CASL-U-2016-1054-000

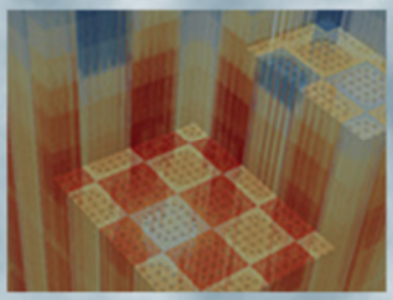
Data Assimilation and Uncertainty Quantification Using VERA-CS for a Core Wide LWR Problem with Depletion

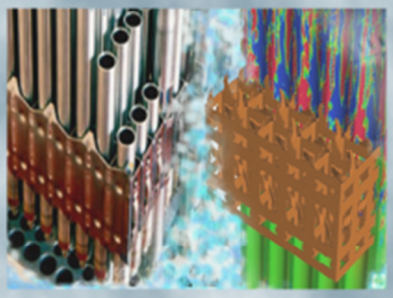
L2:VMA.P12.01

Bassam A. Khuwaileh and Paul J. Turinsky

North Carolina State University, Department of Nuclear Engineering, Raleigh, NC.

This research was supported by the Consortium for Advanced Simulation of Light Water Reactors (www.casl.gov), an Energy Innovation Hub (http://www.energy.gov/hubs) for Modeling and Simulation of Nuclear Reactors under U.S. Department of Energy Contract No. DE-AC05-00OR22725.





Oak Ridge National Laboratory

in partnership with

Electric Power Research Institute
Idaho National Laboratory
Los Alamos National Laboratory
Massachusetts Institute of Technology
North Carolina State University
Sandia National Laboratories
Tennessee Valley Authority
University of Michigan
Westinghouse Electric Company

and individual contributions from

Anatech Corporation
ASCOMP GmbH
CD-adapco, Inc.
Core Physics, Inc.
City University of New York
Florida State University
Notre Dame University
Imperial College London
Pacific Northwest National Laboratory
Pennsylvania State University
Rensselaer Polytechnic Institute
Southern States Energy Board
Texas A&M University
University of Florida
University of Tennessee
University of Wisconsin

REVISION LOG

Revision	Date	Affected Sections	Revision Description
Rev. 0	02/22/2016	All	Initial release
Rev. 1	03/11/2016	All	Draft # 2
Rev. 3	03/16/2016	All	Draft # 3
Rev. 4	03/22/2016	All	Draft # 4
Rev. 5	03/22/2016	All	Draft # 5
Rev. 6	03/23/2016	All	Draft # 6

Abstract

In this report, an uncertainty quantification and data assimilation study is performed on CASL Progression Problem Number 9, a quarter-core, three-dimensional model representative of Watts Bar Nuclear Unit 1 Cycle 1. Simulations are completed using the VERA-CS core simulator. The uncertainty is quantified before and after a data assimilation study is performed to calibrate the cross-sections and a few thermal-hydraulics parameters.

First, the uncertainties of several quantities of interest are quantified via three different methods proposed in a previous milestone report [1]. The performance of the proposed methods is compared and analyzed. Second, a surrogate based data assimilation method is proposed and tested in conjunction with QUESO-DRAM. The data assimilation study is used to update the uncertainty distribution for the many group cross-sections and thermal-hydraulic parameters utilizing responses of interest based upon synthetic experimental data appropriate for CASL Progression Problem Number 9. Finally, these updated parameter uncertainties were used to re-quantify the uncertainties of the quantities of interest for CASL Progression Problem Number 9 and then compared with the uncertainties quantified prior to the data assimilation.

1. Introduction and Theory

In a previous milestone report, efficient uncertainty quantification algorithms were proposed and tested with CASL Progression Problem Number 6 [1], a single fuel assembly represented in three dimensions, utilizing the VERA-CS core simulator and ROMUSE uncertainty quantification software. The first part of this report applies the previously proposed methods to quantify the uncertainties of the quantities of interest for CASL Progression Problem Number 9, a quarter-core, three-dimensional model that is representative of Watts Bar Nuclear Unit 1 Cycle 1 (WB1C1); therefore, it is important for the reader to refer to Ref. [1]. The second part of this report introduces an efficient surrogate based data assimilation method. The method is tested via synthetic data and then the results of this part are used in the third part to exemplify the effects of utilizing the data assimilation results to re-quantify the uncertainties of the quantities of interest for CASL Progression Problem Number 9.

Data assimilation is a mathematical methodology used for establishing a connection between experimental and operational data and simulation completed employing mathematical models. Any mathematical model is an approximate representation of the real phenomenon of interest. The purpose of data assimilation is to improve the performance of the mathematical model by calibrating the model's parameters. Model-based assimilation of experimental measurements (i.e. data assimilation) has been used in various engineering fields (including nuclear engineering) for the enhancement of the predictions made by mathematical models during simulations [2-3].

Overall, there are two major problems facing the data assimilation practice. First, the computational burden associated with running the high fidelity models (reactor core simulators). Second, the curse of dimensionality associated with the number of model parameters that will need to be calibrated. In contrast to uncertainty quantification, the computational cost of data assimilation increases with the number of parameters to be calibrated. In nuclear reactor simulation, the number of neutronics parameters is large due to the fact that the nuclear data cross-sections libraries are detailed for high fidelity modeling and simulation.

Data assimilation utilizing the long operational experience with light water reactors could improve simulation fidelity. In this report the Delayed Rejection Adaptive Metropolis algorithm (DRAM) [4] will be used in conjunction with reduced order modeling, such that the end result is a feasible and applicable algorithm for data assimilation for large-scale reactor simulation applications.

Reduced order modeling can facilitate the two major problems noted earlier. Surrogates can address the first problem (the computational burden associated with running the high fidelity reactor core simulators). On the other hand, reducing the dimensionality of the parameters of interest using the algorithms such as those proposed in Ref. [1,5,6] will address the second problem (the curse of dimensionality). Ref. [2] introduced a high-order predictive model calibration algorithm and applied it to relatively large-scale applications, while Ref. [3] performed data assimilation for a few thermal-hydraulics parameters using a lower-order surrogate to replace the actual thermal-hydraulics simulator (i.e. COBRA-TF). This report will use polynomial surrogate models to substitute for the original coupled models (MPACT, COBRA-TF and ORIGEN) making up VERA-CS (see Figure 1) to calibrate cross-sections (high dimensional parameters) along with the few thermal-hydraulics parameters considered here.

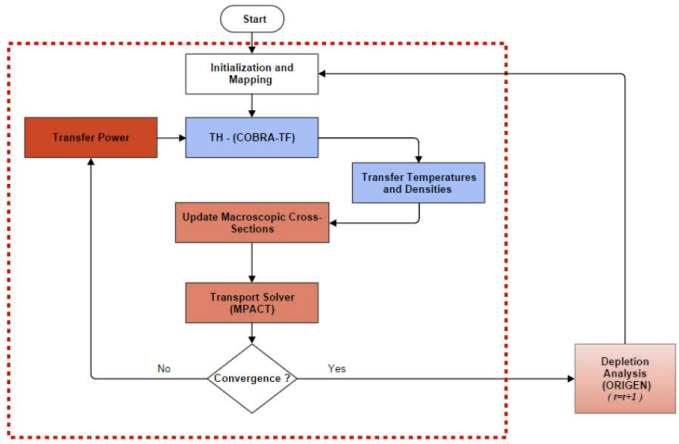


Figure 1.Flow diagram of the VERA-CS computer code.

A subspace-based surrogate model ($\tilde{f}$) will be used to replace the model of interest (f). Ref. [2, 3, 7] introduced a methodology previously developed for goal-oriented surrogate modeling. In this report a gradient free approach is used to construct the basis of the lower dimensional subspace approximation. A 2nd order goal-oriented subspace can be constructed as follows:

$$f \approx \tilde{f} = \bar{S}_1^T \Delta\bar{x} + \left(\bar{S}_2^T \Delta\bar{x}\right)^2 + \varepsilon$$

where f is the quantity of interest (e.g. multiplication factor, maximum fuel pin power and maximum fuel pin temperature), and $\Delta\bar{x}$ is the variation in the parameters of interest from the reference values (e.g. cross-sections, gap conductivity and grid spacer loss coefficient).

The goal of the surrogate here is to perform data assimilation analysis. The uncertainty and mean of each parameter will be updated; therefore, for the surrogate construction the perturbations ($\Delta\bar{x}$) will be generated randomly within an interval of interest. Once the basis of the subspace is determined ($\mathbf{U}$), projecting into the lower dimensional subspace, an efficient goal-oriented surrogate can be constructed as follows:

$$\begin{aligned} f \approx \tilde{f} &= \overbrace{\bar{\beta}_1^T \mathbf{U}}^{\bar{\beta}_{1,r}^T} \overbrace{\mathbf{U}^T \Delta\bar{x}_{rand}}^{\Delta\bar{\alpha}} + \left(\overbrace{\bar{\beta}_2^T \mathbf{U}}^{\bar{\beta}_{2,r}^T} \overbrace{\mathbf{U}^T \Delta\bar{x}_{rand}}^{\Delta\bar{\alpha}} \right)^2 + \varepsilon \\ &= \bar{\beta}_{1,r}^T \Delta\bar{\alpha} + \left(\bar{\beta}_{2,r}^T \Delta\bar{\alpha}\right)^2 + \varepsilon \end{aligned} \tag{1}$$

where $\Delta\bar{x}_{rand}$ is a random input vector sampled to generate a random sample of the input parameters. Given that $\mathbf{U} \in \mathbb{R}^{nxr}$ and $\Delta\bar{\alpha} = \mathbf{U}^T \Delta\bar{x}_{rand} \in \mathbb{R}^r$ then $\bar{\beta}_{1,r}^T = \bar{\beta}_1^T \mathbf{U} \in \mathbb{R}^r$ and $\bar{\beta}_{2,r}^T = \bar{\beta}_2^T \mathbf{U} \in \mathbb{R}^r$. Hence in order to determine the unknown elements of $\bar{\beta}_{1,r}$ and $\bar{\beta}_{2,r}$ VERA-CS

needs to be run at least 2r times from which the coefficients are determined by least squares estimation.

Moreover, in the calculation of the posterior Probability Density Function (PDF) for the parameters, the computational cost is dependent on the dimensionality of the parameter space $\mathbb{R}^p$ (i.e. p). Hence, by recasting the problem into a lower dimensional subspace approximation (i.e. $\mathbb{R}^r$), the number of calibrated parameters is reduced from p to r, where the uncertainty information can be mapped from the reduced space to the full dimensional space as follows:

$$\mathbf{C}_x = \mathbf{U}\mathbf{C}_\alpha \mathbf{U}^T \tag{2}$$

where $\mathbf{C}_x$ is the covariance matrix in the full dimensional parameter space and $\mathbf{C}_\alpha$ is the covariance in the reduced space as determined by the assimilation process. Note that $\mathbf{U}\mathbf{U}^T \approx \mathbf{I}$ where this approximation is justified in Ref. [5, 8, 9]. In addition, running the data assimilation analysis using the surrogate model has a negligible computational cost compared to the computational cost associated with simulating using the original model (high fidelity simulator).

The following is a summary of the algorithm for Subspace-Based Data Assimilation (**SBDA**):

1. Construct the basis of the lower dimensional subspace approximation of the parameter space (**U**).
2. Construct the goal-oriented surrogate model ($\tilde{f}$).
3. Run DRAM with the surrogate model $\tilde{f}$ and replace the input parameter space ($\mathbb{R}^p$) with the reduced space $\mathbb{R}^r$.

In the following sections, the **SBDA** will be used to perform parameter calibration. The parameters of interest are the nuclear data cross-sections (high dimensional parameter) and two thermal-hydraulics parameters (gap conductivity and grid spacer loss coefficients). In this report, we deal with the data assimilation algorithm (i.e. DRAM) as a black box; the real contribution of the work reported herein is the replacement of the original high fidelity model with a surrogate that has negligible computational cost to run.

2. Prior Uncertainty Quantification

Progression Problem Number 9 denotes a full core, 3D representation of WB1C1 with neutronics coupled to thermal-hydraulics and depletion feedbacks. In this problem both fuel and burnable absorbers are depleted throughout the first cycle. However, due to computational resource limitations, this section will perform the uncertainty quantification for a portion of Cycle 1 using several depletion steps. Specifically, the core is depleted to 160 effective full power days (EFPD) via 9 depletion steps (0, 9, 32, 45, 60, 80, 100, 120, 160 EFPD). Table 1 shows a summary of the core problem features while Figure *2* represents the layout and enrichment regions of the core.

Table 1. Progression Problem Number 9 features and design properties.

Property	Value	Details
Fuel Assemblies	193	Modeled in quarter symmetry
Enrichment	3.1%, 2.6%, 2.1%	17 Assemblies with 3.1%, 19 Assemblies with 2.6%, 20 Assemblies with 2.1%
Control Rods	8 banks of B_4C	Total of 18 rods in quarter symmetry.
Burnable Inserts	6,12,4,3,2 inserts of Pyrex burnable poison rodlet clusters.	These burnable poison inserts consist of 24, 20, 16, 12, 8 rod-lets, respectively.

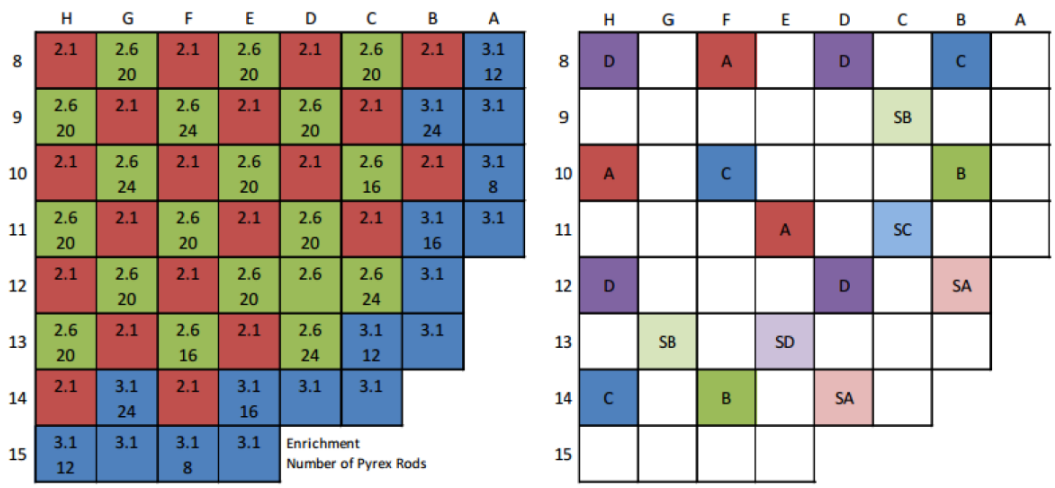


Figure 2. Core layout in quarter core symmetry. Enrichment and Pyrex rods (left) and core bank positions (right). For more information about the banks' worth (A, B, C, D, SA, SB and SD) refer to [11].

In this section, two approaches for multi-physics uncertainty propagation will be introduced, used and compared. The first approach is a linear approach that utilizes the obtained subspace along with a Karhunen-Loève (KL) expansion-based approach to quantify the uncertainty in quantities of interest calculated by multi-physics coupled models (Multi-Physics Efficient Uncertainty Quantification – MP-EUQ). The second approach, Surrogate Based Uncertainty Quantification (SBUQ), is a surrogate model based approach that replaces the original complex model with a simple surrogate form (e.g. polynomial, Gaussian process surrogate) and then quantifies the uncertainties via a Monte Carlo approach. Both of these approaches are summarized in Ref. [1]. In this report VERA-CS models (MPACT- COBRA-TF- ORIGEN) will be replaced by surrogates. Surrogates will be constructed separately for each of these models and then coupled to formulate an approximation to VERA-CS.

As in Ref. [1], the gradient free approach is used to obtain the reduced dimensional input parameter space in the form of the basis of the lower dimensional subspace, which is then used to perform linear uncertainty propagation via the MP-EUQ. In addition, the basis is used to construct the surrogate necessary for SBUQ.

In general, the surrogate quality can be evaluated by examining the residual errors of the quantity of interest, determined by taking the difference of predicted values from the surrogate and the original model. For example, the current surrogate is constructed for the purpose of UQ; therefore, 30 extra random samples are generated employing the parameters' PDFs and Monte Carlo sampling. The perturbations are created in all the parameters (i.e. the gap conductivity (h_{gap}) uncertainty of $\pm$50%, grid spacer loss coefficient uncertainty of $\pm$4%, and with cross-section perturbations generated employing the "44groupcov" covariance library distributed with SCALE6.1 [10]). For the gradient free approach, a subspace of rank 60 (60 degrees of freedom) was found sufficient to capture the variations of the parameters in the depletion cycle of interest with an L_2 – norm error upper bound of < 1% (for more details on the calculation of this error upper bound please consult Appendix A.1), so used in the results to now be reported upon. Table 2 shows the results of the error test of the 2nd order surrogates in terms of the surrogate form-related uncertainties: Root Mean Square (RMS) and the residuals distribution. Figure 3 through Figure 5 shows the quantity of interest residuals associated with the surrogate computed for different parameter sets generated by varying all the parameters simultaneously. For each sample the maximum error point occurring at all depletion steps is plotted, therefore we have only 30 points per plot. First, the surrogates were built separately for each component (MPACT COBRAT-TF and ORIGEN) and then the performance of thee coupled surrogate models was compared to the original VERA-CS sequence via 30 extra samples. The figures show that the 2nd order surrogates of the form represented by Eq. (1) can predict the multiplication factor (k_{eff}), maximum fuel pin power, and maximum fuel pin temperature with sufficient accuracy relative to the uncertainty estimation application.

Just like earlier work reported in [1], these surrogates will be used to quantify the uncertainty in the quantity of interest (including the addition of the surrogate form - related uncertainty which is defined in Appendix A.2) and then compared with the uncertainties estimated via a forward Monte Carlo approach (MCUQ) and finally compared with the uncertainty estimated by the MP-EUQ. Statistical samples are drawn using VERA-CS original models' sequences and then compared to samples drawn from the surrogate model replacing the original VERA-CS. Due to computer resource limitations, Monte Carlo sampling via the original model was limited to 50 samples which resulted in large statistical uncertainties associated with the limited sampling.

Table 2. Features of the Surrogate.

Surrogate Order	RMS	Number of Construction Data Points	Number of Extra Validation Data Points	Residuals Distribution	Surrogate Form – Related Uncertainty (Appendix A.2*
2nd order	$\varepsilon_{k_{eff}}$ → 20.07 pcm	120	30	iid	35.16 pcm
	$\varepsilon_{P_{\max}}$ → 0.01 W/cm			iid	0.01 W/cm
	$\varepsilon_{T_{\max}}$ → 3.0661 C^o			iid	1.12^o

*The numbers here represent the uncertainty in the surrogate response due to the uncertainty in the surrogate coefficients estimated by Eq. (5) which are then used in a Monte Carlo process to estimate the surrogate-form related uncertainty.

Table 3 shows the contribution of each of the parameter's variance into the overall variance in the quantities of interest (Q). Note that the values reported in this table are computed as follows:

$$\frac{\sigma^2_{parameter \to Q}}{\sigma^2_{joint \to Q}} \times 100\% \tag{3}$$

Table 3. Parameter contributions to the overall uncertainty (Refer to Eq.(3)).

Parameter	QoI	0 EFPD	9 EFPD	32 EFPD	45 EFPD	60 EFPD	80 EFPD	100 EFPD	120 EFPD	160 EFPD
Cross-Sections	k_{eff}	94.1%	94.1%	92.2%	92.2%	90.3%	92.2%	92.2%	94.1%	94.1%
	mp	8.4%	9.0%	10.2%	9.0%	10.2%	10.9%	12.3%	11.6%	11.6%
	mt	0.4%	0.4%	0.5%	0.3%	0.4%	0.4%	0.5%	0.4%	0.5%
Gap Conductivity	k_{eff}	5.8%	6.3%	6.3%	5.8%	6.8%	6.3%	6.3%	6.8%	6.3%
	mp	96.0%	98.0%	96.0%	96.0%	96.0%	98.0%	94.1%	96.0%	98.0%
	mt	86.5%	84.6%	86.5%	82.8%	84.6%	84.6%	84.6%	84.6%	84.6%
Grid Spacer Loss Coefficient	k_{eff}	10^{-4}%	10^{-4}%	10^{-4}%	10^{-4}%	10^{-4}%	10^{-4}%	10^{-4}%	10^{-4}%	10^{-4}%
	mp	10^{-6}%	10^{-6}%	10^{-6}%	10^{-6}%	10^{-6}%	10^{-6}%	10^{-6}%	10^{-6}%	10^{-6}%
	mt	10^{-4}%	10^{-4}%	10^{-4}%	10^{-4}%	10^{-4}%	10^{-4}%	10^{-4}%	10^{-4}%	10^{-4}%

Details of the uncertainties estimated via the MP-EUQ (linear Karhunen Loeve-based approach), SBUQ (Monte Carlo sampling 2nd order surrogates) and forward MCUQ (Monte Carlo sampling via original model) are shown in Table 4 through Table 6. In these tables σ_Q denotes the standard deviation in the quantities of interest and σ_{σ_Q} denotes the standard deviation in σ_Q due to Monte Carlo sampling (Q = k_{eff}, maximum fuel pin power (*mp*), maximum fuel pin temperature (*mt*)). Both σ_Q and σ_{σ_Q} are reported relative to the nominal value.

Whenever the surrogate was used the uncertainty was estimated via 500 samples divided into 5 groups, while for the brute force MC calculations a total of 50 samples were used and divided into 5 groups. One can notice that for the multiplication factor (k_{eff}) the uncertainty estimated via the MP-EUQ is within one standard deviation (± σ) of the uncertainties estimated via the Monte Carlo approaches (MCUQ and SBUQ); however, for the maximum fuel pin power and the maximum fuel pin temperature some of the uncertainty values estimated by the MP-EUQ are much greater separated than one standard deviation. The reason for that might be due to the fact that the MP-EUQ is a linear approach assuming that the response of interest is a linear function of the parameters. The results in Table 4 indicate that the linear assumption is sufficient for the multiplication factor, while Table 5 and Table 6 indicate that the linear assumption is not sufficient for the maximum fuel pin power and maximum fuel pin temperature. This indicates that the future uncertainty quantification analysis will be performed via a surrogate-based approach in order to account for the non-linearity.

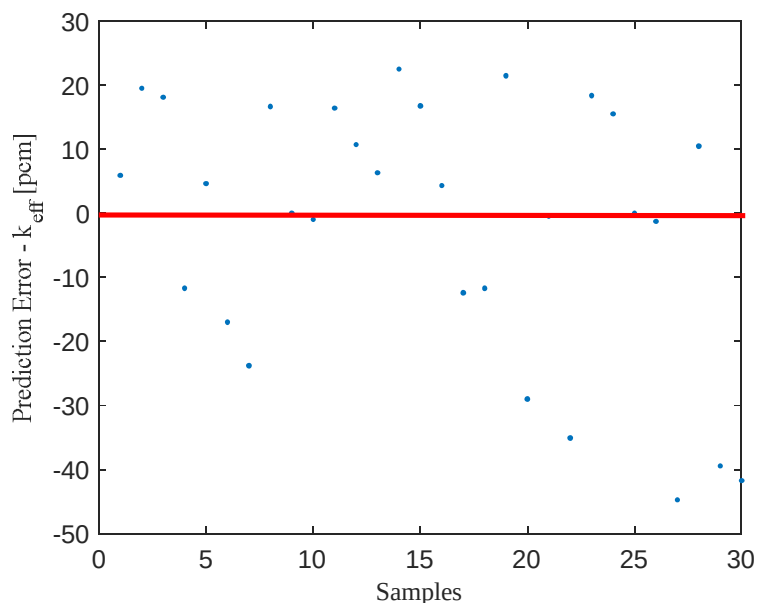


Figure 3. Residuals in predicting the k_{eff} ($\varepsilon_{k_{eff}}$) for a range of the gap conductivity and cross-sections (h_{gap} and $\overline{\Sigma}$) for 30 samples (surrogate vs. VERA-CS).

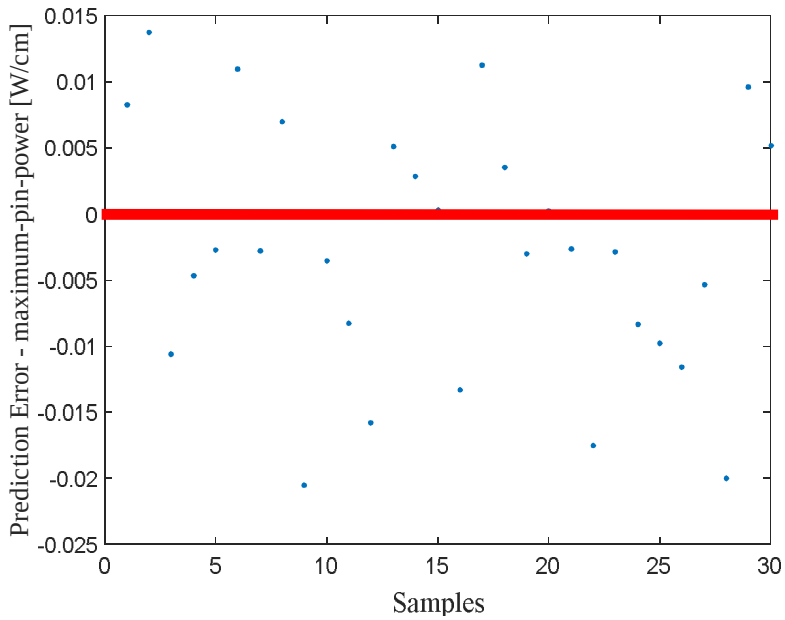


Figure 4. Residuals in predicting the maximum fuel pin power ($\varepsilon_{P_{\max}}$) for a range of the gap conductivity and cross-sections (h_{gap} and $\overline{\Sigma}$) for 30 samples (surrogate vs. VERA-CS).

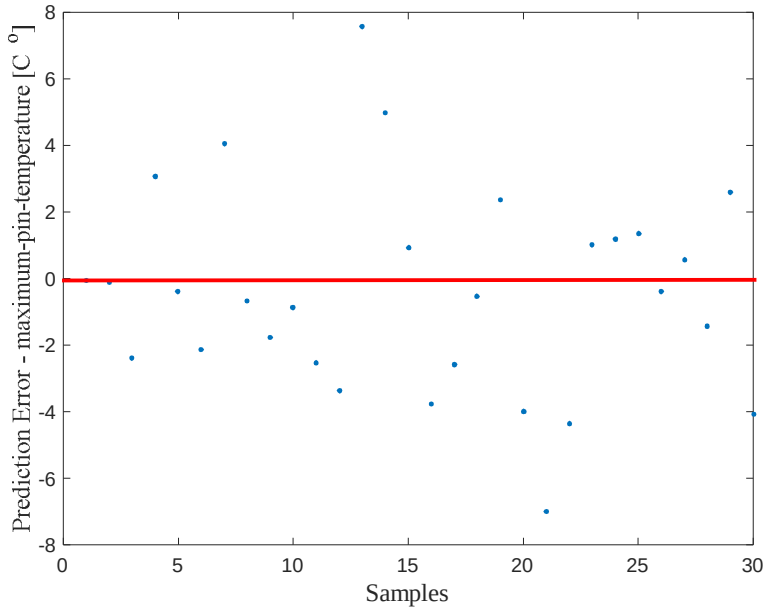


Figure 5. Residuals in predicting the maximum fuel pin temperature ($\varepsilon_{T_{\max}}$) for a range of the gap conductivity and cross-sections (h_{gap} and $\overline{\Sigma}$) for 30 samples (surrogate vs. VERA-CS).

Table 4. Summary of the k_{eff} uncertainty results – joint parameters (MCUQ vs. MP-EUQ vs. SBUQ)

Depletion (irradiation)	STD	MP-EUQ r=60	2nd Order Surrogate 500 samples	Monte Carlo 50 samples
0.0 EFPD	σ_k	0.389 %	0.425 %	0.401 %
	σ_{σ_k}	-	0.019 %	0.063 %
9 EFPD	σ_k	0.399 %	0.395 %	0.417 %
	σ_{σ_k}	-	0.021 %	0.066 %
32 EFPD	σ_{σ_k}	0.390 %	0.415 %	0.399 %
	σ_{σ_k}	-	0.025 %	0.074 %
45 EFPD	σ_k	0.380 %	0.420 %	0.398 %
	σ_{σ_k}	-	0.019 %	0.052 %
60 EFPD	σ_k	0.379 %	0.399 %	0.391 %
	σ_{σ_k}	-	0.032 %	0.071 %
80 EFPD	σ_k	0.357 %	0.398 %	0.392 %
	σ_{σ_k}	-	0.023 %	0.065 %
100 EFPD	σ_k	0.351 %	0.394 %	0.390 %
	σ_{σ_k}	-	0.023 %	0.062 %
120 EFPD	σ_k	0.349 %	0.391 %	0.387 %
	σ_{σ_k}	-	0.023 %	0.058 %
160 EFPD	σ_k	0.341 %	0.389 %	0.381 %
	σ_{σ_k}	-	0.026 %	0.046 %

Table 5. Summary of the Maximum Fuel Pin Power uncertainty results – joint parameters (MCUQ vs. MP-EUQ vs. SBUQ).

Depletion (irradiation)	STD	MP-EUQ r=60	2nd Order Surrogate 500 samples	Monte Carlo 50 samples
0.0 EFPD	σ_{mp}	1%	1.3 %	1.1 %
	$\sigma_{\sigma_{mp}}$	-	0.02 %	0.19%
9 EFPD	σ_{mp}	0.6 %	1.1 %	1.04 %
	$\sigma_{\sigma_{mp}}$	-	0.08 %	0.18 %
32 EFPD	σ_{mp}	0.9 %	1.5 %	1.4 %
	$\sigma_{\sigma_{mp}}$	-	0.011 %	0.2 %
45 EFPD	σ_{mp}	2%	2.2 %	2.9 %
	$\sigma_{\sigma_{mp}}$	-	0.2 %	1 %
60 EFPD	σ_{mp}	2.11 %	1.77 %	1.14 %
	$\sigma_{\sigma_{mp}}$	-	0.14 %	0.58 %
80 EFPD	σ_{mp}	2.45 %	1.64 %	1.00 %
	$\sigma_{\sigma_{mp}}$	-	0.20 %	0.55 %
100 EFPD	σ_{mp}	3.75 %	1.79 %	1.50 %
	$\sigma_{\sigma_{mp}}$	-	0.1 %	0.50 %
120 EFPD	σ_{mp}	2.50 %	1.62 %	1.29 %
	$\sigma_{\sigma_{mp}}$	-	0.11 %	0.33 %
160 EFPD	σ_{mp}	1.72 %	1.45 %	1.41 %
	$\sigma_{\sigma_{mp}}$	-	0.07 %	0.16 %

Table 6. Summary of the Maximum Fuel Pin Temperature uncertainty results – joint parameters (MCUQ vs. MP-EUQ vs. SBUQ).

Depletion (irradiation)	STD	MP-EUQ r=60	2nd Order Surrogate 500 samples	Monte Carlo 50 samples
0.0 EFPD	σ_{mt}	0.05 %	7.23 %	6.53%
	$\sigma_{\sigma_{mt}}$	-	0.07%	1.00%
9 EFPD	σ_{mt}	1.92%	5.10%	6.91%
	$\sigma_{\sigma_{mt}}$	-	0.02%	0.40%
32 EFPD	σ_{mt}	1.82%	4.73%	4.53%
	$\sigma_{\sigma_{mt}}$	-	0.03%	0.33%
45 EFPD	σ_{mt}	2.02%	5.15%	5.02%
	$\sigma_{\sigma_{mt}}$	-	0.03%	0.31%
60 EFPD	σ_{mt}	2.35%	5.59%	5.58%
	$\sigma_{\sigma_{mt}}$	-	0.08%	0.41%
80 EFPD	σ_{mt}	2.44%	5.56%	5.44%
	$\sigma_{\sigma_{mt}}$	-	0.04%	0.39%
100 EFPD	σ_{mt}	3.00%	5.11%	5.00%
	$\sigma_{\sigma_{mt}}$	-	0.09%	0.43%
120 EFPD	σ_{mt}	2.11%	5.46%	5.22%
	$\sigma_{\sigma_{mt}}$	-	0.06%	0.38%
160 EFPD	σ_{mt}	2.33%	5.87%	5.67%
	$\sigma_{\sigma_{mt}}$	-	0.07%	0.31%

3. Data Assimilation for CASL Progression Problem Number 9

In this case study the cross-sections and two thermal-hydraulic parameters are calibrated using the **SBDA** methodology. The core simulator of interest is VERA-CS which is used to model Progression Problem Number 6, denoting a three dimensional fuel assembly, and Progression Problem Number 9, denoting a full core, 3D representation of WB1C1. Progression Problem Number 6 was completed for the purpose of developing, implementing in software, and verifying the SBDA methodology; in this report Progression Problem Number 6 will not be reported upon and the reader is referred to Ref. [1,8]. Figure 1 presents the constituent models of VERA-CS. Referring to the figure; the thermal-hydraulic parameters of interest - in this case study – are the gap conductivity (h_{gap}) and the grid spacer loss coefficient (g_{loss}). The nuclear data parameters of interest are the cross-sections, specifically the fission, absorption and scattering cross-sections for 18 isotope-reaction pairs. The number of energy groups in the VERA-CS cross-sections library is 47 (refer to Table 7); resulting in 846=18×47 cross-section parameters and two thermal-hydraulics parameters to be calibrated. The DRAM – QUESO algorithm (encoded in DAKOTA 6.2 [11]) will be employed to solve the inverse problem associated with data assimilation using 100,000 samples per Markov chain. In the case of nuclear data cross-sections, a Gaussian prior distribution based on the covariance library (44groupcov) [10] is used while uniform distributions are used for the thermal hydraulics parameters (h_{gap} (±50%), g_{loss} (±4%)).

First, the goal-oriented surrogate is constructed in the form represented by Eq. (1). The surrogate is constructed as described in Section 1 and evaluated via examining the norm of residuals and their distribution. In this case study, the measured responses of interest are the multiplication factor (k_{eff}), and the three dimensional distribution for the relative fission reaction rate ($\mathbf{FR}$) within the instrument thimble located in each of the 56 fuel assemblies in the quarter core and assumed measured at 49 axial positions. Such measurements are consistent with core instrumentation using the moveable in-core detector system and when quarter core symmetry applies as is the case for WB1C1.

The surrogates will be constructed similarly as done for the uncertainty quantification analysis with the difference that the perturbations are generated randomly within the intervals of interest for data assimilation (h_{gap} (±50%), g_{loss} (±4%) , and cross-sections (±5%)). Refer to Table 8 for details about the surrogate error analysis. Note that the RMS for the **FR** reflects the RMS of a vector formed from the 49×56 matrix mentioned earlier; that is:

$$\varepsilon_{\mathbf{FR}} = RMS\left(vec\left\{\mathbf{FR}\right\}\right),$$

where *vec* is an operator that stacks the columns of a matrix into a vector. Due to computational resource limitations, 2nd order polynomial surrogates are used; hence the surrogate form-related uncertainty is higher than that of a 3rd order surrogate (which was used for data assimilation of Progression Problem Number 6 as reported upon in Ref. [8]). The effect of a higher surrogate form-related uncertainty is a larger error in calibrating the parameters of interest along with higher uncertainty. Therefore, in this section the data assimilation problem will be solved twice; once with including the surrogate form-related uncertainty and then without including the surrogate form-related uncertainty. Comparing the results of the two cases will indicate the effect of the surrogate errors on the calibration study. Figure *6* presents the k_{eff} surrogate residual errors

along with their distribution. Figure *7* shows the surrogate residual error distribution for the fission rate in the core's center fuel assembly. For the surrogate models to be useful, the residual errors should be an order of magnitude smaller than the experimental uncertainties.

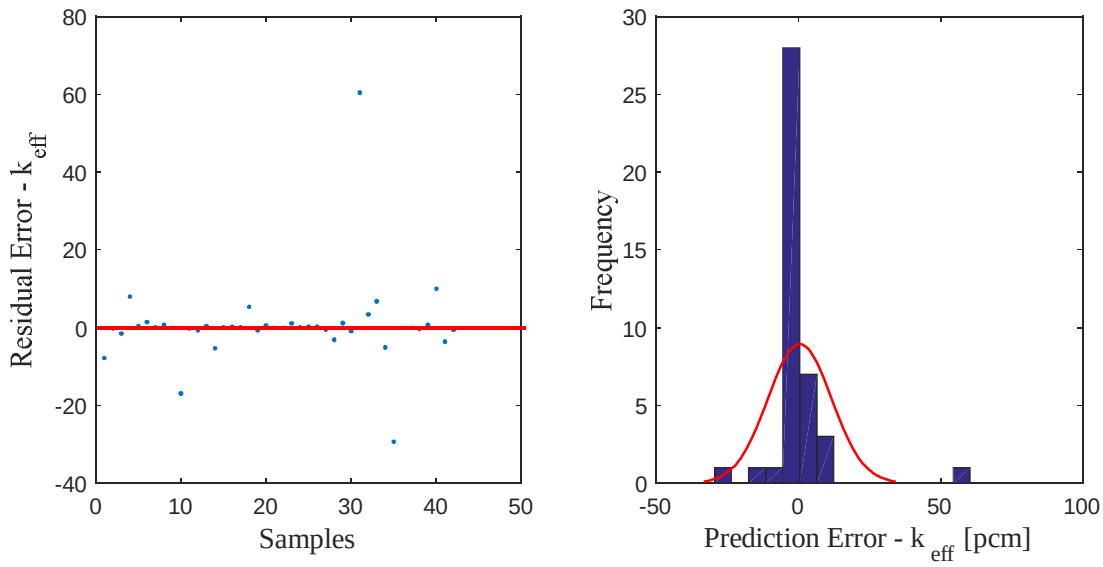


Figure 6. Residual errors along with their distribution - k_{eff}

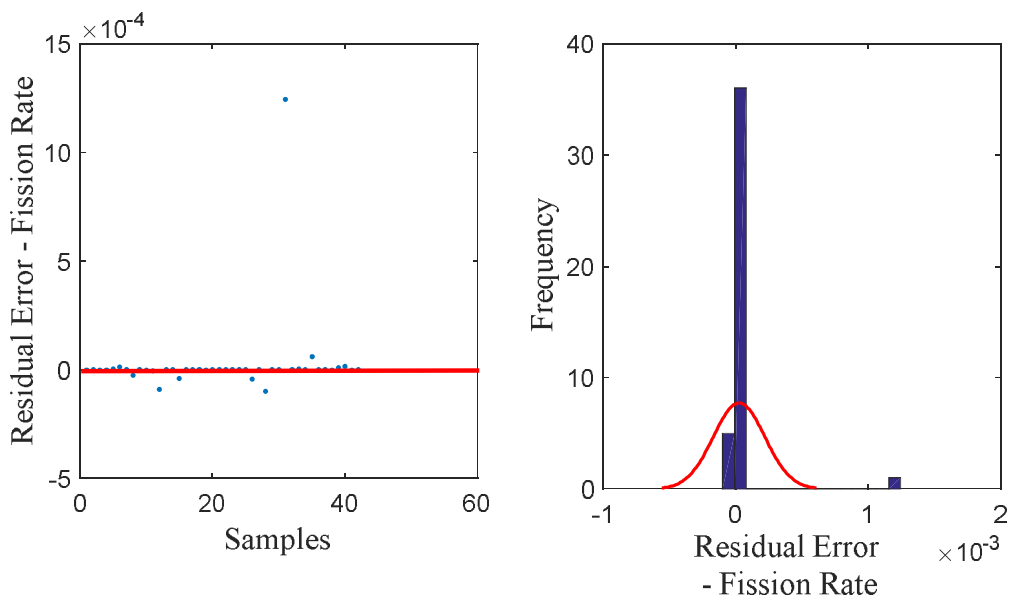


Figure 7. Residual errors along with their distribution – Fission rate in the core's center fuel assembly.

Synthetic data was used instead of real WB1C1 data for the responses of interest [i.e. experimental data to be used in the data assimilation]. Using the synthetic data implies that the actual solution of the data assimilation problem is known a priori, enabling the data assimilation method to be verified. Six depletion steps are used to generate the synthetic responses of interest. Hence, each response is measured at each of the depletion steps (0, 9, 32, 45, 120, 160 EFPD). Table 9 summarizes the measurements and their uncertainties. The synthetic measurements are generated via the high fidelity simulator (VERA-CS) where specific known parameter perturbations are used. Magnitudes of uncertainties associated to each of these measurements are based on the uncertainties that would be experienced in actual experimental measurements. The surrogates are then used for simultaneous calibration of the thermal-hydraulics parameters along with the cross-sections of interest. Therefore, as noted above the ideal solution to the data assimilation problem is known a priori and the performance of the **SBDA** can be evaluated by comparing the known perturbations and the mean variations generated by DRAM in conjunction with the surrogate (**SBDA**).

Figure 8 and Figure 9 shows the Monte Carlo Markov chains and their distributions generated by DRAM for the thermal-hydraulics parameters (grid spacer loss coefficient and gap conductivity). Figure 10 indicates that there is no correlation between the gap conductivity (h_{gap}) and the grid spacer loss coefficient (g_{loss}). It was noted that the multiplication factor is not sensitive to the grid spacer loss coefficient. Moreover, it was found that the effect of the grid spacer loss coefficient is the weakest among all other parameters regarding the fission rate responses at the fuel assemblies' instrument thimble location. Therefore, the variations in the grid spacer loss coefficient are not constrained further by using the current responses of interest.

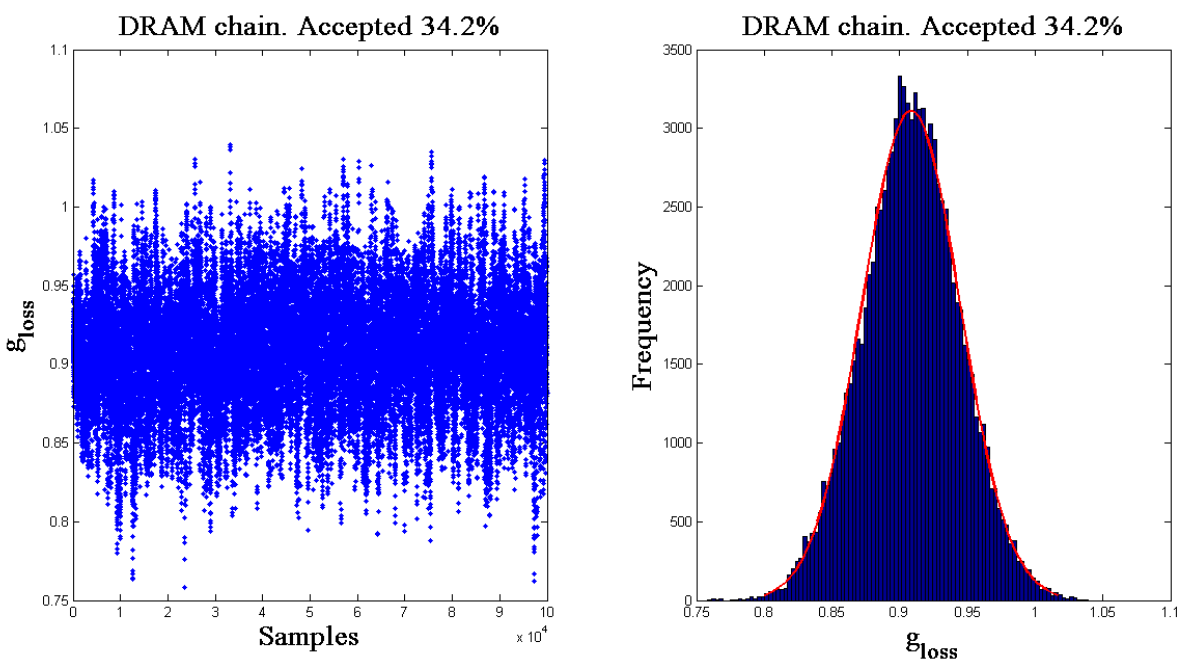


Figure 8. Chain and posterior distribution of the grid spacer loss coefficient (g_{loss}). [Including $\mathbf{C}_{surrogate \to f}$].

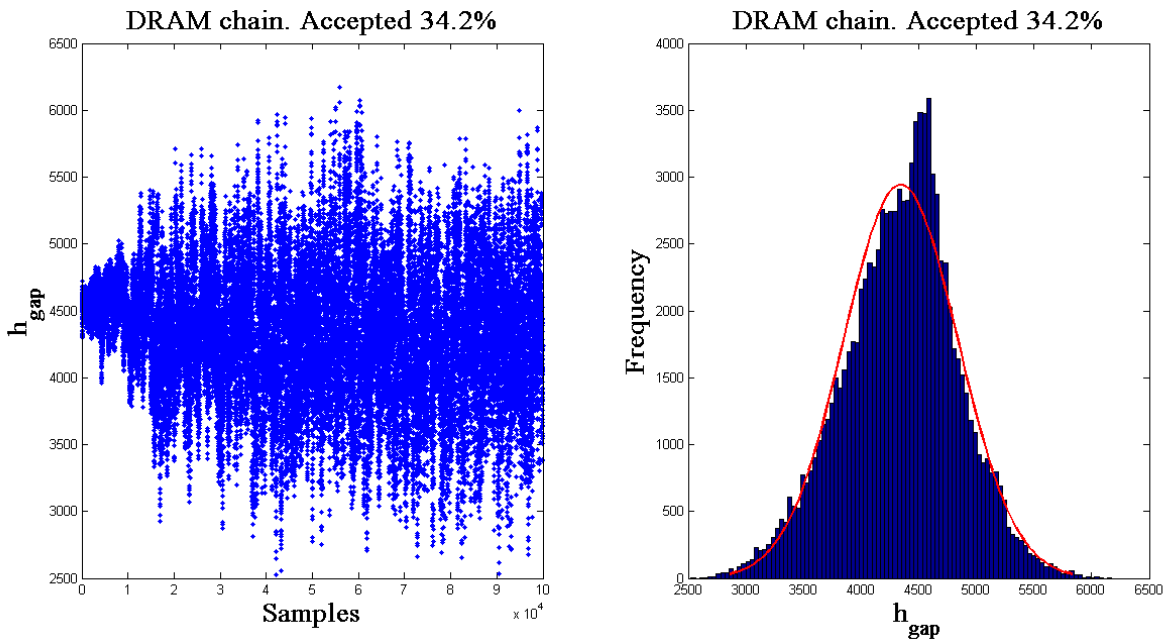


Figure 9. Chain and posterior distribution of the gap conductivity (h_{gap}). [Including $\mathbf{C}_{surrogate \to f}$].

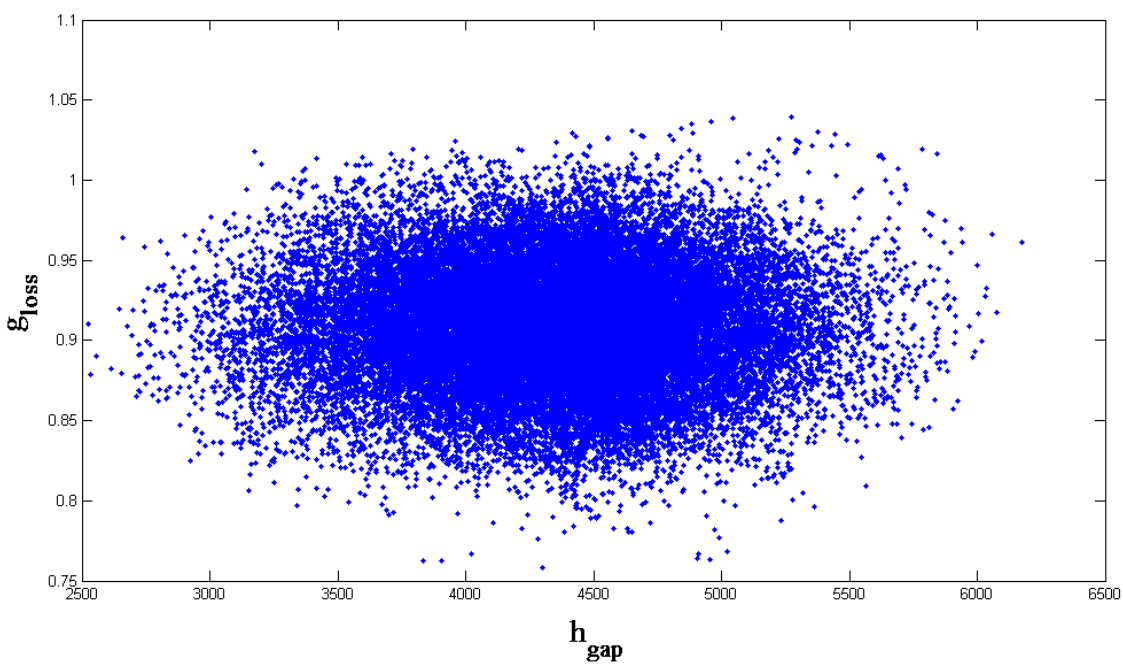


Figure 10. Correlation between the gap conductivity (h_{gap}) and the grid spacer loss coefficient (g_{loss}). [Including $\mathbf{C}_{surrogate \to f}$].

Table 7. The 47 group structure.

g	Energy Boundary	g	Energy Boundary	G	Energy Boundary	G	Energy Boundary
1	20 MeV	13	78.9 eV	25	2.3824 eV	37	0.5032 eV
2	6.0653 MeV	14	47.8512 eV	26	1.8554 eV	38	0.35767 eV
3	3.6788 MeV	15	29.023 eV	27	1.4574 eV	39	0.2705 eV
4	2.2313 MeV	16	13.71 eV	28	1.2351 eV	40	0.18443 eV
5	1.3534 MeV	17	12.099 eV	29	1.1664 eV	41	0.14572 eV
6	0.8208 MeV	18	8.3153 eV	30	1.1254 eV	42	0.11157 eV
7	4.9787 MeV	19	7.33822 eV	31	1.0722 eV	43	0.08197 eV
8	0.1832 MeV	20	6.47602 eV	32	1.0137 eV	44	0.0569 eV
9	67.38 KeV	21	5.715 eV	33	0.97100 eV	45	0.0428 eV
10	9.119 KeV	22	5.04348 eV	34	0.9099 eV	46	0.0306 eV
11	2.0347 KeV	23	4.4509 eV	35	0.7821 eV	47	0.0124 eV
12	0.13 KeV	24	3.9279 eV	36	0.62506 eV	-	-

Table 8. Surrogate features.

Surrogate Order	RMS	Construction Data Points	Validation Points	Residuals Distribution	Surrogate Form – Related Uncertainty (Appendix A.2)*
2nd order	$\varepsilon_{k_{eff}}$ → 14.4 [pcm]	120	30	k_{eff} → i.i.d	35.0 [pcm]
	$\varepsilon_{\mathbf{FR}}$ → 0.0332			FR → i.i.d	0.93% [Maximum]

*The numbers here represent the uncertainty in the surrogate response due to the uncertainty in the surrogate coefficients estimated by Eq. (5) which are then used in a Monte Carlo process to estimate the surrogate-form related uncertainty.

Table 9. Measurements and their uncertainties.

Measurement	0.0 EFPD	9.0 EFPD	32.0 EFPD	120.0 EFPD	160.0 EFPD
k_{eff}	1.000443 ±0.00060	1.000121 ±0.00060	1.00013 ±0.00060	0.999951 ±0.00060	0.99991 ±0.00060
FR (in core's central fuel assembly center)	1.9321 ±0.02	1.8962 ±0.02	1.8212 ±0.02	1.7612 ±0.02	1.7312 ±0.02

In order to evaluate the performance of the **SBDA**, the actual perturbations (which are known) are compared with the perturbations determined by the **SBDA**. A consolidated metric is defined to evaluate the performance of the **SBDA** with calibrating the cross-section parameters. Table 10 reports the energy group averaged difference ($\varepsilon_{x,i}^{DA}$) along with the maximum difference across all energy groups between the actual perturbation and the one estimated by the **SBDA**. The maximum average error difference is 6.1% occurring for the Pu-239 fission cross-sections (when the surrogate form uncertainty is included (i.e. $\mathbf{C}_{surrogate \to f}$). In the case of ignoring the surrogate form–related uncertainty, the maximum average difference is 5.8% occurring for the Pu-239 fission cross-sections. Table 10 summarizes the data assimilation results for a few important parameters including the two thermal-hydraulics parameters. Table 10 and Table 11 indicate that the surrogate model-form error is sufficiently small to not influence the data assimilation results.

Table 10. Assimilation performance measure for the various cross-sections parameters being calibrated.

Parameter	Energy Group Average $\sum_{g=1}^{G}\left\|\Delta\Sigma_{x,g}^{i,syn}-\Delta\Sigma_{x,g}^{i,DA}\right\| \Big/ \sum_{g=1}^{G}\left\|\Delta\Sigma_{x,g}^{i,syn}\right\|$		Maximum Estimation Difference $\left\|\Delta\Sigma_{x,g}^{i,syn}-\Delta\Sigma_{x,g}^{i,DA}\right\| \Big/ \left\|\Delta\Sigma_{x,g}^{i,syn}\right\|$	
	With $\mathbf{C}_{surrogate\to f}$	Without $\mathbf{C}_{surrogate\to f}$	With $\mathbf{C}_{surrogate\to f}$	Without $\mathbf{C}_{surrogate\to f}$
$\varepsilon_{f,U\text{-}235}^{DA}$	2.1%	2.0%	2.9% [g=2]	2.3% [g=8]
$\varepsilon_{f,U\text{-}238}^{DA}$	1.9%	1.7%	3.9% [g=6]	3.3% [g=6]
$\varepsilon_{f,Pu\text{-}239}^{DA}$	6.1 %	5.8%	5.1% [g=41]	6.2% [g=34]
$\varepsilon_{s,O\text{-}16}^{DA}$	5.9%	5.2%	12.1%[g=11]	3.1%[g=8]
$\varepsilon_{s,H\text{-}1}^{DA}$	1.9%	1.8%	6.1%[g=37]	4.0% [g=40]
$\varepsilon_{f,Pu\text{-}240}^{DA}$	4.1%	4.0%	4.2%[g=25]	3.5% [g=18]
$\varepsilon_{a,Xe\text{-}134}^{DA}$	2.6%	2.5%	4.5%[g=26]	2.5% [g=26]
$\varepsilon_{a,B\text{-}10}^{DA}$	1.5%	1.3%	4.0% [g=10]	2.6% [g=12]
$\varepsilon_{a,Xe\text{-}130}^{DA}$	1.1%	0.92%	5.2%[g=8]	3.8%[g=10]
$\varepsilon_{a,Y\text{-}90}^{DA}$	3.0%	2.9%	4.2%[g=12]	2.9%[g=10]
$\varepsilon_{a,Y\text{-}91}^{DA}$	3.3%	3.0%	6.0%[g=15]	4.8%[g=16]
$\varepsilon_{a,Zr\text{-}90}^{DA}$	3.3%	3.0%	3.9%[g=3]	4.5%[g=4]
$\varepsilon_{s,Fe\text{-}56}^{DA}$	2.7%	1.3%	4.8%[g=41]	2.6%[g=40]
$\varepsilon_{s,Mn\text{-}55}^{DA}$	2.5%	1.4%	4.9%[g=30]	3.8%[g=29]
$\varepsilon_{f,Pu\text{-}241}^{DA}$	4.8%	2.8%	7.4%[g=41]	5.2%[g=43]
$\varepsilon_{f,U\text{-}236}^{DA}$	1.9%	3.3%	4.8%[g=13]	2.6%[g=15]

Table 11. Data assimilation results for a few important parameters.

Parameter	Reference Value	Actual Perturbed Value	SBDA Estimated Perturbation With $\mathbf{C}_{surrogate \to f}$	SBDA Estimated Perturbation Without $\mathbf{C}_{surrogate \to f}$
h_{gap}	4500 ± 2250	4359.91	4109.2 ± 504.1	4008.3 ± 419.2
g_{loss}	0.907 ± 0.03628	0.9123	0.9088 ± 0.035	0.9088 ± 0.035
Σ_f^{U-235} [g=47]	1125.219 ±2.66	1258.435	1267.01 ± 1.11	1281.45 ± 1.01
Σ_f^{U-238} [g=27]	0.8927011±0.00528	0.93246	0.9270 ± 0.00243	0.9213 ± 0.00181
Σ_f^{Pu-239} [g=47]	1361.297± 16.03	1441.313	1436.2 ± 6.11	1443.92 ± 5.14
Σ_s^{H-1} [g=47]	75.346183±0.07494	82.47610	78.72 ± 0.0382	80.21 ± 0.0291
Σ_s^{O-16} [g=47]	4.329821±0.04336	4.629821	4.611 ± 0.0211	4.601 ± 0.01800

4. Posterior Uncertainty Quantification

In this section, the importance of the data assimilation study is highlighted and its effect on uncertainty quantification is assessed. The uncertainty study performed in Section 2 is repeated using the updated uncertainties obtained from the data assimilation presented in Section 3 (after burning the first 30,000 samples in the chain and sing the remaining 70,000 samples). In order to obtain the uncertainties efficiently, surrogates will be used to perform the Monte Carlo Uncertainty Quantification (MCUQ), so surrogate form-related uncertainty will be accounted for.

Table 12 through Table 14 compare the uncertainties in the quantities of interest prior and posterior to the data assimilation study. The only difference between the uncertainty study presented here and the one presented in Section 2 is the uncertainty distributions. In all cases the posterior uncertainties for the quantities of interest are reduced from their prior values. Note that the uncertainties (σ_R) are generally reduced along with the uncertainty in those uncertainties (σ_{σ_R}) up to 75% for maximum fuel pin power, 82% for maximum fuel pin temperature and 55%

for the multiplication factor (k_{eff}). This is due to the fact that the updated uncertainty intervals are reduced and this affects the uncertainty in the quantity of interest and the fluctuation in that uncertainty.

Despite the great benefits of data assimilation noted above, its computational cost is really huge unless the original model (i.e. VERA-CS) is replaced with a lower fidelity one (the surrogate model). Therefore, the premise of this work is to replace the original high fidelity and expensive model with a surrogate model that we can run with negligible computational cost. In the current milestone report this hypothesis has been tested in section 3 by placing known perturbations on the original model's parameters and trying to retrieve them with the surrogate approach.

Table 12. k_{eff}: joint

Depletion	Standard Deviation	2nd order Surrogate – Monte Carlo - Original	2nd order Surrogate Monte Carlo – Updated
0 EFPD	σ_k	0.425 %	0.210 %
	σ_{σ_k}	0.019 %	0.012 %
9 EFPD	σ_k	0.395 %	0.189 %
	σ_{σ_k}	0.021 %	0.011 %
32 EFPD	σ_k	0.415 %	0.188 %
	σ_{σ_k}	0.025 %	0.014 %
45 EFPD	σ_k	0.420 %	0.192 %
	σ_{σ_k}	0.019 %	0.009 %
60 EFPD	σ_k	0.399 %	0.193 %
	σ_{σ_k}	0.032 %	0.008 %
80 EFPD	σ_k	0.398 %	0.199 %
	σ_{σ_k}	0.023 %	0.011 %
100 EFPD	σ_k	0.394 %	0.200 %
	σ_{σ_k}	0.023 %	0.018 %
120 EFPD	σ_k	0.391 %	0.201 %
	σ_{σ_k}	0.023 %	0.017 %
160 EFPD	σ_k	0.389 %	0.205 %
	σ_{σ_k}	0.026 %	0.015 %

Table 13. Maximum Fuel Pin Power: joint

Depletion	Standard Deviation	2nd order Surrogate – Monte Carlo - Original	2nd order Surrogate Monte Carlo – Updated
0 EFPD	σ_{mp}	1.3 %	0.3 %
	$\sigma_{\sigma_{mp}}$	0.02 %	0.011 %
9 EFPD	σ_{mp}	1.1 %	0.32 %
	$\sigma_{\sigma_{mp}}$	0.08 %	0.025 %
32 EFPD	σ_{mp}	1.5 %	0.31 %
	$\sigma_{\sigma_{mp}}$	0.011 %	0.008 %
45 EFPD	σ_{mp}	2.2 %	0.4 %
	$\sigma_{\sigma_{mp}}$	0.2 %	0.1 %
60 EFPD	σ_{mp}	1.77 %	0.43 %
	$\sigma_{\sigma_{mp}}$	0.14 %	0.05 %
80 EFPD	σ_{mp}	1.64 %	0.47 %
	$\sigma_{\sigma_{mp}}$	0.20 %	0.09 %
100 EFPD	σ_{mp}	1.79 %	0.45 %
	$\sigma_{\sigma_{mp}}$	0.1 %	0.07 %
120 EFPD	σ_{mp}	1.62 %	0.42 %
	$\sigma_{\sigma_{mp}}$	0.11 %	0.064 %
160 EFPD	σ_{mp}	1.45 %	0.37 %
	$\sigma_{\sigma_{mp}}$	0.07 %	0.043 %

Table 14. Maximum Fuel Pin Temperature: joint

Depletion	Standard Deviation	2nd order Surrogate – Monte Carlo - Original	2nd order Surrogate Monte Carlo – Updated
0 EFPD	σ_{mt}	7.23 %	1.9 %
	$\sigma_{\sigma_{mt}}$	0.07%	0.06%
9 EFPD	σ_{mt}	5.10%	2.0%
	$\sigma_{\sigma_{mt}}$	0.02%	0.04%
32 EFPD	σ_{mt}	4.73%	1.9%
	$\sigma_{\sigma_{mt}}$	0.03%	0.01%
45 EFPD	σ_{mt}	5.15%	1.8%
	$\sigma_{\sigma_{mt}}$	0.03%	0.02%
60 EFPD	σ_{mt}	5.59%	1.7%
	$\sigma_{\sigma_{mt}}$	0.08%	0.05%
80 EFPD	σ_{mt}	5.56%	1.78%
	$\sigma_{\sigma_{mt}}$	0.04%	0.02%
100 EFPD	σ_{mt}	5.11%	1.8%
	$\sigma_{\sigma_{mt}}$	0.09%	0.02%
120 EFPD	σ_{mt}	5.46%	1.9%
	$\sigma_{\sigma_{mt}}$	0.06%	0.01%
160 EFPD	σ_{mt}	5.87%	2.0%
	$\sigma_{\sigma_{mt}}$	0.07%	0.02%

5. Summary and Conclusion

In this milestone report an uncertainty quantification and data assimilation study performed for CASL Progression Problem Number 9 is reported upon. In a previous milestone report, efficient uncertainty quantification algorithms were proposed and tested with CASL Progression Problem Number 6. The same methods were applied here to CASL Progression Problem Number 9. Further, an efficient surrogate based data assimilation methodology was proposed, implemented and tested using CASL Progression Problem Numbers 6 (reported upon in Ref. [8]) and 9 (reported upon here). Finally, the data assimilation results were used to re-quantify the uncertainty for CASL Progression Problem Number 9 and the effect of calibrating the parameters along with uncertainties in the quantities of interest uncertainties was highlighted. The computer allocation used in completing the work reported upon here was 2.5 million core hours provided by the Texas Advanced Computing Center (TACC). In addition, North Carolina State University – High Performance Computers (NCSU-HPC) were used as a supplementary resource.

Based on the study presented here, data assimilation can be performed using a high fidelity core simulation in reactor core problems with demanding but achievable computational cost. Once the surrogate is constructed the computational cost of the remaining analysis is negligible. The computational cost of constructing the surrogate model was reduced via the usage of efficient reduced order modeling algorithms developed and presented in a previous milestone report [1]. Once the model parameters are calibrated, new uncertainty intervals can be used to re-quantify the uncertainties in the quantities of interest. Uncertainties in the quantities of interest were determined to drop significantly (up to 75% for maximum fuel pin power, 82% for maximum fuel pin temperature and 55% for the multiplication factor). The effect of the data assimilation on re-quantified uncertainties is encouraging for further investigation and implementation of data assimilation methods and algorithms.

Future work will be focused on using real plant data to calibrate VERA-CS models. Moreover, different types of surrogates might be employed and used in the data assimilation study. Expectations are that once real plant data are used, model parameters can be realistically calibrated and therefore quantities of interest can be predicted with reduced uncertainties.

Overall, the reduced order modeling approaches presented in this report and the previous one [1] aim to introduce efficient capabilities for data assimilation and uncertainty quantification for VERA. Moreover, the surrogates can be seen as a reduced fidelity level available within VERA for other intended usage whenever needed.

Acknowledgment

The authors would like to express their thanks and gratitude towards those who contributed significantly in reviewing this report: Dr. Brian Williams (Los Alamos National Laboratory), Dr. Jess C. Gehin (Oak Ridge National Laboratory), Dr. Vincent A. Mousseau (Sandia National Laboratory), and Mr. [soon to be Dr.] Douglas Burns (Idaho National Laboratory).

Moreover, the authors would like to thank Texas Advanced Computing Center (TACC) and North Carolina State High Performance Computing (NCSU-HPC) for providing the required computer allocation.

References

1. Khuwaileh, B. A., Hooper, R., Turinsky, P. J., and Mousseau, V. A. (2015). Uncertainty Quantification Analysis Using VERA-CS for a PWR Fuel Assembly with Depletion. CASL-I-2015-0328-000.
2. Proctor, W. C. (2012). Elements of high-order predictive model calibration algorithms with applications to large-scale reactor physics systems. PhD Dissertation, North Carolina State University.
3. Hite, J. M., Abdel-Khalik, H. S., Smith, R. C., Wentworth, M., Prudencio, E., and Williams, B. (2013). Uncertainty Quantification and Data Assimilation (UQ/DA) Study on a VERA Core Simulator Component for CRUD Analysis CASL-I-2013-0184-000.
4. Haario, H., Laine, M., Mira, A., and Saksman, E. (2006). DRAM: efficient adaptive MCMC. Statistics and Computing, 16(4), 339-354.
5. Khuwaileh, Bassam A., Wang C., Bang Yangsuk and Abdel-Khalik, Hany S. (2014). Efficient Subspace Construction for Reduced Order Modeling In Reactor Analysis. PHYSOR 2014-The Role of Reactor Physics Toward a Sustainable Future.
6. Khuwaileh, Bassam A., Hite J. and Abdel-Khalik, Hany S. (2014). Subspace Methods For Multi-Physics Reduced Order Modeling In Nuclear Engineering Applications. PHYSOR 2014-The Role of Reactor Physics Toward a Sustainable Future.
7. Bang, Y. (2012). Hybrid Reduced Order Modeling Algorithms for Reactor Physics Calculations.
8. Khuwaileh, B. A. A. (2015). Scalable methods for uncertainty quantification, data assimilation and target accuracy assessment for multi-physics advanced simulation of light water reactors. PhD Dissertation, North Carolina State University.
9. Halko, N., Martinsson, P., & Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. SIAM Review, 53(2), 217-288.
10. Scale (2011): 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. Available from Scale: A Comprehensive Modeling and Simulation Suite for Nuclear Safety Analysis.
11. Godfrey, A., VERA Core Physics Benchmark Progression Problem Specifications, Revision 4, CASL Technical Report: CASL-U-2012–0131–004, August 29, 2014.
12. Adams, B.M., Ebeida, M.S., Michael S. Eldred, M.S., Jakeman, J.D., Swiler, L.P., Stephens, J.A., Vigil, D.M., Wildey, T.M., Bohnhoff, W.J., Dalbey, K.R., Eddy, J.P., Hu, K.T., Bauman, L.E., Hough, P.D. (2014). DAKOTA, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: version 6.1 user's manual. Sandia National Laboratories, Tech. Rep. SAND2014-4633.
13. Smith, R. C. (2013). Uncertainty Quantification: Theory, Implementation, and Applications (Vol. 12). SIAM

Appendix A

A.1: The Error Upper Bound

The following equation can determine the error upper bound ε_{upper} when approximating the physical quantity $\Delta\overline{\psi}$ using the basis spanned by the columns of matrix $\mathbf{U}$:

$$\varepsilon_{upper} = 10\sqrt{\frac{2}{\pi}}\max_{j=1,\ldots p}\left\|\left(\mathbf{I}-\mathbf{U}\mathbf{U}^T\right)\Delta\overline{\psi_j}\right\|_2 \tag{4}$$

This upper bound is guaranteed with a success probability of $1\text{-}10^{-p}$ where p denotes the number of additional values of physical quantity $\Delta\overline{\psi}$ beyond those used to construct the subspace and obtained as perturbed solutions to the governing equations when parameters are randomly perturbed [9]. For more information about the theory behind this error upper bound estimation and its application please refer to Ref. [8, 9].

A.2: Surrogate Form Related Uncertainty

The surrogate parameters are statistical parameters that are subjected to variance and standard deviation which can be approximated by Eq.(5) assuming unbiased residual error and an **i**ndependent-**i**dentically **d**istributed random error (**i.i.d**) [1, 8, 13]:

$$V\left(\overline{q}\right) \approx \sigma^2\left[\chi^T\left(\overline{q}\right)\chi\left(\overline{q}\right)\right]^{-1}, \tag{5}$$

where,

$$\chi\left(\overline{q}\right) = \frac{\partial\Delta y^{surr}}{\partial\overline{q}} \text{ and } \sigma^2 = \frac{1}{n-p}R^T R ,$$

$\overline{q}$ is the parameters set, Δy^{surr} is the surrogate response and R is the residual obtained by comparing the responses obtained by the surrogate and the responses generated by VERA-CS (the original model). On the other hand, $\chi\left(\overline{q}\right)$ is the sensitivity matrix (if the surrogate generates multiple responses) or a vector (for single response case) that can be calculated analytically or efficiently using a finite difference method. Finally, n-p is the number of degrees of freedom. For more details on this please refer to Ref. [1, 8, 13]. Once the variance $V\left(\overline{q}\right)$ is determined, the uncertainty in quantities of interest due to the surrogate's coefficients can be estimated via a Monte Carlo approach.