	.5765 ± .0001	.5765 ± .0002	.5758 ± .0002	.5765 ± .0001	.5776 ± .0002	.5781 ± .0002
Γ_2	48.15 ± 4.75	46.50 ± 0.49	46.41 ± 0.49	43.77 ± 0.52	46.41 ± 0.49	108.73 ± 0.85	117.09 ± 0.63
Γ_{2f}	.6000 ± .0447	.4455 ± .0335	.4450 ± .0335	.4417 ± .0337	.4450 ± .0335	.8667 ± .0464	.9810 ± 0.050
E_3	1.270 ± .356	1.2696 ± .0001	1.2695 ± .0001	1.2695 ± .0001	1.2696 ± .0001	1.2695 ± .0008	1.2693 ± .0001
Γ_3	47.69 ± 4.67	42.91 ± 0.33	42.88 ± 0.33	42.33 ± 0.34	42.41 ± 0.33	57.21 ± 0.42	59.77 ± 0.42
Γ_{3f}	.8001 ± .0800	.5674 ± .0474	.5671 ± .0474	.5772 ± .0488	.5627 ± .0481	.7389 ± .0542	.7627 ± .0481
χ^2	46.	1.5	1.4	1.0	1.0	6.9	22.

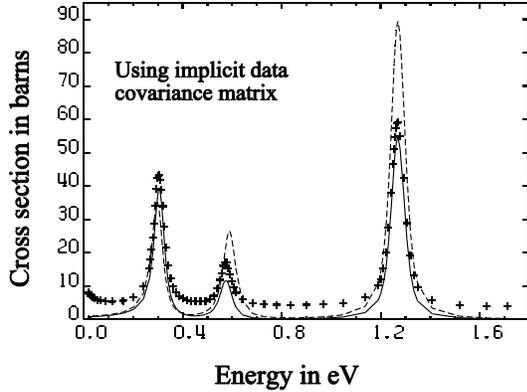


Figure 3. Fitting the reduced data using the implicit data covariance matrix. Crosses represent the reduced data, dashed curve the calculation using prior values of the parameters, and solid curve the calculation using posterior parameter values.

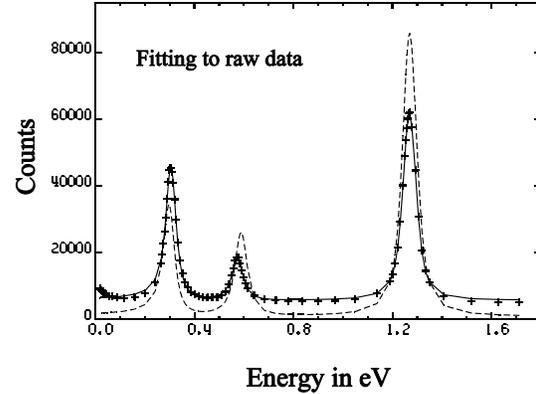


Figure 4. Fitting the raw data directly. Crosses represent the raw data, dashed curve the calculation using prior values of the parameters, and solid curve the calculation using posterior parameter values.

4.3. Fitting to Raw Data

While both the explicit and the implicit covariance methods can provide correct values for resonance parameters and covariance matrix, an analyst may find the results somewhat unsatisfying because the plotted curves do not appear to agree with the experimental data (as in Figs. 1 or 3, because the assumed background is incorrect). In our simple example, adjusting the theoretical curve to agree with the data (as in Fig. 2) is straightforward because the difference is due only to a constant background. In general, such manual adjustment is not possible.

A more satisfying method might be to include the data-reduction parameters within the analysis procedure. That is, rather than transforming from directly-measured values r_i to reduced values d_i as in Eq. (3), the analyst would fit directly to the raw data. The theoretical calculation must then include the (reverse) transformation from cross section to counts, analogous to Eq. (3),

$$T_i = A\sigma_i + B \quad , \quad (15)$$

and data-reduction parameters A and B may be treated on equal basis with the R-matrix parameters (i.e., included in the fitting process). Resonance parameter values obtained using this procedure are given in Table I; plots are shown in Fig. 4. Initial values for A and B used in this run were 950 ± 50 and 1000 ± 1000 respectively; posterior values are 965.90 ± 37.65 and 573.59 ± 17.61 .

The argument can be made that this method (fitting to raw data) is mathematically more rigorous than the methods involving the data covariance matrix, because the data covariance matrix is implicitly (erroneously) assumed to be linear. The derivation of the covariance matrix, Eq. (8), ignores second- and higher-order terms. Fitting to raw data completely eliminates that potential problem; R-matrix parameter values extracted by fitting to raw data should, in general, be more accurate than those found using other methods. For this example, differences between this method and the other two are small but noticeable; see Table I.

4.4. Hybrid Method

The method of fitting to raw data, while seemingly useful and rigorous, also has drawbacks: Analysis codes such as SAMMY would be required to include mathematical descriptions of all data-reduction processes for all types of experiments. Experimentalists would have to provide analysts with both the raw data and the details of the data reduction process. Neither situation is likely to occur.

There is, however, yet another possibility, which we will call the hybrid method. Here the analyst will fit to the reduced data, including only the statistical uncertainties for the data covariance matrix. (These statistical uncertainties correspond to the square root of v_{ii} of Eq. (9), so-called because they are derived directly from the statistical distribution of the raw data.) Data-reduction parameters such as normalization and background will be included in the theoretical calculation. Prior values for these parameters will be set so that the effect is zero (that is, normalization would be set to one and background to zero), and prior values of the uncertainties on these data-reduction parameters will correspond to the measured uncertainties on the parameters used to reduce the data. For our example with only normalization and background, the theoretical calculation is

$$T_i = A \sigma_i + B \quad , \tag{16}$$

with $A = 1.0$ and prior uncertainty $\Delta A = \Delta a / a \sim 0.001$, and $B = 0.0$ with prior value for uncertainty $\Delta B = \Delta b \sim 0.1$. Using both A and B (along with the R-matrix parameters) as fitting parameters in a SAMMY run produces the results reported in Fig. 5 and Table I.

With this hybrid method (as with the fit-to-raw-data method), posterior values for the data-reduction parameters are included in the SAMMY output. The posterior value for A is 1.000 ± 0.001 and for B is 4.471 ± 0.016 . [If the analyst wished, he could modify the data set by this value of B to simplify further analyses, so that the “new” experimental data would be more closely related to the cross section of interest.]

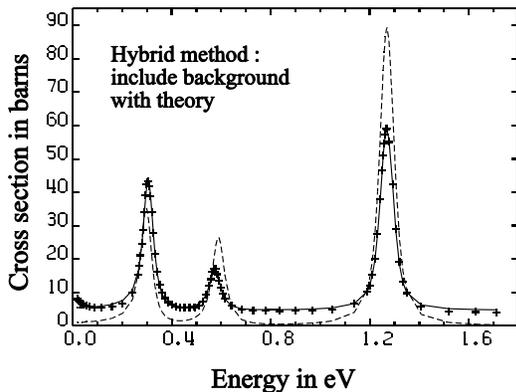


Figure 5. Fitting the reduced data using the hybrid method. Crosses represent the reduced data, dashed curve the calculation using prior values of the parameters, and solid curve the calculation using posterior parameter values.

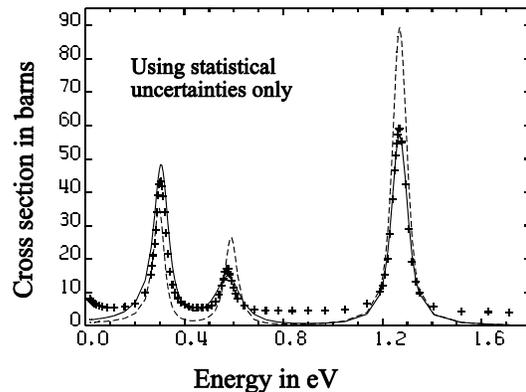


Figure 6. Fitting the reduced data using only the statistical uncertainties. Crosses represent the reduced data, dashed curve the calculation using prior values of the parameters, and solid curve the calculation using posterior parameter values.

4.5. Diagonal Data Covariance Method

One additional method should be mentioned here, as it is commonly used but can lead to inaccurate results. This method involves fitting to the reduced data, but using only the diagonal elements of the data covariance matrix, ignoring all off-diagonal contributions. Two possibilities exist for this method: the first is to use only the statistical contribution (v_{ii} of Eq. (10)) for the data covariance matrix; the second is to include so-called “systematic” or “common” errors (the second term in Eq. (10)) but neglect the terms with i not equal to j .

In a situation such as our example, where there were errors in the data-reduction process, this method will produce erroneous results. In the last two columns of Table I, we find that the resonance energies are reasonably well determined by this method, but the widths are quite poorly determined. The reason becomes clear when Fig. 6 is examined: because this method makes no provision for an incorrect background, the best-fit curve is one which goes nearly to zero between resonances while attempting to fit the resonance peak. Hence, the shape is badly distorted.

5. CONCLUSIONS

In this report we have discussed several different practical alternatives for properly including measured uncertainties into the data analysis process. Alternatives include the use of either explicit or implicit data covariance matrices, fitting to raw data, and a hybrid method (in which some of the data-reduction parameters are treated as search parameters). Each of the alternatives is available in the SAMMY R-Matrix analysis code (though only to a limited extent, e.g., for fitting to raw data). The common practice of ignoring off-diagonal data covariances was shown to be incorrect, in that its use can lead to erroneous results.

The mathematical methods underlying the implicit data covariance method were presented in some detail, as this approach is not yet in common use.

ACKNOWLEDGMENTS

The author gratefully acknowledges the cooperation and encouragement of H. Derrien and L. Leal, whose SAMMY analyses have made extensive use of both the implicit data covariance and the hybrid method. Thanks go also to H. Leeb and A. Plompen, whose helpful suggestions during SAMMY workshops have greatly improved my understanding and presentation of this material.

This work was sponsored by the Office of Environmental Management, U. S. Department of Energy, under contract DE-AC05-00OR22725 with UT-Battelle, LLC.

REFERENCES

1. N. M. Larson, "Updated Users' Guide for SAMMY: Multilevel R-Matrix Fits to Neutron Data Using Bayes' Equations," ORNL/TM-9179/R5, Oak Ridge National Laboratory, Oak Ridge, TN (2000).
2. N. M. Larson, "Introduction to the Theory and Analysis of Resolved (and Unresolved) Neutron Resonances via SAMMY," *Proceedings the IAEA Workshop on "Nuclear Reaction Data and Nuclear Reactors: Physics, Design and Safety" held at the International Centre for Theoretical Physics, Trieste, Italy, 23 February - 27 March 1998*, pp. 1-45, World Scientific, (1999).
3. H. Jeffreys, *Theory of Probability*, Third Edition, Oxford University Press, London (1961).
4. H. Derrien, N. M. Larson, and L. C. Leal, "Covariances Matrices for Use in Criticality Safety Predictability Studies," ORNL/TM-13492, Oak Ridge National Laboratory, Oak Ridge, TN (1997).
5. L. C. Leal, H. Derrien, N. M. Larson, R. Q. Wright, "R-Matrix Analysis of ^{235}U Neutron Transmission and Cross Sections in the Energy Range 0 to 2.25 keV," ORNL/TM-13516 (1997). Also *Nucl. Sci. and Eng.* **131** 230 (1999).