ANS MC2015 - Joint International Conference on Mathematics and Computation (M&C), Supercomputing in Nuclear Applications (SNA) and the Monte Carlo (MC) Method • Nashville, TN • April 19-23, 2015, on CD-ROM, American Nuclear Society, LaGrange Park, IL (2015)

METRICS FOR DIAGNOSING UNDERSAMPLING IN MONTE CARLO TALLY ESTIMATES

Christopher M. Perfetti* and Bradley T. Rearden
Oak Ridge National Laboratory
Reactor and Nuclear Systems Division
P.O. Box 2008, Bldg. 5700
Oak Ridge, TN 37831-6170, USA
* perfetticm@ornl.gov

ABSTRACT

This study explored the potential of using Markov chain convergence diagnostics to predict the prevalence and magnitude of biases due to undersampling in Monte Carlo eigenvalue and flux tally estimates. Five metrics were applied to two models of pressurized water reactor fuel assemblies and their potential for identifying undersampling biases was evaluated by comparing the calculated test metrics with known biases in the tallies. Three of the five undersampling metrics showed the potential to accurately predict the behavior of undersampling biases in the responses examined in this study.

Key Words: Monte Carlo, tally biases, undersampling, convergence metrics, SCALE

1 INTRODUCTION

Monte Carlo methods for calculating the eigenvalues of fissile systems represent the fission source by simulating multiple batches, or generations, of fission neutrons, where the fission sites created during one generation serve as the birth sites for neutrons in the next generation. Failure to simulate enough particles in each generation can result in a phenomenon known as "undersampling," where neutrons do not interact sufficiently with all regions in the problem during each generation. This underrepresentation of regions in a model has been shown to impact the accuracy of tally response and uncertainty estimates in Monte Carlo calculations [1] [2]. As reported previously by Brown [1] and Perfetti and Rearden [3], and as shown in Figure 1, undersampling can result in significant biases in Monte Carlo eigenvalue estimates (up to several percent) and even larger biases in flux tally estimates (up to several tens of percent).

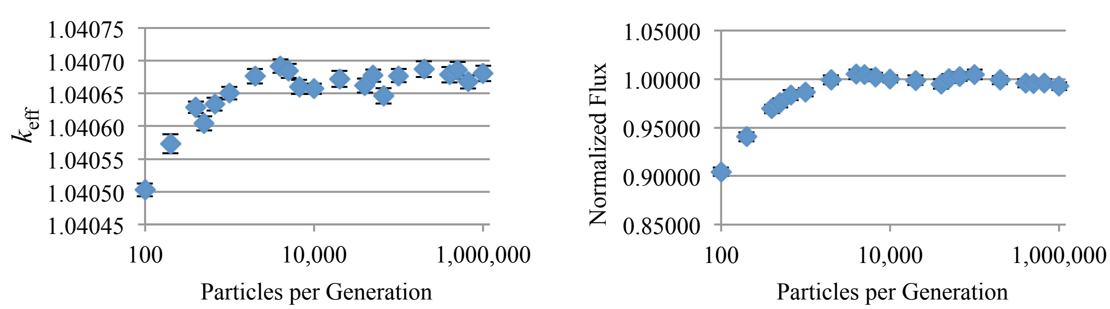

Figure 1. Undersampling in eigenvalue estimates (left) and flux tally estimates in an axial segment of a fuel pin (right) in an infinitely reflected model of a fuel assembly as a function of the number of particle histories simulated per generation [3].

The Organisation for Economic Co-operation and Development Nuclear Energy Agency Working Party on Nuclear Criticality Safety's Expert Group on Advanced Monte Carlo Techniques (AMCT) was formed to advance the knowledge base regarding Monte Carlo criticality calculations that rely on obtaining accurate flux and reaction rate estimates, such as Monte Carlo depletion calculations for burnup credit applications [3] [4]. The long-term goal of the AMCT collaboration is to understand the magnitude, prevalence, and impact of biases in eigenvalue estimates, reaction rate tallies, and tally variance estimates and to create a set of best practices to maximize the reliability of Monte Carlo calculations by mitigating the effect of undersampling. In previous work in the AMCT collaboration, Perfetti and Rearden observed significant biases in flux tally and fission rate estimates in models of pressurized water reactor (PWR) fuel assemblies and spent fuel shipping casks [3]. The magnitude of these biases was larger than those previously observed for eigenvalue estimates in similar systems and was surprisingly prevalent in models of relatively simple systems (models of single, infinitely reflected fuel assemblies). This study builds upon the previous work by Perfetti and Rearden by investigating statistical metrics that can be implemented in Monte Carlo codes to predict the occurrence of biases in Monte Carlo simulations. These metrics are applied to the previously observed tally biases and are evaluated for their reliability at predicting the onset and magnitude of undersampling biases in Monte Carlo eigenvalue and flux tally estimates.

2 TALLY CONVERGENCE METRICS

The goal of developing tally convergence metrics is to provide tools for Monte Carlo analysts to ensure the fidelity of simulation results. Similarly to how Shannon Entropy is used to verify fission source convergence [5], analysts might, for example, guarantee that a Monte Carlo tally is accurate within a 1% undersampling bias so long as their convergence metric of choice is smaller than some threshold value. When developing metrics to predict undersampling in Monte Carlo simulations, one strives to satisfy two criteria: metrics should be able to diagnose undersampling "on the fly" (i.e., while the calculation is still in progress), and they should be universally applicable. Observing undersampling biases in Monte Carlo simulations traditionally requires performing multiple simulations of the same model using different random seeds and identifying differences in tally estimates that exceed the statistical uncertainty of the estimates. This process is effective at identifying undersampling in longer-term investigative studies, such as the AMCT collaboration, but it typically imposes a significant computational burden, both because of the usually large runtimes associated with repeated Monte Carlo simulations of complex systems and because the biases can be small and thus require a high degree of convergence. Therefore, to effectively predict undersampling and to provide guidance to Monte Carlo analysts in practical applications, metrics must be able to diagnose undersampling in a single calculation, ideally on the fly, so that the simulation parameters can be adjusted (or corrected) during simulations when a metric detects that responses of interest are being significantly undersampled. Secondly, the wide range of Monte Carlo tally responses and the even wider range of Monte Carlo applications demand that tally convergence metrics be universally applicable. Metrics should be response- and system-independent; in other words, they should be able to consistently predict the behavior of undersampling biases for various Monte Carlo tally responses, such as eigenvalue, neutron flux, reaction rate, and sensitivity tally estimates (all scored with and without energy bins), in systems with vastly different neutron spectra.

This study evaluated the potential of several tally convergence metrics by calculating them for several Monte Carlo tally responses in the systems included in the AMCT study and comparing them to the previously observed tally biases [3]. Described in more detail in Ref. [3], the responses of interest spanned system eigenvalue estimates and energy-integrated flux tallies in axial segments of PWR fuel pins. Two systems were examined in this study: an infinitely reflected fuel assembly in a PWR (the R2 case) and an infinitely reflected PWR assembly in a spent fuel shipping cask (the S2 case). Both systems were previously found to produce significant (tens to hundreds of percent) flux tally biases, despite the relative geometric simplicity of the infinitely reflected models [3].

In this study the magnitude of the undersampling biases was quantified by examining the "fraction of undersampling," or the fraction by which tallies differed from their reference values. The reference value for each tally was obtained by performing multiple Monte Carlo simulations with different random seeds and using 10,000 particle histories per generation for 10,000 active generations, a value that was previously found to produce no observable flux tally biases [3]. As shown in Figure 1, tally biases were observed by varying the number of particle histories simulated in each generation while fixing the total number of active histories used in the simulations at 100 million. Each case was simulated 30 times using different random number seeds to estimate the true variance of the responses of interest. In this study, data points for the convergence metrics were calculated using between 100 and 10,000 particle histories per generation, resulting in a spectrum of data points for the same sets of tallies with various degrees of undersampling. All calculations in this study were performed using the KENO Monte Carlo code within the SCALE code package [6].

An ideal convergence metric will have a one-to-one relationship with the magnitude of the undersampling bias observed in tallies in different systems, thereby allowing analysts to anticipate the degree of undersampling that may occur for a tally estimate, given the value of its convergence metric. This study examined the potential for the following five metrics to diagnose and correlate to the magnitude of undersampling biases:

1. Contributing Particles per Generation (see Sect. 2.1)
2. The Heidelberger-Welch Relative Half-Width (RHW) (see Sect. 2.2)
3. The Geweke Z-Score (see Sect. 2.3)
4. The Gelman–Rubin Scale Reduction Factor ($\hat{R}_c$) Diagnostic (see Sect. 2.4)
5. Tally Entropy (see Sect. 2.5)

Three of the methods (the Heidelberger-Welch RHW, the Geweke Z-score, and the Gelman–Rubin diagnostics) are commonly used for diagnosing the convergence of Markov chains [7]; two of the methods (the Contributing Particles per Generation and Tally Entropy) are new convergence metrics that were developed in the course of this study.

2.1 Contributing Particles per Generation

The first tally convergence metric examined in this study was the Contributing Particles per Generation metric, which simply describes the average number of particles within a single generation that contribute nonzero scores to a tally estimate. Because undersampling occurs when too few particles interact with the tally region and when too few particles are used to sample the fission source of a system, the degree of undersampling observed in a tally should be inversely proportional to the average number of particles that create tally scores in that region in each generation. Figure 2 shows the relationship between the fraction of undersampling and the contributing particles per generation for each of the eigenvalues and flux tallies in the R2 and S2 cases. Tallies that produced biases containing more than 75% relative uncertainty were omitted from Figure 2 and from all other figures in this study.

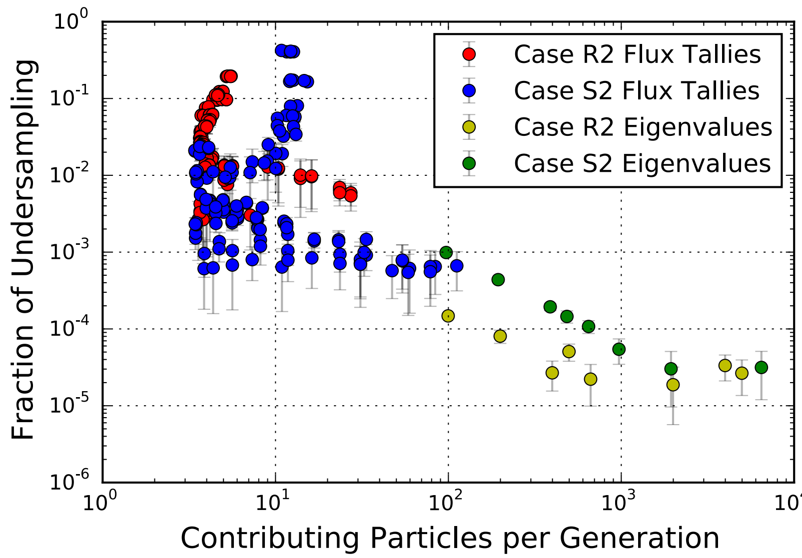

Figure 2. Effectiveness of the Contributing Particles per Generation metric for predicting undersampling.

As shown in Figure 2 and as expected, the fraction of undersampling in the Monte Carlo tallies generally decreased as the contributing particles per generation metric increased. While this trend is more apparent over the entire span of the tally data, it is less apparent for the flux tallies, especially those that saw ~10 contributing particles per generation. In this range the R2 flux tally data curves backward before decreasing and a significant portion of the S2 tallies actually see an increased prevalence of undersampling as the number of contributing particles per generation increased. Therefore, the Contributing Particles per Generation metric was observed to predict undersampling with some general degree of accuracy, but it did not effectively predict undersampling biases in all tallies.

2.2 Heidelberger-Welch Relative Half-Width

The Heidelberger-Welch RHW metric examines whether the sample size within a Markov chain is sufficient to provide accurate estimates for the mean value of a parameter by testing whether tally scores within the Markov chain vary significantly outside the margin of error of the confidence interval, α, of the chain. The statistic for the Heidelberger-Welch RHW test is given by [8]

$$RHW = \frac{z_{(1-\alpha/2)}\sqrt{\hat{s}_n/n}}{\theta_n}, \tag{1}$$

where $z_{(1-\alpha/2)}$ represents the *Z*-score (the number of standard deviations from the mean of normally distributed data) of the $100(1-\alpha/2)^{\text{th}}$ percentile, n is the length of the Markov chain, and θ_n and $\hat{s}_n$ are the estimated mean and variance, respectively, of the members in the chain.

The SAS statistical package, a software suite that offers a plethora of statistical analysis tools, uses a default RHW statistic of less than 0.1 to indicate a sufficiently sampled Markov chain [7].

In this application the elements of the Markov chain were assumed to be scores for a tally that were produced by individual particle histories within a single generation; therefore, rejection by the Heidelberger-Welch RHW test indicated that additional particle histories needed to be simulated within each generation to produce an accurate estimate for the response of interest. An α value of 0.05 was assumed in this study, and $\hat{s}_n$ was calculated assuming that particle scores within a single generation were completely uncorrelated.

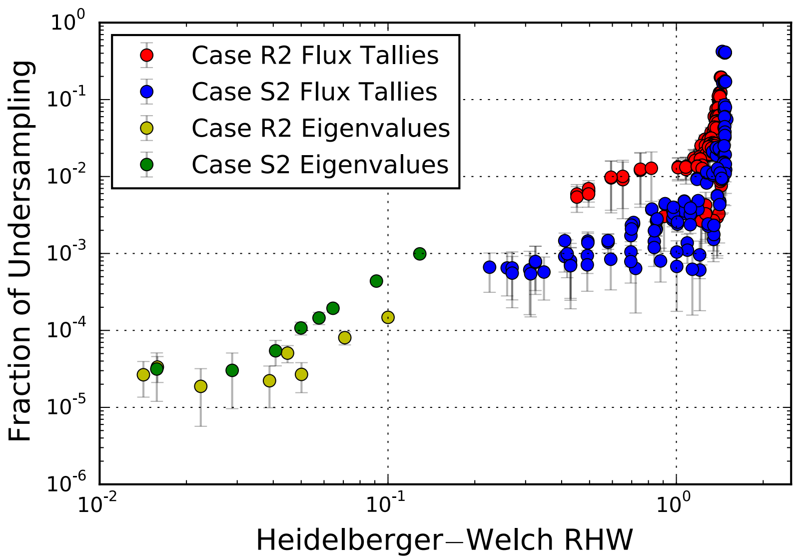

Figure 3. Effectiveness of the Heidelberger-Welch RHW metric for predicting undersampling.

As shown in Figure 3, the RHW metric effectively predicted the onset an magnitude of undersampling, and there appears to be a much stronger relationship between the RHW values and the magnitude of the undersampling bias than was observed for the Contributing Particles per Generation metric. The previously recommended SAS Heidelberger-Welch RHW acceptance value (0.1) appeared too to be rather stringent in these cases, corresponding to an undersampling bias of approximately 0.05%; ensuring a less than 1% undersampling bias in this application required metric values of approximately 0.5 or less.

Several tallies with relatively small RHW values encountered larger undersampling biases than were expected. This type II error will not be discussed in detail here, but was confined almost entirely to the most severely undersampled flux tallies with the largest statistical uncertainties – primarily those at the bottom of the S2 assembly. This behavior was also observed for the Tally Entropy metrics (to a greater degree), and the Gelman–Rubin diagnostics (to a lesser degree). These anomalous data points were filtered from the figures in this study using the "less than 75% bias uncertainty" filtering that was mentioned previously. Some large RHW scores also produced smaller biases than were expected, but this type I error is preferable to type II error because it ensures that Monte Carlo analysts will err on the side of caution when accounting for undersampling biases.

2.3 Geweke Z-Score

The Geweke Z-Score tests for Markov chain convergence by examining whether the tally contributions from the first half of the Markov chain differ significantly from those in the second half [9]. This comparison treats each half of the Markov chain as an independent estimate of the chain's half-mean of the tally and computes a Z-score to test whether the two means are equivalent. The Geweke Z-Score that is calculated is given by

$$z = \frac{\theta_1 - \theta_2}{\sqrt{\frac{\hat{s}_{n_1}}{n_1} + \frac{\hat{s}_{n_2}}{n_2}}}, \tag{2}$$

where θ_1 and θ_2 represent the tally means from the first and second halves of the Markov chain, respectively, $\hat{s}_{n_1}$ and $\hat{s}_{n_2}$ represent the variance of the first and second halves of the Markov chain, respectively, and n_1 and n_2 represent the number of samples in the first and second halves of the Markov chain, respectively. In this application the Geweke Z-score was calculated by comparing the nonzero tally scores produced from particle histories in the first half of the generation to the scores from the second half; a second Geweke Z-score was also calculated by including the particle histories that produced tally scores of zero, but this metric showed very little correlation to the undersampling biases and is not discussed in detail in this study. As with the Heidelberger-Welch RHW and all other metrics in this study, the variance of particle scores within a single generation was assumed to be completely uncorrelated.

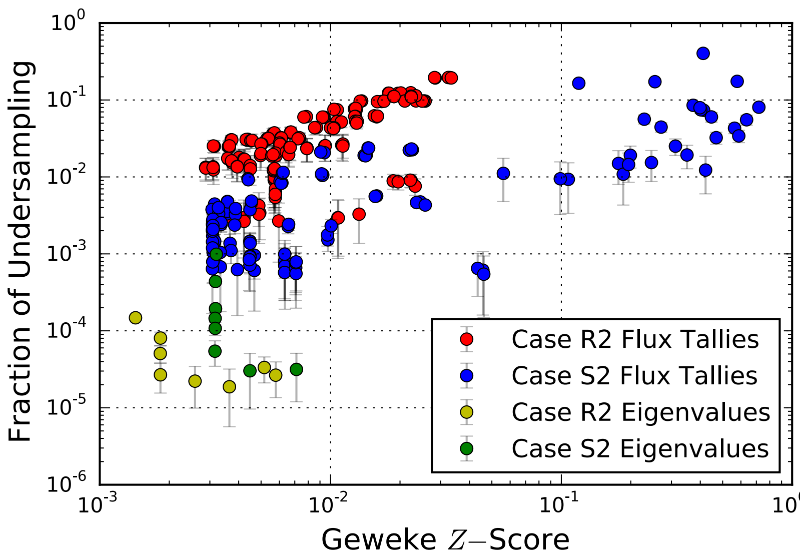

Figure 4. Effectiveness of the Geweke Z-Score for predicting undersampling.

Figure 4, which shows the Geweke Z-Scores that were calculated for the R2 and S2 problems, indicates that the Geweke Z-Score was somewhat effective at predicting the undersampling biases. The R2 and S2 flux tallies produced Geweke Z-Scores that showed some broad correlation with the undersampling bias, but the eigenvalue estimates produced Z-Scores that showed no (and sometimes inverse) correlation to the undersampling bias. The Geweke Z-Score may be effective in predicting undersampling in the tally estimates with larger undersampling biases (more than 1%), but it could not effectively predict undersampling in eigenvalue estimates and was therefore determined to be an ineffective metric for predicting the onset and magnitude of undersampling biases.

2.4 Gelman-Rubin $\hat{R}_c$ Diagnostic

Gelman–Rubin diagnostics assess convergence of Markov chains by splitting the chains into subchains and testing whether the tally variance within the subchains differs significantly from the variance between the subchains [10] [11]. In this application the master Markov chain was the "chain" of tally scores created by particle histories within each single generation, and these scores were split into three subchains, a common number of subchains when using Gelman–Rubin convergence diagnostics [7]. The subchains of tally scores were used to calculate the corrected Scale Reduction Factor (SRF), $\hat{R}_c$, the Gelman–Rubin metric for assessing convergence of Markov chains. A thorough description of how to calculate $\hat{R}_c$ is available in Ref. 11. In general, $\hat{R}_c$ values that are close to 1.0 indicate Markov chain convergence, and in practice $\hat{R}_c$ values that are less than about 1.2 or 1.1 are considered acceptable [7] [11].

As shown in Figure 5, the Gelman–Rubin $\hat{R}_c$ values that were calculated for the R2 and S2 case tallies were able to accurately predict the undersampling biases. In the figure the undersampling bias grows rapidly for small $\hat{R}_c$ values (<1.02), flattens into a plateau region for $\hat{R}_c$ values between 1.05 and 1.2, and then again grows rapidly. The historically recommended minimum $\hat{R}_c$ values for ensuring convergence (less than about 1.1 or 1.2) corresponded to an undersampling bias of approximately 1%.

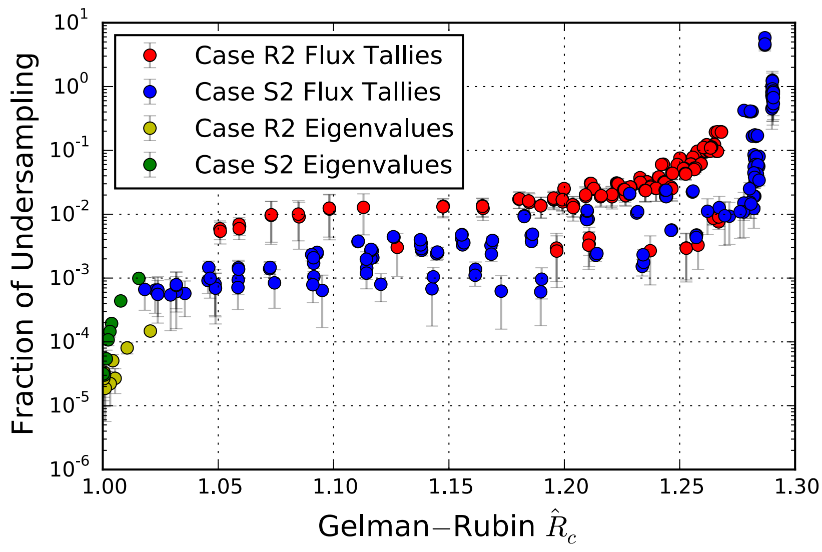

Figure 5. Effectiveness of the Gelman–Rubin $\hat{R}_c$ Diagnostic for predicting undersampling.

The Gelman–Rubin $\hat{R}_c$ has a minimum value of 1, and, as shown in Figure 6, subtracting 1 from the calculated $\hat{R}_c$ values and plotting the results on a log scale gives further insight into the behavior of the undersampling biases. This plot looks very similar to the Heidelberger-Welch RHW data plotted in Figure 3, which is promising because it suggests that the two different metrics are detecting the same (and hopefully true) trend in the undersampled tallies, and the relationship between the Gelman–Rubin ($\hat{R}_c - 1$) values and the undersampling bias is possibly stronger than the relationship observed for the Heidelberger-Welch RHW.

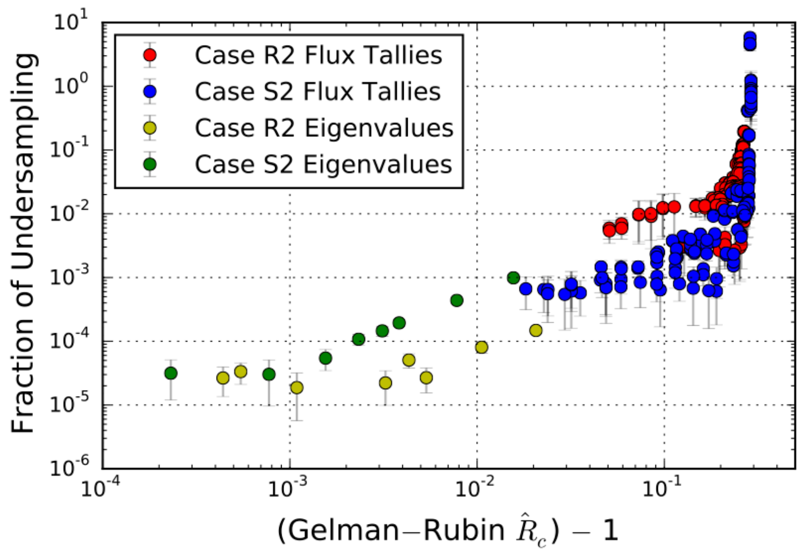

Figure 6. Effectiveness of the Gelman–Rubin $\hat{R}_c$ – 1 Diagnostic for predicting undersampling.

2.5 Tally Entropy

The last metric examined in this study, Tally Entropy, is a new metric that was developed using the information theory concept of Shannon Entropy. The Shannon Entropy, H, of an information signal with N messages is defined as [12]

$$H = -\sum_{n}^{N} p_n \ln{(p_n)}, \tag{3}$$

where p_n is the probability that a signal is received in the n^{th} message. Shannon Entropy has been used previously by Brown and Ueki to detect unconverged fission sources in Monte Carlo simulations [5]. In their application, Brown and Ueki calculated the Shannon Entropy of the fission source by imposing a spatial mesh over the model and calculating the fraction of fission sites that occur in each mesh interval (i.e., the probability that a fission site occurs in a mesh interval). Shannon Entropy that has not yet converged to an average value indicates that the fission source is still iterating toward the true distribution of fission sites in the problem and that additional inactive generations should be simulated. Unfortunately, Shannon Entropy cannot be used in this way to assess the convergence of Monte Carlo tallies because undersampled tallies may produce falsely converged Shannon Entropy estimates that are different but indistinguishable (a priori) from the entropy that would be produced by a converged set of tallies. Therefore, in this work an alternative approach has been developed for using the concept of Shannon Entropy to diagnose undersampling in Monte Carlo tally estimates.

The Shannon Entropy of a signal containing N messages can produce a minimum entropy of zero and a maximum entropy of $ln\ (N)$; the signal will produce an entropy of zero if all of the signal is received in only one of the N messages, and will produce maximum entropy when

1. The number of messages in the signal, N, becomes very large and
2. Each message contributes an equal amount of information ($p_1 = p_2 = p_n$).

These two conditions happen to also be ideal for scoring unbiased Monte Carlo tally estimates: each tally should receive scores from a large number of particle histories in each generation, and each particle history should contribute a similarly sized score to the tally estimate. Therefore, the Tally Entropy convergence metric predicts undersampling biases by calculating how much the Shannon Entropy of the tally estimate differs from its maximum entropy. The entropy of each tally is determined by calculating p_x, which is the probability that the message (the particle history) produces a signal (a tally score); p_x is therefore interpreted as the fractional contribution of the particle x to a tally estimate within generation j and is calculated by dividing the tally score produced by particle x by the sum of the tally scores produced in generation j:

$$p_x = \frac{Tally\ Score\ of\ Particle\ x}{Sum\ of\ all\ Tally\ Scores\ in\ Gen.\ j}. \tag{4}$$

After the p_x values are calculated for the particles within a generation, the following equation is used to calculate the entropy of the scores for tally i in the generation:

$$H_{i,j} = -\sum_{Particle\ x}^{N_{i,j}} p_x \ln(p_x), \tag{5}$$

where $N_{i,j}$ is the number of particle histories in generation j that produced nonzero scores for tally i. The Tally Entropy test statistic for tally i is then calculated by the following equation:

$$Tally\ Entropy_i \equiv \frac{\langle \ln(N_{i,j}) \rangle - \langle H_{i,j} \rangle}{\langle \ln(N_{i,j}) \rangle}, \tag{6}$$

where the $\langle\ \rangle$ operator denotes the average of a value over all active generations.

Figure 7 shows the tally entropy values that were calculated for the tallies in the R2 and S2 cases. Like the Heidelberger-Welch RHW and Gelman–Rubin diagnostics, the Tally Entropy metric generally seems to predict the onset and magnitude of undersampling biases. When plotted in a log-log scale, the Tally Entropy values scale much more linearly than the Heidelberger-Welch RHW or Gelman–Rubin diagnostics, possibly indicating a more straightforward relationship between the metric and the magnitude of the undersampling bias; furthermore, the eigenvalue and flux tally data points show a greater degree of overlap for the Tally Entropy metric than was observed for the other two metrics, indicating that it may be a more tally-independent metric for diagnosing undersampling.

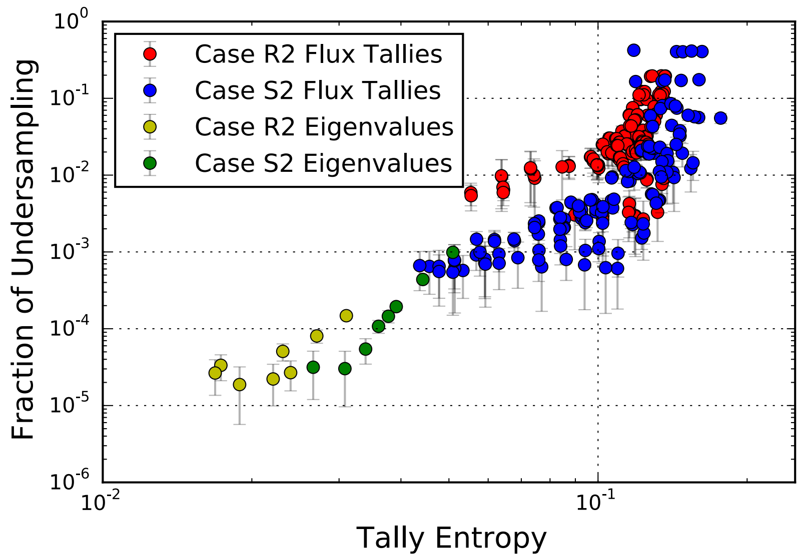

Figure 7. Effectiveness of Tally Entropy for predicting undersampling.

3 METRIC-TO-METRIC COMPARISON

Having identified the Heidelberger-Welch RHW, the Gelman–Rubin $\hat{R}_c$, and the Tally Entropy diagnostics as potential metrics for predicting undersampling in Monte Carlo tally estimates, this study then investigated whether or not the different metrics would agree on the degree of undersampling present in a given tally. This was investigated by comparing the convergence metrics that were produced for each tally data point, as plotted in Figure 8, and by determining whether the metrics predict a similar degree of undersampling for each data point. Figure 8 indicates that there is a strong one-to-one relationship between each of the tally convergence metrics, which suggests that the three metrics agree on the incidence and magnitude of the undersampling biases. The convergence metrics stray from this strong one-to-one relationship for the eigenvalue estimates, and it is yet to be determined whether one metric is superior at predicting undersampling for eigenvalue estimates.

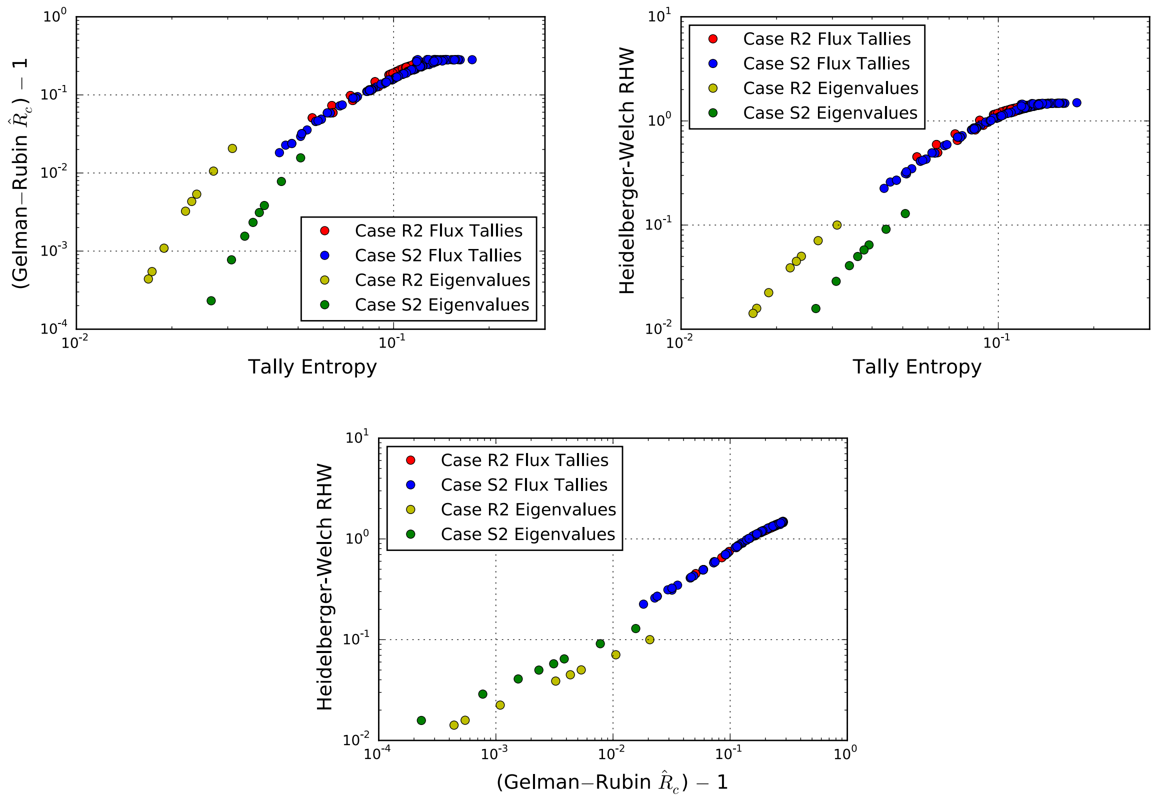

Figure 8. Correlation between undersampling metrics: Tally Entropy vs the (Gelman–Rubin $\widehat{R}_c$) – 1 (upper left), Tally Entropy vs the Heidelberger-Welch RHW (upper right), and the (Gelman–Rubin $\widehat{R}_c$) – 1 vs the Heidelberger-Welch RHW (bottom).

The Tally Entropy metric was developed using concepts from Information Theory, but it lacks the mathematical rigor of the Heidelberger-Welch RHW or the Gelman–Rubin diagnostics, which produce test statistics that can be tested on their statistical significance. The strong one-to-one relationship between the these metrics and the Tally Entropy metric suggests that it may be possible to subject the Tally Entropy metric to a statistical test and obtain a more mathematically rigorous conclusion on whether undersampling has been observed.

4 CONCLUSIONS

This study explored the potential applicability of several statistical metrics for predicting the occurrence and magnitude of undersampling biases in Monte Carlo eigenvalue and flux tally estimates. This study examined the prevalence of undersampling biases in eigenvalue and energy-integrated flux estimates in the models of fuel assemblies in PWRs and shipping cask cases, and, of the five undersampling metrics that were examined, the Heidelberger-Welch RHW, Gelman–Rubin $\widehat{R}_c$, and Tally Entropy metrics were observed to correlate strongly with the observed undersampling biases. This study has demonstrated proof-of-principle for the use of these metrics to predict undersampling. The next phase of this work (and of the AMCT study) is to repeat this analysis for a much broader set of system responses (including reaction rate tallies, multigroup flux estimates, and possibly sensitivity coefficient estimates) in a broad range of applications to determine whether these metrics can truly predict the prevalence of undersampling biases in Monte Carlo simulations.

Based on the preliminary results observed in this study, to estimate tally responses with an undersampling bias of less than 1%, it is recommended that Monte Carlo analysts ensure that the undersampling metrics for the tallies fall below these threshold values:

- Heidelberger-Welch RHW ≤ 0.50
- Gelmen-Rubin $\hat{R}_c \leq 1.05$
- Tally Entropy ≤ 0.05

5 ACKNOWLEDGMENTS

This manuscript has been authored by UT-Battelle LLC under Contract No. DE-AC05-00OR22725 with the US Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

6 REFERENCES

1. F. B. Brown, "'K-effective of the World' and Other Concerns for Monte Carlo Eigenvalue Calculations," *Progress in Nuclear Science and Technology*, **2**, pp. 738–742 (2011).
2. B. T. Mervin, S. W. Mosher, J. C. Wagner, and G. I. Maldonado, "Uncertainty Under-Prediction in Monte Carlo Eigenvalue Calculations," *Nuclear Science and Engineering*, **173**, pp. 276–292 (2013).
3. C. M. Perfetti and B. T. Rearden, "Quantifying the Effect of Undersampling in Monte Carlo Simulations using SCALE," Proc. PHYSOR 2014, Kyoto, Japan, September 28 – October 3, 2014, American Nuclear Society (2014).
4. B. T. Rearden, W. J. Marshall, C. M. Perfetti, J. Miss, Y. Richet, "Quantifying the Effect of Undersampling Biases in Monte Carlo Reaction Rate Tallies," *Organisation for Economic Co-operation and Development Nuclear Energy Agency Expert Group on Advanced Monte Carlo Techniques Benchmark Proposal*, July 2013.
5. T. Ueki and F.B. Brown, "Stationarity Modeling and Informatics-Based Diagnostics in Monte Carlo Criticality Calculations," *Nuclear Science and Engineering*, **149**, 38 (2005).
6. *SCALE: A Comprehensive Modeling and Simulation Suite for Nuclear Safety Analysis and Design*, ORNL/TM-2005/39, Version 6.1, Oak Ridge National Laboratory, Oak Ridge, Tennessee (June 2011). Available from Radiation Safety Information Computational Center at Oak Ridge National Laboratory as CCC-785.
7. *SAS/STAT(R) 9.2 User's Guide, Second Edition*, SAS Institute, Cary, North Carolina (2010). Available at: http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_introbayes_sect008.htm
8. P. Heidelberger and P. D. Welch, "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations," *Simulation Modeling and Statistical Computing*, **24**(4), pp. 233−245 (1981).

9. J. Geweke, "Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments", in J. M. Bernardo, J. O. Berger, A. P. Dawiv, and A. F. M. Smith, *Bayesian Statistics*, volume 4, Clarendon Press, Oxford, UK (1992),

10. A. Gelman and D. B. Rubin, "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, **7**(4), pp. 457–472 (1992).

11. S. P. Brooks and A. Gelman, "General Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational and Graphical Statistics*, **7**(4), pp. 434−455 (1997).

12. C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, **27(**3), pp. 379−423 (1948).