Monte Carlo Source Convergence and the Whitesides Problem

Roger N. Blomquist
Reactor Analysis Division
Argonne National Laboratory*
9700 S. Cass Ave.
Argonne, IL 60439
RNBlomquist@anl.gov

ABSTRACT

The issue of fission source convergence in Monte Carlo eigenvalue calculations is of interest because of the potential consequences of erroneous criticality safety calculations. In this work, we compare two different techniques to improve the source convergence behavior of standard Monte Carlo calculations applied to challenging source convergence problems The first method, superhistory powering, attempts to avoid discarding important fission sites between generations by delaying stochastic sampling of the fission site bank until after several generations of multiplication. The second method, stratified sampling of the fission site bank, explicitly keeps the important sites even if conventional sampling would have eliminated them. The test problems are variants of Whitesides' "Criticality of the World" problem, in which the fission site phase space was intentionally undersampled in order to induce marginally intolerable variability in local fission site populations. Three variants of the problem were studied, each with a different degree of coupling between fissionable pieces.

Both the superhistory powering method and the stratified sampling method were shown to improve convergence behavior, although stratified sampling is more robust for the extreme case of no coupling. Neither algorithm completely eliminates the "loss" of the most important fissionable piece, and if coupling is absent, the lost piece cannot be recovered unless its sites from earlier generations have been retained.

Finally, criteria for measuring source convergence reliability are proposed and applied to the test problems.

*. Work supported by the U. S. Department of Energy Office of Environmental Management under Contract W-31-109-Eng-38.

1. INTRODUCTION

The issue of convergence of Monte Carlo eigenvalue calculations is of interest because of the potential consequences of erroneous criticality safety calculations. The “Criticality of the World” problem, reported by Whitesides[1], identified poor convergence as a pitfall in Monte Carlo eigenvalue calculations of an array of very weakly coupled fissionable pieces. At issue was the eigenvalue of a postulated storage building with concrete walls (modeled with water) surrounding a 9x9x9 array of identical, widely separated ^{239}Pu spheres. The eigenvalue of this system is ~ 0.93, but when the central sphere was replaced with one with twice the volume of a regular sphere and which is critical by itself, the eigenvalue estimate remained unchanged. A large uncertainty is normally estimated for unconverged problems, but in this case, the uncertainty estimate was typical of a well-converged Monte Carlo. Only the analyst’s a priori knowledge of the eigenvalue aroused suspicion toward the results. In the codes tested in this work, both automated convergence checking and methods aimed at convergence reliability make the chances of convergence failures more remote.

We wish to discriminate between two forms of fission source convergence problems. The first, deterministic slow convergence in both Monte Carlo and deterministic calculations, occurs in loosely coupled systems even if the number of neutron histories in each generation is infinite (the deterministic limit), and is remedied in deterministic transport calculations by a variety of acceleration methods. The second, the “Criticality of the World” problem, afflicts only Monte Carlo calculations and is a symptom of stochastic undersampling. Each of these problems is perhaps best attacked with a different solution.

2. THE METHODS

In this work, we compare two methods aimed at improving Monte Carlo source convergence reliability. The main differences between the two methods as implemented are (1) the kind of event that generates a potential fission site, and (2) the number of generations that passes between stochastic sampling of the fission site vector, (3) whether additional effort is made to keep fission sites in all fissionable pieces, and (4) whether fission sites from earlier generations are retained. The superhistory method is aimed at not discarding important fission sites during sampling of the fission site vector until enough generations have transpired to permit natural production of relatively many progeny sites. In contrast, the fission source stratified sampling method is aimed at preventing the sampling process from removing all of the sites from any single fissionable piece from the source vector during the renormalization at the end of each generation. The methods are applied to Monte Carlo calculations with control parameters (number of histories per generation, number of generations applied to source convergence) that are intentionally marginalized to make convergence reliability differences between methods discernible.

2.1 STRATIFIED SAMPLING

Gelbard and Roussell[2] proposed a stratified fission source sampling method that prevents fissionable pieces from being “lost” during a Monte Carlo calculation. Such losses occur due to the variability in the local fission site population if the average number of particles per fissionable piece is so small (i.e., several) that the statistical variation in the local fission site population is similar in magnitude to the local site population. Of course, it is always possible to guarantee preservation of the critical piece in the fission source sample by specifying a large number of neutrons per generation, but such an approach suffers from several drawbacks. The appropriate number of neutrons may be difficult to determine if there are a great many loosely coupled fissionable pieces; and it may be a very large number, requiring unacceptably long computing times. Stratified sampling tests[2] were conducted on one-dimensional slab diffusing systems using a greatly simplified Monte Carlo algorithm that simulates only the birth and absorption of neutrons. The fissionable pieces were made very thick compared to a migration length, nearly eliminating the coupling between the pieces, and one of the slabs was made much more reactive than the others. As a result of the poor coupling, the standard fission site algorithm converged the fission source spatial distribution to a random eigenvector. The stratified sampling method, however, more reliably converged the fission source distribution to the fundamental mode, even for quite small numbers of neutrons per fissionable piece. The dramatic improvement in the reliability of convergence for the diffusion problem indicated promise for less idealized systems.

Mohamed and Gelbard[3] then applied both a conventional algorithm and the stratified sampling method implemented in VIM[4] to the Whitesides problem to assess the resulting reliability improvement. They discovered that although the Whitesides problem's convergence might have been affected by the fission site variability problem, it's primary difficulty stemmed from the slow convergence inherent in loosely coupled systems. (In fact, for these kinds of problems, convergence is slow even if deterministic methods are used.) To determine the influence of neutronic coupling on source convergence, they also analyzed two more extreme variants of the Whitesides problem -- one with lower coupling due to the removal of the water reflector surrounding the array, and another with coupling between spheres completely eliminated by introduction of thick pure absorber between the spheres.

In the production fission site algorithm in VIM, a potential fission site weight, $(\nu\sigma_f)/(k\sigma_t)$ is calculated at each collision, where k is an input parameter usually set to slightly below k_{eff} to ensure a full fission site bank. Each unit of site weight results in a selected site of weight 1.0, and the remaining fraction is selected by roulette, with the survivor's weight set to 1.0. Sites that are selected are added to the site bank unless the bank is full, in which case roulette is played, with the survivors randomly placed in the bank and the losers discarded. Because each history includes a varying number of collisions, and each collision occurs at a different energy, this process introduces variability in the local fission source population. Finally, at the end of each generation, the next generation of starters is constructed from the site bank; the site bank represents a probability density function. If the site bank contains exactly the right number of starters, each site is used once. If the site bank contains more or fewer sites than there will be starters, it is randomly sampled. This sampling process adds to the variability; and the end result is the non-zero probability of losing a fissionable piece from the source vector.

The fission site bank stratified sampling algorithm, described in detail by Gelbard and Roussell[2], alters this procedure in order to keep at least one site in each fissionable piece. The fission production rate tally in each piece, based on an estimate of the flux, is less variable than the probability that at least one fission site is produced because every track through or collision in the fissionable piece contributes to the tally. This rate, the expected value of the fission production rate, is used to apportion fission site weights among the fissionable pieces at the end of a generation, and it is the sampling of this distribution that is a variety of the generic stratified sampling method. If only one potential fission site happens to be produced in a particular piece during that generation, it is retained with its weight adjusted according to the expected value. Only if the resulting weight is extremely small, e.g., 10^{-6} , is the site discarded by roulette. An important fissionable piece in which the histories are uncharacteristically unproductive during a generation is thus retained in the source, and during subsequent generations its fission site population will probably grow.

2.2 SUPERHISTORY POWERING

The code used to test the superhistory powering method[5] is MONK[6]. Its conventional fission source algorithm is collision-based for the 69-group WIMS library is used here (but more purely analog for the DICE quasi-continuous energy library). At collision time, a fraction of the coliding neutron is absorbed (with weight W_a), and when fission is selected as the result of the fractional absorption, a site with weight νW is placed in the site bank. The surviving fraction continues its random walk after the scattering until its weight falls below a cutoff, when roulette is used. Sites are retained in the bank for about three source iterations (in this case of one neutron generation each) to provide a larger site population and to reduce correlation between one stage and the next. During the settling generations, the sites from the earliest two superhistory stages in the reserve store are underweighted by a factor of ten to avoid impeding source convergence. After completion of a user-specified number of settling stages, the underweighting is removed.

The superhistory powering method[5,6] has been implemented in the production versions of MONK to decrease eigenvalue bias and to improve the source convergence behavior of Monte Carlo eigenvalue calculations. This method is similar to the conventional one, except that a series of generations (normally ten) are included within a stage, and only after the stage is complete is the fission site bank stochastically sampled. During all but the final generation in a stage, when fission is selected, the one daughter neutron is produced with a weight of $(\nu W_a)/k$, where k is the eigenvalue estimate. Each of these neutrons and their progeny are tracked until the last generation in a superhistory stage. On average, the fission site population in a more reactive piece in an array of fissionable pieces will be able to increase generation by generation. During the tenth generation, a fission results in one neutron with weight, νW_a, placed in the site bank ("reserve store") for future sampling. Although the sampling process at the beginning of the next stage will generally omit some sites, the more reactive piece in an array will have had more generations to accumulate sites, so the probability that it will be "lost" from the source vector is correspondingly reduced. As in the conventional algorithm, the reserve store contains sites from three superhistory stages. If there are ten generations per superhistory, this means that the fission site bank draws on fission from 10, 20, and 30 generations before, so a barely unconverged source will be set back when tallying begins. The presence of sites from past generations also provides some possibility that a reactive fissionable piece that unluckily fails to produce a site will be recovered when the reserve store is sampled.

3. EMPIRICAL COMPARISONS

This preliminary study consists of two comparisons of algorithms using several different criteria for convergence reliability. The goal is to quantify the probability of a source convergence failure, and this requires a definition of failure. Several are used in this work, and the comparison of definitions is useful, apart from the comparison of source convergence methods themselves. Although most work on Monte Carlo methods includes computing time as an important metric, e.g., the figure of merit, $\sigma^2 T$, the interest here is in convergence reliability. Accordingly, running times have been ignored here. Mohamed and Gelbard reported that stratified source sampling method running times were as much as 100% larger than those for conventional sampling. We do not expect there to be a significant running time penalty for the superhistory powering method.

The improvement in convergence reliability of MONK with superhistory powering and conventional fission source algorithm is measured, as is the improvement in convergence reliability of VIM with stratified sampling and its conventional algorithm, using results reported earlier[3]. Sets of Whitesides problem replica calculations matching those performed by Mohamed were run using MONK, and their state of convergence compared to the conventional method, using the same numbers of starting neutrons per sphere, the same number of generations skipped before accumulating tallies, and the same number of tally generations (150) as in the previous VIM work. By replica, we mean a calculation that is identical except for the pseudorandom number sequence used. For each of the three test problems, the number of starting particles per sphere for each configuration had been determined empirically[3] as that which produces a 5% failure rate for VIM with the stratified sampling method. There were 6 starters per sphere and 80 neutron generations skipped for the reference case, 3 starters per sphere and 50 generations skipped for the ureflected case, and 6 starters per sphere and 30 generations skipped for the fully decoupled case. In the earlier work, any calculation that produced an eigenvalue that was more than 0.01 below the correctly converged eigenvalue was deemed to have failed. While no one would knowingly perform Monte Carlo calculations on the ragged edge of adequate source convergence, the marginal convergence in these cases was thought to be useful for comparing the effectiveness of fission site generation and selection algorithms in converging the fission source. This matching of control parameters in the MONK calculation to those from the VIM calculations was followed strictly for purposes of comparison; consequently, the generally unconverged eigenvalues from MONK or VIM reported here should not be taken as representative.

It should be reiterated here that the calculations performed for this study suffered from both convergence problems: the “criticality of the world” problem, and the more common mathematical slow convergence.

The reliability comparisons are somewhat complicated by several issues reflecting differences in the codes that we could have eliminated only by modifying the codes. First, VIM produces potential fission sites at each collision, but in MONK fission must be the reaction selected in the random walk. Second, the MONK calculations were performed using a 69-group library, while VIM used its continuous energy cross sections. Third, the initial VIM fission sites were deterministically stratified, but MONK sampled a uniform fission source. The rest of the starting particles were randomly distributed among the remaining spheres, not stratified. (Large numbers of MONK calculations with different pseudorandom number seeds were screened to produce a set with the same number of starting particles in the central sphere as those used in VIM.) Fourth, only a few superhistory stages (but the same number of generations as with VIM) are skipped by MONK, so the original source guess is still included in the source vector reserve store when the tally stages begin. At that point, the underweighting of the earlier stages is stopped, suddenly increasing the portion of the sites from the unconverged distribution. Except for the reference eigenvalue computations, in which all of the initial source is in the central sphere, none of these calculations can be regarded as fully converged.

The replica eigenvalues[3] from VIM with conventional and stratified source sampling are reproduced in Figure 1. The corresponding results from MONK conventional source sampling and superhistory powering are shown in Figure 2. Note that the distribution of eigenvalue estimates is clearly not normal.

Tables I - III summarize the results for the three configurations and provide some additional insight into algorithm performance and convergence reliability criteria. No effort was made with MONK to use cross sections appropriate for such an unusual system, so all of the MONK replica eigenvalues have been normalized to the reference eigenvalues for comparing states of convergence. Comparisons of average eigenvalues and of three distinct types of failure rates were made. In the first row, the average k_{eff}s and especially their uncertainty estimates show the combined effects of variability and slow convergence. Each replica’s individual eigenvalue statistical uncertainty was estimated to be about 0.0010 for conventional sampling and about 0.0017 for superhistory powering, but the uncertainties in the ensemble means, i.e., the variance among replicas, were much larger for nearly all of the conventional calculations (up to 0.0146). Note that the uncertainty estimates in the tables are based on a normal distribution, so the usual inferences and confidence limits based on normal distributions should not be given too much weight.

The second row shows the "convergence failure rate", where we apply the criterion used by Mohamed and Gelbard, deeming eigenvalues more than 0.01 below the true eigenvalue to have failed to converge. This criterion is unfair when applied to superhistory powering because of the inadequate number of settling stages artificially applied to these replicas. In nearly every set of replicas, both advanced methods reduced the failure rate. The one exception is the reference configuration, where the superhistory method slightly increased the failure rate using this criterion. This criterion fails to discriminate between slow convergence and a criticality of the world failure.

The third row shows the incidences of true "criticality of the world" failures, i.e., $k_{eff} \sim 0.82$. This sort of failure only occurred for the extreme decoupled case (in 8% of the replicas). Apparently the coupling in the other more realistic cases is sufficient to permit recovery from the "loss" of a fissionable piece from the source vector, irrespective of fission source sampling algorithm.

Rows 4-6 compare the normalized eigenvalues at several confidence limits. Due to the non-normal distribution of eigenvalue estimates, we use empirical confidence limits, ordering the eigenvalue estimates from highest to lowest, commonly know as "order statistics". By definition, 95% of the eigenvalue estimates are larger than or equal to the 95^{th} percentile eigenvalue, which means there is, on average, a 5% chance of an eigenvalue lower than the 95^{th} percentile eigenvalue if another replica were run (a failure). The gains tabulated are the percent of error removed by the advanced method when compared to the conventional one, in percent. The 90^{th} and 95^{th} percentile eigenvalues were improved substantially by both methods, but stratified sampling was more robust for the decoupled system. For all three configurations, the 70^{th} percentile eigenvalues are well above nearly all "criticality of the world" failures, so both the stratified sampling and the superhistory methods produce diminished improvement over the conventional algorithms when the 70^{th} percentile is the failure criterion.

The progression from Table I to Table III shows the reliability performance trends as a qualitative function of neutronic coupling. Since most of the MONK calculations were barely converged, the contamination of the fission source distribution from earlier stages (the reserve store) set back the final convergence during the tally stages of the MONK calculations. This has been confirmed by repeating some of the replicas, each with two additional settling stages to force the elimination of the unconverged source information from the reserve store. In these replicas, the eigenvalues were converged, except when a criticality of the world failure had occurred. Obviously, real criticality calculations would include enough settling stages, so the generally unconverged eigenvalues from MONK reported here should not be taken as representative.

The source convergence reliability problem has several aspects, only one of which, i.e., the incidence of convergence failures, is examined here. The second aspect, not studied in this work, is the possibility that the analyst will not detect a convergence failure. One feature of MONK that performed very well was the suite of statistical tests designed to indicate the quality of the eigenvalue estimates. The statistical test for adequate source convergence in MONK nearly always signalled insufficient convergence, except when a true "criticality of the world" failure occurred. Furthermore, the MONK output includes a list of pieces which are not sampled. Consequently, care should be taken when drawing inferences about which code or algorithm treats convergence more robustly.

CONCLUSIONS

Both the superhistory powering method and the stratified sampling method were shown to improve convergence behavior for the Whitesides problem and its variants. Neither algorithm completely eliminates the probability of losing the critical sphere, and if coupling is absent, there is no way for the source iteration to recover. Stratified sampling is more robust in the extreme case of no coupling, but at the cost of some additional running time due to its conditional intervention in the random walk. Superhistory powering does have the advantage of simplicity, since the fissionable pieces do not have to be identified and given special treatment during the random walk process. Presumably, this simplicity provides the advantage of no additional computational effort, so running times should not be greater when using superhistory powering. For systems where stratified source sampling is not substantially more effective, the computational penalty would place it at a disadvantage compared to superhistory powering.

The difference in convergence reliability performance between the two codes with conventional source sampling needs to be resolved. A “cleaner” comparison with a fission-based algorithm in VIM would eliminate doubts fostered by the different underlying fission site generation schemes, as would using the more generally applicable cross sections available with MONK. Further investigation ought to improve our understanding of fission site algorithm source convergence performance. Furthermore, there are more advanced techniques for statistical analysis[8] of converging series that are useful for slowly converging problems not suffering from the “criticality of the world” problem.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the advice and assistance of Nigel Smith and AEA Technology in detailed description of the fission site algorithm in MONK.

REFERENCES

1. G. E. Whitesides, “A Difficulty in Computing the k-effective of the World,” Trans. Am. Nucl. Soc., 14, 680 (1971).

2. E. M. Gelbard and B. Roussell, “Proposed Solution to the ‘k_{eff} of the World’ Problem,”, Trans. Am. Nucl. Soc., 73, 201 (1995).

3. A. Mohamed and E. M. Gelbard, “Stratified Source-Sampling Techniques for Monte Carlo Eigenvalue Analysis,” Proc. Int. Conf. Physics of Nuclear Science and Technology, Hauppage, NY, October 5-8, 1998.

4. R. N. Blomquist, "VIM-A Continuous Energy Monte Carlo Code at ANL," A Review of the Theory and Application of Monte Carlo Methods, Proceedings of a Seminar-Workshop, ORNL/RSIC-44 (April 1980).

5. R. J. Brissenden and A. R. Garlick, “Biases in the Estimation of K-eff and its Error by Monte Carlo Methods,” Annals of Nuclear Energy, 12 No 2 (1986).

6. N. R. Smith, et. al., “The Unification of MONK – Extending the Monte Carlo Horizon”, Proceedings of the Sixth International Conference on Nuclear Criticality Safety, Versailles, France, September 20-24, 1999.

7. N. R. Smith, “The MONK Superhistory Powering Algorithm”, http://www.aeat.co.uk/ answers/PAGE50A.HTM.

8. L. Demaret, et. al., “Accurate Determination of Confidence Intervals in Monte Carlo Eigenvalue Calculations”, Proceedings of the Sixth International Conference on Nuclear Criticality Safety, Versailles, France, September 20-24, 1999.

Table I: Reference (Original Whitesides) Configuration Eigenvalues

	MONK (20 replicas)			VIM (20 replicas)		
	Superhist.	Conv.	Gain	Strat. Samp.	Conv.	Gain
avg k-eff (uncert)	0.9931 (0.0012)	0.9882 (0.0040)	0.0049 (0.0042)	0.9955 (0.0008)	0.9914 (0.0014)	0.0041 (0.0016)
low by > 0.01	40%	35%	-5%	5%	40%	35%
low by ~ 0.18	0	0	0	0	0	0
95^{th} percentile k	0.9812	0.9310	75%	0.9911	0.9779	60%
90^{th} percentile k	0.9891	0.9508	78%	0.9862	0.9808	28%
70^{th} percentile k	0.9902	0.9889	12%	0.9933	0.9869	49%

Table II: Unreflected Configuration Eigenvalues

	MONK (46 replicas)			VIM (20 replicas)		
	Superhist.	Conv.	Gain	Strat. Samp.	Conv.	Gain
avg k-eff (uncert)	0.9824 (0.0050)	0.9560 (0.0076)	0.0264 (0.0091)	0.9993 (0.0011)	0.9858 (0.0062)	0.0135 (0.0063)
low by > 0.01	28%	72%	46%	5%	30%	25%
low by ~ 0.18	0	0	0	0	0	0
95th percentile	0.8791	0.8427	23%	0.9812	0.9086	79%
90th percentile	0.9363	0.8464	59%	0.9919	0.9208	90%
70th percentile	0.9924	0.9340	88%	0.9986	0.9808	93%

Table III: Decoupled Configuration Eigenvalues

	MONK (34 replicas)			VIM (20 replicas)		
	Superhist.	Conv.	Gain	Strat. Samp.	Conv.	Gain
avg k-eff (uncert)	0.9677 (0.0099)	0.9447 (0.0117)	0.0230 (0.0153)	0.9965 (0.0006)	0.9684 (0.0146)	0.0281 (0.0146)
low by > 0.01	68%	98%	30%	0	15%	15%
low by ~ 0.18	12%	21%	98%	0	0	0
95th percentile	0.8283	0.8279	0.2%	0.9907	0.8168	95%
90th percentile	0.8286	0.8286	0	0.9917	0.8164	96%
70th percentile	0.9987	0.9743	95%	0.9952	0.9912	45%

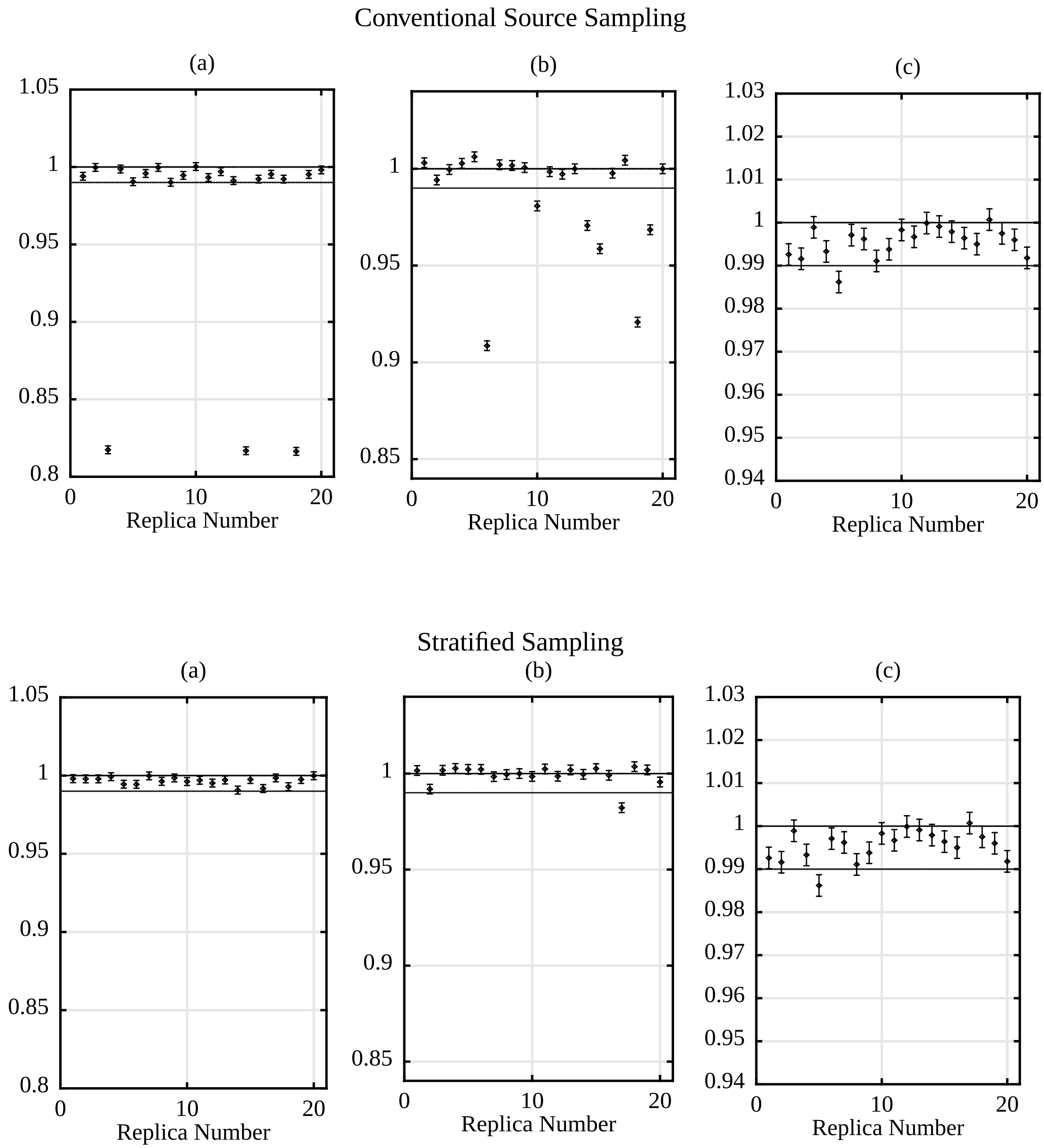

Figure 1: Eigenvalue Estimates for VIM Replicas: Conventional Source Sampling vs. Stratified Sampling for (a) Completely Decoupled Array, (b) Unreflected Array, and (c) Whitesides Configuration.

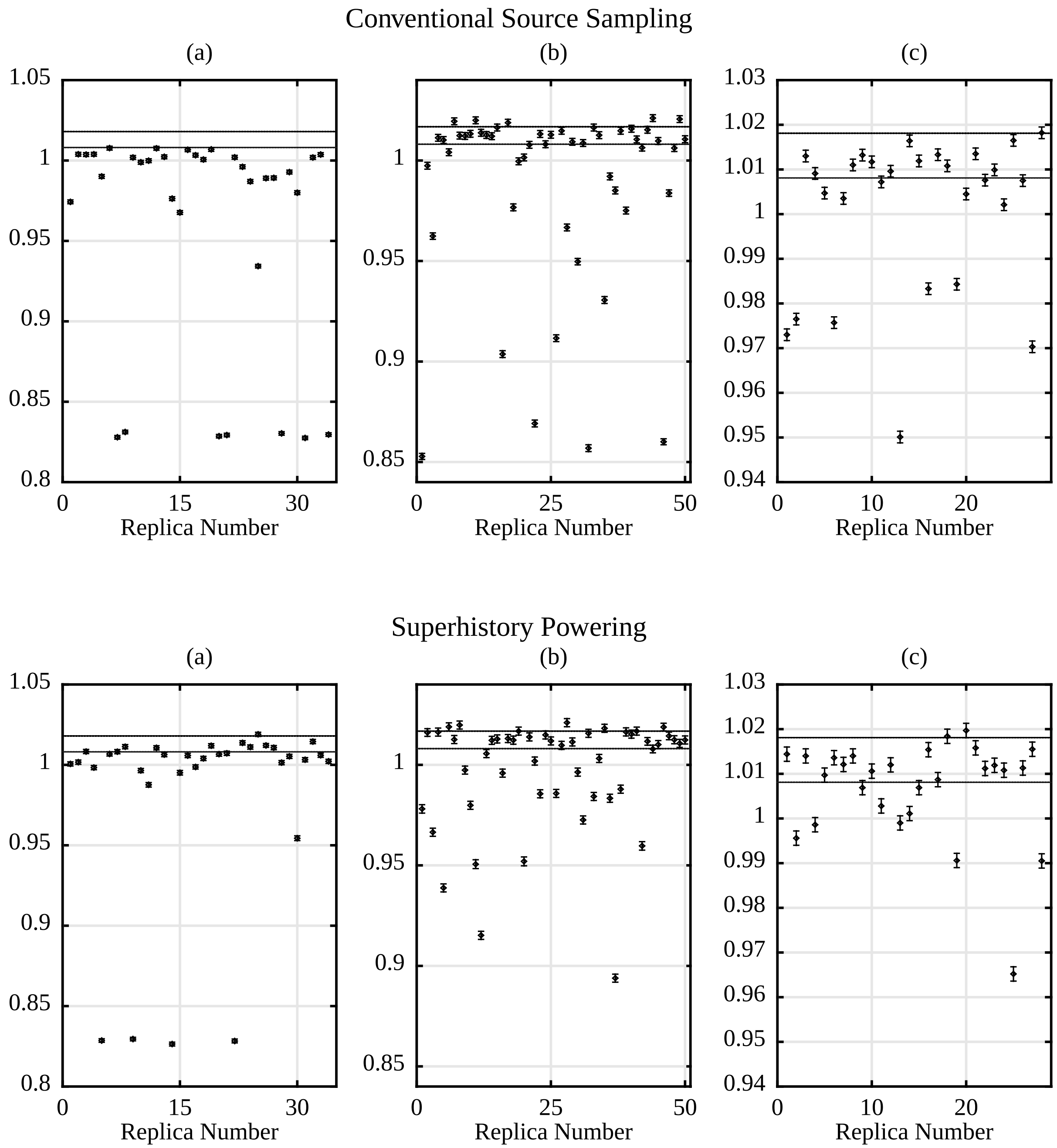

Figure 2: Eigenvalue Estimates for MONK Replicas: Conventional Source Sampling vs. Superhistory Powering for (a) Completely Decoupled Array, (b) Unreflected Array, and (c) Whitesides Configuration.