Applied Radiation and Isotopes 72 (2013) 83–88

Contents lists available at SciVerse ScienceDirect

Applied Radiation and Isotopes

journal homepage: www.elsevier.com/locate/apradiso

Hybrid biasing approaches for global variance reduction

Zeyun Wu, Hany S. Abdel-Khalik*

Department of Nuclear Engineering, North Carolina State University, Raleigh, NC 27695, United States

HIGHLIGHTS

- Hybrid Monte Carlo Deterministic Method based on Gaussian Process Model is introduced.
- Method employs deterministic model to calculate responses correlations.
- Method employs correlations to bias Monte Carlo transport.
- Method compared to FW-CADIS methodology in SCALE code.
- An order of magnitude speed up is achieved for a PWR core model.

ARTICLE INFO

Article history:
Received 30 June 2011
Received in revised form
5 September 2012
Accepted 5 September 2012
Available online 17 October 2012

Keywords:
Hybrid Monte Carlo - DT approach
FW-CADIS
Gaussian process
Global variance reduction

ABSTRACT

A new variant of Monte Carlo—deterministic (DT) hybrid variance reduction approach based on Gaussian process theory is presented for accelerating convergence of Monte Carlo simulation and compared with Forward-Weighted Consistent Adjoint Driven Importance Sampling (FW-CADIS) approach implemented in the SCALE package from Oak Ridge National Laboratory. The new approach, denoted the Gaussian process approach, treats the responses of interest as normally distributed random processes. The Gaussian process approach improves the selection of the weight windows of simulated particles by identifying a subspace that captures the dominant sources of statistical response variations. Like the FW-CADIS approach, the Gaussian process approach utilizes particle importance maps obtained from deterministic adjoint models to derive weight window biasing. In contrast to the FW-CADIS approach, the Gaussian process approach identifies the response correlations (via a covariance matrix) and employs them to reduce the computational overhead required for global variance reduction (GVR) purpose. The effective rank of the covariance matrix identifies the minimum number of uncorrelated *pseudo responses*, which are employed to bias simulated particles. Numerical experiments, serving as a proof of principle, are presented to compare the Gaussian process and FW-CADIS approaches in terms of the global reduction in standard deviation of the estimated responses.

Published by Elsevier Ltd.

1. Introduction

Interest in Monte Carlo (MC) models for particle transport applications, such as nuclear reactor criticality calculations, has increased in the past decade. On the one hand, a MC simulation can model the complex physics of radiation transport in reactor analysis applications more accurately. This is because a MC simulation does not make as many simplifying assumptions as deterministic models, such as resonance calculations, do. Also, it can handle more complex reactor and fuel assembly geometries. However, MC simulation requires much longer computer run time than deterministic simulation in order to accumulate sufficient number of histories to reach statistically significant results. Although the continuing growth of computer power is expanding the range of MC applications, there is still a need for advanced approaches to speed up MC convergence. In our context, convergence means that the standard deviation of the quantity of interest diminishes below a user-defined value. Accelerating the convergence of Monte Carlo Simulation implies the ability to reach convergence with fewer number of MC-simulated particles. In this manuscript, we focus on global variance reduction techniques for reactor physics models, where tallies are required everywhere in the phase space.

Several hybrid DT-MC approaches (Haghighat and Wagner, 2003; Densmore and Larsen, 2003; Becker et al., 2007) have been developed recently for global variance reduction (GVR). The basic idea is to employ deterministic models (both forward and adjoint) to bias source particles and assign appropriate particle importance maps to MC models. If this is done properly, hybrid approaches can accelerate the convergence of MC simulation,

* Corresponding author. Tel.: +1 919 515 4600; fax: +1 919 515 5115.
E-mail address: abdelkhalik@ncsu.edu (H.S. Abdel-Khalik).

0969-8043/$ - see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.apradiso.2012.09.026

that is, obtain acceptable reduction in the statistical deviations of the responses of interest with less computational overhead. The term "variance reduction" is often used for approaches intended to accelerate Monte Carlo simulations. We focus here on techniques that are based on response biasing via so-called weight window maps. In these techniques, MC-simulated particles are biased in the phase space (energy and space) to maximize the number of particles contributing to a specific response of interest, which results in reducing the associated variance according to the MC Law of large numbers, i.e.,$\sigma \propto 1/\sqrt{N}$.

On the assumption that the statistical deviation of responses resulting from mathematical modeling and simulation of the radiation transport can be treated as a Gaussian process (GP) (Kennedy and O'Hagan, 2001), a new MC-DT hybrid approach for GVR was recently developed and denoted GP approach (Zhang and Abdel-Khalik, 2011a). Here, we describe the principle behind this approach and compare it with the Forward-Weighted Consistent Adjoint Driven Importance Sampling approach (FW-CADIS) (Wagner et al., 2009), which is currently implemented in the SCALE package from Oak Ridge National Laboratory (SCALE, 2009). FW-CADIS method employs both the forward and the adjoint solutions from deterministic transport calculations to produce efficient biased source map and particle weight window map and, thus, to accelerate the Monte Carlo simulation. The GP method presented here is similar to the FW-CADIS method, but it uses a new way to construct the required weighting factors, which is based on the Gaussian process theory. In the FW-CADIS method, these weighting factors are simply the reciprocals of the response quantities.

A central requirement for GP methods is construction of a covariance matrix that describes the correlations between the responses. In the initial phases of this work (Zhang and Abdel-Khalik, 2011a), the covariance matrix was constructed using analog MC simulation results. In this paper, we introduce a new approach using deterministic calculation. This is paramount to realistic applications because the covariance matrix must be estimated inexpensively.

Moreover, several variants of the GP approach are introduced and compared with the FW-CASIS approach in performance. In particular, we analyze the effects of employing a combined subspace-GP approach and a combined GP-FW-CADIS approach. The subspace approach was introduced recently as a different way to bias MC-simulated particles (Zhang and Abdel-Khalik, 2011b).

Finally, in contrast to the earlier work, the GP approach is applied here to a higher-dimensional problem to investigate the scalability of the approach. A representative prototypical PWR core model is employed. The MAVRIC sequence (a Monte Carlo simulation code with the capability of biasing particles with deterministic models) in SCALE package is employed in this study to facilitate the comparison of the GP and the FW-CADIS approaches.

2. Mathematical description

GP theory assumes that random fluctuations of a response follow a process that can be described by a normal distribution. For GP-based models with many responses, the correlations between the responses are described by a covariance matrix. The covariance matrix is symmetric positive definite, with diagonal entries equal to the variances of the responses and off-diagonal entries representing the covariance between the responses. A diagonal and well-conditioned matrix implies that all responses are statistically independent of one another, i.e., the number of independent degrees of freedom is equal to the number of responses and, hence, no single biasing technique can be used to reduce all of them simultaneously. However, if the matrix is dense and poorly conditioned (i.e., numerically singular), then the responses are highly correlated. Moreover, the effective number of degrees of freedom determined by the numerical rank of the matrix is smaller than the number of responses. If these correlations are known a priori, one can elect to bias MC-simulated particles directly towards the independent degrees of freedom rather than the original responses. In such case, the associated computer overhead will be reduced. The lower the effective rank of the covariance matrix, the more correlations between the responses, and the lower the computational overhead needed to reduce the variances globally. The key to this approach is the ability to estimate the covariance matrix before initiating MC simulation.

Let $\vec{r} = [r_1 \quad r_2 \quad \dots \quad r_n]^T \in \Re^n$ be a vector of n responses of interest representing n random Gaussian processes. Consider N realizations of these random processes denoted by

$$\left\{r_i^k\right\}_{i=1}^{n}, \quad k = 1,2,\dots,N.$$

An unbiased estimate of the covariance between the two responses r_i and r_j can be given by

$$\mathrm{cov}\,(r_i,r_j) = \lim_{N\to\infty} \frac{1}{N-1} \sum_{k=1}^{N} \left(r_i^k - \hat{r}_i\right)\left(r_j^k - \hat{r}_j\right) \tag{1}$$

where $\bar{r}_i$ is the unbiased estimate of the mean of the response r_i, i.e.,

$$\hat{r}_i = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} r_i^k.$$

The covariance information between all pairs of responses can be represented compactly by a symmetric covariance matrix $\mathbf{C} \in \Re^{n\times n}$ such that $\mathbf{C}_{ij} = \mathrm{cov}\,(r_i,r_j)$. This matrix can be rewritten using singular value decomposition as follows (Golub and Van Loan, 1996):

$$\mathbf{C} = \mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^T = \sum_{i=1}^{n} \sigma_i \vec{w}_i \vec{w}_i^{\,T}, \tag{2}$$

where $\mathbf{W} = [\vec{w}_1 \quad \vec{w}_2 \quad \dots \quad \vec{w}_r] \in \Re^{n\times r}$ is a matrix of n orthonormal singular vectors and $\boldsymbol{\Sigma} = \mathrm{diag}\{\sigma_1,\sigma_2,\dots,\sigma_n\}$ is a diagonal matrix of nonzero singular values.

In an earlier work (Zhang and Abdel-Khalik, 2011b), a concept of a pseudo response was introduced to alter the biasing for MC-simulated particles. We recall the definition here for the sake of a complete discussion.

Using elements of probability theory, one can show that all responses of the form

$$\xi_i = \vec{w}_i^{\,T} \vec{r},$$

denoted hereinafter as pseudo responses, are uncorrelated, i.e.,

$$\mathrm{cov}\,(\xi_i,\xi_j) = 0 \quad \text{for } i \neq j, \text{ and } \mathrm{cov}\,(\xi_i,\xi_i) = \sigma_i.$$

Each pseudo response represents a linear combination of the original responses; the weights are determined by the elements of the singular vectors of the covariance matrix.

The idea here is that, if the rank of the covariance matrix is significantly smaller than its dimensions, as observed in the earlier work, the number of pseudo responses will be much smaller than the number of original responses. One can take advantage of this situation by designing weight windows to bias MC-simulated particles towards the pseudo responses rather than the original responses. A pseudo response with very small singular value implies that the associated linear combination of the original responses has a very small statistical variation and, hence, no MC-simulated particles need to be sampled to calculate it.

To provide a basis for comparison, we contrast the GP approach to the FW-CADIS approach (Wagner et al., 2009). In both of these approaches, the weight window-based biasing is determined by solving an adjoint equation of the form (demonstrated first for a single response)

$$\mathbf{L}^*\left(\vec{\phi}_i^*\right)=\frac{\partial r_i}{\partial \vec{\phi}},$$

where $\boldsymbol{L}^*$ is the adjoint transport operator with appropriate boundary conditions, $\vec{\phi}$ is the forward flux, and $\vec{\phi}_i$ is the adjoint flux corresponding to the i^{th} response r_i. In nuclear engineering, forward flux (the solution obtained from standard neutron transport equation) is a quantity corresponding to the total length traveled by all neutrons per unit time and volume (it works as an important indicator of the intensity of neutron activities within the nuclear reactor), and adjoint flux is, as the name implies, the solution obtained from the equation in adjoint format of neutron transport. Adjoint flux is an important quantity in nuclear reactor analysis because its magnitude physically indicates the importance of the neutron in its phase space.

If $\vec{\phi}_i$ is employed to determine weight windows, one can reduce the variance of the response r_i only, implying generally high variances for the other responses. To achieve GVR, the FW-CADIS approach formulates a single response of the form

$$\xi^{\text{FW-CADIS}}=\sum_{i=1}^{n}\frac{1}{\phi_i}r_i \tag{3}$$

with an associated adjoint model of the form

$$\mathbf{L}^*\left(\vec{\phi}_i^*\right)=\sum_{i=1}^{n}\frac{1}{\phi_i}\frac{\partial r_i}{\partial \vec{\phi}} \tag{4}$$

In this technique, more MC-simulated particles are sent to regions in the phase space where the flux is low. The rationale behind this is that regions with lower flux receive fewer MC-simulated particles and their associated responses are therefore expected to have higher variances. To render variances uniform across the phase space, more particles are encouraged to go to lower flux regions.

Employing the terminology introduced for the GP approach, one can think of the FW-CADIS approach as also utilizing a pseudo response where the weights for the linear combination are now determined by the reciprocal of the flux. In the GP approach, however, the pseudo responses are calculated from the singular vectors of the covariance matrix as shown above.

The covariance matrix describing the statistical correlations between different responses can be estimated using an inexpensive deterministic model. To simulate statistical uncertainties, recall that the interaction between a neutron and a target nucleus is a random process whose probability is characterized by cross-sections. By sampling cross-sections randomly, one can estimate the response correlations.

Based on this idea, the covariance matrix $\boldsymbol{C}$ is constructed by performing l deterministic forward model executions. Denote the n responses generated in the j^{th} model execution by a vector $\vec{r}_j$, where $j=1,2,\ldots,l$. After l executions, the estimator for the covariance matrix can be constructed as follows:

$$\mathbf{C}=\frac{1}{l-1}\sum_{j=1}^{l}\left(\vec{r}_j-\hat{\vec{r}}\right)\left(\vec{r}_j-\hat{\vec{r}}\right)^T, \tag{5}$$

where $\hat{\vec{r}}$ is a vector of estimated means of all responses, i.e., $\vec{r}=\frac{1}{l}\sum_{j=1}^{l}\vec{r}_j$.

In addition to Eq. (2), the covariance matrix $\boldsymbol{C}$ can also be approximated with truncated SVD approach as follows:

$$\mathbf{C}_r=\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^T=\sum_{i=1}^{r}\sigma_i\vec{w}_i\vec{w}_i^T, \tag{6}$$

where r is known as the effective rank of the covariance matrix. The effective rank is estimated using the inequality

$$\|\mathbf{C}-\mathbf{C}_r\|<\varepsilon, \tag{7}$$

where ε is a user-defined tolerance. As mentioned before, the GP pseudo responses are described by

$$\xi_j^{\text{GP}}=\vec{w}_j^T\vec{r},\quad j=1,2,\ldots,r. \tag{8}$$

The MAVRIC sequence is then executed to bias the MC solution to minimize the variance of the pseudo responses. In its basic implementation, the GP approach executes the MAVRIC sequence r times with r adjoint model evaluations of the form

$$\mathbf{L}^*\left(\vec{\phi}_i^*\right)=\frac{\partial \xi_i^{GP}}{\partial \vec{\phi}}i=1,\ldots,r. \tag{9}$$

The premise is that reducing the variance of the r pseudo responses simultaneously reduces the variances of all original responses because of the correlations.

Note that, in the FW-CADIS technique, only one adjoint and one forward model executions are required. The GP approach, however, requires at least r forward flux executions to construct the covariance matrix. Moreover, it requires r different adjoint calculations to bias the solution towards the various pseudo responses. These additional adjoint runs have to be considered when one compares figures of merit, which are usually provided in Monte Carlo techniques to assist users in assessing the statistical behavior of the obtained results. In this work, the unbiased estimators for the response's mean and variance determined from all r simulations are calculated as follows (Papoulis and Pillai, 2002):

$$\bar{x}=\frac{\sum_{j=1}^{r}\frac{1}{\sigma_j^2}x_j}{\sum_{j=1}^{r}\frac{1}{\sigma_j^2}}\quad\text{and}\quad\frac{1}{\sigma^2}=\sum_{j=1}^{r}\frac{1}{\sigma_j^2}, \tag{10}$$

where x represents a given response, while x_j and σ_j are the mean value and the standard deviation of the response of interest evaluated in the j^{th} run. These formulas assume that all evaluations are statistically independent, which follows from the independence of the pseudo responses. Moreover, the formulas ensure that simulations with high variance will have little impact on the unbiased estimate of the mean value. The variance formula implies that the overall variance gets smaller as more simulations are executed, which is consistent with the law of Monte Carlo sampling.

It is important to note that the potential value of the GP approach can be realized for models with high dimensional phase space and many responses. Like any projection approach, the potential for reduction increases with increasing dimensionality of the model. When only few responses are required, however, it is better to design weight window maps that target the responses of interest only. Our primary interest focuses on two applications. One of them is to use MC models to generate the few-group cross-sections for core wide calculations. In this case, thousands of assembly calculations must be completed. Given the correlations between these models, a huge reduction in the computational overhead may be possible. The other application is core wide MC models, where the flux is required everywhere in the phase space.

We note that the approach described above is the simplest one to take advantage of response correlation. In effect, the GP approach has reduced the number of original responses to r pseudo responses, which have been reduced separately using

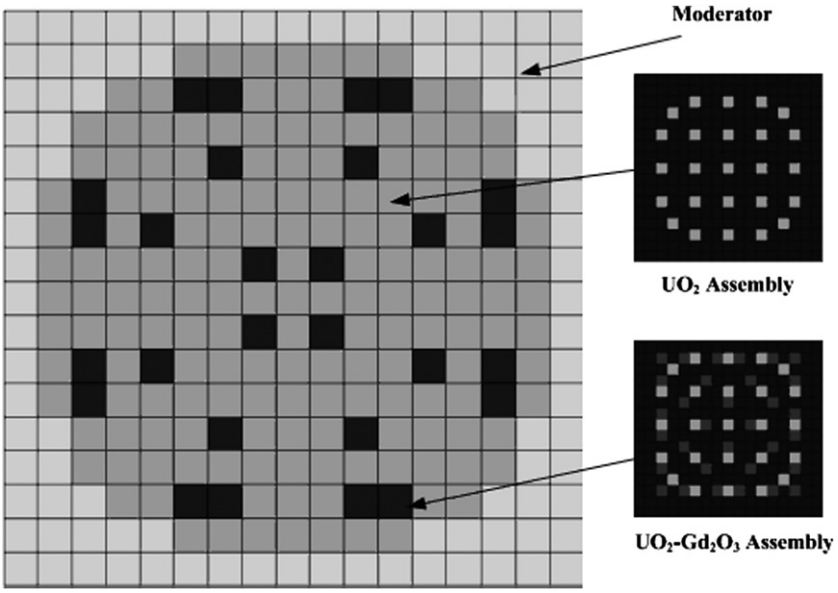

Fig. 1. PWR core model.

adjoint-based weight windows. Here we propose two other biasing approaches: one is combining the pseudo responses by scaling each one based on its associated singular values, the other is coupling the GP and FW-CADIS approaches, that is, combining the pseudo responses by scaling each pseudo responses by an appropriately selected pseudo flux. The mathematical representatives of the pseudo responses for these two approaches are as follows:

(1) combined GP approach

$$\begin{aligned} &\xi_j = \vec{w}_j^T \vec{r}, j = 1, \cdots, r \\ &\xi_{\text{com}} = \sum_{j=1}^{r} \sigma_j \xi_j \end{aligned} \quad ; \tag{11}$$

(2) coupled GP and FW-CADIS approach

$$\begin{aligned} &\psi_j = \vec{w}_j^T \vec{\phi}, j = 1, \cdots, r \\ &\xi_j = \vec{w}_j^T \vec{r}, j = 1, \cdots, r \\ &\xi_{\text{com}} = \sum_{j=1}^{r} \frac{1}{\psi_j} \sigma_j \xi_j \end{aligned} \tag{12}$$

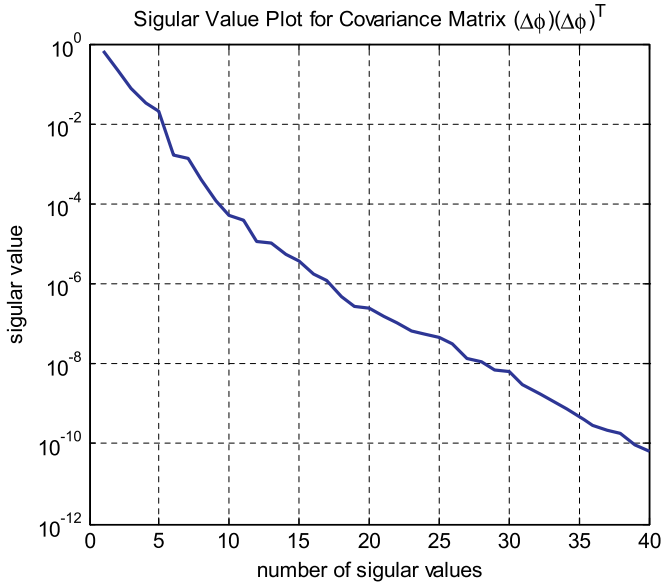

Fig. 2. Covariance matrix singular values.

3. Numerical experiments

A prototypical PWR full-core model has been developed employing the standard SCALE geometry utility. It is designed as a slight variation of the published benchmark problems (Yamamoto et al., 2002; Douglass et al., 2010). The full-core model consists of 193 fuel assemblies laid out in a 17×17 grid scheme and surrounded by light water. The cubic volume of the whole active core is $365.6 \times 365.6 \times 335.3$ cm^3. The cubic volume of each assembly is $21.505 \times 21.505 \times 335.28$ cm^3. Two types of fuel assemblies are modeled: a UO_2 and a UO_2–Gd_2O_3 fuel assembly. Fig. 1 shows the loading pattern of the full-core. Each assembly consists of a 17×17 grid of pin cells with each pin cell measuring 1.265×1.265 cm^2 in the X–Y plane.

The Denovo code, a deterministic neutron transport solver with both forward and adjoint calculation capabilities in the MAVRIC sequence, was employed to estimate the covariance matrix. Fig. 2 shows the normalized singular values of the covariance matrix. Note that the singular values decline quickly over several orders of magnitude in fewer than 20 model runs. This indicates that the corresponding responses of interest (thermal flux in this case) are highly correlated.

Fig. 3 shows the procedure for determining the effective rank of covariance matrix. A cutoff of $r=12$ (solid horizontal line) is selected so that the error in Eq. (7) would not exceed the desired statistical deviation of the responses.

To compare the various proposed hybrid approaches, we employed a simple metric, which describes the relative reduction in standard deviation of quantities of interest defined as follows:

$$\varepsilon = \frac{\sigma^{\text{Approach}} - \sigma^{\text{FW-CADIS}}}{\sigma^{\text{FW-CADIS}}} \times 100\%, \tag{13}$$

where σ^{Approach} is the standard deviation of quantities of interest for the approach considered, and all comparisons are done with the FW-CADIS approach.

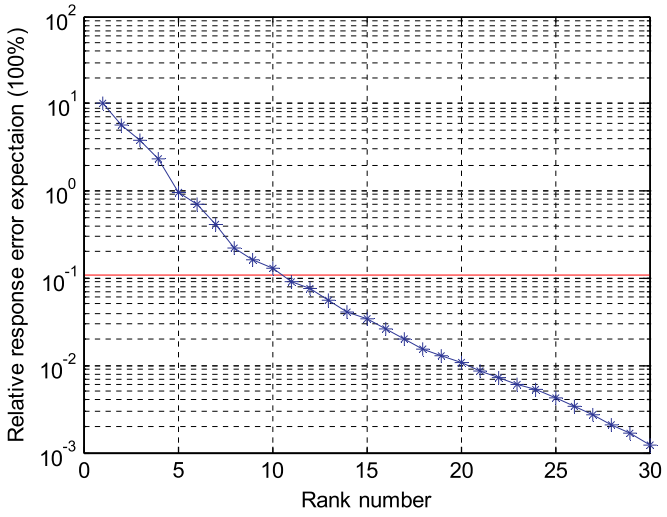

Fig. 3. Estimate of the effective rank of the covariance matrix.

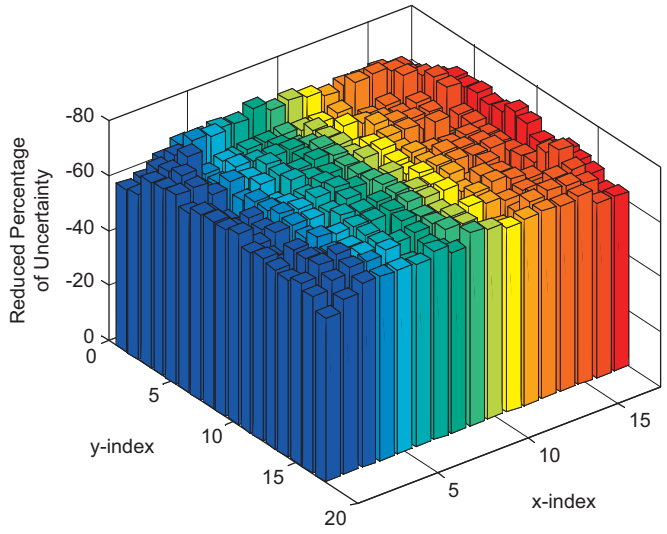

Fig. 4. Comparison of standard deviations of the thermal flux responses provided by GP and FW-CADIS.

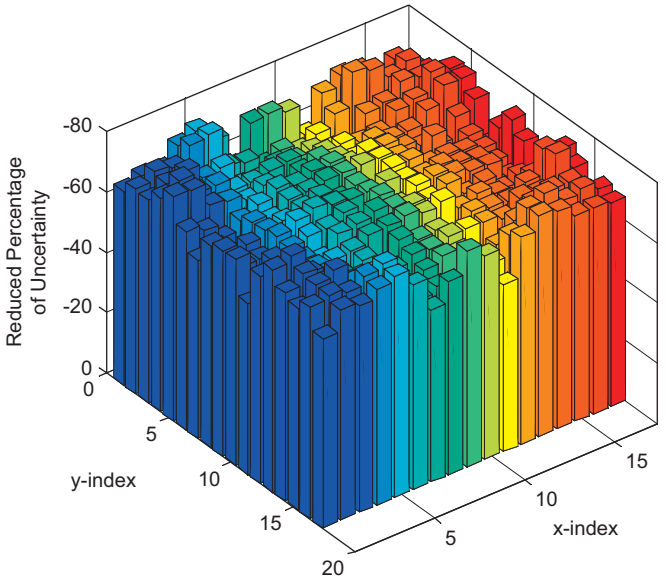

Fig. 5. Comparison of standard deviations of the thermal flux responses provided by combined GP and FW-CADIS.

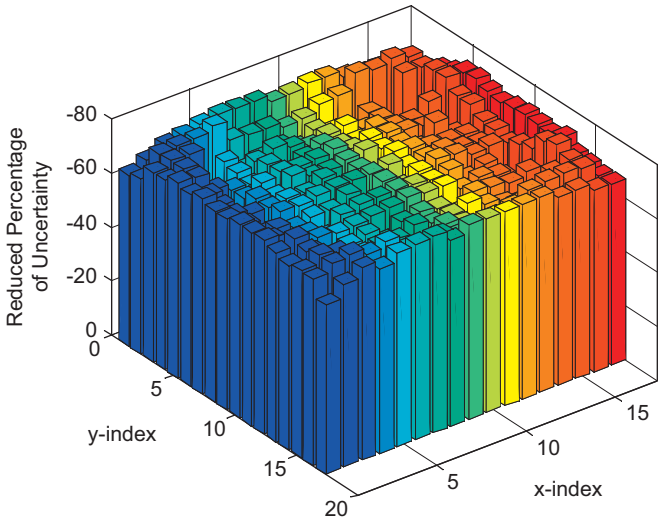

Fig. 6. Comparison of standard deviations of the thermal flux responses provided by Coupled GP&FW-CADIS and FW-CADIS.

No figure-of-merit comparisons were done in this study, as MC calculations dominated the total computational time (the MC computer time is in the order of hours, while deterministic time is in the order of seconds). In the future, a more complete comparison based on the figure-of-merit type metric will be performed to help compare the different approaches. This will be done as our future work will focus on the eigen-value problem, where deterministic models are expected to take more time, especially for high dimensional problems.

Results in Fig. 4 show that the standard deviations of spatial thermal fluxes in the simple reactor are reduced on average by a factor of four as compared with the FW-CADIS approach. Regarding the combined GP and GP&FW-CADIS hybrid approaches, Figs. 5 and 6 show similar levels of variance reduction. Slightly better performance is observed with the GP&FW-CADIS approach.

On average, all approaches show a reduction in the standard deviation of thermal flux response by a factor between three and four, which translates into fewer histories by a factor of 10–16 to reach the same variance as with the FW-CADIS approach. An earlier paper on FW-CADIS reported a potential 6-10-fold gain in speed over the analog approaches (Wagner et al., 2009). The results suggest that, when the FW-CADIS and GP approaches are combined, one can increase speed by a factor of up to 100.

4. Conclusion

A new approach to global variance reduction is presented, which is based on the Gaussian process theory, and its results are compared with the results of the FW-CADIS approach. The Gaussian process approach takes advantage of correlations between the variances of the various responses, which are described by the covariance matrix. Deterministic calculations are employed to construct an estimate of the covariance matrix by randomly sampling cross-sections. A full-core model is implemented and analyzed. A significant rank reduction is observed for the covariance matrix, implying that only few adjoint evaluations are needed to perform global variance reduction. Current work has focused on the fixed-source problem only. The ultimate goal

of our work is to apply the global variance reduction to eigen-value problems.

References

Becker, T.L., Wollaber, A.B., Larsen, E.W., 2007. A hybrid Monte Carlo-deterministic method for global particle transport calculations. Nucl. Sci. Eng. 155 (2), 155–167.

Densmore, J.D., Larsen, E.W., 2003. Variational variance reduction for particle transport eigenvalue calculations using Monte Carlo adjoint simulation. J. Comput. Phys. 192 (2), 387–405.

Douglass, S., Rahnema, F., Margulies, J., 2010. A stylized three dimensional PWR whole-core benchmark problem with gadolinium. Ann. Nucl. Eng. 37 (10), 1384–1403.

Golub, G.H., Van Loan, C.F., 1996. Matrix Computations, 3rd ed. Johns Hopkins University Press, Baltimore, MD.

Haghighat, A., Wagner, J.C., 2003. Monte Carlo variance reduction with deterministic importance functions. Progr. Nucl. Energy 42 (1), 25–53.

Kennedy, M.C., O'Hagan, A., 2001. Bayesian calibration of computer model. J. R. Stat. Soc. B 63 (3), 425–464.

Papoulis, A., Pillai, S., 2002. Probability, Random Variables, and Stochastic Processes. McGraw Hill, New York.

SCALE, 2009. A modular code system for performing standardized computer analyses for licensing evaluations, ORNL/TM-2005/39, Version 6, vols. I–III. Available from Radiation Safety Information Computational Center at Oak Ridge National Laboratory as CCC-750.

Wagner, J.C., Blakeman, E.D., Peplow, D.E., 2009. Forward-weighted CADIS method for variance reduction of Monte Carlo calculations of distributions and multiple localized quantities. In: Proceedings of International Conference on Advances in Mathematics, Computations, and Reactor Physics. (M&C 2009). Saratoga Spring, NY. May 3–7.

Yamamoto, A., Ikehara, T., Ito, T., Saji, E., 2002. Benchmark problem suite for reactor physics study of LWR next generation fuels. J. Nucl. Sci. Technol. 39 (8), 900–912.

Zhang, Q., Abdel-Khalik, H.S., 2011a. Gaussian process approach for global variance reduction. Trans. Am. Nucl. Soc. 104, 805–808.

Zhang, Q., Abdel-Khalik, H.S., 2011b. Adjoint-based global variance reduction approach for reactor analysis problems. In: Proceedings of International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering (M&C 2011). Rio De Janeiro, Brazil. May 8–12.