XI.A. STRATEGY FOR DATA EVALUATION WITH SAMMY

In the "best of all possible worlds," data analysis with SAMMY (or any other analysis code) would follow a "cookbook procedure" and produce perfect results on the first attempt. However, the real world is far too complicated for that. Nevertheless, an outline is presented here, as a guide to the user in designing his/her own stratagem for data analysis. The discussion below applies primarily to analyses in the resolved resonance region.

(The author is greatly indebted first to Claire Perey for drafting the original version of this guide, during the early days of SAMMY development, based on her extensive experience using SAMMY to analyze transmission data, and also to current members of the ORNL Nuclear Data Group for their advice and improvements to the 2006 version.)

Step 1. Preparation of DATa files

For guidance in choice and preparation of experimental data to be used in the SAMMY analysis, the reader is referred to the many papers describing particular evaluations. Examples can be found in [LL99, RS00, and others]. In general, one should include all available data that is considered to be reliable and consistent. Ideally, many different types of data should be included (transmission, capture, fission, differential elastic, etc.).

Whenever possible, transmission data should be used rather than total cross section, because analyses of total cross section data often incur numerical problems due to blacking-out resonances. In addition, experimental resolution applies to transmission, not to total cross section; for this reason, cross sections derived by taking the logarithm of experimental transmissions are considered "effective total cross sections" and are highly dependent on the thickness of the sample. For best results, always fit transmission data rather than total cross section.

The formats for experimental data in SAMMY are discussed in Section VI.C.

Care must be taken to determine the conditions under which the experimental data were measured. Effective temperature (for Doppler broadening), components of the resolution function (burst width, channel width, flight-path length, etc.), background, and normalization information should be noted, as should nuclide abundances within the sample(s). Uncertainties should be determined or estimated for all parameters. When information is not available for older data, reasonable assumptions should be made; values for these parameters can be adjusted (fitted) during the analysis process.

Step 2. Preparation of INPut and PARameter files

Before the INPut and PARameter files (Sections VI.A and VI.B, respectively) can be created, quantum number information (spin groups and channels) must be established; see Section II.C.1 for details. For help in determining the appropriate spin group quantum numbers, use program SAMQUA [OB03, OB04] as described in Section X.J.

In many cases, results of earlier evaluations for some of the relevant nuclides are available in the evaluated nuclear data files (ENDF). ENDF File 2 (which contains resonance parameters) can be used as input to SAMMY, greatly simplifying the SAMMY setup procedure. (See Section IX of this manual for details concerning the use of ENDF files as input to SAMMY.) Do not, however, rely entirely on the information in the ENDF files, for the following reasons:

- Older files sometimes did not give proper descriptions of the spin groups and channels.
- Except for the new LRF = 7 format, ENDF formats do not have provision for defining parity. *Do not assume that the spins and parities extracted from ENDF files are correct.*
- Many older files used one of the Breit-Wigner approximations, which should not be used for new evaluations. (When reading an ENDF file, SAMMY produces a comparable SAMMY-style INPut file called SAMNDF.INP; be sure to remove any command lines referring to Breit-Wigner from this file before using it as input to future SAMMY runs.)
- Small resonances observable in new measurements may not have been seen in the early measurements.
- New measurements may permit extension of the resolved resonance region to higher energies, requiring the use of higher *l*-values.

Therefore, even when preliminary information is available from ENDF File 2, it is recommended that SAMQUA be used to ensure that all possible spin groups and channels are considered for inclusion in the analysis.

If R-matrix parameters are not available from ENDF, results from resonance parameter analyses may be found in the EXFOR files [EXFOR]. Users should be aware that these are not full evaluations but are closely tied to individual measurements.

Information concerning individual resonances may be available in the literature. See, for example, the Mughabghab compilation [SM06].

Information from data bases or the literature will probably need to be augmented with information available only in the new data you are preparing to analyze. Program SUGGEL (Section X.P) was designed to provide guidance in making spin-group assignments for resonances. Program RSAP [RS03] also has features that aid in PARameter-file preparation. One such feature is a peak search procedure that "finds" resonances in new data and generates starting PARameter-file entries. A certain amount of trial and error on the analyst's part will also be needed but will pay dividends in the future.

The R-matrix is a complicated function, involving long-range interference. At any given energy, the "background" R-matrix has contributions from nearby small resonances, from large resonances within the region being analyzed, and from resonances outside the region of analysis. The effect of the outside resonances is approximated either by fictitious or "dummy" resonances or by a parameterized external R-function. (For more on this topic, see Section II.B.1.d and references [LL99] and [FF01].) Be sure to include either dummy resonances or another description of the external levels in your PARameter file, as it will not be possible to fit the data without this.

Step 3. Initial SAMMY calculations

Once preliminary INPut and PARameter files have been created, and DATa files are available, it is time to make preliminary SAMMY runs. Begin with the "best" data, which generally means the highest-resolution (longest flight path, lowest temperature, etc.) transmission data.

The first runs will be "no-Bayes" runs, which means that no fitting of theory to data is to be attempted. Instead, the purpose is to debug your input, that is, to determine whether the input is sensible and to make any necessary adjustments. For these initial runs, the following commands should be included in the INPut file:

```
DO NOT SOLVE BAYES Equations
GENERATE PLOT FILE AUtomatically
PRINT ALL INPUT PARAmeters
```

and possibly

```
PRINT INPUT DATA
```

Only statements that are not defaults are listed above; to see exactly what these commands will accomplish, see Table VI A1.2. Depending on the particular experiment being calculated, other commands may be needed as well.

When the initial run is completed, look carefully at the results. Probably you will wish to start with a plot, to visually compare the measured data with the calculated cross section. (See Section VII.C for a discussion of plotting options.) Next, especially if the calculated curve is very different from what you expected, carefully compare parameter values listed in the SAMMY.LPT file (Section VII.A) with those from the PARameter and INPut files; this will help you locate input mistakes such as misplaced decimal points, typographical errors, misunderstood units, and other mishaps.

You will also want to look over the experimental cross sections (or transmissions) and uncertainties carefully (either from the plot file or as printed in the LPT file, if you chose to print that information). This will help you discover and correct/discard/adjust any oddities. Among the possibilities here are the following:

- Uncertainties that are unreasonably small; often these are associated with very small cross section values. One way to spot these is to look at the weighted residuals (using the `PRINT WEIGHTED RESIDUALS` command) for values that are orders of magnitude higher than most.
- Large discontinuities, which may indicate a formatting problem.
- For total cross section data (which you will use only if absolutely necessary), the peaks of large resonances may appear "chopped off" (because the cross section values were taken from blacking-out resonances in transmission experiments). The uncertainties on these data points will need to be increased significantly.

Corrections for experimental conditions can have major effects on the calculated cross sections, so reasonable estimates for Doppler- and resolution-broadening parameters will probably need to be included even at this early stage. Unless you have other information about the conditions

under which these data were measured, initial attempts should assume room temperature (300 K) for the Doppler broadening (Section III.B.1), and the simple Gaussian resolution function with reasonable parameter values (Section III.C.1.a).

By performing a few successive no-Bayes runs, you will eventually reach the stage where the calculated cross sections are roughly consistent with the measured data. That is, you will find

- good approximations for the values of the energy parameters for the large resonances seen in the region you are planning to analyze and
- a combination of parameters for your external R-function (or outside dummy resonances) and for the widths of the large resonances in your energy range to give reasonable estimates of the effect of outside resonances.

For some of the data sets, you may also discover energy-scale misalignments and/or normalization or background problems. Energy re-alignment of all the experimental data sets should be performed at this stage, using the transmission experiment with the longest flight path as the standard. See Section III.E.8 for guidance for correcting the energy scale, and Section III.E.3 for normalization and background options.

Step 4. Preliminary analysis: gross features

Perform a preliminary analysis on each individual data set (again beginning with the "best" data set), using Bayes' theorem to adjust the most relevant parameters. At this stage it is recommended that you assign

- small but realistic uncertainties on the energy parameters of the large resonances inside your energy range and
- fairly large uncertainties on the widths of these resonances and on the parameters outside your range of analysis.

Table VI B.2 provides many options for input of prior uncertainties on resonance and other parameters.

In a few runs you will have generated a theoretical cross section that roughly agrees with your data. (In earlier years, when computation time was prohibitive, these runs were often made using severely averaged data sets to speed up the calculation. With today's high-speed computers, averaging is almost never necessary.)

For these runs, you will need to modify your INPut files by disabling the command that prohibits fitting. This command line can either be removed, or it can be commented-out with a pound sign (#) in the first column, as shown here:

```
#DO NOT SOLVE BAYES Equations
```

One default command should be noted:

```
DO NOT SUPPRESS ANY intermediate printout
```

When things go wrong, the additional printout (in the SAMMY.LPT file, as described in Section VII.A) provided by this command is helpful in detecting where, and possibly why, the trouble started. If everything goes right, use of this option allows you to follow the path of the parameters' adjustments.

Step 5. Further analysis: small energy regions

Small regions should now be studied individually, with all the necessary details, progressively including narrow resonances. At this stage you will want to refine the resolution function for each data set, and include other corrections (multiple scattering, background, etc.); see Section III for details.

Care should be taken to include enough points in the auxiliary energy grid to properly describe the unbroadened cross section, in order to perform the Doppler and resolution integrations accurately. (See the discussion in Section III.A.1 regarding the density of points in the auxiliary grid.) One simple test you can make to see whether your grid is sufficiently dense is to double the number of points used (i.e., set NXTRA = 1 in card set 2 of the INPut file). If values for the Doppler- and resolution-broadened cross sections change appreciably, then the grid is not sufficiently dense. If values remain approximately the same, then the original grid is acceptable. Remember that it is the analyst's responsibility (not the SAMMY author's) to ensure that the auxiliary grid is adequate for your purposes.

Step 6. Final fitting runs

At the end of this final step you will obtain the complete and consistent parameter and covariance files for all data sets included in the analysis. Therefore, none of the input parameter values or their uncertainties, nor the data points or their uncertainties, should be manually changed at any time during this final step. If changes are needed, the entire process should begin anew.

All relevant parameters should be varied in this step: set the flag = 1 in the PARameter file.

Some parameters may be held fixed: (1) Values for some parameters are well determined from other experiments but have little effect for the current data; these should be held fixed or perhaps PUPped (see Section IV.D.2) so that their uncertainty contributes to the data covariance matrix. (2) If necessary to conserve computer resources, some parameters may be held fixed at the values found in earlier steps. Care should be taken to hold fixed only those parameters whose effect is short range. (3) In some cases, a parameter may be essentially irrelevant for the data being analyzed and therefore need not be flagged.

All measurement uncertainties should be included at this stage. Statistical uncertainties are defined in the DATa file, but systematic uncertainties require different treatments. The simplest way for the user to take care of systematic measurement uncertainties is to PUP the data-reduction parameters (normalization, background, sample thickness, etc.) by setting the flag = 3 in either the PARameter or (more likely) the INPut file. When additional systematic uncertainties are available,

these should be included via the user-supplied implicit data covariance option (see Section IV.D.3 for a discussion and VI.C.3.b for input details).

Two options are now available in SAMMY for analyzing several data sets together. The first (SAMMY's original) method involves sequential analyses, one data set after another, using the output parameter covariance matrix (PCM) from one run as input to the next run. The other is to analyze all data sets simultaneously. There are advantages and disadvantages to each.

Sequential analyses: Final results from sequential analyses would be exactly equivalent to final results from simultaneous analysis, provided the theory were linear with respect to the varied parameters. Since R-matrix theory is exceedingly non-linear, exact equivalence is not possible. Further, the ordering of the data sets in the sequence can be important. Generally, one would include the best data sets first, to firmly establish values for the most important parameters, before moving on to less reliable data.

Often the analyst will take the final PARameter and COVariance file from the entire sequence of runs, and repeat the sequence using those files as initial input. This process may be repeated several times, until convergence is reached. Generally, the criterion for convergence is that parameter values (and hence chi-squared values) no longer change very much. (Caution: When this kind of repetition is used a total of N times through the sequence, the final parameter covariance matrix is diminished by a factor of N. That is, the uncertainty on any parameter must be multiplied by $\sqrt{N}$ to obtain the "true" uncertainty as determined by this sequential analysis.)

A major advantage of sequential analysis is the ability to locate data sets which are inconsistent with the other data. When such a data set is encountered in the sequence, parameter values will always be modified, so it is unlikely that convergence will be reached. By studying the sequence of runs, the analyst will be able to determine which data set is causing the problem, and make adjustments as needed.

Simultaneous analysis: It is now possible, although still a bit awkward, for SAMMY to do simultaneous analysis of even very large data sets. (Details are presented in Section IV.E.1.) A series of runs is made, one for each data set. The contributions to the Y and W matrices (needed for Bayes' equations) are generated in those runs and saved in binary files. An additional SAMMY run then reads those files and solves Bayes' Equations to determine the updated parameter values and associated parameter covariance matrix.

Iteration for non-linearities is also possible with simultaneous analysis, although each iteration requires another complete series of K+1 runs (where K is the number of data sets). Details are given in Section IV.E.1.

With simultaneous analyses, it is possible to calculate the true least-squares solution, in which the prior parameter covariance matrix is assumed to be diagonal and infinite. (See Section IV.E.3 for details.) This feature was added primarily to facilitate comparison with other codes, most of which use least squares rather than the more general Bayes' equations used in SAMMY.

Step 7. Obtaining the parameter covariance matrix

The procedure described in step 6 produces both a "good" set of parameter values and the associated parameter covariance matrix (PCM). Ideally, this is the PCM which would be included in the analyst's final report. In practice, this PCM is widely regarded as too small, which means that the uncertainties (the square root of the diagonal of the covariance matrix) are deemed to be smaller than expected. In reality, not all sources of uncertainty are included in the analysis, even when the analyst goes to great pains to include all known sources of experimental uncertainty. (See Section IV.E.6 for further discussion.)

Analysts therefore usually wish to adjust (increase) the SAMMY-produced PCM in some fashion, to better conform to their expectations regarding the correct PCM. A variety of techniques have been developed to assist in this process; these are described in Section IV.E.6.

Given the recent surge in perception of the importance of uncertainty information in all aspects of nuclear data, it is important that anyone reporting the results of data analysis should include information about the PCM. Do not neglect this aspect of your analysis.

Step 8. Statistical studies

Most reports regarding resonance parameters also include discussion of statistical properties of those parameters. See, for example, references [HD90], [OB97], [LL99], [RS06]. The code SAMDIS (Section X.F) can be used for some of those studies.

Step 9. Reporting results

To create a table of resonance parameters, including uncertainties, use SAMMY's command

```
PUBLISH
```
or
```
CREATE PUBLISHABLE List of parameters
```

which creates a file called SAMMY.PUB. This file can be ported to a spreadsheet to be formatted as needed for publication. See Section VII.F for details.

To write the resonance parameters and covariance matrix into an ENDF file, see Section IX.

The SAMMY author would appreciate receiving preprints or reprints of any publications describing work using the SAMMY code. Kindly e-mail these to LarsonNM@ornl.gov.