SAND96-0461C
CONF-960543--20

Verification of unfold error estimates in the UFO code

D. L. Fehl and F. Biggs
Sandia National Laboratories
Albuquerque, N. M. 87185
(505) 845-7822

ABSTRACT

Spectral unfolding is an inverse mathematical operation which attempts to obtain spectral source information from a set of tabulated response functions and data measurements. Several unfold algorithms have appeared over the past 30 years; among them is the UFO (UnFold Operator), code written at Sandia National Laboratories. In addition to an unfolded spectrum, the UFO code also estimates the unfold *uncertainty* (error) induced by estimated random uncertainties in the data. In UFO the unfold uncertainty is obtained from the error matrix. This built-in estimate has now been compared to error estimates obtained by running the code in a Monte Carlo fashion with prescribed data distributions (Gaussian deviates). In the problem studied, data were simulated from an arbitrarily chosen blackbody spectrum (10 keV) and a set of overlapping response functions. The data were assumed to have an imprecision of 5% (standard deviation). 100 random data sets were generated. The built-in estimate of *unfold uncertainty* agreed with the Monte Carlo estimate to within the statistical resolution of this relatively small sample size (95% confidence level). A possible 10% bias between the two methods was unresolved. The Monte Carlo technique is also useful in underdetermined problems, for which the error matrix method does not apply. UFO has been applied to the diagnosis of low energy x rays emitted by Z-pinch and ion-beam driven hohlraums.

I. Introduction

Spectral unfolding is an inverse mathematical operation which attempts to obtain information about a source spectrum $F(E)$ from a set of measurements (data) $\{D_i\}$, $i = 1,\ldots,M$. Often the connection between $\{D_i\}$ and F can be written in terms of response functions R_i and a set of M Fredholm integral equations (first kind)[1]:

$$D_i = \int_{E_1}^{E_2} F(E) R_i(E) dE + \epsilon_i \ . \quad (i=1,2,\ldots,M) \qquad (1)$$

The interpretation of Eqs. (1) is that the datum D_i in the i-th "channel" has two contributions: one due to the sum of spectral components $F(E)dE$, each weighted by the response function of that channel; the other due to a random variable ϵ_i, representing uncertainty and perturbation in the measurement process. Such integral equations arise in plasma physics[2,3] and radiation hardness testing[4,5] for filtered-detector, x-ray diagnostics and for magnetic ion-beam spectrometers[6].

Given only a set of integral equations and data, it is not possible, in general, to reconstruct $F(E)$ at every point in its domain $[E_1,E_2]$ because (1) the number of integral equations M is finite, and (2) such problems are "ill-posed"[1] That is, spectral averaging within the integrals admits wildly oscillating and physically unacceptable solutions *in addition* to the desired solution; such unacceptable solutions can be found no matter how precisely the measurements are made[1]. Yet, some *partial* reconstruction of the source may be possible if the integrals in Eqs. (1) are reduced to matrix form. For example, suppose the desired solution can be approximated as a linear combination of known basis functions $B_j(E)$ $(j=1,2,\ldots,N)$ with unknown coefficients F_j:

$$F(E) \approx \sum_{j=1}^{N} F_j B_j(E) \quad . \qquad (2)$$

Basis functions include polynomials, weighted delta-functions, or contiguous histograms (first-order B-splines)[7]. Substituting Eq. (2) into the integrals of Eqs.(1) then yields the matrix approximation

$$D_i \approx \sum_{j=1}^{N} R_{ij} F_j + \epsilon_i \ , \quad (i=1,2,\ldots,M) \qquad (3)$$

where

$$R_{ij} = \int_{E_1}^{E_2} R_i(E) B_j(E)\, dE \quad . \tag{4}$$

Reformulating Eqs. (1) as matrix equations, however, does not guarantee acceptable numerical solutions, since the system may still be "ill-posed" and may even yield an ill-conditioned[1] matrix R.

If an acceptable formulation and a useful unfold algorithm for a given problem have been found, it is appropriate to inquire how random uncertainties ϵ_i in the data affect the spectral estimates F_j obtained in the unfold process. In this paper we report the results of two independent methods of estimating unfold uncertainty with the UFO (UnFold Operator)[8] computer code. One method is based on transformations of the error matrix, and the other utilizes the Monte Carlo technique.

II. The UFO solution and its built-in error propagation method

Several unfold computer codes have appeared over the past 30 years (*e.g.*, UNSPEC[9], STAYSL[10], YOGI[5], UFO) to estimate F_j from Eqs. (3). The UFO code, chosen for this work, is a matrix manipulation code. Instead of inverting Eqs. (3) directly, UFO minimizes the least squares residual $\chi^2(F_1, \ldots, F_N)$,

$$\chi^2(F) = \sum_{i=1}^{M} \frac{\left\{ D_i - \sum_{j=1}^{N} R_{ij} F_j \right\}^2}{\sigma_i^2} = (D - R{\cdot}F)^T W^{-1} (D - R{\cdot}F) \quad , \tag{5}$$

between the data D_i and data predictions $\Sigma_j R_{ij} F_j$ by solving the normal equations $R^T W^{-1} D = R^T W^{-1} R F$ for the components F_j. The solution method, due to Lawson and Hanson[11], is similar to Singular-Value-Decomposition[12], in which the inverse $A = (R^T W^{-1} R)^{-1} R^T W^{-1}$ is not needed. Even underdetermined systems ($M < N$) can be considered. In Eq. (5, right) the spectral components F_j and the data D_i have been written, respectively, as $1 \times M$ and $1 \times N$ (column) matrices. The $M \times M$ matrix W^{-1} is diagonal with elements $1/\sigma_i^2$, where σ_i^2 is the estimated variance of ϵ_i. These random variables for the data are assumed to be independent and normally distributed with zero mean. Additional fixed and inequality constraints, plus weighted curve fits and smoothing equations, can also be added to the UFO formulation.

For nonsingular problems, UFO propagates data uncertainties σ_i into its unfold as $\sigma(F) = AWA^T$, where the matrix W is the covariance matrix[13,14] for the data (here by assumption diagonal with elements σ_i^2), A is the inverse matrix (which can be constructed unless the normal equations are singular), and $\sigma(F)$ is the covariance matrix for the unfold. Diagonal elements of $\sigma(F)$ are the variances σ_j^2 in the components F_j,, while the j-j' off-diagonal elements are the covariances between the F_j and $F_{j'}$. Unfold components are usually correlated (*i.e.*, have non-zero covariances), even if the data are uncorrelated.

III. Monte Carlo Comparison Method

UFO is most often used as a "black box" operator, and so it is important to verify its built-in, error propagation estimates independently. Since the Monte Carlo technique has successfully been applied to similar uncertainty propagation problems[4,6,15,16], we have applied it to a non-trivial unfold problem and compared the resulting unfold uncetainty to the built-in (nominal) UFO estimates, predicted with error matrices.

The Monte Carlo method was formulated as follows: (1) nominal data $\{\Delta_i\}$ for the unfold problem were simulated from a known spectral function; (2) it was assumed that the simulated data were the population means for normal, independent *distributions* of possible measurements with variances σ_i^2 that could be obtained in hypothetical experiments; (3) using the GASDEV[14] algorithm, we constructed $N_{sample} = 100$ sets of perturbed data $\{D_i\}$, each drawn from the assumed distributions; (4) each perturbed data set was then inverted by the UFO algorithm to obtain a set of perturbed spectral estimates $\{F_i\}$ (The distributions of data thus produce distributions of spectral unfolds.); and (5) the averages $\langle F_i \rangle$ and sample variances $s(F_i)^2$ of the unfolded distributions were computed. These parameters correspond, respectively, to the simple unfold F_j^{nom} of the simulated data and to diagonal terms σ_j^2 of the propagated error matrix. While 100 data sets is not large by Monte Carlo standards, this number is sufficient to estimate variances to within 14% at the 95% confidence level[17]. In this study each iteration of UFO took a couple of minutes on a 3600 series VAX computer.

IV. A Test Problem for UFO: Results and Discussion

For the UFO comparison, input data were simulated from an arbitrarily chosen, black body spectrum ($T = 10$ keV) and a set of similar, but overlapping, response functions. Fig. 1 shows the known spectrum $F(E)$, the set of response functions $\{R_i(E)\}$, and individual spectrum-response products $R_i(E)F(E)$. The response functions are non-zero only over a finite energy range. Simulated data $\{\Delta_i\}$ were obtained as in Eqs. (1) by integrating the spectrum-response products from $E_1 = 0$ to $E_2 = 300$ keV; these data vary by about 8 orders of magnitude across the 10 channels. The data were also assumed to be uniformly uncertain by 5% (i.e., $\sigma_i = 0.05\Delta_i$). The *unfold spectrum* was approximated by 10 contiguous, histogram basis functions (not shown) of unit height, the boundaries of which coincided with the overlap points of the response functions and the domain of the problem. These boundaries define a partition of the domain into unfold bins which were numbered like the response functions.

This problem was purposely chosen to be difficult for most unfold algorithms. For example, numerical precision is expected to be a problem. That is, on many computers the number of significant figures for single precision, floating point arithmetic is ~ 7. For this problem, the least significant figure in data channel Δ_1 corresponds to the first significant figure in Δ_9, and linear combinations of the data may show round-off effects. The second hurdle in this problem is the extreme skewness of the spectrum for bins j = 3 - 10 (tail region). Here a significant fraction of the integral in each data channel comes from the spectrum in the preceding bin, making off-diagonal terms significant.

Figures 2 and 3 show the result of solving this problem with the built-in UFO algorithms. The histogram unfold F_j^{nom} of the simulated data and its estimated bounds ($F_j^{nom} \pm \sigma_j$) are illustrated in Fig. 2. The original spectrum is also shown for comparison. One observes that the nominal unfold closely matches the known black body spectrum out to bin j = 9. This result is probably due to the numerics issue, noted above, an interpretation supported by the relatively large error bounds. (We show below that the location of such disagreements depends on the unfold algorithm.) Fig. 3 shows the relative unfold uncertainty σ_j / F_j^{nom} in per cent (solid line). For the first few bins (where the spectrum is largest) 5% data uncertainties translate into 5-10% unfold uncertainties; but as the bin energy increases, the unfold uncertainty rapidly increases. Such behavior is due to overlap of the response functions and the declining tail region of the source spectrum. An anomalous break in this pattern is again visible in bin j = 9.

The results of running UFO with Monte Carlo perturbations are also shown in Figs. 2 and 3. The solid dots in Fig. 2 represent the averages $\langle F_j \rangle$ over the distribution of unfolds in the j-th energy bin. One notes good agreement with the nominal UFO unfold, even in the seemingly anomalous bin j=9. There are no normalizations between the curves. Fig. 3 shows the corresponding ratios $s_j / \langle F_j \rangle$. Again reasonable agreement is shown between the two error propagation methods, both using the UFO unfold algorithm.

Table I compares directly the results of the two error propagation methods. The nominal unfold and error matrix uncertainties agree with the corresponding Monte Carlo estimates to within the expected resolution of a 100 element statistical sample (95% confidence level), but there appears to be an unresolved bias of $\sim 10\%$ between the two uncertainty estimates. Similar agreement (not shown) exists between the error matrix and Monte Carlo estimates of the *covariance* between the spectral unfold components F_j and $F_{j'}$.

To see if the unfold and error estimates are algorithm dependent, the sample problem was also unfolded by direct inversion. That is, an inverse ρ was found for the response matrix R in Eqs. (3) and (4), using the SVD algorithm given by Press[14]. The basis functions were the same as in the UFO unfold. The unfold components for the direct inversion method were then found by back-substitution: *i.e.*, $F = \rho D$, where the matrix notation has again been used. Since R is tri-diagonal and non-negative, its inverse ρ is also *nearly* diagonal with elements which decrease in (absolute) value away from the diagonal. The unfold error matrix was computed as $\rho W \rho^T$ for the same data error matrix W as above.

Applying this direct inversion algorithm to the nominal data in this problem, one obtains the unfold (Direct Inverse) in Fig. 4 and the unfold uncertainties ("X") in Fig. 5. The results of the two UFO methods are also shown for comparison. In the direct inversion algorithm, the effects of round-off can be studied directly. For example, in Fig. 4 the direct inverse unfold drifts away from the source spectrum without recovery at bin $j = 8$, as sums and differences of large numbers rattle about. Similarly, the unfold uncertainty for the direct inverse (Fig. 5) *decreases* in bin $j = 8$ because the elements of ρ for this and higher energy bins no longer significantly couple in contributions from the relatively large initial data channels. In this example, all three unfold methods agree in unfold value and uncertainty up to about the 6-th bin, in which the spectrum is down from peak by about 4 orders of magnitude.

These comparisons suggest that the nominal, unfold uncertainty predictions in the UFO code are reasonable. At the 95% confidence level there is no reason to believe that the results of the error matrix and Monte Carlo methods differ, although a possible 10% bias was unresolved and may be algorithm dependent. The Monte Carlo method is also useful in underdetermined problems, for which the error matrix does not exist. Efforts are underway to speed up such computations.

ACKNOWLEDGEMENTS

This wok performed by Sandia National Laboratories is supported by the U. S. Department of Energy under contract DE-AC04-94AL85000.

REFERENCES:

1. I. J. D. Craig and J. C. Brown, *Inverse Problems in Astronomy* (Adam Hilger, Boston, 1986).
2. H. N. Kornblum and R. L. Kauffman, Rev. Sci. Instrum. **57(8)** 2179 (1986).
3. G. A. Chandler, *et. al.*, Rev. Sci. Instrum. **63(10)** 4828 (1992).
4. G. A. Carlson, D. L. Fehl, and L. J. Lorence, Jr.,Nucl. Instrum. and Methods **B62** 264 (1991).
5. S. G. Gorbics and N. R. Pereira, Rev. Sci. Instrum. **64(7)** 1835 (1993).
6. D. L. Fehl, R. J. Leeper, and R. P. Kensek, Rev. Sci. Instrum. **63(10)** 4786 (1992).
7. C. De Boor, *A Practical Guide to Splines*, (Springer-Verlag, New York, 1978).
8. L. Kissel, F. Biggs, and T. Marking, "UFO (UnFold Operator): Command Descriptions," Version 4.0, Sandia National Laboratories Report SAND82-0396 (Alququerque, New Mexico, Jan. 15, 1990), unpublished.
9. H. F. Finn, "UNSPEC Reference Manual", Version 10/29/82, Lawrence Livermore National Laboratory Report UCID-19616 (Livermore, California, 1 December 1982), unpublished.
10. F. G. Perey, "Least Squares Dosimetry Unfolding: The Program STAY'SL," Oak Ridge National Laboratories Report ORNL/TM-6062 (Oak Ridge, Tenn., Oct 1977), unpublished.
11. C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (Prentice-Hall, Englewood Cliffs, New Jersey, 1974)
12. G. E. Forsythe, M. A. Malcolm, and C. B. Moler, *Computer Methods for Mathematical Computations* (Prentice-Hall, Englewood Cliffs, New Jersey, 1977).
13. S. L. Meyer, *Data Analysis for Scientists and Engineers* (Wiley, New York, 1975).
14. W. H. Press, *et. al.*, *Numerical Recipes* (Cambridge University Press, New York, 1992).
15. D. G. Shirk and N. M. Hoffman, Rev. Sci. Instrum. **56(5)** 809 (1985).
16. D. L. Fehl, "Error analysis of an underdetermined, spectral unfold problem as applied to an x-ray absorption spectrometer," in the Proceedings of the Eighth IEEE International Pulsed Power Conference (San Diego, 16-19 June 1991), edited by R. White and K. Prestwich, p. 598.
17. M. G. Natrella, National Bureau of Standards and Technology, Report #PB93-196038 (National Technical Information Service, Springfield, VA), originally published as Experimental Statistics, NBS Handbook 91, reprinted with corrections in October 1966.

FIGURE CAPTIONS:

1. Known spectrum and response functions for the UFO test problem. The original black body spectrum, from which data were simulated, is shown as a solid line. The response functions $R_i(E)$ are shown as dashed lines at the bottom [$i = 1,2,\ldots, 10$ from lowest to highest energy]. Spectrum-response products $R_i(E)F(E)$ are shown as thin dotted lines.

2. Comparison of the original and unfolded spectra. The black body spectrum $F(E)$ is shown as a continuous dotted line, the nominal unfolded spectrum $\Sigma_j F_j^{nom} D_j(E)$ as a piecewise-constant solid line, and the average of the Monte Carlo unfolds $\langle F_j \rangle$ as a solid dot at the center of each unfold bin. Also shown as piecewise-continuous dashed lines are the unfold uncertainty bounds $\Sigma_j(F_j^{nom} \pm \sigma_j)D_j(E)$, due to data uncertainties (5%) and estimated with UFO's built-in error matrix method. The logarithmic scale makes these error bounds appear asymmetric about the nominal unfold; lower bounds are not visible in the last couple bins (highest energy) because $\sigma_j > F_j^{nom}$.

3. Comparison of the relative unfold uncertainty for the two error propagation methods. The piecewise-constant solid line indicates the ratio σ_j / F_j^{nom} (in %) from the error matrix estimate, and the solid dots represent the ratio $s(F_j)/\langle F_j \rangle$ (in %) from the Monte Carlo analysis, with 100 samples. The data were assumed to be uncertain by 5%.

4. Effect of a different unfold algorithm on the unfolded spectrum. Shown are $F(E)$ (continuous dotted line), the nominal unfolded spectrum from the UFO algorithm (piecewise constant solid line, as in Fig. 2), and the unfold (piecewise constant dashed line) from a direct inversion of Eqs. (3) and (4).

5. Comparison of relative unfold uncertainty for different unfold algorithms. Relative unfold uncertainties from UFO are shown as in Fig. 3. The relative unfold uncertainty estimates with the direct inverse algorithm (error matrix) are shown as "X".

Table I. Comparison of unfold estimates and uncertainties[a] derived by the Monte Carlo and Error Matrix methods and shown as corresponding ratios.

Bin	$\langle F_j \rangle / F_j^{nom}$	res.[b]	$s(F_j)/\sigma_j$	res.[c]	Bin	$\langle F_j \rangle / F_j^{nom}$	res.[b]	$s(F_j)/\sigma_j$	res.[c]
1	1.0014	0.012	1.009	0.140	6	0.979	0.073	1.100	0.140
2	1.0007	0.015	1.130	0.140	7	1.021	0.201	1.100	0.140
3	1.0043	0.027	1.090	0.140	8	0.990	0.164	1.110	0.140
4	0.9993	0.032	1.040	0.140	9	1.413	3.903	1.120	0.140
5	1.0073	0.060	1.060	0.140	10	0.972	0.303	1.120	0.140

Notes:

a. The unfold estimate $\langle F_j \rangle$ with associated uncertainty $s(F_j)$ is obtained by applying the Monte Carlo method to UFO. F_j^{nom} is the nominal UFO unfold from the simulated data; σ_j is then estimated by the error matrix method, built into UFO.

b. Resolution here means the estimated uncertainty in the ratio $\langle F_j \rangle / F_j^{nom}$ due to statistical fluctuations in $\langle F_j \rangle$ for 100 element samples and is determined by $s(F_j)$ and Student's t distribution (99 degrees of freedom, 95% confidence level, Ref. 17).

c. Resolution here means the estimated uncertainty in the ratio $s(F_j)/\sigma_j$ due to statistical fluctuations in $s(F_j)$ for 100 element samples and is determined by the χ^2 distribution (99 degrees of freedom, 95% confidence level, Ref. 17).

Fig 1

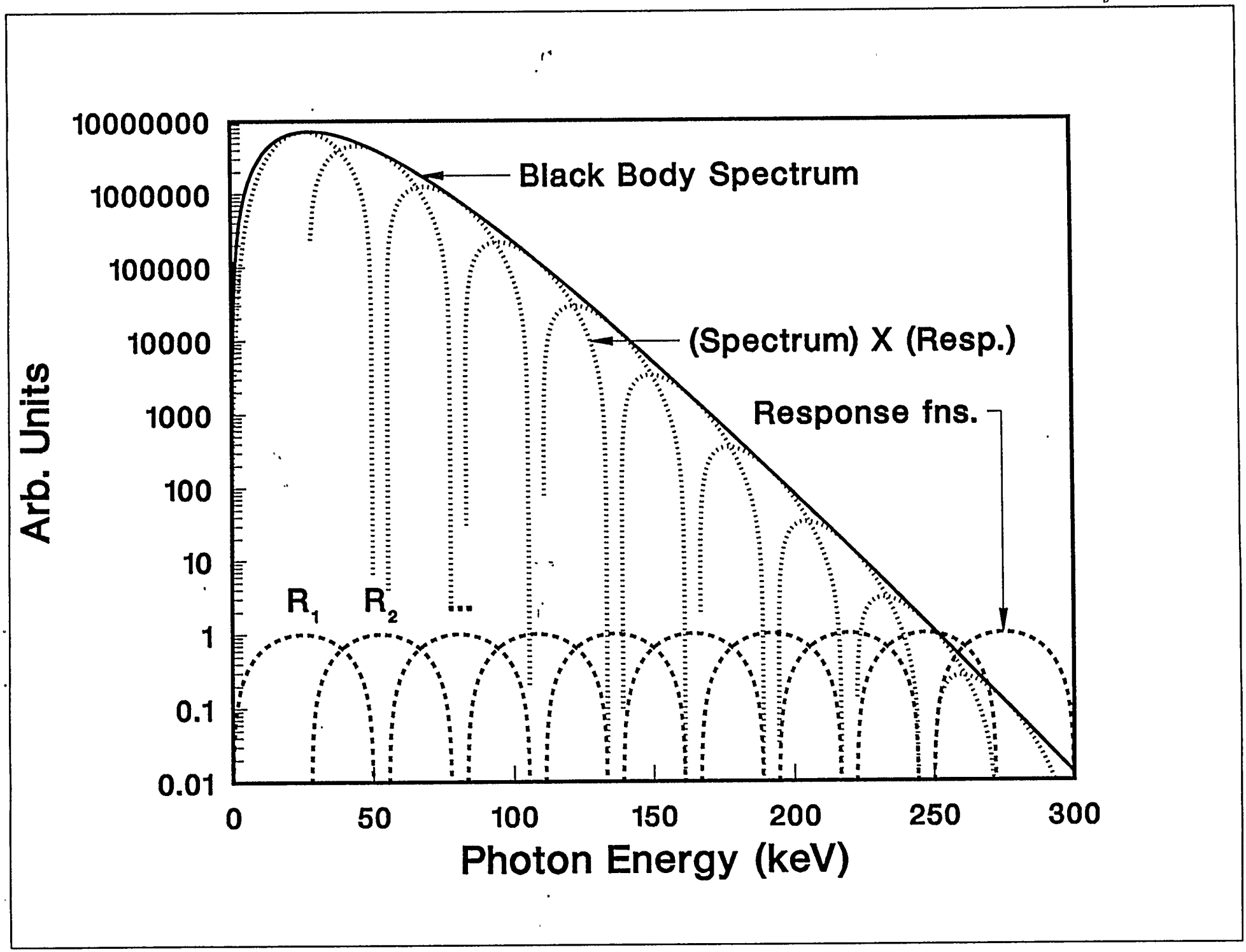

Black Body Spectrum
(Spectrum) X (Resp.)
Response fns.
R1
R2
...
Arb. Units
10000000
1000000
100000
10000
1000
100
10
1
0.1
0.01
0
50
100
150
200
250
300
Photon Energy (keV)

Fig 2

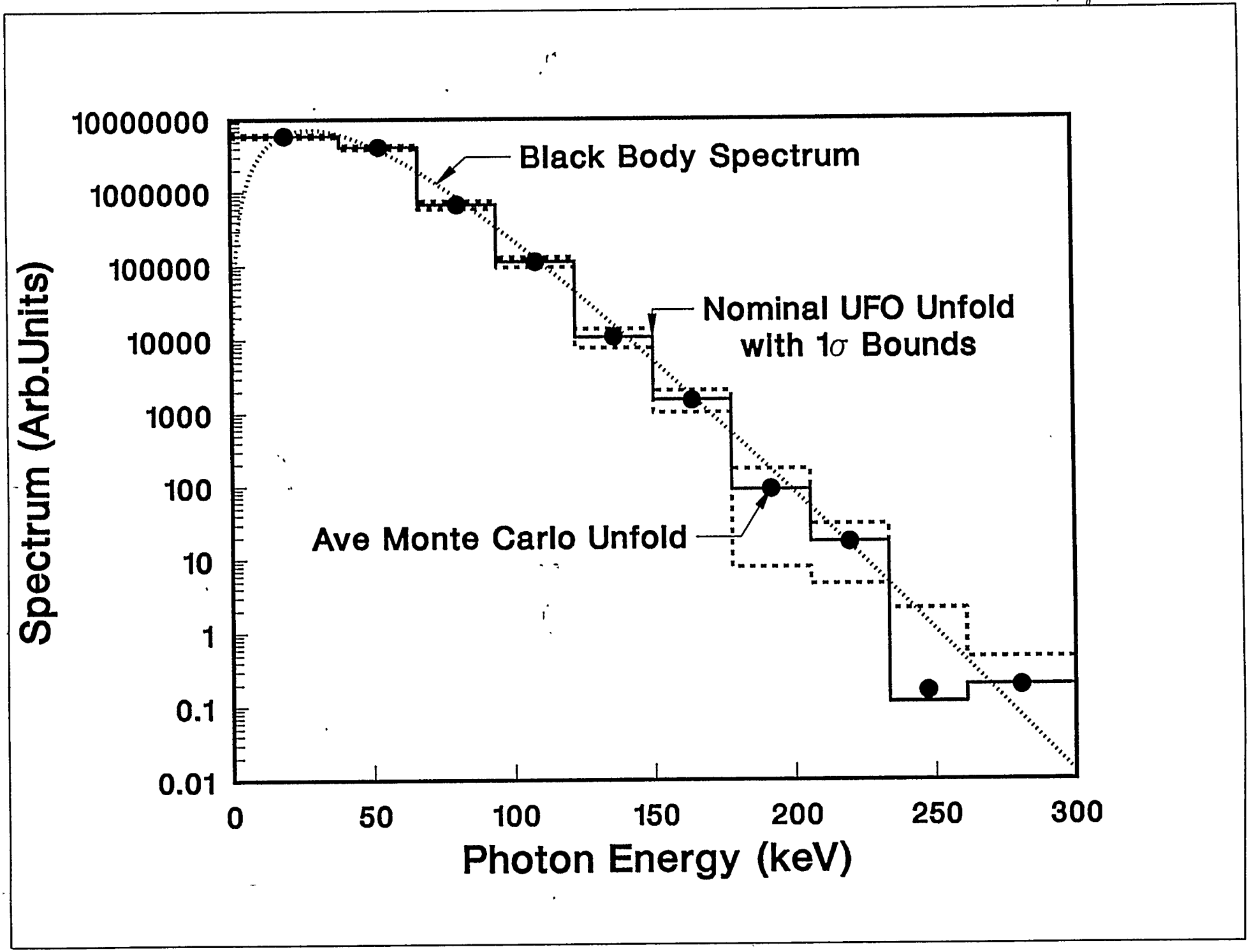

Black Body Spectrum
Nominal UFO Unfold
with 1σ Bounds
Ave Monte Carlo Unfold
Spectrum (Arb.Units)
10000000
1000000
100000
10000
1000
100
10
1
0.1
0.01
0
50
100
150
200
250
300
Photon Energy (keV)

Fig 3

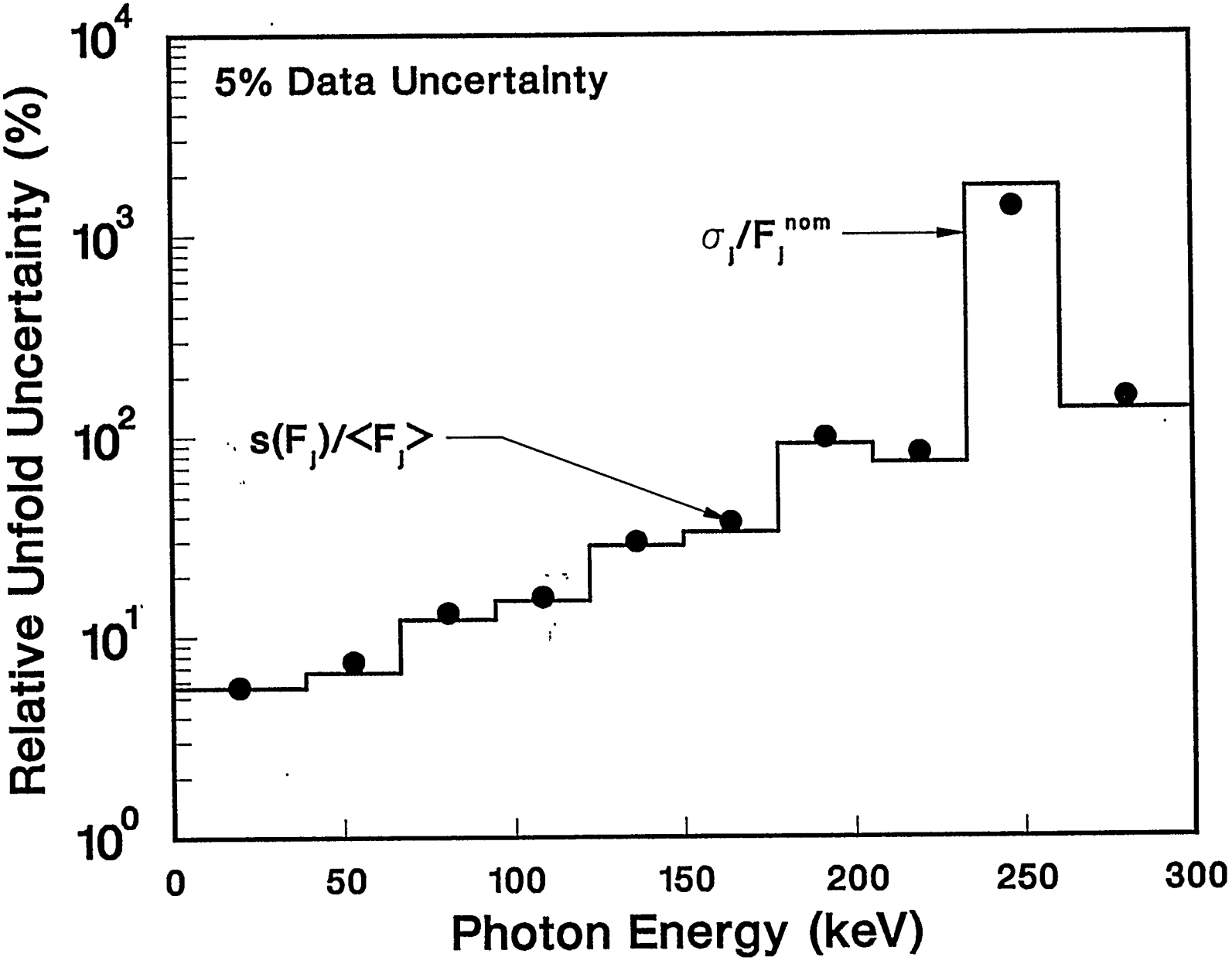

5% Data Uncertainty
σ_j/F_j^{nom}
$s(F_j)/\langle F_j \rangle$
Relative Unfold Uncertainty (%)
10^4
10^3
10^2
10^1
10^0
0
50
100
150
200
250
300
Photon Energy (keV)

Fig 4

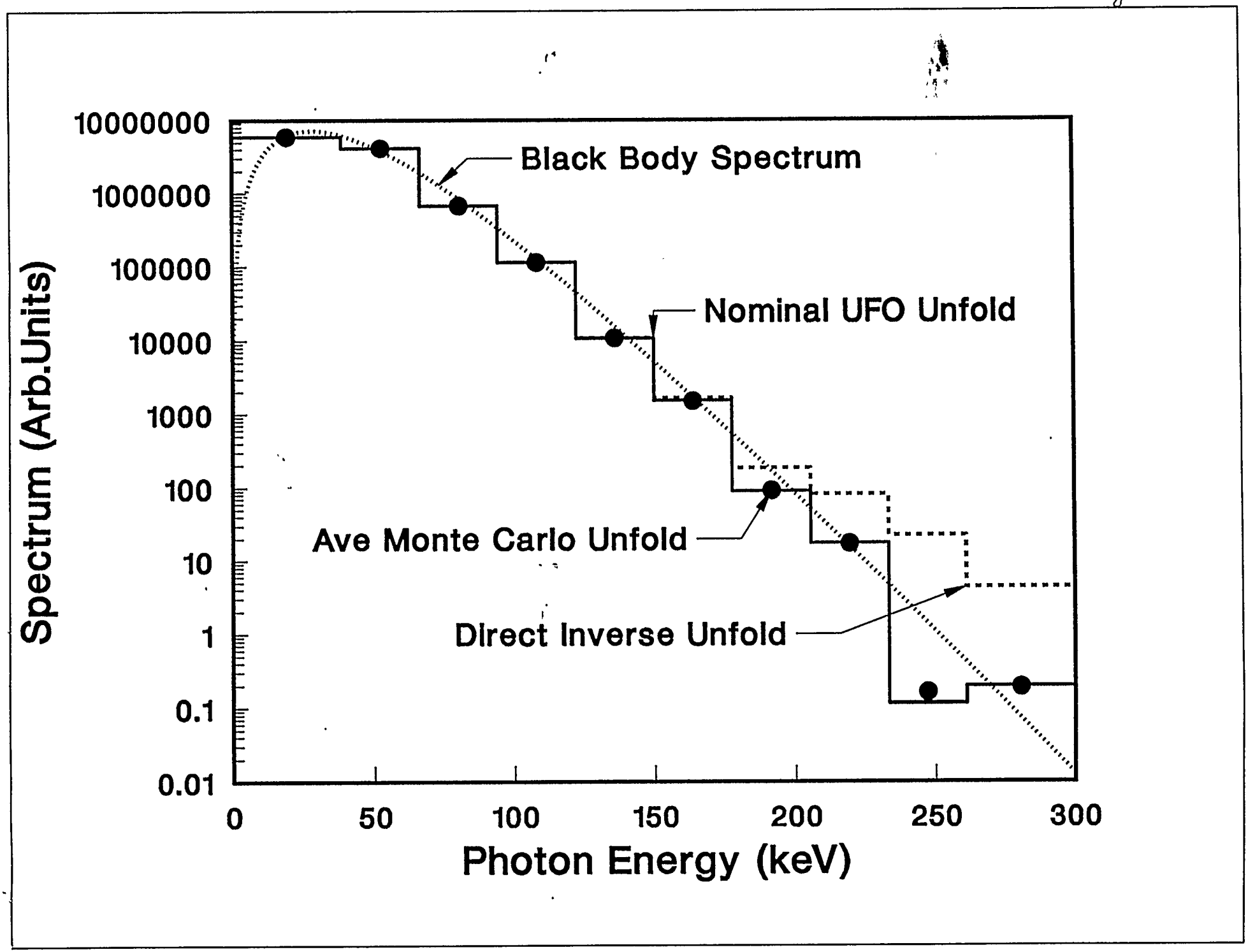
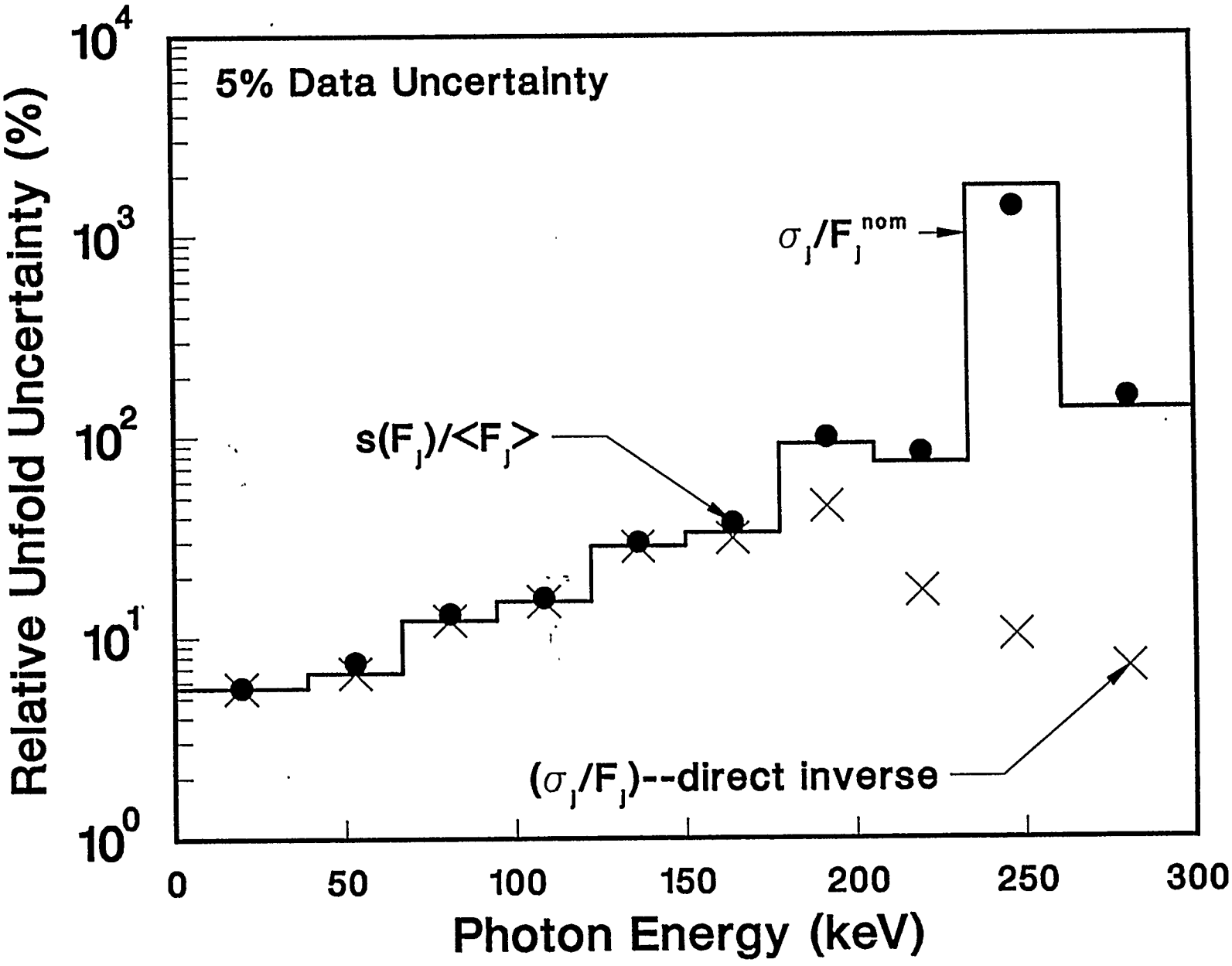

Spectrum (Arb.Units)
10000000
1000000
100000
10000
1000
100
10
1
0.1
0.01
0
50
100
150
200
250
300
Photon Energy (keV)
Black Body Spectrum
Nominal UFO Unfold
Ave Monte Carlo Unfold
Direct Inverse Unfold

Fig 5


5% Data Uncertainty
Relative Unfold Uncertainty (%)
10^4
10^3
10^2
10^1
10^0
σ_j/F_j^{nom}
$s(F_j)/\langle F_j \rangle$
(σ_j/F_j)--direct inverse
0
50
100
150
200
250
300
Photon Energy (keV)