

Title:

**MONTE CARLO ERROR ESTIMATION APPLIED
TO NONDESTRUCTIVE ASSAY METHODS**

Author(s):

Robert J. Estep, David Miko, and Sheila Melton

Submitted to:

<http://lib-www.lanl.gov/la-pubs/00357103.pdf>

MONTE CARLO ERROR ESTIMATION APPLIED TO NONDESTRUCTIVE ASSAY METHODS

Robert J. Estep, David Miko, and Sheila Melton
Los Alamos National Laboratory
P.O. Box 1663
Los Alamos NM 87545
restep@lanl.gov, dmiko@lanl.gov, smelton@lanl.gov

ABSTRACT

Monte Carlo randomization of nuclear counting data into N replicate sets is the basis of a simple and effective method for estimating error propagation through complex analysis algorithms such as those using neural networks or tomographic image reconstructions. The error distributions of properly simulated replicate data sets mimic those of actual replicate measurements and can be used to estimate the std. dev. for an assay along with other statistical quantities. We have used this technique to estimate the standard deviation in radionuclide masses determined using the tomographic gamma scanner (TGS) and combined thermal/epithermal neutron (CTEN) methods. The effectiveness of this approach is demonstrated by a comparison of our Monte Carlo error estimates with the error distributions in actual replicate measurements and simulations of measurements. We found that the std. dev. estimated this way quickly converges to an accurate value on average and has a predictable error distribution similar to N actual repeat measurements. The main drawback of the Monte Carlo method is that N additional analyses of the data are required, which may be prohibitively time consuming with slow analysis algorithms.

INTRODUCTION

Analysis algorithms used in nondestructive assay (NDA) have grown in complexity in recent years, with the introduction of tomography, neural networks, fuzzy logic, expert systems, list-mode neutron counting, and other innovations. Finding analytical expressions for the propagation of error through such algorithms can be exceedingly difficult, if not impossible. We have had some success in applying Monte Carlo simulation and other approximate methods to error estimation in the tomographic gamma scanner (TGS) and the combined thermal/epithermal neutron (CTEN) analysis algorithms. In this report we document these methods and their effectiveness and illustrate their use with some example cases.

APPROXIMATE METHODS FOR ERROR ANALYSIS

We are interested in multivariate problems with a set \mathbf{x} of m measured uncorrelated independent variables $\{x_1, x_2, \dots, x_m\}$ and n dependent variables $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$, related by a set \mathbf{f} of m operators $\{f_1(), f_2(), \dots, f_m()\}$ that may or may not be true functions, such that

$$y_i = f_i(x_1, x_2, \dots, x_M) \quad . \quad (1)$$

When the $f_i()$ are ordinary functions, an analytical expression for the error can be obtained using the well known formula¹

$$\sigma_{y,i}^2 = \sum_j \sigma_{x,j}^2 \left(\frac{\partial f_i}{\partial x_j} \right)^2 \quad . \quad (2)$$

For all cases discussed here, the $\{x_j\}$ values used as input are radioactive counts with no background subtraction, so that

$$\sigma_{x,j} = \sqrt{x_j} \quad . \quad (3)$$

The $\{y_i\}$ values will generally include the final mass or activity of at least one isotope. In addition, the errors in correction factors or in other intermediate quantities are often of interest and would then be included in the set. In some TGS analyses masses are determined independently for more than one energy or isotope, in which case the $\{y_i\}$ values correspond to the different masses.

Approximation of Partial Derivatives

For well-behaved functions eqn. (2) can be approximated by using

$$\sigma_{y,i}^2 \cong \sum_j \sigma_{x,j}^2 \left(\frac{\Delta y_i}{\Delta x_j} \right)^2 \quad , \quad (4)$$

where Δy_i is the change in y_i caused by substituting $x_j \rightarrow x_j + \Delta x_j$ one j at a time. This method, which we will refer to here as the "dy/dx" method, is well suited for passive list counting, where the $\{x_j\}$ values are the counts in the Feynman and signal-triggered histograms for the background and foreground counts. The Δx_j values are set to σ_{x_j} . We typically use 11 channels in each histogram, for a total of $m=44$ input values. Thus, the analysis of the histograms must be run 44 times to estimate the error. This is a fast analysis, as the lists are only sorted once to produce the unrandomized histograms, so no noticeable delay is caused by these repetitions. This would not be the case for TGS, which is a relatively slow analysis with a much larger number of input values.

The TGS analysis cannot use the approximation in eqn. (4) for another important reason, which is that the $f_i()$ are not well-behaved functions for most analysis options. The NNLS algorithm, for example, does a constrained least squares fit that removes some image voxels from the problem and forces their values to zero. A change Δx_j in a single input variable can alter the selection of which voxels to suppress, which would result in a stepped response in the output Δy_i rather than a continuous response. This is a threshold effect; a slightly smaller change might not alter the selection of voxels to suppress, so the response Δy_i can be discontinuous for changes in Δx_j . This means that one must use several values of Δx_j in the analysis, weighted by their probabilities, to get a good estimate of $\Delta y_i/\Delta x_j$. Many reconstruction algorithms display this kind of discontinuous behavior, as do some neural networks and fuzzy logic analyses. For this reason, of the NDA methods discussed here, only the passive list analysis can use the approximation in eqn. (4) without refinements.

The Monte Carlo Replicate (MCR) Method

In the Monte Carlo replicate (MCR) method one generates N randomized replicate input data sets $\{^k x_j\}$, $k = 1, \dots, N$, and analyzes each replicate set as if it were an independent measurement. Since they are counts, the replicates $\{^k x_j\}$ are generated using a Poisson random number generator, $P()$, which takes the presumed mean value of a count and returns a pseudo-random number with a Poisson distribution about that mean. That is,

$$^k x_j = P(x_j) \quad , \quad (5)$$

where the k means simply that this was the k 'th Poisson deviate generated in this way. Note that the $\{x_j\}$ values used in eqn. (7) are not mean values, but are themselves Poisson deviates (generated by nature) about the true mean values. This same approximation, of course, is made when applying any error estimation method to real data.

The $\{y_i^k\}$ are given by applying eqn. (1) to each replicate set,

$$y_i^k = f_i(x_1^k, x_2^k, \dots, x_M^k) \quad . \quad (6)$$

This yields N sets of output values $\{y_i^k\}$ that can be used to estimate std. dev. (σ) values using

$$\sigma_{y,i} \cong \sqrt{\frac{\sum_{k=1}^N (y_i^k - \bar{y}_i)^2}{N-1}} \quad , \quad (7)$$

where

$$\bar{y}_i = \frac{\sum_{k=1}^N y_i^k}{N} \quad . \quad (8)$$

The σ values estimated in eqn. (7) can in principle be made arbitrarily accurate, in the sense that they approach the best estimate that can be made from the data, by using a large enough number N of replicates. For analyses that are relatively slow, however, it is important to know the smallest number of iterations that gives a usable error estimate. The rule for actual repeat measurements is that the error in $\sigma_{y,i}$, (i.e., $\sigma(\sigma_{y,i})$) can be estimated from¹

$$\sigma(\sigma_{y,i}) = \frac{\sigma_{y,i}}{\sqrt{2(N-1)}} \quad , \quad (9)$$

where N is the number of repeated measurements. In principle this rule should apply as well to the MCR estimate in eqn. (7).

RESULTS AND DISCUSSION

In this section we describe the application of the MCR and dy/dx methods to a few specific cases. These are simple counting, TGS,² and passive list-mode neutron counting using the gray barrel methods of Brunson, et al..³ In addition we present selected data from measurements and simulations that demonstrate the effectiveness of these methods.

Example Case: Simple Counting

Counting a source for a fixed interval with no background subtraction is a useful example because of its simplicity and illustrates both the strengths and weaknesses of the MCR method. In this case there is one input and output value, $y = x$ counts, so $\sigma_y = \sigma_x$ is estimated using

$$\sigma_y = \sqrt{x} \quad . \quad (10)$$

We used a Poisson random number generator to simulate 20 separate measurements of the counts x for means values $x_{\text{avg}} = 10$ and 10000 counts, ranging from $\sigma = 32\%$ to $\sigma = 1\%$. The results for our analysis of all 20 of the $x_{\text{avg}} = 10$ measurements are illustrated in fig. (1), which shows the ratio of the estimated error σ_y from the MCR method to the root-x estimate of eqn. (10) as a function of the number of replicates performed, both for individual cases (A) and for the average of the 20 cases (B, dark circles). Also shown in (B), as open squares, is $\sigma(\sigma_y)$ for the MCR estimates, with the predicted value from eqn. (9) shown as the solid line. Fig. (2) shows the same plots for the $x = 10,000$ counts case.

What we see in comparing fig. (1) and (2) is that they are very similar in terms of convergence to the correct average value with increasing N . The average of the 20 separate measurements converges to the correct value very quickly, within 5 to 10 iterations. What's more, the std. dev. $\sigma(\sigma_y)$ for the MRC error estimate agrees very well with the predicted value from eqn. (9) in both cases.

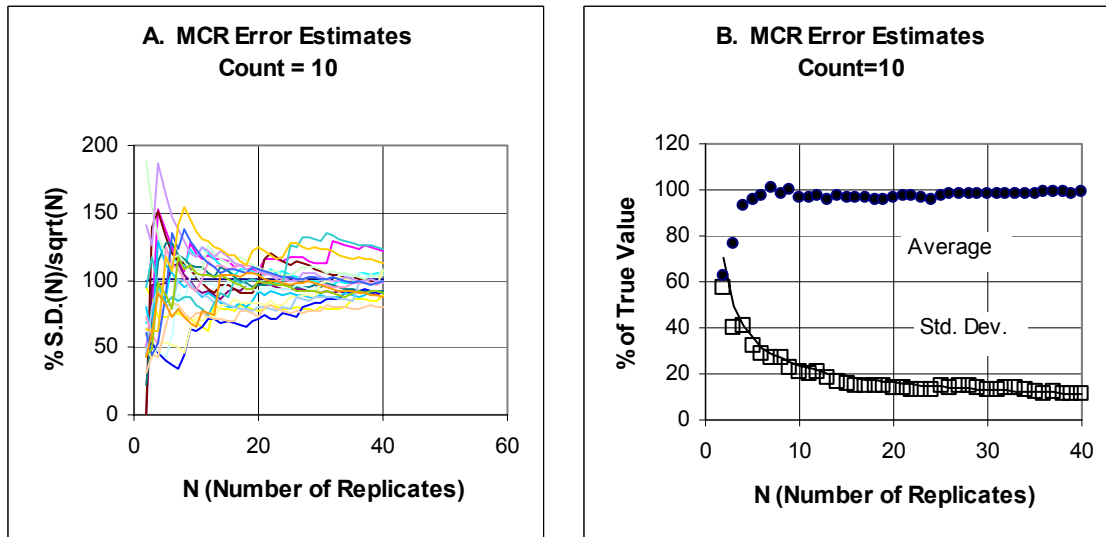


Figure 1. Performance of the MCR method for $x(\text{avg}) = 10$ counts. (A) The ratio of the estimated error from the MCR method to the root- x estimate as a function of N , the number of replicates performed. (B) The same ratio averaged over all 20 measurements (dark circles), compared with the std. dev. in the MCR error estimate relative to the root- x estimate. The line is the predicted error from eqn. 9.

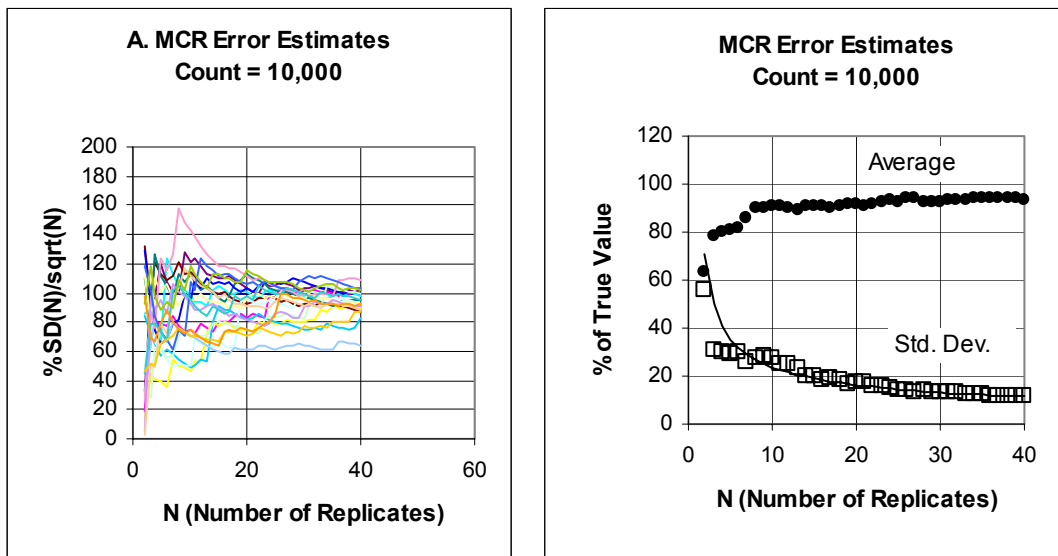


Figure 2. Performance of the MCR method for $x(\text{avg}) = 10,000$ counts. (A) The ratio of the estimated error from the MCR method to the root- x estimate as a function of N , the number of replicates performed. (B) The same ratio averaged over all 20 measurements (dark circles), compared with the std. dev. in the MCR error estimate relative to the root- x estimate. The line is the predicted error from eqn. 9.

TGS

The TGS_FIT software^{4,5} uses all of the region-of-interest sums generated by the WIN_TGS⁶ acquisition code (a peak sum and two background sums for each gamma-ray peak) for both the transmission and emission scans as the input Poisson variable set $\{x_j\}$. In the usual 16-layer, 150-views/layer scan, this generates $m = 3*2400*(n_e+n_t)$ values to be randomized, where n_e is the number of emission energies and n_t is the number of transmission energies. For example, with three transmission and three emission energies in the analysis, the number of measurements m in the set $\{x_j\}$ is 43,200. The output $\{y_i\}$ values are the n_e emission peak total mass values, each of which is the sum of the individual voxel masses in the emission image. The number of MCR method repetitions to perform can be set in software on the TGS_FIT command line. We have used 10 MCR iterations in the past to keep the analysis time under three minutes. With the much faster PC clone computers available today, a typical multiple-energy TGS analysis with 10 replicates could be done in well under a minute, so it should be feasible to go to a higher number of replicates.

The effectiveness of the MCR for TGS can be judged from the results published by Bonner, et al.,⁷ on Pu-239 assays performed by several instruments on 25 real waste drums from TA-55, including the original TA-18 prototype TGS scanner.² Eight replicate measurements were performed on each drum, which allowed an estimate to be made of the "true" std. dev. in the Pu-239 masses. Although eight replicates is a small number from which to get a good estimate of the true $\sigma_{y,i}$, we can use eqn. (9) to estimate the std. dev. in that value (i.e., $\sigma(\sigma_{y,i})$). For 8 replicates, $\sigma(\sigma_{y,i}) = 0.267*\sigma_{y,i}$. Thus, when the agreement between the measured std. dev. and that estimated using the MCR method is not obvious, one can at least say whether it appears to be statistically significant. Table I compares the measured and estimated std. dev. values of the Pu-239 mass measured using the TA-18 prototype. Only 24 cases are listed, as one of the drums was too heavy to place on our old prototype (newer TGS systems have no problem with drums of that weight).

Examination of the estimated and measured errors in Table I reveals good agreement, especially considering the difficulty of the problem. The error distributions in TGS assays are unusual, due the use of variance reduction strategies at lower masses and to the "zero-clipping" effect inherent in constrained fits, which do not allow negative values in the reconstructed emission image. The

**Table I. Comparison of measured and MCR errors in TGS^a :
Assays of Real Waste Drums From TA-55 (LANL)**

Drum ID	Description	Pu-239 mass	Measured std. dev. ^b	+/- σ (σ)	MCR std. dev. ^c	difference/ σ (σ)
LA56086	plastic/kimwipes	4.908 g	0.029 g	0.008 g	0.066 g	4.77
LA56038	salt, chloride, MgO	96.2	1.51	0.404	1.16	0.87
LA56147	non-actinide metal	0.664	0.113	0.030	0.061	1.72
LA56088	non-actinide metal	16.37	0.28	0.075	0.21	0.94
LA56159	<i>plastic/kimwipes</i>	<i>(0.096)</i>	<i>0.19</i>	<i>0.051</i>	<i>0.012</i>	<i>3.50</i>
LA56144	<i>graphite</i>	<i>(0.048)</i>	<i>0.017</i>	<i>0.005</i>	<i>0.018</i>	<i>0.22</i>
LA56140	glass	7.81	0.08	0.021	0.12	1.87
LA56031	HEPA filters	3.43	0.093	0.025	0.069	0.97
LA56072	non-actinide metal	25.02	0.27	0.072	0.33	0.83
LA56151	plastic/kimwipes	10.8	0.05	0.013	0.11	4.49
LA56118	plastic/kimwipes	44.1	0.13	0.035	0.19	1.73
LA56158	<i>plastic/kimwipes</i>	<i>(0.065)</i>	<i>0.02</i>	<i>0.005</i>	<i>0.022</i>	<i>0.37</i>
LA56089	non-actinide metal	9.47	0.19	0.051	0.15	0.79
LA56154	<i>plastic/kimwipes</i>	<i>(0.021)</i>	<i>0.01</i>	<i>0.003</i>	<i>0.016</i>	<i>2.24</i>
LA56087	plastic/kimwipes, rubber, paper/wood	21.47	0.17	0.045	0.19	0.44
LA56061	<i>plastic/kimwipes</i>	<i>(0.027)</i>	<i>0.021</i>	<i>0.006</i>	<i>0.014</i>	<i>1.25</i>
LA56130	glass ^d	4.2	0.33	0.088	0.24	1.02
LA56162	<i>rubber</i>	<i>(0.016)</i>	<i>0.015</i>	<i>0.004</i>	<i>0.015</i>	<i>0.00</i>
LA56156	<i>plastic/kimwipes</i>	<i>(0.007)</i>	<i>0.009</i>	<i>0.002</i>	<i>0.008</i>	<i>0.42</i>
LA56138	plastic/kimwipes	25.88	0.28	0.075	0.19	1.20
LA56078	<i>plastic/kimwipes</i>	<i>(0.066)</i>	<i>0.03</i>	<i>0.008</i>	<i>0.021</i>	<i>1.12</i>
LA56123	non-actinide metal	68.27	1.04	0.278	0.87	0.61
LA55925	MgO, salt	108	2.55	0.682	0.99	2.29
LA56098	plastic/kimwipes	112 g	0.49 g	0.131 g	0.42 g	0.53 g

a. Entries in italics, with the masses in closed parentheses, are below the nominal .5 g sensitivity of TGS

b. Std. Dev. in the eight measured masses

c. 10 MCR replicates were used in the TGS analysis

d. This drum was identified by CTEN and TGS counting as containing Curium

result is that SNM masses well below the nominal TGS sensitivity can exhibit rel. std. dev. values in the neighborhood of 10%, which could lead one to think that an assay is valid when the predicted mass value is in fact not reliable. That is, the usual "3-sigma" rule for the minimum detectable activity should not be used on TGS mass values. Those assays showing less than 0.5 g Pu-239 are indicated in the table with italic type. There are three approx. 4-sigma outliers in the table, as indicated by the last column, which shows the difference between the measured and estimated errors as a multiple of σ ($\sigma_{y,i}$). However, 50% of the values were within one times σ ($\sigma_{y,i}$) and 75% were within 2 times σ ($\sigma_{y,i}$), which is consistent with the predicted 23.6% std. dev. value for $N=10$ MCR iterations. Values of 67% and 95%, respectively, would be expected if the MCR estimates were in complete agreement with the measurements and exhibited no scatter.

Passive List-Mode Neutron Counting

As was mentioned above, we sort passive list-mode foreground and background data only once into signal-triggered and Feynman histograms. These histograms counts are Poisson-distributed, and are used as the input $\{x_j\}$ values in computing errors using both the MCR and dy/dx approaches. The latter is expected to be much more accurate than the MCR approach when it can be used, particularly when the number of MCR replicates is small. There are several output variables of interest in the gray barrel methodology, which actually comprises at least three different analysis methods. These are combined in different ways in the different analysis options. Two parameters of interest are termed ***nmr0*** and ***ratio2***, which are the mean of the net cluster probability distribution and the ratio of the triples to doubles rates, respectively. We examined the error levels in these as a function of the number of MCR iterates for 20 simulated list-mode measurements generated by the Los Alamos MULT_SIM program, which creates pseudo-random list data with user-specified properties. The simulated list data were analyzed using the Los Alamos CTEN_FIT^{8,9} software.

Two cases that we examined use list data simulated for a counter dieaway-time of 40 μ s, a singles counting efficiency of 0.15, and neutrons rates of 1000 nps (neutrons per second) from true fission neutrons and 500 nps from uncorrelated (α , n) neutrons. The first set of simulations, "Set 1," uses a 300-s count time, while the second set, "Set 2," uses a 30 second count. Figure (3) shows the average of the MCR error and its std. dev. for the Set 1 ***nmr0*** and ***ratio2*** parameters as a function of the number of MCR iterates. Once again the eqn. (9) predicted error in the std. dev. is shown as a solid line. Figure (4) shows similar results for Set 2. The dashed lines at the tops and bottoms in all four plots are the average and std. dev. of the errors computed with the dy/dx approach (eqn. 4), which do not depend on the number of MCR iterates.

Fig. (5) shows the average 20-iteration MCR and dy/dx method error estimates plotted vs. the true error for the five Set1 list mode parameters ***nmr0***, ***nmr1***, ***nmr2***, ***ratio1***, and ***ratio2***. The excellent accuracy of the (average) error estimates over the range 1% to 100% error is apparent. Also, note the very good agreement between the MCR and dy/dx methods, which, as has also been apparent in the previous figures, generally give the same average result after 10 to 20 iterations of the MCR method, although it takes many more iterations before the scatter in the MCR estimate converges to that of the dy/dx method.

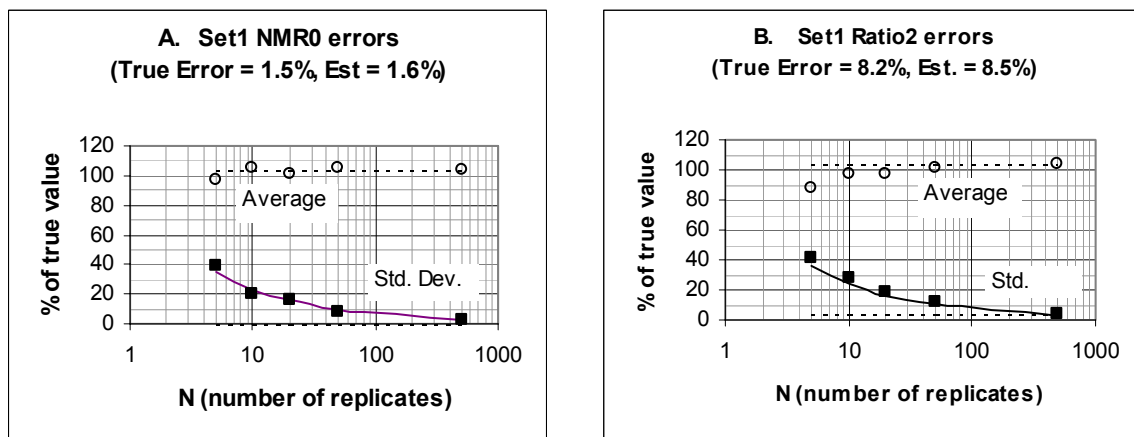


Figure 3. (A) Ratio of the true and estimated MCR errors for the Set 1 **nmr0** passive list counting parameter as a function of the number of MCR replicates N , averaged over twenty (simulated) measurements (open circles). Std. Dev. in the MCR error estimates are also plotted (dark squares), with a solid line showing the predicted value. The dashed lines at the top and bottom are the dy/dx method estimates. (B) Similar graph for the Set 1 **ratio2** parameter. The estimated errors (Est) in the caption are for the dy/dx method.

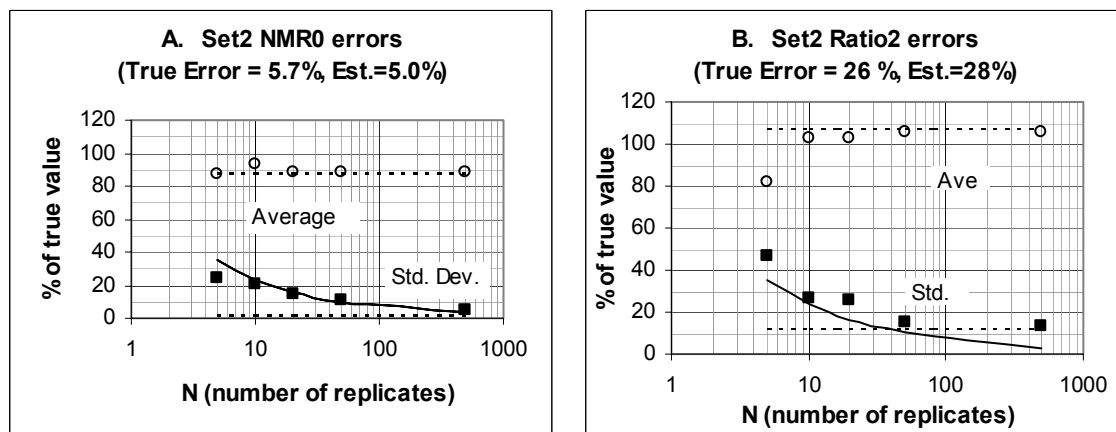


Figure 4. (A) Ratio of the true and estimated MCR errors for the Set 2 **nmr0** passive list counting parameter as a function of the number of MCR replicates N , averaged over twenty (simulated) measurements (open circles). Std. Dev. in the MCR error estimates are also plotted (dark squares), with a solid line showing the predicted value. The dashed lines at the top and bottom are the dy/dx method estimates. (B) Similar graph for the Set 2 **ratio2** parameter. The estimated errors (Est) in the caption are for the dy/dx method.

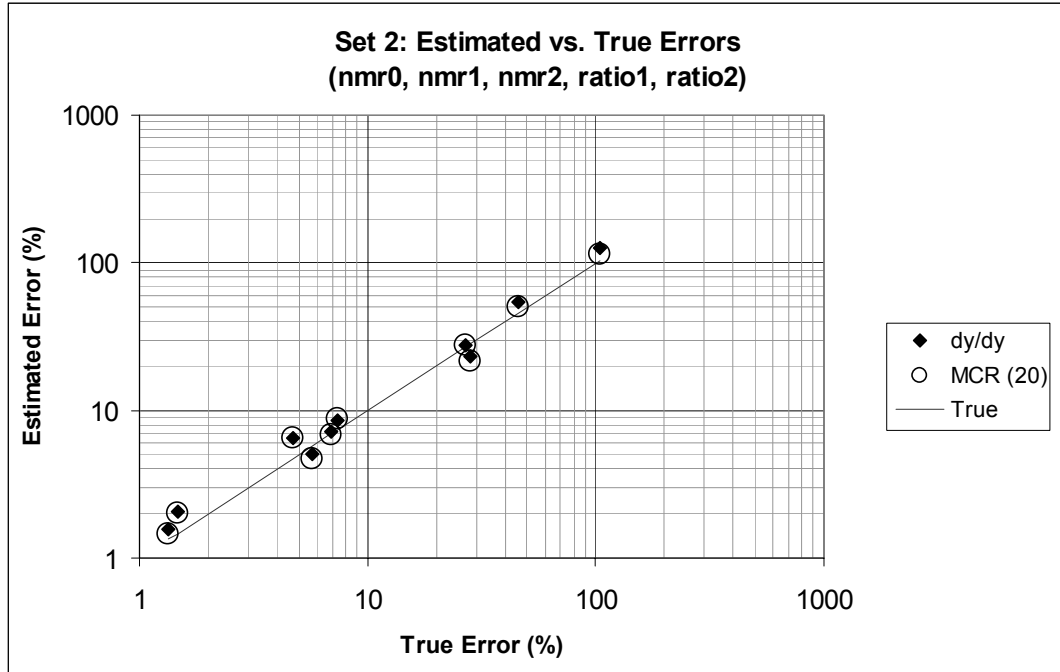


Figure (5). The average 20-iteration MCR and dy/dx method error estimates plotted vs. the true error for the five Set1 list mode parameters ***nmr0***, ***nmr1***, ***nmr2***, ***ratio1***, and ***ratio2***.

CONCLUSIONS

Our conclusion is that the MCR method is a good alternative for estimating errors in difficult situations where an analytical expression cannot be found and the dy/dx method is not practical. It gives good estimates on average and the scatter in the estimates is predictable. In practice, the MCR approach is very similar to making actual replicate measurements in order to measure the error, and is almost certainly easier and faster. The drawback is that, as with actual replicate measurements, because of the scatter a large number of repetitions are needed to ensure an truly accurate result on every case, and this may not be feasible.

There are questions that arise in connection with using MCR error estimation when some limit or cutoff is involved, say, when a 3-sigma pass/fail test is performed on an assay result. In such situations it might be advisable to add the scatter ($\sigma(\sigma_{y,i})$) in quadrature with the error estimate ($\sigma_{y,i}$) in order to get a conservative value of the error.

The dy/dx method is a better approach than MCR in those situations where it can be used. Although we ruled it out for all but one of our applications, when only a few input variables are involved it might be worth the effort to refine the method for use with badly behaved analyses, such as neural networks that make classifications. As mentioned above, a reasonable refinement would be to use several values of Δx_j , weighted by their probabilities, for each input variable x_j .

ACKNOWLEDGEMENT

This research was funded by the Mixed Waste Focus Area, Idaho National Engineering and Environmental Laboratory, and by the DOE, Environmental Management Science and Technology (EM-50).

REFERENCES

1. Leo, W. R., *Techniques for Nuclear and Particle Physics Experiments*, Chapter 4, Springer-Verlag, Berlin, Germany (1987).
2. R. Estep, et al., "Tomographic Gamma Scanning to Assay Heterogeneous Radioactive Waste," Nucl. Sci. and Eng., 118 (1994) p 145-152.
3. G. Brunson and N.J. Nicholas, "Shift-Register Neutron-Coincidence Counting and the Gray Barrel Problem," Los Alamos Rept. LA-12414-MS (Oct. 1992).
4. R. Estep, "TGS_FIT: Image Reconstruction Software for Quantitative, Low-Resolution Tomographic Assays," Los Alamos Rept. LA-12497-MS, Jan. 1993.
5. R. Estep, "TGS_FIT 3.1 User's Manual," to be published.
6. R. Estep and J. Cavender, "The WIN_TGS software package for tomographic gamma scanner systems," Proc. 35th INMM meeting, Naples, FL, (July 17-20, 1994).
7. C. Bonner, et al., "Preliminary report on the comparison of multiple non-destructive assay techniques on LANL plutonium facility waste drums," Proc. 39th INMM meeting, Naples, FL, (July 26-30, 1998).
8. R. Estep, C. Buenafe, and S. Melton, "User's Manual for the CTEN_FIT Software," Los Alamos Rept. (in publication).
9. R. Estep, C. Buenafe, and S. Melton, "Integration of TGS and CTEN assays Using the CTEN_FIT Analysis and Databasing Software," presented at this conference.