Feasibility Study of An Integrated Framework for Characterization of Uncertainties with Application to CANDU Steady State and Transient Reactor Physics Simulation

Prepared for:

Canadian Nuclear Safety Commission
280 Slater St., Ottawa ON K1P 5S9, Canada

Under Contract No. 87055-14-0184.

Prepared by:

Professor Hany S. Abdel-Khalik

School of Nuclear Engineering, Purdue University
400 Central Drive, West Lafayette, IN 47906, United States
Email: abdelkhalik@purdue.edu Voice: 1.919.749.9717

June 16th, 2015

Table of Contents

List of Tables

List of Figures

Acknowledgement

List of Acronyms

BE	Best-Estimate
BEPU	Best Estimate Plus Uncertainty
CRANE	Complexity Reduction Algorithms for Nuclear Engineering calculations
DAKOTA	Design Analysis Kit for Optimization and Terascale Applications
DOF	Degree of Freedom
DR	Dimensionality Reduction
E_PGPT	Exact-to-Precision Generalized Perturbation Theory
GPT	Generalization Perturbation Theory
GPMSA	Gaussian Process Model/Sensitivity Analysis
INL	Idaho National Laboratory
LANL	Los Alamos National Laboratory
NRC	Nuclear Regulatory Commission
ORNL	Oak Ridge National Laboratory
RAVEN	Reactor Analysis Virtual control Environment
RFA	Range Finding Algorithms
ROM	Reduced Order Modeling
PDF	Probability Density Function
PSUADE	Problem Solving for Uncertainty Analysis and Design Exploration
SA	Sensitivity Analysis
TMC	Total Monte Carlo
UC	Uncertainty Characterization
UCF	Uncertainty Characterization Framework
UQ	Uncertainty Quantification
XSUSA	Cross Section Uncertainty and Sensitivity Analysis

1.0 <u>PROPOSED SCOPE</u>

Quantification and understanding of uncertainty sources is an essential requirement of best-estimate (BE) reactor analysis simulation as it provides a reliable metric by which the quality of the predictions can be assessed. Although direct comparison against measurements provides the ultimate evidence that simulation predictions are reliable, the true value of any BE simulation lies in its ability to analyze reactor conditions for which measurements are unavailable. Therefore, there is a clear need to characterize, i.e., to propagate and prioritize, all sources of uncertainties in order to reliably use the results of BE calculations in the various aspects of reactor design, operation, and safety.

This study investigates the feasibility of the development of a first-of-a-kind integrated framework for uncertainty characterization (UC) with primary application to CANDU neutronics calculations. The goal is to provide a comprehensive and scientifically defendable methodology for characterizing uncertainties in all BE reactor analysis calculations, including both steady state and transient simulations. The framework will be based on open-source libraries for standard UC process algorithms as well as novel algorithms, to be developed and implemented by this project, to enable efficient execution for CANDU reactor analysis applications.

The UC framework (UCF) is to accomplish four primary functions. First, it will identify all sources of uncertainties resulting from modeling assumptions, numerical approximations, nuclear data uncertainties, and technological parameters uncertainties. Second, it will propagate the identified uncertainties to the responses of interest such as the core eigenvalue, power distribution, bundle enthalpy rise, etc. Third, it will map the propagated uncertainties to the wide range of operating conditions. Fourth, it will generate a priority identification and ranking table (PIRT) which identifies and ranks according to importance the dominant sources of uncertainties.

Despite the importance of uncertainties, the nuclear simulation codes have always lacked an integrated framework for their characterization. This is primarily due to: a) the complex nature of nuclear models, i.e., based on a multi-level homogenization strategy where a number of models are linked together in a sequential or circular manner to account for the wide range of physical phenomena involved, the large variations in energy and length scales, and the various forms of feedback mechanisms; b) the individual simulations codes requiring long execution times and being associated with voluminous size of input and output data streams; and c) the recent advances in uncertainty algorithms have been primarily demonstrated for modern software platforms, i.e., new codes; it is however difficult to incorporate these advances in some of the legacy codes used extensively in the design and regulatory spaces.

Given these challenges, the proposed UCF must employ efficient techniques to reduce the computational cost required to propagate and prioritize uncertainties, which is otherwise

intractable with conventional UC techniques. The reduction algorithms selected in this study are based on recent advances in reduced order modeling (ROM) techniques which have been recently applied with success to other LWR reactors physics simulation, e.g., BWRs and PWRs [Bang,, 2012]. Our objective is to propose a number of viable approaches based on ROM to propagate the various sources of uncertainties, including parameters and modeling uncertainties, in typical CANDU reactor analysis calculations. We also discuss the expected challenges and needed resources to construct a UCF that can provide comprehensive characterization of uncertainties in an affordable manner that can be scientifically defended.

This report is organized as follows. Section 2.0 provides a short background on UC and its basic process algorithms for characterization of parameters and modeling uncertainties. Section 3.0 overviews the sources of uncertainties in CANDU reactor physics calculations, and discusses some of the challenges of employing basic UC process algorithms for their characterization. Section 4.0 provides an overview of ROM techniques/algorithms, which represent the enabling engine for the UCF application to CANDU calculations.

Section 5.0 proposes a preliminary implementation plan of the UC basic process algorithms and their ROM rendition to characterize the various sources of uncertainties outlined in section 3.0. Section 6.0 describes both the UCF hardware and software requirements. Section 7.0 overviews existing UC toolkits that may be leveraged in support of the proposed UCF. A pilot study is conducted in section 8.0 to illustrate the application of the UC algorithms and provide an initial demonstration of the contribution of the various sources of uncertainties considered in this project. The UCF's milestones and deliverables are listed in section 9.0 along with proposed timeline for their completion over a five-year performance period.

Five appendices are appended to the supplement the discussions whenever needed. Appendix A provides additional background on UC; Appendix B discusses a short literature review of sensitivity analysis methods; Appendix C discusses the range finding algorithm, the basic ingredient of ROM techniques; Appendix D explains the differences between the two major ROM approaches, the dimensionality reduction and the surrogate model construction; and Appendix E provides an overview of the CRANE and E_PGPT tools which are developed by the author and his collaborators to render reduction in support of the pilot study.

2.0 <u>UC BACKGROUND</u>

Uncertainties are unavoidable in any simulation, since all constructed physical and/or engineering models representing the scientific bases for any simulation, are approximations of reality. For high consequence systems, such as nuclear reactors, the uncertainties must be well characterized to ensure safe and reliable operation during normal operation and accident scenarios. Characterization of uncertainties refers to all engineering analyses conducted to provide scientifically-defendable quantities that measure the reliability of the reactor simulation predictions.

Traditionally, uncertainties in reactor calculations have been addressed using a conservative bounding analysis. Later, the best estimate plus uncertainty (BEPU) method has been sought as a more realistic alternative to the conservative bounding analysis, contingent upon the proper quantification of uncertainties of BE simulation results. To fully realize the benefits of the BEPU method, one must explicitly account for all sources of uncertainties, including initial and boundary conditions, modeling and numerical sources of uncertainties, as well as those originating from the individual codes used in the overall coupled simulation.

The UC involves two primary processes. The first process, referred to as uncertainty quantification (UQ), propagates all known sources of simulation uncertainties in order to understand their impact on the reactor behavior. The second process, referred to as sensitivity analysis (SA), acts in conjunction with UQ to help identify the key sources of uncertainties. We discuss these processes here but more details are given in Appendix A on the basic UC definitions, and Appendix B on SA.

The UQ process devises a metric that is used to measure uncertainties. The measurement process may be thought of as a hypothetical, i.e., virtual, experiment that determines all possible outcomes/states/results/values of the phenomenon under consideration [Hubbard, 2010]. An important part of this process is to assign probabilities to the various possibilities. The combined possibilities and their associated probabilities are often described by a probability density function (PDF) or a histogram which may be viewed as the metric by which uncertainty are measured, i.e., quantified. Other metrics that are functions of the resulting PDF or histogram could also be employed to measure uncertainties. Examples include the mean, the standard deviation, kurtosis, etc., of the PDF or the histogram; and tail probability that is the probability to exceed a certain value, commonly referred to as failure probability.

We distinguish here between two types of UC exercises, experimentally-based and code-based, with the latter being the focus of this project. An experimentally-based UC exercise involves a direct comparison between simulation predictions and real experimental measurements, typically referred to as *bias*. The advantage of this exercise is that it provides a clear-cut quantification of the bias between simulation and reality. Its disadvantage is that it does not explain how the bias magnitude relates to the various sources of uncertainties inherent in the

simulation. Because of that, it is difficult to map this bias to other conditions not covered by the available experiments.

A code-based UC exercise, however, implies a self-assessment of the uncertainty sources, i.e., the code propagates its own known uncertainties, and provides them as part of the standard output along with BE results. In doing so, one can apportion the propagated uncertainties to the various sources of uncertainties, and can devise methods for their mapping to other operating conditions. The primary disadvantage of code-based UC techniques is that it propagates the "*known*" sources of uncertainties only, implying that all modeling inadequacies that are unknown to the modeler cannot be properly quantified. Ideally, one should employ a combination of the two exercises to fully characterize uncertainties and devise methods for their mapping to all operational conditions of interest.

The UCF investigated in this study will be focusing on the code-based UC only and will be designed to accomplish the following four tasks: a) Identification of Uncertainty Sources; b) Propagation of Uncertainty Sources; c) Mapping of Uncertainty Sources; and d) Prioritization of Uncertainty Sources. The first three tasks may be referred to as part of an UQ analysis, while the last one is commonly known as SA. These tasks are discussed in the next four subsections.

2.1 Identification of Uncertainty Sources

From a high level, any BE simulation code developed to predict the behavior of a given engineering system may be viewed as a set of procedures/rules that are used to manipulate the model input data in order to generate a set of output responses that can be used directly or after some further manipulation to measure the performance of the engineering system. The set of procedures/rules are collectively referred to as a model of the engineering system. The model reflects our understanding of the physical phenomena governing the behavior of the engineering system. The model is developed in the form of mathematical equations whose solution can be used to calculate the output responses. These equations must be manipulated further to enable their solution on a digital computer. This overall process contains several sources of uncertainties which may be grouped into three categories, modeling, numerical, and parameter uncertainties.

2.1.1   Modeling uncertainties

Any modeling scheme introduces two fundamental sources of modeling uncertainties, one originating from the modeler's incomplete knowledge about the physics governing the system, and the other resulting from the modeling simplifications and approximations introduced to achieve computational efficiency. Conceptually, the first source is the most difficult one to quantify since it requires one to make quantitative statements about one's lack of knowledge. The second source can, in principle, be quantified. This requires the development of algorithms

to capture the effect of the modeling simplifications and how they propagate and interact with other sources of uncertainties.

The first source can only be quantified if one has access to experimental data. In our proposed implementation, we will focus on code-based UC only, therefore our inherent assumption here is that uncertainties resulting from the modeler's lack of knowledge are extremely negligible. Since our focus is on neutronics calculations only, this assumption is not inadequate since the physics behavior of neutrons transport inside nuclear reactor cores are well understood and have been extensively validated over the years.

Specifically, we will assume that the Monte Carlo pointwise cross-section models provide perfect representation of the physics of neutron transport inside the CANDU core. While this assumption can be challenged for specific models, it will be assumed valid for the range of problems of interest to us in this project. Accordingly, the UCF will focus on capturing the second source only; that is resulting from the use of modeling approximations, such as the use of diffusion theory, multi-group approximations, cell lattice reflective boundary conditions, numerical discretization errors, etc.

2.1.2 Numerical uncertainties

Further, one must discretize the model's equations, often developed in an abstract mathematical form involving continuous integral and/or differential operators, to generate a set of algebraic equations that can be manipulated further by computers. This process is referred to as numerical discretization and is also expected to contain numerical uncertainties, commonly referred to as errors (the difference between uncertainties and errors is described in section 2.1.4).

For deterministic radiation transport, the UCF will lump both modeling and numerical uncertainties under the name of modeling uncertainties. In doing so, we will assume that the Monte Carlo simulation contains no numerical uncertainties. This is not a perfect assumption, because even in Monte Carlo simulation, one has to discretize the angular variable used to describe the neutron direction of travel into angular bins, which gives rise to numerical uncertainties. In thermal reactors, we expect this source of numerical uncertainty to be negligibly small given the short mean free path of the neutrons. In fast reactors however, this source of uncertainty needs to be adequately quantified.

2.1.3 Parameter uncertainties

All input data to a computational model, typically referred to as parameters, could contain uncertainties which propagate throughout the model and give rise to

response uncertainties. Input data uncertainties are unavoidable because input data are either experimentally evaluated or generated using pre-processor codes (cf., section 2.1.6 for the various classes of input data). The objective is to determine all possible responses variations, often described in the form of a probability distribution function (PDF), due to all possible input data variations within their known ranges of uncertainties.

Conceptually, this source of uncertainties is the easiest to quantify via a standard UQ sampling-based approach, cf. section 2.2.1. Prioritizing the dominant sources of uncertainties however using a conventional SA approach is computationally infeasible for typical CANDU reactor models due to the large number of input data. Therefore, an important requirement for the UCF is the development of ROM algorithms capable of propagating and prioritizing uncertainties in a computationally efficient manner.

2.1.4 Uncertainty Types

There are fundamentally two types of uncertainties, aleatory (i.e., random or irreducible) and epistemic (i.e., systematic or reducible) [Jaynes, 2009]. Aleatory uncertainties originate due to inherent randomness in the physical model itself which renders them irreducible, i.e., they cannot be minimized/reduced with additional measurements. For example, the number of counts registered in a detector is inherently random - the associated standard deviation of the counts will depend on the nature the radiation interactions inside the detector volume and therefore cannot be reduced even if the counting experiment is repeated an infinite number of times.

Epistemic uncertainties refer to biases in the calculations which result from approximations, assumptions, or lack of knowledge about the true value of model input data. For example, if the nominal value of v, neutron fission yield, used in BE calculations is lower than its true value, the associated reactor analysis model will consistently under-predict the core's critical eigenvalue.

Epistemic uncertainties are reducible in principle with additional measurements. For example, by repeating the experiments used to measure v, better estimates of its true value can be evaluated. The distinction between the two types is important and must be done carefully. By way of an example, in the above detector example, the number of counts measured is subject to an aleatory uncertainty; however the mean value of the counts measured over a given time interval suffers from an epistemic uncertainty, implying that a better estimate of the mean can be obtained with additional measurements.

In reactor analysis, both aleatory and epistemic uncertainties are present and must be carefully quantified. Examples of aleatory uncertainties include the dimensions and compositions of the various materials comprising the reactor core, all subject to manufacturing tolerances. This follows because any manufacturing process cannot render the exact engineering specifications, and hence a level of uncertainty is to be expected. Epistemic uncertainties include nuclear data, e.g., microscopic cross-sections, thermal-hydraulics data, e.g., heat transfer coefficients, and systematic errors resulting from numerical, e.g., discretization and iterative techniques, and modeling, e.g., use of homogenization theory, approximations.

It is instructive to mention that the terms 'error' and 'uncertainty' have been traditionally used to distinguish between systematic (i.e. epistemic) and random (i.e. aleatory) sources of uncertainties. This distinction is not needed in our context, and hence the two terms will be used interchangeably.

2.1.5 *Uncertainty Representation*

Both aleatory and epistemic uncertainties are described using PDFs. Let $p(x)$ be a PDF which defines the following quantity [Jaynes, 2009]:

$$\Pr\left(x_1 \le x \le x_2\right) = \int_{x_1}^{x_2} p(x)dx$$

Depending on the type of uncertainty, the above quantity has two different interpretations. Recall that aleatory uncertainty implies that x (representing a component of the input data) does not take on a single value because it is random in nature. In this case, the above definition measures the frequency of finding x in the interval $[x_1, x_2]$.

Contrary to that, epistemic uncertainty implies that x takes on a single value which is unknown. However, based on the evidence available (i.e., from experiments, expert judgment, etc.), one can speak of the likelihood (or the probability) of finding the true value in a certain interval. Note that these two interpretations are distinctly different, however they are both referred to as probabilities. We will distinguish these differences whenever needed.

The most common PDF is the normal Gaussian distribution, which is described by a mean value x_μ, representing the nominal value for x used in BE calculations, and a standard deviation describing either the modeler's confidence in the mean value for epistemic uncertainties, or the inherent randomness of aleatory uncertainties,

$$p(x) = A \exp\left(-\frac{1}{2}\left[x - x_\mu\right]^T \mathbf{C}_x^{-1}\left[x - x_\mu\right]\right)$$

This representation is used to denote the PDF of multiple input data, aggregated in a vector x with mean x_μ such that the ith component $[x_\mu]_i$ is the mean value of the ith input data $[x]_i$, and $\mathbf{C}_x$ is a matrix that characterizes the correlations between the data.

Input data correlations are unavoidable whether the data are directly measured, obtained from a fitting procedure, or pre-calculated using preprocessor codes. For Gaussian parameters, a covariance matrix is sufficient to describe all correlations. Its diagonal elements represent the variance (square of standard deviation) of the individual data, and the off-diagonal elements are measures of the correlations. A diagonal matrix with zero off-diagonal elements implies a set of input data that are uncorrelated. The covariance matrix is symmetric and of the form:

$$\left[\mathbf{C}\right]_{ij} = \rho_{ij}\sigma_i\sigma_j$$

where $[*]_{ij}$ denotes the element at the intersection of the ith row and j^{th} column. The ρ_{ij} is the standard correlation coefficients between the ith and j^{th} input data; it ranges between -1 and $+1$. A $\rho_{ij} = 1$ implies perfect positive correlation, i.e., both parameters move in the same direction. The σ_i is the standard deviation of the ith parameter.

The covariance matrix may also be decomposed using rank revealing decompositions [Meyer, 2001]:

$$\mathbf{C} = \mathbf{W}\mathbf{\Sigma}\mathbf{W}^T$$

where $\mathbf{\Sigma} \in \mathbb{R}^{r \times r}$ is a diagonal matrix $\left[\mathbf{\Sigma}\right]_{ii} = \hat{\sigma}_i^2$, and $\mathbf{W} \in \mathbb{R}^{n \times r}$ is an orthonormal matrix, i.e., its columns are mutually orthogonal and of unit norm, i.e., $\mathbf{W} = \begin{bmatrix} w_1 & w_2 & .... & w_r \end{bmatrix}$ where $w_i^T w_j = \delta_{ij}$, and δ_{ij} is the standard Kronecker delta function which is equal to zero for all i and j pairs except when $i = j$, it is equal to one. The implication of this decomposition is that one can transform the originally correlated input data into an uncorrelated set defined by: $\hat{x} = \mathbf{W}^T x$, where the ith uncorrelated input data $\left[\hat{x}\right]_i$ has a standard deviation of $\hat{\sigma}_i$.

Notice that the number of uncorrelated input data is r, which can be anywhere between 1 and n. The case of $r = n$ means that the original input data are correlated, but not perfectly. With $r < n$ however, the implication is that there are $n - r$ perfect

correlations between the data, which allows one to reduce the effective number of input data from n down to r.

In this case, one can recast the UC problem as follows. Note that in the original UC formulation, one is given x and their prior PDF, and wishes to determine the PDF of y such that: $y = f(x)$. In the recast formulation, one first transforms the input data into an uncorrelated set $\hat{x}$ and determine their associated PDF, then applies UC to the transformed model $y = f(\hat{x})$. This represents one of the core ideas required to render feasible the application of UC techniques to CANDU reactor problems. As will be discussed later, ROM techniques achieve significant dimensionality reduction by identifying all the correlations between the various variables, e.g., model parameters, flux solution, power distribution, etc.

2.1.6 Uncertainty Sources Classification

To describe how the various sources of uncertainties are propagated using the proposed UCF, we employ the following representation of a typical reactor analysis simulation model. Let the responses of interest, e.g., critical eigenvalue and bundle enthalpy, etc., be described by:

$$y = f(x, \alpha, \eta)$$

where y is a vector whose components represent the responses of interest. The vectors x, α, η represent the input data to the model classified into three groups as shown in Fig. 2.1.

- The first group x is taken to represent all input model physics parameters, e.g., nuclear cross-sections, heat transfer coefficients, friction factors, etc., referred to hereinafter as "*model parameters*" for short.
- The second group α represents all the parameters that can be controlled by the operator, referred to as "*control parameters*", e.g., fuel design including composition and geometry details, coolant flow rate, soluble boron content, control rod position, etc.
- The third group η represents all the decisions made by the modeler, e.g., use of deterministic vs. probabilistic model, use of a specific resonance treatment model, a specific numerical iterative technique, etc., which will be referred to as "*modeling decisions*".

Note that the model and control parameters are typically continuous variables, whereas the modeling decisions are generally symbolic, e.g. *$\eta=0$* or *$\eta=A$* for Monte Carlo continuous cross-section model, and *$\eta=1$* or *$\eta=B$* for multi-group deterministic model, etc.

Most practitioners refer to model parameters and control parameters uncertainties, as simply parameter uncertainties, whereas uncertainties resulting from modeling decisions are typically referred to as modeling errors/uncertainties. We will employ these terminologies in our discussion

Model parameters x are either experimentally measured/evaluated or calculated from preprocessor models and therefore must contain uncertainties. This follows because any experimentally-measured value contains uncertainties due to the systematic errors and noise in the measurement procedure; and any preprocessor model is, as mentioned before, a form of approximation of reality and hence contains uncertainties. By way of example in neutronics calculations, the continuous ENDF cross-sections are based on the fitting of experimental differential cross-section measurements to well-established nuclear models, cf. section 3.1. The multi-group cross-sections however are calculated based on a preprocessor model in which an assumed or calculated flux shape is used to collapse the continuous cross-sections into multi-group format.
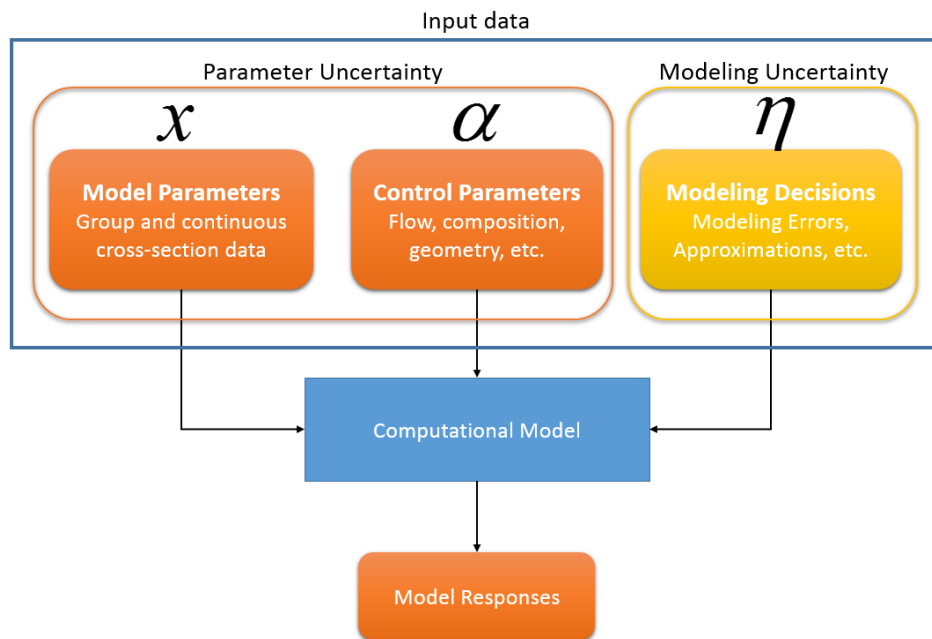
Fig. 2.1 Uncertainty Sources Classification

Control parameters α are set by the operator/analyst/designer, and their associated uncertainties are typically a result of the engineering mechanism by which the respective parameters are controlled. For example, the enrichment of the fuel depends on the accuracy of the manufacturing process at the fuel enrichment plant. Just like model parameters, control parameter uncertainties are described by PDFs. In practice, uniform PDFs, as opposed to normal distributions, are often used with

control parameters. This is because maximum and minimum limiting parameter values can be defined with uniform PDFs, which allows the modeler to assign zero probabilities to extreme values that are non-physical – not possible with the long tails of the normal distribution.

The modeling decisions η are symbolic, e.g., alpha-numeric, and therefore cannot be described using PDFs. Their values are arbitrarily assigned by the modeler to denote the type of assumptions and approximations employed to simplify the analysis models. Special algorithms must therefore be devised to propagate these uncertainties, and enable their mapping to the intended range of operating conditions.

2.2 Propagation of Uncertainty Sources

There are fundamentally two different UQ approaches: the sampling-based approach and the reduced order modeling (ROM)-based approach, both depicted in Fig. 2.2. The sampling approach emulates the basic definition of UQ; it tries to determine all possible responses variations resulting from parameters uncertainties by executing the model as many times as possible. The results of the many executions are combined statistically to derive quantities of interest, such as mean, standard deviation, failure probabilities, parameters sensitivities, etc. The ROM approach replaces the original model with a model of reduced complexity, often referred to as a surrogate model. The premise is that the surrogate model can be executed much more efficiently and as many times as required for both the UQ and SA analyses. Details on the ROM approach are described section 4.0.

The basic algorithm behind all sampling-based methods is to randomly sample all input parameters from their prior PDFs and execute the forward model [Helton, 2006]. After each execution, the responses variations are recorded and the procedure is repeated with different random samples until a reliable estimate of responses PDFs is obtained. This method is advantageous because of its simplicity and ability to obtain detailed (i.e., all moments) PDFs for all responses. This is primarily important for a general nonlinear model and general input parameters PDFs where the responses PDFs are expected to deviate considerably from the Gaussian shape. Another advantage is that the number of samples is independent of the number of input parameters n.

Two notable examples of this methodology in the nuclear engineering communities are the Total Monte Carlo (TMC) method developed by the NRG group in the Netherlands [Rochman, 2014], and the SAMPLER super-sequence developed by ORNL in the United States [Williams, 2013]. TMC is designed to propagate basic nuclear data uncertainties starting at the ENDF level, cf. section 3.1, to macroscopic core attributes of interest using Monte Carlo continuous cross-section simulation. The SAMPLER super-sequence propagates the uncertainties starting with the multi-group cross-sections, cf. section 3.2.

(a) Sampling-based UQ   (b) ROM-based UQ
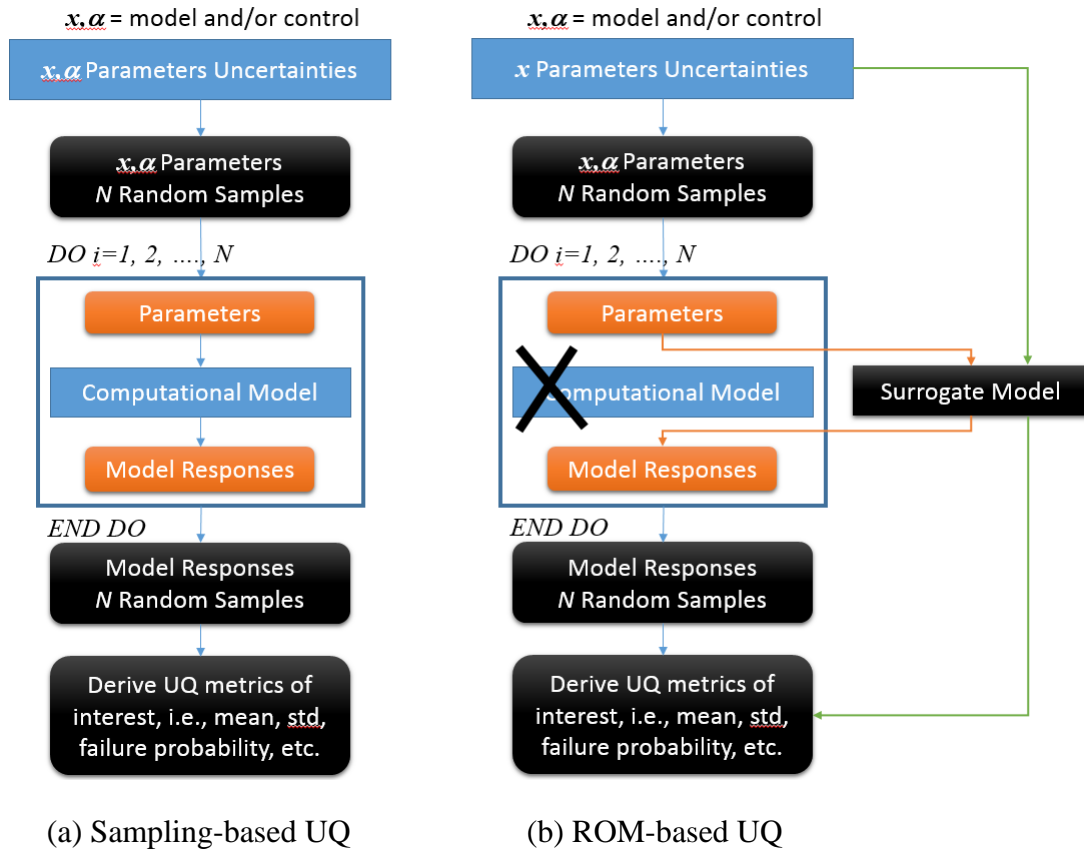
Fig. 2.2 Fundamental UQ Approaches

The primary challenge facing sampling-based methods is that it is very difficult to infer the importance of the various sources of uncertainties, i.e., difficult to determine sensitivity information, when the number of afforded random samples is much smaller than the number of uncertainty sources. One can show that the number of required samples is proportional to n^s, where s is the order of sensitivity information, i.e., $s=1$ denotes second-order derivatives, and n is the number of uncertainty sources. This renders the sampling-based UQ approach infeasible for the full characterization of uncertainties in typical CANDU reactor physics calculations.

The rest of this section will focus on presenting the basic UC algorithms used for the propagation and prioritization of parameter and modeling uncertainties. After ROM techniques are introduced in section 4.0, and the sources of uncertainties specific to CANDU reactor physics calculations are fully exposed in section 3.0, we will explain how the basic UC algorithms can be modified using ROM techniques to render efficient execution for CANDU applications.

The mathematical rendition of the basic UC algorithm may be described as follows. Let's say one is interested in calculating the first moment of the response PDF due to uncertainties in the model parameters x, this may be written as follows:

$$y_\mu = \int f(x, \alpha, \eta) p(x) dx$$

This expression implies an averaging of the response function y over the range of possible values for x, with the weights determined using the PDF of x, i.e., $p(x)$.

In the late 19th century, it was proved that this integral may be evaluated numerically:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} f(x_i) = \int f(x) p(x) dx$$

This integral implies that one needs to evaluate the function f at random samples x_i that are drawn from the PDF $p(x)$. More importantly, the theorem guarantees convergence of the summation to the true value of the integral as the number of samples approaches infinity.

Moreover, it can be shown that the convergence speed is proportional to the square root of the number of samples. This theorem represents the basis for all sampling techniques. The most powerful feature about this theorem is that the convergence of the summation to the true integral depends only on the number of samples, implying no dependence on the number of model parameters. The next two sub-sections apply this algorithm to propagate parameters and modeling uncertainties.

It is important to note here that while this algorithm is very efficient in capturing the main features, i.e., moments, of a general PDF, it is not very effective in capturing extreme (low probability) events, i.e., which typically happen at the tail of the PDFs and referred to as failure events. In this case, one must rely on more sophisticated sampling strategies to ensure that enough samples are produced in the failure region.

2.2.1 Parameters uncertainty propagation

Propagation of parameters uncertainties begins with the determination of the PDFs that describe their prior uncertainty. Regarding control parameters, their prior uncertainties are based on the engineering tolerance of the associated parameters, e.g., fuel enrichment, diameter of the fuel pin, etc. Regarding model parameters, their prior uncertainties are typically determined using a combined experimental and evaluation procedure, which requires careful propagation of uncertainties from both the experiment and the evaluation procedure. An example of that is the evaluation of the ENDF pointwise continuous cross-sections, cf. section 3.1.

Once the prior uncertainties in the form of PDFs are available, the model responses uncertainties can be determined using the following sampling procedure, referred to as UQ^p, with the superscript p denoting the parameters to distinguish it from the algorithm discussed in the next section for modeling uncertainties propagation.

- Generate N samples of the parameters that are consistent with their prior PDFs.
- Execute the model N times and record and/or histogram responses variations.
- Determine the responses PDFs and/or moments thereof.

The most common moments include the mean (which represents the average value for the response of interest) and the standard deviation (which measures the analyst's confidence in the mean value or its inherent randomness).

The number of samples required to get adequate statistics (i.e., first and second moments) on the responses of interest is typically in the range of few to several hundred samples for a general model. If the model is perfectly linear, one needs only two samples to fully characterize the response PDF, one sample at the nominal parameter value, and another at a perturbation thereof. One can show that for a linear model, the uncertainties in the responses of interest may be given by the following formula:

$$\mathbf{C}_y = \mathbf{S}\mathbf{C}_x\mathbf{S}^T$$

where $\mathbf{S}$ is denoted as the sensitivity matrix, which contains the first order derivatives of the responses with respect to the parameters, i.e.,

$$\left[\mathbf{S}\right]_{ij} = \frac{\delta[y]_i}{\delta[x]_j} \text{ , or compactly } \mathbf{S} = \frac{dy}{dx} \in \mathbb{R}^{m \times n}$$

In neutronics calculations, the derivatives are typically determined using an adjoint approach, however they can also be determined using forward finite differencing approach:

$$\frac{\delta[y]_i}{\delta[x]_j} = \frac{f\left(x_0 + \Delta[x]_j e_j\right) - f\left(x_0\right)}{\Delta[x]_j} \text{ , } i=1,\ldots,m, \text{ and } j=1,\ldots,n$$

This expression implies that in the forward finite differencing approach one has to perturb each input parameter once and execute the code to calculate the derivatives of all responses with respect to the given parameter. The computational cost is therefore proportional to the number of parameters, which is prohibitive for typical reactor neutronics models.

The adjoint approach however requires a single execution of the forward model and a single execution of the adjoint model to calculate the derivatives of a single response with respect to all model parameters. If one is interested in calculating the derivatives of all responses, the cost of adjoint calculations also becomes prohibitive. In the proposed UCF, both the numbers of responses and parameters

are large, since we are interested in propagating uncertainties throughout all the codes comprising a standard CANDU computational sequence. For example, in cell lattice calculations, the parameters represent the multi-group cross-sections and the responses are the few-group cross-sections functionalized in terms of a wide range of core conditions. To overcome this problem, we employ ROM techniques as described in section 4.0 to minimize the number of required forward and/or adjoint model executions to a manageable number.

2.2.2 *Modeling uncertainty propagation*

Different from model and control parameters, modeling decisions responsible for the modeling uncertainties are symbolic in nature, they describe a known change in the modeling assumptions or the numerical approximation techniques which cannot be described using a continuous variable with an associated prior PDF. To address this situation, the uncertainties are propagated by conducting two UQ^p studies, each with a different modeling decision. We will denote this approach by UQ^m, where m denotes modeling uncertainties.

By way of an example, consider the uncertainties resulting from the use of continuous cross-sections Monte Carlo model vs. a deterministic multi-group model, referred to in terms of their associated decision variables $\eta=0$ and $\eta=1$, respectively, as depicted in Fig. 2.3. In this case, the response denotes the discrepancies (i.e., bias) between the models predictions for the quantities of interest, e.g., core's critical eigenvalue. Propagation of the uncertainties resulting from these different modeling decisions is done as follows:

- Generate N samples for the control and model parameters
- Execute the model N times with a user-defined modeling decision, e.g., $\eta = 0$
- Re-execute the model N times with a different modeling decision, e.g., $\eta = 1$
- Record the differences in the responses between the two above cases, and their variations over the N model runs, and histogram the results.
- Determine the responses PDFs and/or moments thereof.

Given that the number of modeling decisions is problem- and/or analyst-dependent, we will lump all decisions together and focus only on their aggregated effect. In particular, we will consider the following cases:

- $\eta = 0$: The most accurate neutronic model available, assumed to be a Monte Carlo model with explicit representation of spatial heterogeneity and continuous cross-sections.
- $\eta = 1$: Deterministic model with multi-group cross-sections, and homogenized spatial regions. In section 5.0, we describe how modeling

uncertainties resulting from different levels of homogenization can be quantified, with each level denoted by a different value for η
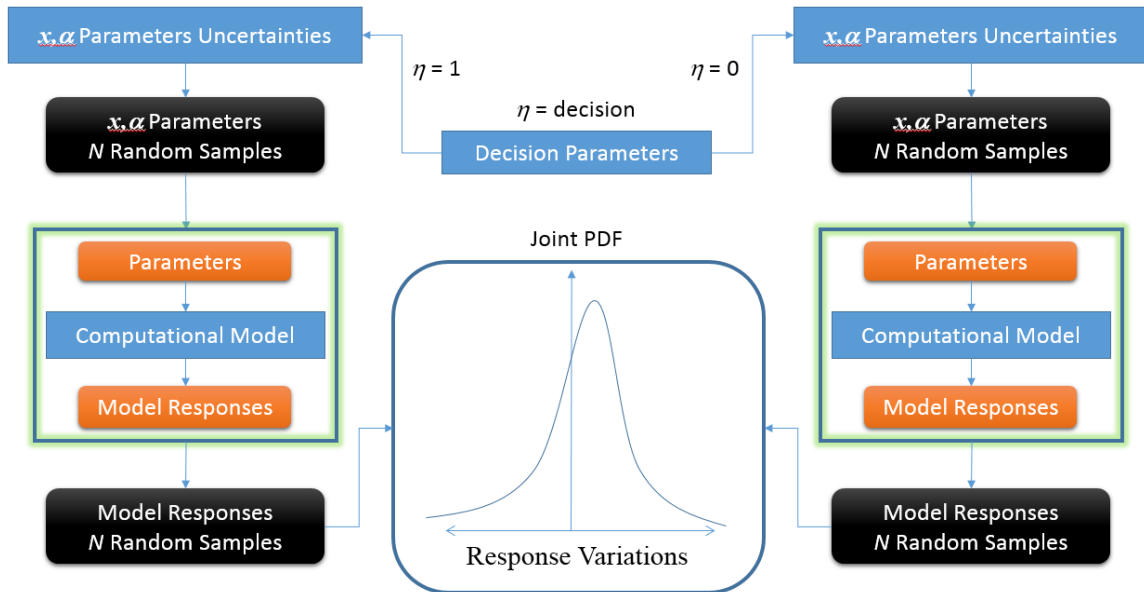
Fig. 2.3 Propagation of Modeling Uncertainty

2.3 Mapping of Uncertainty Sources

Mapping of uncertainties implies the ability to determine how uncertainties are scaled between the different operational conditions, i.e., how do uncertainties in the fresh fuel map to the discharged fuel? This is an important capability for any UCF to support model validation needs. This follows because in model validation, one needs to identify the range of operational conditions over which the model is considered an adequate representation of reality. Adequacy is measured in terms of uncertainty, where areas in the operational phase space are identified such that their associated uncertainty is below given preset tolerance on the maximum uncertainty allowed.

Code-based uncertainties are easier to map because they simply involve repeating the UC exercise with different operational conditions. Experimentally-based uncertainties however are much more difficult to map, because as discussed earlier, they aggregate all sources of uncertainties, offering almost no insight on their sources, and how they may be mapped.

We focus in this project on code-based uncertainties. As discussed earlier, we split the sources of uncertainties into two groups, control and model parameters, (collectively referred to as parameter uncertainties), and modeling decisions (referred to as modeling uncertainties or errors). Parameter uncertainties, as discussed in the previous section, are

propagated using a straightforward sampling approach. To render their mapping, the UC exercise can be repeated for all operational conditions of interest.

This straightforward approach is not adequate for mapping modeling uncertainties. This is because modeling uncertainties require the execution of an additional model, that's the high fidelity model, which forms the basis for estimating modeling inadequacies. The execution of the high fidelity model is typically very expensive, and cannot be repeated for all possible operational conditions of interest. If it was practical, one would not need to execute the lower fidelity model at all. Therefore, one must devise an effective mapping strategy that uses limited number of high fidelity model executions to map the uncertainties for all conditions of interest, for which no high fidelity model predictions are available. An example of that is the ability to characterize the modeling errors resulting from homogenization theory in core-wide depletion calculations using a limited number of Monte Carlo simulations for small sections of the core. To achieve that, we explain below how the UQ^m is extended to map uncertainties.

The idea of the UQ^m mapping algorithm is to establish a mapping between two sets of conditions (as defined by the control parameters), the first represents the intended domain of operation, and the second represents the conditions for which the high fidelity model predictions are available. The assumption is that the UQ^m algorithm described in section 2.2.2 has already been employed to propagate modeling uncertainties for the second set of conditions, and one desires to map these uncertainties to the first set of conditions where only the lower fidelity model predictions are available.

The UQ^m mapping algorithm employs the lower fidelity model to generate a joint PDF using the standard sampling-based UQ approach applied to both sets of conditions. This joint PDF describes how modeling uncertainties are expected to be correlated between the two sets of conditions, which facilitates the mapping of modeling uncertainty.

The UQ^m mapping algorithm is described as follows by way of an example: consider that one is interested in estimating the modeling uncertainties for the few-group constants used in downstream CANDU core-wide calculations. Few-group constants have to be functionalized in terms of a wide range of conditions, including different burnup, different fuel temperatures, coolant voiding and temperature, soluble boron content, with and without interstitial control and reactivity devices, etc. This requires the execution of the lattice code many times. We assume that a high fidelity model, e.g., Monte Carlo with continuous cross-sections, denoted by $\eta = 0$, is available for execution for only a small number of the above noted conditions, denoted by α_0, where α_0 specifies the control parameters such as fuel temperature, soluble boron content, etc. Let α_1 represent the conditions for which high fidelity model predictions are not available, and to which

modeling uncertainties are to be mapped. Symbolically, one can perform the following operations

High Fidelity Model: $\quad y = f\left(x, \alpha = \alpha_0, \eta = 0\right)$

Lower Fidelity Model: $\quad y = f\left(x, \alpha = \alpha_0, \eta = 1\right)$ and $y = f\left(x, \alpha = \alpha_1, \eta = 1\right)$

Note that the lower fidelity model can be executed readily for both sets of conditions, α_0 and α_1, whereas the high fidelity model is only available at the α_0 conditions. The UQ^m mapping algorithm is split into three steps.

1. Determine the modeling uncertainties at conditions α_0
   - Using the algorithm in section 2.2.2, calculate the PDF of y describing the modeling uncertainties at the conditions α_0. Denote this PDF by $p(y, \alpha_0)$. Recall that this requires $2N$ model executions, N times using the high fidelity model, and N times with the lower fidelity model.
2. Generate Joint PDF between α_0 and α_1 conditions (Fig. 2.4)
   - Identify all sources of uncertainties, and their associated prior PDFs.
   - Generate N samples for all parameters.
   - Execute the lower fidelity model twice for both sets of conditions α_0 and α_1.
   - Record the responses variations for both conditions.
   - Produce a scatter plot for the responses at α_0 (x-axis) vs. the responses at α_1 conditions.
3. Mapping of Modeling Uncertainties from α_0 to α_1 conditions (Fig. 2.5)
   - Plot the $p(y, \alpha_0)$ on the x-axis.
   - Using parametric (e.g., fitting techniques [Box, 1987]) or nonparametric (e.g., kernel density estimators [Silverman, 1986]) techniques, map $p(y, \alpha_0)$ from the x- to the y-axis, and denote the mapped PDF by $p(y, \alpha_1)$.
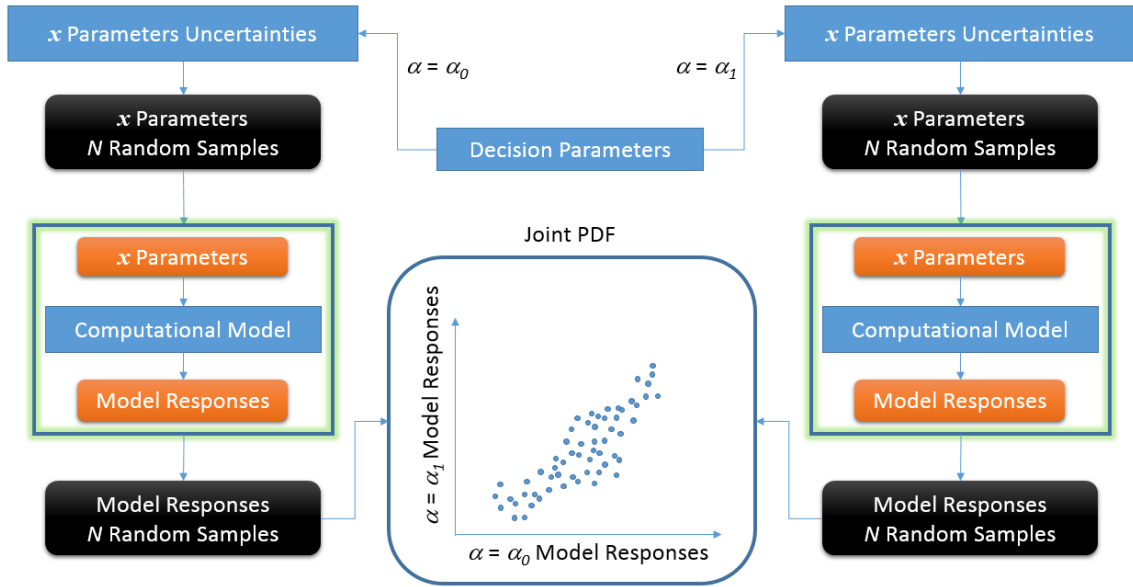
Fig. 2.4 Generation of Joint PDF of Two Operating Conditions
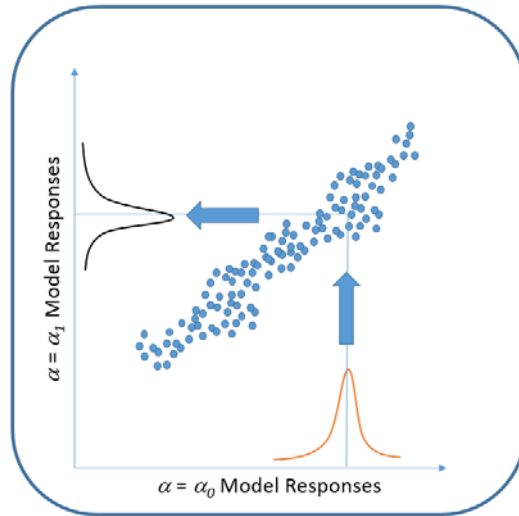
Fig. 2.5 UQ^m Mapping Algorithm

2.4 Prioritization of Uncertainty Sources

This step is referred to as sensitivity analysis (SA). It aims to determine in a quantitative manner the importance of the various sources of uncertainties to the propagated response uncertainty. This step is important as it provides guidance on modeling improvements that are required to reduce uncertainties. The basic UC-based SA algorithm is based on the following theorem, denoted by variance-based decomposition theorem [Saltelli, 2000].

$$Var(y) = Var_x\left(E\left[y \mid x\right]\right) + E_x\left[Var(y \mid x)\right]$$

This theorem states that the variance of a given response may be split into two components, one due to the input model parameter x (the red term), and the other (the blue term) due to all other model parameters. The red term is referred to as the variance of the conditional expectation, whereas the blue term is the expectation of the conditional variance. Appendix A contains a general overview of SA methods.

This theorem can be used to split the propagated uncertainty into many different ways depending on the goal of the analysis. For example, for simple linear models with n parameters, one can split the RHS into n different terms. The implication is that the computational cost becomes dependent on the number of parameters. With nonlinear models, the nonlinear interactions between the parameters must also be taken into account, which, depending on the model, could grow exponentially with the number of model parameters (referred to as the curse of dimensionality).

To render practical implementation, the UCF must employ recent advances in ROM to reduce the computational cost required to identify the most dominant sources. Particularly, in support of SA, the UCF will split the uncertainties into three primary components, one from the model parameters, another from control parameters, and one last from the modeling decisions. The individual sources associated with each of these three components will be identified and ranked using ROM. The reason for this split is that control parameters uncertainties are typically aleatory uncertainties which cannot be reduced, whereas model parameters are typically epistemic uncertainties, and can be reduced with additional measurements. Finally, modeling uncertainties will be identified since they are directly related to the decisions made by the modeler. This will help give insight on the modeling decisions contributing the most to the propagated uncertainties.

Note that from an algorithmic viewpoint, both model and control parameters are propagated using the same approach, i.e., the UQ^p approach, whereas the modeling uncertainties are propagated using the UQ^m approach, and mapped using the UQ^m mapping algorithm.

3.0 UNDERSTANDING UNCERTAINTY SOURCES IN CANDU CALCULATIONS

3.0 UNCERTAINTY SOURCES IN CANDU CALCULATIONS

In neutronics simulation, uncertainties arise from many sources, including uncertainties from basic nuclear parameters, i.e., point-wise cross-sections, manufacturing tolerances on geometry, burnt fuel isotopics, etc. Uncertainties also originate from modeling approximations/assumptions often introduced to render practical execution, e.g., multi-group approximation, reflective boundary conditions assumption in cell lattice calculations, etc. Finally, numerical approximations, resulting from discretizing the continuous equations into algebraic forms amenable for computer manipulation, also introduce uncertainties in the simulation results, see section 2.1 for more details.

In this section, we explain how uncertainties are introduced into the CANDU neutronic computational sequence, divided up into four stages as shown in Fig. 3.1.
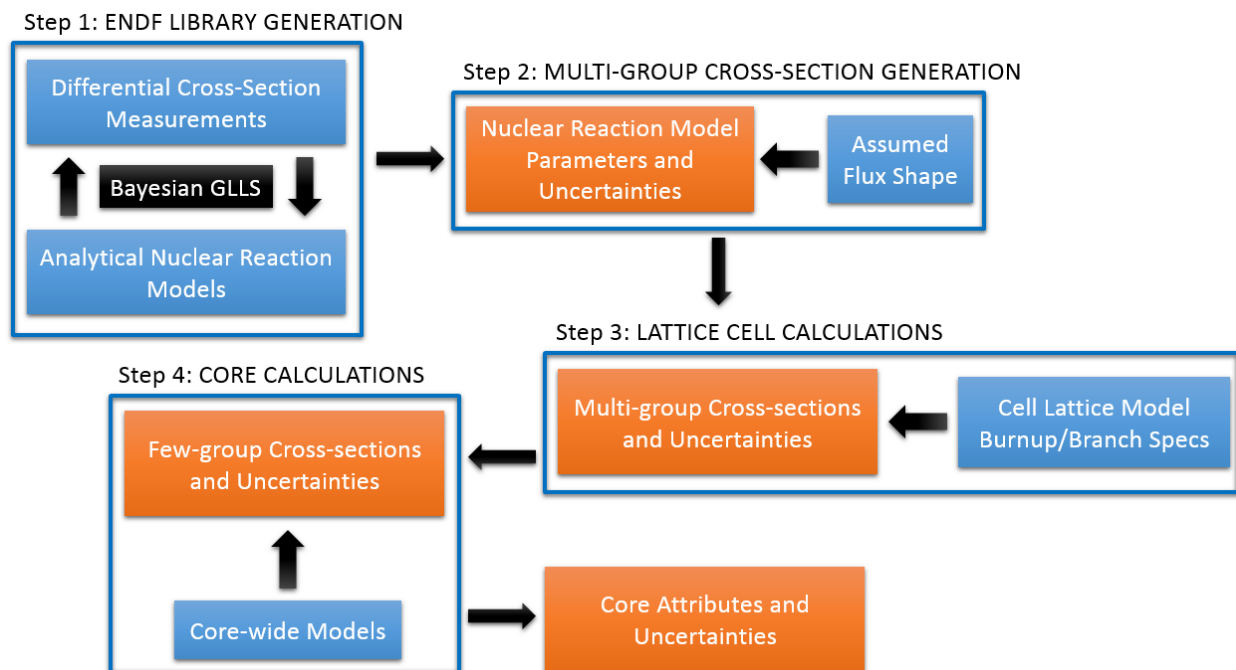
Fig. 3.1 CANDU Computational Sequence

3.1 ENDF Pointwise Cross-sections Generation

The first stage is common to calculations of all different types of reactors. It is done once by cross-section experimentalists/evaluators, and is typically repeated only when new cross-section measurements become available. The product is an ENDF library that can be used by reactor analysts for all reactor types. The raw cross-section measurements are referred to as differential measurements (shown in Fig. 3.2 as the noisy data). These measurements are fitted to analytical models (shown as orange contour), which contain a number of undetermined coefficients, referred to as nuclear reaction model parameters. An example of nuclear reaction model parameter is the resonance width and center when the

fitting is done under a resonance. These analytical models are based on nuclear theories such as the R-matrix theory. An example of a code that performs these calculations is the SAMMY code of ORNL [Larson, 1984]. The fitting procedure is based on a generalized least-squares procedure, also known as Bayesian Estimation, which requires some initial guesses for the nuclear reaction model parameters. The measurements are fed in a sequential manner to the algorithm, wherein the model parameters determined from a previous iteration (i.e., with a given set of measurements) are used as the initial guess for the next iteration (i.e., the next batch of measurements). The process is repeated a number of times until all available measurements are employed. In doing so, it is left to the evaluator(s) to discard measurements which are believed to contain outliers, i.e., measurements with high uncertainties that are not consistent with the rest of measurements. This process of selecting or discarding measurements relies solely on the common sense judgments exercised by the evaluator(s). An important result of the least-squares fitting procedure is the uncertainties (in the form of covariance matrices) of the reaction model parameters.

The uncertainties in this stage originate from two sources, first the differential cross-section measurements uncertainties, and the nuclear reactor model forms used to describe the continuous cross-sections. Given the maturity of nuclear models and the evaluation procedure, the cross-section fitting results are considered by most practitioners to be satisfactory for all reactor analysis calculations. Accordingly, the basic assumption in most neutronic UQ studies is that nuclear reaction model parameters represent the main source of uncertainty for all downstream calculations. To our knowledge, this assumption has never been validated using a rigorous methodology, and hence we suggest that it should be assessed as part of the proposed UCF application. Possible ideas for the quantification of this source of uncertainty are discussed in section 5.1.
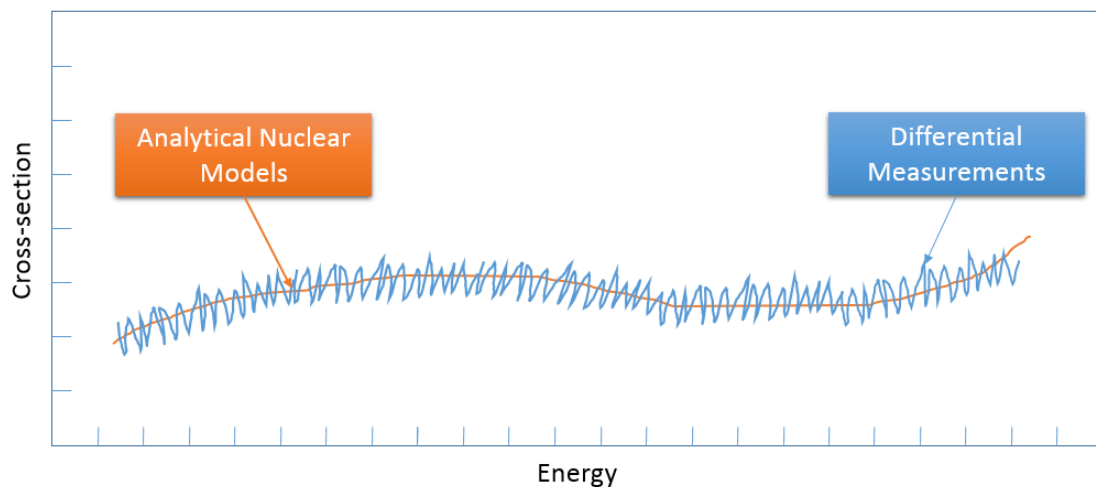
Fig. 3.2 Evaluation of Point-wise Cross-Sections

3.2 Multi-group Cross-section Generation

This stage collapses the pointwise cross-sections into a multi-group format using assumed flux shapes. The number of groups are typically selected based on a combination of expert judgment and a trial-and-error approach that attempts to resolve all aspects of the flux spectrum, expected to affect the integral quantities of interest such as eigenvalue, reaction rates, etc. This stage is specific to the flux spectrum expected in the reactor, and hence must be repeated for different reactor types. Since our interest in this work is on CANDU reactors only, this stage needs to be done only once.

Regarding the uncertainties of the multi-group cross-sections, they can be estimated by propagating the uncertainties of the nuclear reaction model parameters using the standard sandwich relationship described in section 2.2.1. A typical computer code that performs this stage is the PUFF code of the SCALE's code package, developed by ORNL [Wiarda, 2006].

The multi-group cross-sections contain essentially two different sources of uncertainties, one originating from the nuclear reaction parameters used to construct the continuous cross-sections, and the other from the assumed flux shape. Currently, a tool like PUFF accounts only for the first source of uncertainty (under the constraints of linearity assumption), implying that the flux shape is assumed to have no uncertainty. This latter source of uncertainty is difficult to estimate because the real flux shape is unknown a priori.

Therefore, assumed flux shape uncertainty must be treated as a source of modeling uncertainty using a decision variable η, see section 2.1.6 for the definition. To estimate this source, one must be able to compare the predictions against a high fidelity model that directly uses the continuous cross-sections, i.e., without any collapsing. The discrepancies between the predictions of the low and high fidelity models can be used estimate the modeling bias, which has to be repeated to take into account its dependence on other modeling conditions, such as composition, temperature, etc., i.e., control parameters. The next section discusses how the UQ^m methodology, cf., section 2.2.2, could be employed to estimate this source of uncertainty.

Finally, notice that it is not clear whether the uncertainties resulting from the assumed flux shape are independent of the nuclear reaction model parameters uncertainties. The UCF should be employed to investigate their dependence. To our knowledge, this investigation has never been attempted before.

3.3 Cell Lattice Calculations

In this stage, one calculates the few-group cross-sections for the cell lattices expected to be loaded in the CANDU core. These calculations must be repeated every time a new cell lattice design is introduced. Cell lattice calculations start with the multi-group cross-

sections, and calculate the few-group cross-sections for a wide range of core conditions. The result is a very large matrix of few-group cross-sections, which are fitted to polynomial expressions to facilitate their interpolation in downstream core-wide calculations.

The brute force application of the forward-based UQ approach would prove to be computationally expensive, even if one is not interested in capturing sensitivity information. This is because the UQ computational cost will be few to several hundred times higher than the cost required to generate the reference few-group cross-sections, which is large considering the wide range of conditions that must be captured to properly functionalize cross-sections.

To provide an idea about the size of the data streams flowing through cell lattice calculations, consider a typical transport code that is used to calculate the few-group cross-sections. The code is to be executed a number of times equal to N_B x N_C x N_L times, where N_B refers to the number of burnup steps, and N_C is the number of branch cases required to functionalize cross-section dependence on core conditions such as fuel and coolant temperature, coolant voiding, boron content, etc, and N_L is the number of cell lattice types in the core which is typically 1 for standard CANDU cores. This product is in the order of 1000 for typical CANDU cores. To propagate uncertainties, one would need to repeat these model executions N times, which is in the order of few to several hundreds. This results in in the order of 10^5 model executions which is prohibitive in practical applications.

If Monte Carlo model is employed to propagate the uncertainties, e.g., using a Total Monte Carlo approach [Rochman, 2014], the cost gets multiplied by another factor representing the ratio of the computational cost of executing Monte Carlo to that of an equivalent deterministic code. This factor is typically in the order of 100, and could be more if responses include space and energy-resolved data.

Variance reduction techniques, used to accelerate Monte Carlo convergence, could reduce this factor to be comparable to or slightly higher than deterministic calculations. Variance reduction techniques are used to bias particles histories from birth to death based on given reference solutions. Since the UC model executions represent small perturbations from a reference case, variance reduction techniques are expected to be very effective in reducing the cost of Monte Carlo executions.

Further, if one is interested in identifying the dominant contributors to the propagated uncertainties via an SA, one needs to execute the sequence a number of times that is proportional to the number of model parameters. In this case, the parameters represent the multi-group cross-sections. For typical multi-group libraries, this number is in the order of 10^5 for typical CANDU calculations to account for tracking about 70 isotopes using few hundred energy groups, and two to four reactions per nuclide, e.g., fission, absorption, scattering, capture, etc. This increases the number of required code runs to be in the order

of 10^6 to 10^9 executions, which is prohibitive despite the expected increase in computer power.

In this stage, two sources of uncertainties are introduced, one from the multi-group cross-sections propagated from the previous stage, and the other resulting from the modeling assumptions, such as the use of reflective boundary conditions, and the use of a deterministic transport solver, and the use of multi-group instead of continuous cross-sections. The first source is again straightforward to account for. The computational cost becomes impractical when sensitivity information is required, i.e., to understand the contribution of the individual multi-group cross-sections on the propagated few-group uncertainties.

The second source depends on the modeling decisions taken and therefore must be treated as a source of modeling errors. First, regarding the use of deterministic transport solver and multi-group cross-sections, this source could be identified by comparing model predictions against a high fidelity continuous cross-section Monte Carlo model. The second source is more difficult to account for because it depends on the type of neighboring bundles in the reactor core. To account for that source, one must emulate the impact of the neighbors via a super cell lattice calculations (i.e., 3x3 array of 9 bundles), see section 5.3.3 for more details on how to accomplish this.

Finally, similar to the previous stage, the correlations between the multi-group uncertainties and the modeling uncertainties (resulting from the transport model, multi-group cross-sections, and neighbors approximations) are to be investigated by the proposed UCF.

3.4 Downstream Core-wide Calculations

The last stage involves the calculation of core-wide power distribution during steady state and transient conditions starting with the few-group cross-section data. For typical CANDU models, the number of few-group cross-sections is equal to N_{FG} x N_B x N_C x N_L where N_{FG} is the number of few-group cross-sections generated per a single transport model execution, which is in the order of 10, representing the thermal and fast absorption cross-sections, transport cross-section, fission cross-sections, prompt neutron yield, and energy release, and Xe and Sm fast and thermal absorption cross-sections. If macroscopic depletion model for core calculations is used, the total number of few-group cross-sections is in the order of 10^4. If microscopic depletion models are used, this number is scaled by the number of nuclides tracked at the core level to account for the individual nuclides' cross-sections.

The sources of uncertainties in core-wide calculations include, uncertainties from the few-group cross-sections; uncertainties from the radiation transport model employed (e.g., nodal diffusion theory assumptions, and two energy-group cross-section representation);

and uncertainties from non-neutronic models, e.g., thermal-hydraulics models, and their associated correlations used to describe the transfer of the heat from the fuel to the coolant, and the corresponding feedback into neutronics calculations, e.g., fuel temperature feedback, coolant and moderator temperature and density feedbacks; and finally any uncertainties from the control parameters such as the lattice dimensions, fuel composition, flowrates, inlet coolant temperatures, etc.

The first source, i.e., few-group cross-section uncertainties, can be treated using a standard UQ^p sampling-based approach as described earlier. The second source, i.e., the radiation transport model, can be estimated using the UQ^m methodology by using core-wide Monte Carlo models. This will also require the UQ^m mapping algorithm to allow the efficient mapping of uncertainties to the wide range of core configurations, since cannot afford to execute core-wide Monte Carlo simulation for all conditions of interest. The third source, i.e., non-neutronic models, can also be propagated using the UQ^p and UQ^m and its associated mapping algorithm, but will be considered to be outside the scope of the current proposal since our focus will be only on neutronic sources uncertainties.

4.0 ROM BACKGROUND

If infinite computer power is available, one could apply the UC algorithms, i.e., UQ^p, UQ^m, and UQ^m mapping algorithm, and SA, cf., section 2.2, in a brute force manner. In reality however, nuclear reactor physics calculations continue to challenge the state-of-the-art computing platforms despite the startling growth in computing power realized over the past few decades. This is mainly due to the incredible detail of heterogeneities characteristic of reactor analysis models.

Moreover, as discussed earlier, the computational cost of UC is expected to depend on the number of uncertain parameters and the model nonlinearity. Therefore, a brute force UC application to a CANDU reactor analysis model is expected to require in the order of 10^6-10^9 model executions, which is not practically possible. This represents the top challenge for any UC practitioner: *How to reduce the number of required model runs and/or how to make them affordable?*

ROM refers to any process that reduces the complexity (i.e., order) of the analysis models. If the complexity is reduced, the number of uncertainty sources could be reduced, which will reduce the number of required model runs. This type of reduction is referred to as dimensionality reduction. Also, reducing the complexity helps reduce the cost of the calculations which renders repeated model execution to be affordable. This type of reduction constructs a model of reduced complexity that can be used in lieu of the original model, hence the common terminology of 'surrogate' model, i.e., it can be used to replace/represent the original model for the sake of completing UC analysis. This section discusses these two types of reduction, i.e., dimensionality reduction (DR), and surrogate model construction (often referred to as function approximation), with more details given in Appendix D.

4.1 Function Approximation or Surrogate Model Construction

Function approximation focuses on replacing the original function f used to describe the relationship between the model input data (i.e., parameters) and responses by a simpler parametric function $\tilde{f}$ in the form of an analytic expression with a number of undetermined coefficients, e.g., polynomial expansion. The coefficients are determined via fitting against a number of executions of the original function. Depending on the manner by which the approximating function is selected, one may categorize function approximation techniques into two categories, mathematical function approximation, and physics-based function approximation.

4.1.1 Mathematical-based Function Approximation

In this category, the functional form is selected in a heuristic manner, based solely on the modeler's familiarity and experience with running the model many times over a wide range of conditions. For example, consider that one is interested in

replacing the transport solver by a surrogate model to investigate the effect of fuel-to-coolant ratio on the core's critical eigenvalue. In this case, the modeler executes the transport solver numerous times with different fuel-to-coolant ratios, and visually establishes a trend. The closest function from a template of pre-determined functions is selected, e.g., second order polynomial with unknown coefficients.

Many methods fall under this category, like polynomial chaos expansion, stochastic collocation techniques, response surface methodologies, polynomial expansions, Lagrange polynomial expansion, etc. The main advantage of this approach is that it can be applied in a black box manner, i.e., non-intrusive manner, precluding the need for code modifications or any direct knowledge of the inner workings of the model.

The mathematical-based function approximation approach faces two primary challenges. First, the choice of the parametric function form is heuristic and is only based on an expert opinion. This renders difficult the quantification of the errors resulting from the approximation, and more importantly the mapping of such errors to other operational conditions. Second, the number of model executions required to determine the unknown coefficients is unrealistically high for reactor analysis problems. This follows because the number of model executions is a function of the number of parameters as well as the order of nonlinearity of the model. At a minimum, for a linear surrogate model with n parameters, one must execute the original complex model n times, which could be prohibitive in CANDU reactor physics applications, see sections 3.3 and 3.4 for typical sizes of the uncertainty spaces encountered in CANDU reactor physics applications.

4.1.2   *Physics-based Function Approximation*

In this category, the functional form is solely determined by the physics model, i.e., no assumptions or selection from pre-determined template of functions is made. The primary advantage of this approach is that one can establish reliable error bounds on the surrogate model predictions. The primary challenge is that one must have intimate knowledge of the inner workings of the model, implying knowledge of the model equations, and the type of solver employed, and assumptions made, etc., and sometimes the availability of the adjoint-solver is required.

An example of a physics-based function approximation technique is the exact-to-precision generalized perturbation theory (EpGPT), developed by the author and his collaborators over the past several years [Wang, 2013]. It has been recently implemented in the SCALE code package as a super-sequence, named CRANE [Mertyurek, 2014]. This capability allows one to replace the transport solver completely by an analytical expression capable of calculating the angular flux and

the eigenvalue to a preset user-defined tolerance, which can be set to match the machine precision of the standard forward transport model. The primary requirement of this approach is the availability of an adjoint solver for the transport code, which may not be available to the practitioners.

It is noteworthy to mention that direct application of either the mathematical or the physics-based function approximation techniques to a high dimensional model, such as a CANDU reactor analysis model, is computationally infeasible. Therefore, it is important to apply dimensionality reduction (discussed in section 0) first prior to the development of either types of surrogates.

4.2 Dimensionality Reduction

Instead of approximating *f*, the DR approach reduces the effective number of degrees of freedom used to describe the input data. In doing so, the function f remains unchanged. An important feature of DR techniques is that one can upper-bound with high confidence the errors resulting from the reduction, which allows one establish a scientific approach that ensures the reliability of the reduced models. This feature is lacking from conventional function approximation techniques such as polynomial chaos, Fourier expansion methods, response surface techniques, etc. The reason for that is explained in Appendix D.

4.2.1 Basic Reduction Idea

ROM-based DR techniques provide a rigorous mathematical approach by which the uncertainty space can be effectively shrunk into a manageable size to enable the practical application of UC techniques. The idea is to identify the so-called active degrees of freedom (DOFs) which are strongly correlated with the response variations. Inactive DOFs denote directions in the parameter space that have negligible impact on the responses of interest. Earlier studies have shown that the number of active DOFs in reactor physics calculations (for a wide range of reactor types, e.g., BWRs, LWRs, SFRs) is in the order of few hundred, which is considerably smaller than the nominal size of the uncertainty space, expected to be in the millions to billions [Bang, 2012, Jessee, 2011, Abdel-Khalik, 2008]. The implication is that one can recast all UC algorithms in terms of the active DOFs which renders the process computationally manageable.

To describe ROM mathematically[1], rewrite the model equation as follows:

$$\mathbf{Q}_y y = f\left(\mathbf{Q}_x x, \alpha, \eta\right)$$

---

[1] We limit the reduction to the *x* parameters only to minimize the notational cluttering. In principle, one can apply the reduction to all input parameters including both model and control parameters.

where $\mathbf{Q}_x \in \mathbb{R}^{n \times n}$ and $\mathbf{Q}_y \in \mathbb{R}^{m \times m}$ are projection operators on the active subspaces, which are mathematical constructs (in the form of rectangular matrices) that describe the active DOFs. Appendix D provides extensive intuitive details on the meaning of these operators.

These matrices have small ranks as compared to the their sizes, i.e., $r_x \ll n$ and $r_y \ll m$; the ranks represent the number of active DOFs. The mathematical interpretation of these matrices may be best described using a visual argument. Consider that x lives in a three dimensional space and y in a two dimensional space. Let $\mathbf{Q}_x$ be a projection onto a two dimensional subspace (a plane that passes through the origin), and let $\mathbf{Q}_y$ be a projection onto a one dimensional subspace (a line that passes through the origin). The reduction implies that while x has three DOFs, only two are important, i.e., the third orthogonal component has a negligible impact on the response of interest. In the y-space, while there are two responses, they are perfectly correlated. One can therefore recast the model above as follows:

$$y^{(r)} = f\left(x^{(r)}, \alpha, \eta\right)$$

where the superscript (r) refers to the reduced parameters. To complete UC, one must use the transformations $\mathbf{Q}_x$ and $\mathbf{Q}_y$ to calculate the PDFs for the reduced variables, which is mathematically trivial to do, since the transformation operators are linear.

In principle, one can also render nonlinear dimensionality reduction, where the reduction operators are nonlinear. This is however considered beyond the scope of the current project, mainly because significant reduction can be attained with linear reduction for reactor analysis problem, thereby precluding the need for sophisticated nonlinear reduction techniques.

Note that linear DR does not mean that one is linearizing the original model. It simply means that the reduced parameters are linearly related to the original parameters, however the functional form of the original model being reduced remains unchanged.

Finally, the errors resulting from the reduction can be constrained by a preset tolerance ε according to the following relationship:

$$\left\| \mathbf{Q}_y f\left(\mathbf{Q}_x x, \alpha, \eta\right) - f\left(x, \alpha, \eta\right) \right\| \le \varepsilon \quad \text{for all } x, \ \alpha, \text{ and } \eta$$

One can show that this upper-bound is satisfied with high probability, i.e., $p = 1 - 10^{-s}$, where s is a small integer that corresponds to an additional number of

model executions. For example, if one wants this bound to be satisfied with probability 0.99999, one needs to execute the original model 5 additional times with randomized parameter values. Details on this are discussed in Appendix C.

Fig. 4.1 illustrate the optimum application of both DR and surrogate construction techniques. First, the active subspaces for both the parameters and responses spaces are identified. Next surrogate model construction techniques are applied assuming that the spaces are now replaced by the respective active subspaces only. This significantly reduces the cost required to build the surrogate model, allowing it to be a useful tool in subsequent UC calculations.
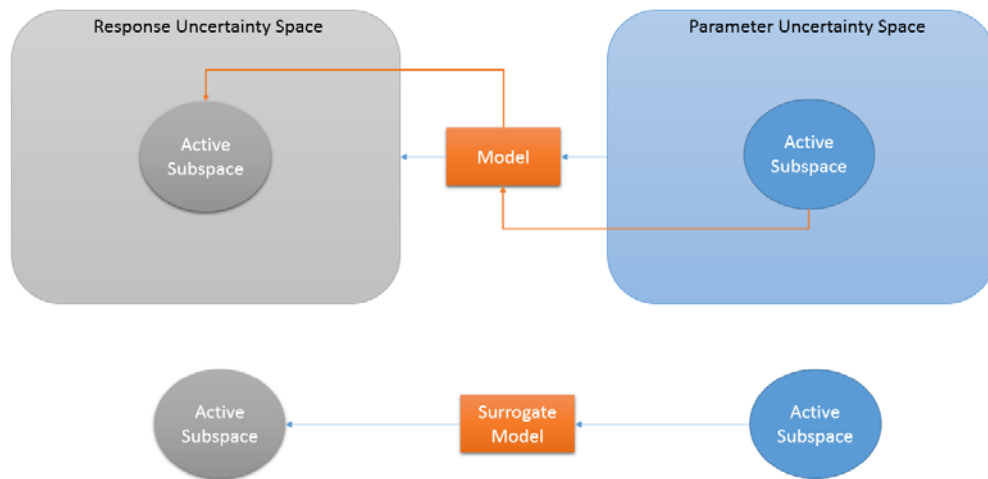
Fig. 4.1 ROM Application

4.2.2 Basic Reduction Algorithms

Fundamentally, there are two core reduction algorithms, a gradient-based and a gradient-free reductions, as shown in Fig. 4.2; the green box implies a loop of N model executions. Gradient-based reduction implies that with access to the derivatives of the responses with respect to the parameters, one can identify an active subspace for the parameters. Gradient-based reduction is currently the only state-of-the-art approach for reducing the parameter space dimensionality for standalone codes when the number of afforded model executions is much smaller than the number of model parameters[2]. Derivatives in neutronics calculations are best calculated using variational (i.e., adjoint) methods. Therefore, the gradient-based reduction approach is mostly useful for neutronic solvers which have an

---

[2] When the number of afforded model executions is much higher than the number of parameters, one can render reduction in the parameter space without direct access to the derivatives [Kramer, 2011]. These methods will not be discussed here because in reactor analysis, the number of model parameters is much higher than the number of afforded model executions.

adjoint capability and can calculate the derivatives of their responses with respect to the nuclear cross-sections and composition data.

Gradient-free reduction implies that only forward model executions are required to render reduction in the response space. While this approach is not effective for performing parameter reduction for standalone codes (see previous footnote), it is very effective for codes connected in a chain, because the output responses of one code are fed as input parameters to the next code in the chain.

(a) Gradient-Free   (b) Gradient-Based
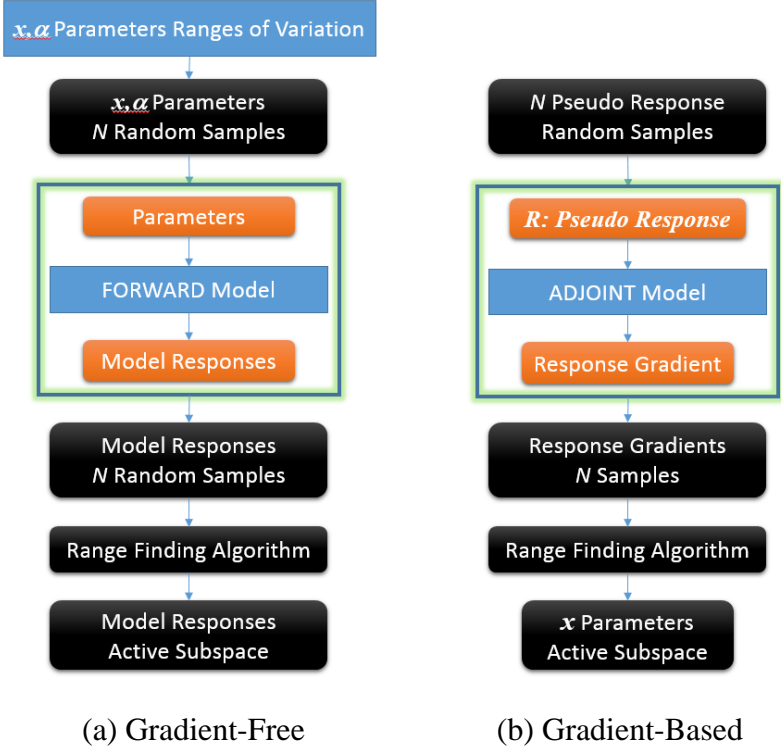
Fig. 4.2 Fundamental ROM DR Approaches

Both the gradient-based and gradient-free reductions rely on an algorithm from matrix theory, referred to as range finding algorithm (RFA) [Halko, 2011]. RFA employs randomized model executions to generate snapshots of the output responses (for gradient-free reduction) or the derivatives (for gradient-based reduction), which are subsequently filtered using rank revealing decompositions [Meyer, 2001] to identify the active DOFs and the associated active subspaces. Details on the RFA are provided in Appendix C.

To explain the distinction between gradient-based and gradient-free reduction, consider the following situation, shown in Fig. 4.3, involving a coupling between a thermal analysis code, a radiation transport code, and a cross-section processing code. The transport code calculates the power distribution P starting with cross-

sections processed at given temperatures T by the cross-section processing code. The power P is fed to the thermal code to calculate the temperature distribution T, which is fed to the cross-section processing code to generate new cross-sections Σ for the transport code.

If one is interested in calculating an active subspace for the cross-section space using only the transport code, one must have access to the derivatives of the transport code responses with respect to the cross-sections. If interested in calculating an active subspace for the power distribution using transport code only, one can employ the gradient-free approach by simply randomizing the cross-sections and executing the code N times as described below.
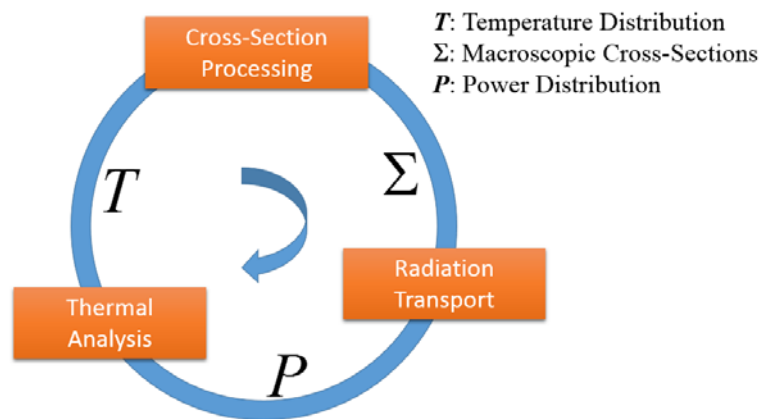
Fig. 4.3 Typical Reactor Analysis Code Chain

Now, if one wants to take advantage of the whole chain, one can render the reduction using gradient-free approach only or a hybrid of both the gradient-free and gradient-based approaches. In the former, the transport code is first used to generate an active subspace for the power distribution. The thermal code is executed next with randomized power distributions that are constrained to the active subspace just determined. The result is an active subspace for the temperature distribution. The process is repeated with the cross-section generation code to produce an active subspace for the cross-section space. This approach has been demonstrated in earlier work [Khuwaileh, 2013], and shown to result in a smaller active subspace for the cross-section space than determined using the transport code alone. The reason for this is that the active subspace determined with the entire chain identifies the active DOFs that are important to the entire chain which are expect to be less than those of individual codes.

In the latter approach, one can generate an active subspace at the intersection of two codes using two different reductions. For example, at the cross-section space, the transport code can generate an active subspace using gradient-based methods,

while the cross-section processing code can generate another active subspace using gradient-free method as just described. These two active subspaces are expected to be different. One can employ intersection techniques [Wang, 2014] to identify a single subspace that combines the active DOFs that are important to both codes. Earlier work has shown that the best results (in terms of the size of the active subspace) are obtained with a hybrid use of gradient-free and gradient-based methods. The performance is degraded slightly when only gradient-free methods are employed to render reduction for the whole chain. The application of this idea has been applied to a number of problems, including thermal-neutronic coupling [Khuwaileh, 2013], and quasi-state depletion calculations [Bang, 2012], where the Bateman equation, resonance calculations, and transport calculation are executed sequentially to deplete the fuel in lattice cell and core-wide calculations.

Fig. 4.4 depicts a typical reduction for a code chain composed of two codes (limited to two codes only for demonstration purposes). The flowchart on the right shows the original (pre-reduction) chain, where a subset of the output of model #1 is passed as input to model #2. In the left flowchart, gradient-free reduction is applied to the model #1, and the resulting active subspace is used to confine the random samples for model #2.

Fig. 4.5 shows how both the gradient-free and gradient-based reductions can be combined when gradient-free reduction is not enough. This approach is typically used when sensitivity information must be evaluated and the sizes of the active subspaces generated using gradient-free approach are too large to render practical implementation. In this case, model #2 is used to generate an active subspace using the gradient-based approach, while model #1 uses a gradient-free approach.

Based on past experience, the gradient-free approach is most effective for thermal and neutronics coupling, and for depletion calculations, both may be referred to as multi-physics models. This is because adjoint-based calculation of the derivatives for a multi-physics model is typically impractical for two reasons. First, one must implement an adjoint capability suited for a multi-physics coupling which requires substantial and intrusive code modifications, and strong familiarity with the source codes of all the coupled physics models. Second, the computational cost required for multi-physics adjoints does not scale linearly with the cost of single-physics adjoints, because now one must obtain achieve iterative convergence across all coupled physics.

In more realistic scenarios, a multi-physics model is equipped with an adjoint capability for the neutronic solver only. In this case, one could use a hybrid approach that combines both the gradient-based and gradient-free reduction to gain

maximum reduction. Fig. 4.5 depicts this situation, where model #2 is equipped with an adjoint capability, whereas model #1 is only available in a forward mode.
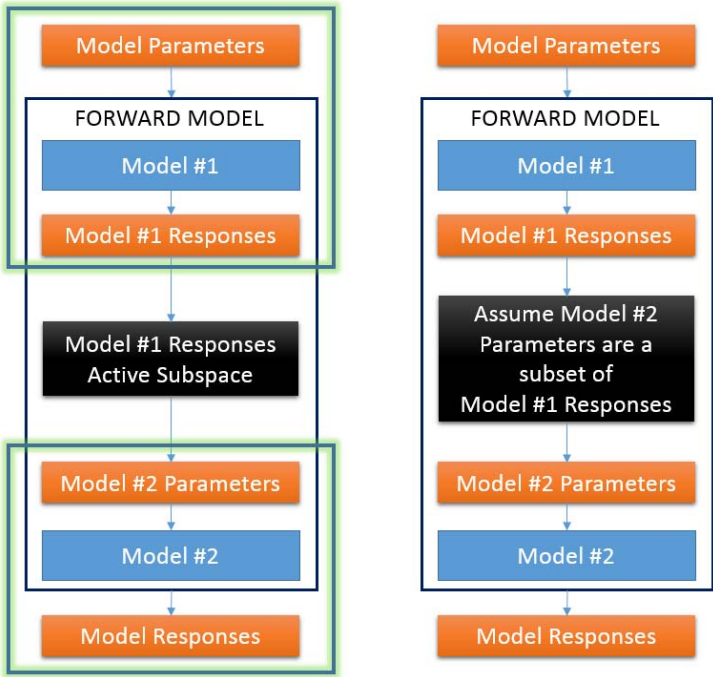
(a) Post-Reduction   (b) Pre-Reduction

Fig. 4.4 Gradient-free Reduction for Code Chain

Fig. 4.5 Combined Gradient-based & free Reductions
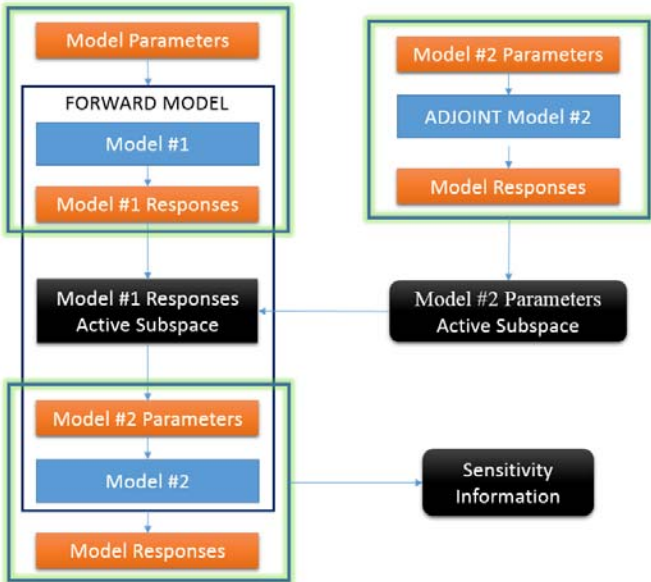
In this case, model #2 is used to identify an active subspace for its input parameters, which are also the output responses of model #1. An additional active subspace can be determined using model #1. The intersection between these two subspaces is used as the basis for reducing the dimensionality of the model #2 parameters. Like before, the errors resulting from this type of reduction can be rigorously assessed to ensure that no important components are erroneously discarded.

Finally, in some situations, none of the codes in the chain have an adjoint capability, implying that one must rely only on the gradient-free approach which may not be adequate enough to render the needed level of computational efficiency.

To address this, we first would like to note that in all previous applications, the active subspaces were determined using the actual codes through which uncertainties are to be propagated. ROM however provides a valuable flexibility in the determination of the active subspace by allowing the user to calculate the reduction errors for a general user-defined active subspace. The implication here is that one could use a different but representative code to determine the active subspace via a gradient-based reduction approach, and then employ such subspace to constrain the parameter perturbations for the computational code of interest, i.e., the one through which uncertainties are to be propagated. This situation is depicted in Fig. 4.6.
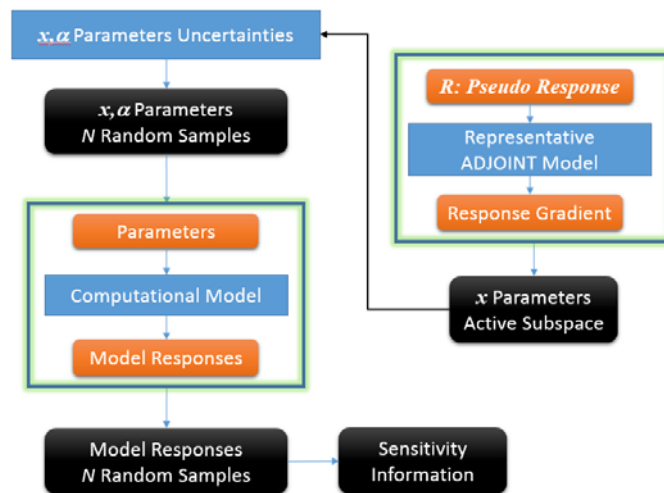
Fig. 4.6 Gradient-based Reduction with Representative Adjoint Model

To accomplish reduction via a different code, one must construct models, say for the cell lattice calculations, and apply the gradient-based reduction as illustrated earlier. The only requirement here is that both the original model and the representative model must use the same data interfaces, e.g., both should use the same number of groups for cross-section representation.

The rest of this section will detail the mathematical description of the above reduction methods, including gradient-based, gradient-free, and their hybridization.

The gradient-free algorithm is implemented as follows:

1. Generate N samples of the parameters according to their expected ranges of variations
2. Execute the (forward) model to calculate the corresponding responses samples $y_{i=1,...,N}$
3. Aggregate the responses samples in a matrix $\mathbf{Y} = \begin{bmatrix} y_1 & y_2 & .... & y_N \end{bmatrix} \in \mathbb{R}^{m \times N}$
4. Using rank revealing decomposition, calculate $\mathbf{Q}_y$.

Qualitatively, this algorithm calculates response variations resulting from perturbations along random directions in the parameter space. The result is the matrix $\mathbf{Q}_y$ which can be used to represent all possible response variations within the tolerance set by the user.

For the gradient-based reduction, we need to define a pseudo response z, which represents a random linear combinations of the model responses, it describes the adjoint-equivalent of executing a model in a forward mode with randomized parameter perturbations.

$$ z = \mu^T y $$

where μ is a random vector that contains the weights used to combine the model responses. Assume that the derivatives[3] of z with respect to x are available, i.e., dz/dx. The algorithm proceeds as follows:

1. Generate N samples of the parameters according to their expected ranges of variations
2. Execute the model (typically the adjoint model) to calculate $dz/dx\big|_{i=1,...,N}$
3. Aggregates the derivatives in a matrix $\mathbf{Z} = \begin{bmatrix} \dfrac{dz}{dx}\Big|_1 & \dfrac{dz}{dx}\Big|_2 & .... & \dfrac{dz}{dx}\Big|_N \end{bmatrix} \in \mathbb{R}^{n \times N}$
4. Using rank revealing decomposition, calculate $\mathbf{Q}_x$

This approach calculates the gradient of the pseudo response at random points in the parameter space. If the model is linear, the gradient remains the same. However for nonlinear model, the gradient variations will identify directions in the parameter

---

[3] This reduction can be applied to both control and model parameters. To minimize the notations clutter, it is described here for model parameters only.

space that are responsible for nonlinear variations. By employing a randomized approach, one can determine the matrix $\mathbf{Q}_x$, whose columns represent all parameter directions to which the response variations are sensitive.

Finally, when rendering reduction for a chain of codes, the reduction results for a given code are employed to generate randomized samples for the next code. Consider the case where the first code in the chain generates via gradient-free reduction the responses y and their active subspace represented by the matrix $\mathbf{Q}_y$. The input parameter perturbations for the next code can be determined as follows:

1. Assume that the next code in the chain reads y directly.
2. Generate N random samples of y_i according to their expected ranges of variations.
3. Project the samples using: $\tilde{y}_i = \mathbf{Q}_y y_i$.
4. Use the projected samples $\tilde{y}_i$ to execute the next code in the chain.

For the general case when both gradient-based and gradient-free reductions are combined, the mathematical details are more involved, and the interested reader may be referred to a previous publication [Wang, 2014].

5.0 UCF IMPLEMENTATION PLAN FOR CANDU REACTOR CALCULATIONS

This section explains how the UCF can employ the UC techniques discussed in section 2.0, and their ROM rendition discussed in section 4.0, to characterize the uncertainties for a typical CANDU reactor analysis sequence which were discussed in section 3.0. Our focus will be on introducing ideas that may be used to capture the various sources of uncertainties, and their algorithmic requirements, and associated implementation challenges. Whenever appropriate, flowcharts and/or some implementation details will be given to provide some perspective on the computational cost required, the computer codes and computational resources; however the information presented is not meant to be comprehensive or to provide a detailed implementation plan.

The discussion is split into four subsections, mirroring the four stages of CANDU reactor analysis calculations that are described in section 3.0. In each subsection, we will refer to the related model and control parameters, representing parameter sources of uncertainties, and the modeling decisions representing the modeling uncertainty sources. References to the basic UC algorithms, i.e., UQ^p, UQ^m, and UQ^m mapping algorithm, described in section 2.0 will also be made.

5.1 ENDF Pointwise Cross-sections Generation

The primary source of uncertainties here is the one originating from the differential cross-section measurements which are used based on a fitting procedure to calculate the nuclear reaction model parameters. This uncertainty analysis exercise is already done by evaluators, and the uncertainties of nuclear model parameters are available in the ENDF format in the covariance file MT=33. An example code that performs these calculations is the ORNL's SAMMY code [Larson, 1984].

These uncertainties are assumed to represent the starting point for all downstream calculations. The inherent assumption here is that modeling uncertainties resulting from the choice of the form of nuclear reaction models are already included in the uncertainties of the reaction model parameters. This assumption is considered by many practitioners to be adequate, although it has never been validated. Using our nomenclature defined in section 2.1.6, we will assume that all uncertainties are embedded in the nuclear reaction model parameters.

1. Model parameters: differential cross-section measurements
2. Model parameters uncertainties: differential cross-section measurements uncertainties
3. UC Algorithm: Bayesian Generalized Least-Squares Fitting
4. Control parameters: none
5. Modeling decisions: choice of the nuclear models, e.g., R-matrix theory, single- or multi-level Breit-Wigner resonance models, etc.

5.1.1 *Propagation of differential cross-section measurement uncertainties*

As described in the earlier section, the nuclear reaction model parameters are generated from a fitting procedure, which is subject to its own errors. For example, the model form used to represent the cross-sections is selected by the analyst. Also, given the wide range of cross-section variations with neutron energy, the entire neutron energy range is split into sub-regions, whose boundaries are determined by the analyst, wherein each region has its own reaction model and its own parameters. One could separate all these selections made by the analyst into a number of modeling decision variables. However, we will assume that all sources of errors and uncertainties are assumed aggregated in the overall residual terms representing the minimized discrepancies between the measured differential cross-sections and the fitted ones which are fully represented by the reaction model parameters uncertainties. The validity of this assumption remains to be assessed.

5.1.2 *Propagation of modeling uncertainties from nuclear reaction models*

To understand the impact of modeling decisions, i.e., choice of nuclear reaction models, an UQ^m algorithm could be employed to generate the ENDF pointwise cross-sections using different nuclear reaction models (which could be done using the ORNL's SAMMY code). The resulting ENDF libraries can be used to estimate the standard reactor physics quantities such as eigenvalue, and lattice power distribution using a continuous cross-section Monte Carlo model for a single cell lattice. This will provide the justification whether the modeling uncertainties can be safely discarded.

5.1.3 *Propagation of overall uncertainties*

If the previous investigation indicates the form of the nuclear reaction models introduces a noticeable level of uncertainty, one could adjust the Bayesian fitting procedure to account for that effect. This is done by artificially increasing the differential cross-section measurement uncertainties to account for the uncertainties resulting from the nuclear reaction models uncertainties. This results in increasing the uncertainties of the nuclear reaction model parameters.

5.2 Multi-group Cross-section Generation

In this stage, the goal is to calculate the multi-group cross-sections and propagate their uncertainties which are the starting point for all downstream CANDU reactor analysis calculations. As mentioned in section 3.2, this stage is done only once using an assume flux shape that is representative of the CANDU thermal spectrum. The uncertainties are propagated from the previous stage in the form of uncertain reaction model parameters. The uncertainty sources and the proposed algorithms for their propagation are described as follows:

1. Model parameters: nuclear reaction model parameters
2. Model parameters uncertainty: covariance matrix in ENDF format
3. UC algorithm: UQ^p, i.e., forward sampling-based UQ approach
4. Control parameters: null
5. Modeling decisions: flux shape used to collapse the continuous cross-sections
6. UC algorithm: UQ^m
7. Correlation between model and decision parameters: UQ^p

The UCF application to this stage will achieve three main milestones:

5.2.1 *Propagation of continuous cross-sections uncertainties*

This milestone represents a straightforward application of the linearized UQ approach that employs the sandwich equation, cf. section 2.2.1. It starts with the nuclear reaction model parameters reference values and their covariance matrices, and for a given flux shape, it calculates the references multi-group cross-sections and their uncertainties in the form of a covariance matrix. This process is already completed by computer codes like NJOY ERRORJ module, and SCALE PUFF code. The basic assumption of the linearized UQ approach is that the cross-sections variations result in linear variations in the multi-group cross-sections, which is considered an acceptable assumption by most practitioners.

To our knowledge, this assumption has not been rigorously verified, which is seamlessly done by propagating the nuclear reaction model uncertainties using a sampling-based UQ^p approach. Comparison of the standard deviations and the correlation structure of the multi-group cross-sections and their impact on integral quantities of interest such as the eigenvalue could help determine whether the linearity assumption is acceptable.

5.2.2 *Propagation of uncertainties resulting from assumed flux shaped*

This milestone needs to establish whether the assumed flux shape impacts the propagated multi-group cross-section uncertainties. Initially, a SA study could be devised to determine the relative importance of this term. This can be done by

repeating milestone in section 5.2.1 with different flux shapes and checking the variations in the multi-group cross-sections uncertainties. These uncertainties can be integrated with sensitivity profiles from cell lattice calculations to determine their impact on important metrics, such as reactivity coefficients, peaking factors, etc. If the effect is small, this source of uncertainties could be discarded. If it is important, we proceed to completing the next milestone.

5.2.3   Propagation of overall uncertainties

Here we focus on propagating the uncertainties resulting from the assumed flux shape, then combine that with those resulting from nuclear reaction model parameters, and finally investigate any correlations between them.

To estimate the uncertainties resulting from the assume flux shape, we employ the UQ^m algorithm, wherein the assumed flux shape is considered one of the modeling decisions that result in modeling errors. Consider a cell lattice model of a fuel bundle modeled via a high fidelity Monte Carlo continuous cross-section model. This model can calculate a detailed flux solution and collapse the continuous cross-sections into multi-group cross-sections as the responses of the Monte Carlo model. These predictions can be compared to those generated directly with an assume flux shape. The discrepancies represent a measure of the errors resulting from the assumed flux shape. To convert these discrepancies into a PDF, we repeat this process for different compositions that are expected throughout depletion. The result is a PDF that describes the uncertainties in the multi-group cross-sections resulting from the assumed flux shape.

To combine the two sources of uncertainties, the milestone in section 5.2.1 is repeated but now with the flux shape uncertainties included as part of the uncertain model parameters set using a standard UQ^p approach.

5.3 Cell Lattice Calculations

In this stage, the goal is to estimate the few-group cross-sections uncertainties resulting from the uncertainty sources described below:

1. Model parameters: multi-group cross-sections
2. Model parameters uncertainty: covariance matrix in multi-group format
3. UC algorithm: ROM-based UQ^p
4. Control parameters: null
5. Modeling decisions: transport solver, multi-group approximation, and reflective boundary conditions.
6. UC algorithm: UQ^m, and UQ^m mapping algorithm
7. Correlation between model and decision parameters: ROM-based UQ^p

The few-group uncertainties must be generated as a function of the wide array of conditions required for downstream core calculations, e.g., fuel temperature, coolant density and temperature, boron content, etc. Here, the UCF will achieve three main milestones:

5.3.1 *Propagation of multi-group cross-sections uncertainties*

These uncertainties can be propagated using a straightforward UQ^p approach like before, however given our interest in estimating the contribution of the various sources of uncertainties, we will have to rely on an ROM-based approach to supplement the standard UQ^p approach. This is because the number of multi-group cross-sections is too large to render a forward SA approach practical. The proposed approach combines both adjoint and forward model executions to propagate and prioritize the dominant sources of uncertainties. The result is a ranking table of all multi-group cross-sections and their contributions to the few-group uncertainties.

The requirement here is to have a cell lattice code capable of calculating both forward and adjoint fluxes. The combined used of forward and adjoint capabilities allows one to reduce the effective dimensionality of the uncertainty space to a manageable size to complete UC.

The algorithm is carried out in two steps. The first step focuses only on the identification of the active subspace of the uncertainty space, which represents the multi-group cross-sections with dominant impact on the few-group cross-section uncertainties. The second step employs a standard UQ^p approach to propagate the dominant uncertainties only. Previous work has shown that the effective dimensionality of the space for typical LWR calculations is only in the order of 10 to 100 dimensions [Abdel-Khalik, 2014].

This approach has two distinct advantages. The errors resulting from restricting the uncertainty space to an active subspace can be rigorously quantified and can be made very small rendering them negligible for all practical reactor analysis applications, as shown by earlier studies [Bang, 2012, Wang 2013].

The second advantage is that the reduction can be carried out with a code that is different from the one used to propagate uncertainties. This is very valuable because the selected cell lattice code may not have the capability to calculate the adjoint solution. In this case, one could use a similar code, e.g., SCALE's NEWT, to do the adjoint calculations. The only requirement here is that the two codes employ the same energy-group structure for the cross-sections to ensure consistency of the reduction results.

5.3.2 *Propagation of uncertainties resulting from the transport solver and multi-group approximation*

This milestone will investigate the impact of the transport solver and multi-group cross-section approximation on the propagated few-group cross-section uncertainties. To our knowledge, this investigation has not been done before. The basis for this will be a high fidelity Monte Carlo model whose predictions will be compared to the standard cell lattice code used in CANDU calculations. The main requirement here is that the Monte Carlo model employed must be able to calculate the few-group cross-sections.

The main obstacle here is computational in nature since Monte Carlo is computationally intensive. The computational cost could be reduced significantly by employing two techniques, first by reducing the effective dimensionality of the uncertainty space, as explained in section 4.2, and by employing variance reduction techniques to accelerate Monte Carlo convergence. As described in section 4.2.2, one can reduce the dimensionality of the uncertainty space using a representative code (i.e., different from the one through which uncertainties are to be propagated). In this case, one could rely on a fast-executing deterministic code to reduce the dimensionality of the uncertainty by determining the active subspace, which is subsequently used to propagate uncertainties using the Monte Carlo code [Abdo, 2015]. The errors resulting from the uncertainty space reductions are very small, and can be quantified rigorously, as described earlier.

Regarding variance reduction, earlier work has shown that if one is interested in executing many Monte Carlo models that are closely similar, one can employ the adjoint solution to draw weight-window maps that bias the particles histories resulting in accelerated convergence [Zhang, 2014]. Noticeable improvements in performance have been noticed by using this approach. In our case, the set of model runs represent perturbations of the multi-group cross-sections around some their nominal values, as determined by the multi-group covariance matrix. The weight-windows can be drawn using a deterministic adjoint solver as described earlier. Clearly, the main requirement here is that the Monte Carlo model employed should have the means to perform variance reduction.

With these two acceleration techniques, we expect uncertainties to be propagated and prioritized in an efficient manner using Monte Carlo models. This will allow one to get an improved estimate of the few-group uncertainties that is not contaminated by modeling errors resulting from the transport solution.

5.3.3 *Propagation of uncertainties resulting from the reflective boundary conditions*

This uncertainty is treated as a source of modeling errors resulting from a modeling decision, and hence can be propagated using the proposed UQ^m approach, cf. section 2.2.2. To achieve that, one has to set up two models, one representing the standard single bundle cell lattice model, and another representing a higher fidelity model that takes into account the effect of the neighbors using a super cell lattice model, i.e., with the surrounding 8 bundles as shown in Fig. 5.1.

In this approach the 8 neighboring cells to be surrounding the reference cell lattice in the core are modeled explicitly to understand their impact on the few-group cross-sections[4]. Given that the neutrons mean free path is shorter than the characteristic length of a single cell lattice in a CANDU reactor, the 3x3 super cell is expected to capture this source of uncertainty.

The UQ^m approach can be used to calculate the discrepancy between the few-group cross-sections calculated using the standard single model cell lattice model and the super cell lattice model. This discrepancy is calculated for a wide range of conditions, i.e., different compositions, temperatures, etc., to generate a joint PDF that relates the high fidelity and lower fidelity model predictions. An UQ^m mapping algorithm, cf. section 2.3, could then be employed to map the uncertainties to all possible neighbor configurations expected in the real core.

(a) 3x3 Super Cell   (b) Standard Cell Lattice
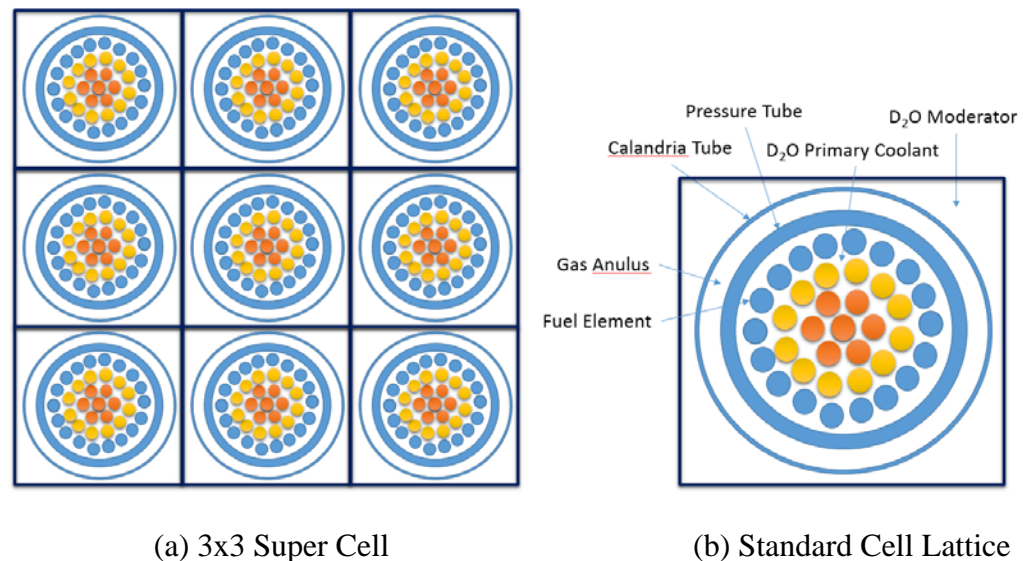
Fig. 5.1 Cell Lattice Models

---

[4] Other variations are also possible in the super cell design to account for the presence of absorber devices between adjacent lattices and lattice cell at the core radial boundary whose adjacent lattice cells are expected to be significantly dissimilar in design.

Also, given that in practice one cannot model all possible super cells expected in the core, the UQ^m mapping algorithm can be employed to map the modeling uncertainties for the specific core loading, without having to repeat super cell calculations for each lattice and its surroundings in the core.

5.3.4 Propagation of overall uncertainties

This step focuses on combining all sources of uncertainties estimated in sections 5.3.1, 5.3.2, and 5.3.3 using a sampling-based UQ^p approach. Depending on the size of the relative contributions of each source, several approaches can be devised. The most accurate approach is to repeat the algorithm in section 5.3.3 but now allow for the perturbation of multi-group cross-sections along with the composition and temperature variations. The simplest approach is to assume these sources of uncertainties are independent and combine them using the standard sandwich quadratic formula.

5.4 Downstream Core-wide Calculations

In this final stage, the goal is to estimate the uncertainties in the core attributes such as core reactivity, bundle enthalpy rise, control rod worth, etc. during steady state and transient calculations which result from the following sources of uncertainties:

1. Model parameters: few-group cross-sections and thermal-hydraulic parameters (e.g., friction factors, heat transfer coefficients, all parameters associated with empirical correlations)
2. Model parameters uncertainty: covariance matrix in few-group format (this format is not standard, and is dependent on the core simulator model employed).
3. UC algorithm: ROM-based UQ^p approach.
4. Control parameters: core conditions, e.g., flow rate, boron content, etc.
5. Modeling Decisions: transport solver, few-group approximation, and non-neutronic models
6. UC algorithm: UQ^m
7. Correlation between model, control, and decision parameters: ROM-based UQ^p

The UCF application over this stage will achieve the following milestones:

5.4.1 Propagation of few-group cross-section uncertainties

As mentioned earlier, given the enormous size of the few-group cross-sections, an initial reduction of the uncertainty space will be done based on the UC results from cell lattice calculations. This reduction does not require any adjoint calculations at the core level. It can be achieved using filtering techniques, i.e., RFA, of the samples generated during the propagation of uncertainties through cell lattice models. This initial reduction helps reduce the size of the few-group uncertainty

space to a manageable size, expected to be in the order of 100, based on previous experience with LWR models [Jessee, 2011, Bang, 2012, Abdel-Khalik, 2008]. At this small size, a standard UC approach could be employed to propagate and prioritize the multi-group cross-section uncertainties and the control parameters uncertainties.

5.4.2   *Propagation of control parameters and non-neutronic parameters uncertainties*

Control parameters uncertainties are addressed here because in reality some of the input data to the core simulator are measured at the plant and hence are subject to measurement uncertainties. These uncertainties can be easily addressed using a standard UQ^p approach since their number is small. Also, notice that the model parameters are expanded to include non-neutronic parameters. We propose not to include these sources to confine the discussion to neutronic analysis only. Based on discussions with CNSC staff, a decision could be made whether these uncertainties should be propagated in the present study. Finally, regarding the decision parameters, as mentioned before, we propose to exclude the propagation of modeling errors resulting from non-neutronic models in the present study, and focus only on errors resulting from the use of few-group nodal diffusion theory employed in standard CANDU core-wide calculations.

5.4.3   *Propagation of uncertainties from few-group nodal theory approximation*

To propagate the uncertainties resulting from the use of diffusion theory and few-group cross-section approximation, we propose to employ a Monte Carlo model of a small region of the core, e.g., one-quarter or one-eighth core model at a wide range of core conditions, including different compositions, coolant and fuel temperature variations, etc. Using the UQ^m approach, the discrepancies between the two model predictions sampled at the range of core conditions generates a PDF that describes the effect of modeling errors on core attributes of interest. Also, like before, all sources of uncertainties, i.e., modeling errors, few-group cross-section uncertainties, and control parameters, can be combined by repeating the UQ^p approach but now with the inclusion of all sources of uncertainties.

6.0 UCF REQUIREMENTS

To realize the UCF benefits, all calculated uncertainties must be subject to rigorous verification and validation exercises such as those employed for model verification and validation. Further, the computational cost must be small enough to render practical routine execution along with BE calculations. This section discusses these two high level requirements, along their associated software and hardware requirements.

6.1 Verification and Validation of the UCF Results

Similar to any new code development effort, rigorous verification and validation exercises must be set in place to ensure credibility of the UCF results. In the standard V&V process applied to a general BE model, the verification exercise asks the question of whether the model has been implemented correctly. In other words, has the analyst made any mistakes while implementing the model? Is the model producing the expected results? If the model can be solved analytically, are the predictions generated by the code matching the analytical results? Validation asks different questions such as: does the model represent an adequate representation of reality? are the model predictions consistent with measurements? The verification questions are mathematical in nature; they compare the predictions of a conceptual model to its computerized version, while the validation questions are physical in nature; they compare a conceptual model predictions to reality.

Both verification and validation questions can be customized to lend credibility to the implementation of the UCF under investigation in this study. With regard to verification, one is interested in providing evidence that the UC algorithms are implemented correctly. For example, one would be interested to verify whether the random samples have been generated in a manner that is consistent with their prior PDFs, and whether the samples been correctly constrained to their respective active subspaces, etc. With regard to validation, one needs to assess whether the assumptions made about the various uncertainty sources are adequate. To do that, one must compare the uncertainties to the actual discrepancies found between real experimental measurements and model predictions. For example, the validation exercise could help determine whether it is acceptable to discard some of the uncertainty sources such as the nuclear reaction model forms used to construct the pointwise cross-sections. It could also assess whether the linearity assumption used to propagate the reaction model parameters uncertainties is adequate. An important overall question that must be answered by the validation exercise is whether the observed discrepancies are statistically consistent with the propagated uncertainties. The statistical consistency can be measured with metrics like chi-square. If the consistency checks fail, it is possible that the uncertainties propagated are under or over-estimating the real uncertainties.

6.2 Routine Execution

The importance of routine characterization of uncertainties via UQ and SA analyses has been heavily emphasized by the nuclear engineering community in order to optimally realize the benefits of advanced modeling and simulation software used for reactor analysis. Currently, the existing methods for UC are computationally inefficient rendering them unsuitable for routine execution along with BE calculations. Therefore, there is a stringent need to devise efficient UC algorithms that can render the process computationally tractable on a routine basis.

To render a successful UCF implementation, one must develop three layers of algorithms to support the application of ROM to UC analysis: a) Basic UC/ROM process algorithms; b) UC CANDU-specific algorithms; and c) Buffer codes for the various computational codes comprising a typical CANDU computational sequence.

6.2.1 Basic UC/ROM algorithms

The first layer, basic UC/ROM process algorithms, refer to all the fundamental operations required to support the algorithms of the second layer. These basic algorithms include statistical packages to compute and process probability density functions and moments, design randomized samples from given distributions, perform decomposition of the uncertainty space, supporting linear algebra decomposition algorithms, and pattern recognition and data reduction algorithms. These process algorithms have been developed by researchers in many scientific communities, and therefore are readily accessible from the public domain.

Most of these process algorithms are already available under open-source licenses in a number of toolboxes for UC and ROM that have been developed by practitioners in a number of scientific communities. With little effort, these process algorithms could be extracted for use in a customized framework for CANDU applications. For example, the NASA UQTools toolkit [NASA 2012], the Lawrence Livermore National Laboratory PSUADE toolkit [PSUADE 2011], the Los Alamos National Laboratory GPMSA Toolkit [GPMSA Website]. All these toolkits are readily available and their associated process algorithms have already received sufficient level of verification to justify their adoption for the proposed framework. A detailed discussion of these tools is given in section 7.0

6.2.2 UC CANDU-specific algorithms

The second layer involves building CANDU-specific algorithms using the basic process algorithms, e.g., UQ^p, an algorithm for the propagation of parameter uncertainties, UQ^m, an algorithm for propagation of modeling uncertainties, UQ^m mapping, an algorithm for the mapping of uncertainties, and ROM-SA, an

algorithm to prioritize the dominant sources of uncertainties. As described earlier, some of these algorithms have been implemented in a generic form in a number of publically available UQ toolkits, cf. section 7.0. However, considerable level of customization is needed to suite CANDU reactor analysis applications, especially in regard to the UQ^m and UQ^m mapping algorithms since they are closely tied to the sources of modeling uncertainties, which are model-specific.

The sequence of CANDU calculations will also dictate the optimum manner in which these algorithms are employed. For example, at the multi-group level, one needs to implement an intersection subspace approach that combines the active subspace generated by gradient-based reduction and the active subspace inferred from the multi-group covariance matrix. At the few-group level, an ROM-SA algorithm needs to be applied prior to the application of the UQ^p algorithm to core-wide calculations. These decisions are be based on the analyst's familiarity with the CANDU reactor analysis sequence, and hence customization is needed to maximize the efficiency of the UC application. Examples of the level of customization of the basic UC/ROM process algorithms are given in section 5.0.

6.2.3   *Buffer codes construction*

Finally, the third layer involves the construction of buffer codes to read and write the interfaces of all the computational codes in the CANDU reactor analysis sequence. They must provide the functionalities to perturb the parameters and read the responses as needed by the basic process algorithms, and allow one to constraint the perturbations to the active subspace as required by the CANDU-specific algorithms. Except for heavily-protected proprietary codes, e.g., HELIOS multi-group cross-section format [Casal, 1992], writing buffer codes is a straightforward process once the codes' input and output files formats are provided. This layer requires the most customization, and hence must be conducted every time a new code is introduced in the computational sequence.

6.3 Computational resources requirements

This section describes the computational resources requirements in terms of parallel code environment, data storage and manipulation by the various UC/ROM algorithms. To determine these needs, one must address two closely related questions. The first one is what kind of computing platforms are required. To ensure UC is applied on a routine basis, one must confront the reality of industrial calculations which largely depend on desktop computers. Access to supercomputing resources is only available to government national laboratories and academic institutions typically on a need basis. Therefore, the developments of UC/ROM process algorithms must take into consideration the limited computer power available to the nuclear practitioners, including fuel vendors, utilities, and regulators.

The second question is how much computational overhead in terms of computer time and storage is needed to complete UC? The computational overhead cannot greatly exceed the overhead needed for BE calculations; otherwise it will not be realistic to expect practitioners to perform these analyses on a routine basis. Assuming that BE calculations are done on desktop computers, the computational overhead for UC must not exceed the computational power provided by a small cluster of few hundred nodes.

Regarding storage requirements, it must be noted that while computer storage is relatively cheap, brute force application of UC algorithms is not advisable, because it could result in massive storage requirements which negatively impacts the efficiency of UCF calculations. By way of an example, a brute force sampling-based UQ algorithm would generate a covariance matrix for the few-group cross-sections, expected to have a size that is proportional to the squared number of few-group cross-sections. This number is extremely large considering that the few-group cross-sections are functionalized in terms of a wide range of core conditions for typical CANDU core-wide calculations. Having to store and manipulate that amount of data is computationally unsound, and therefore ROM techniques must be carefully employed to not only reduce the computational cost of the UCF calculations, but also reduce the data streams that are flowing between the various codes, and are used to characterize uncertainties.

6.3.1 Parallel Code Environment

Access to a small computing cluster consisting of few hundred nodes (a minimum of 100 nodes, each with a dual core) is necessary to render the process of uncertainty characterization computationally feasible. This is because the UC/ROM process algorithms require executing the associated reactor analysis models few hundred times. Employing a small cluster will ensure that the time required for the UC application is of the same order of magnitude or slightly higher than the cost of reference BE calculations. Without a cluster, the computing cost on a single machine would be approximately three orders of magnitude higher than reference BE calculations.

To better explain the value of parallel computing environment, one may split the computational overhead of the overall UCF into two components, a primary component which involves the repeated code executions, and a secondary component which prepares the perturbed input files, and processes the output files using UC/ROM process algorithms.

The first component is embarrassingly parallel, since it requires only coarse-grained parallelization implying that the computational time will scale directly with the number of available processors. This is achieved by directly distributing the number of code executions on all available processors, i.e., one code execution per

processor. The only requirement here is the availability of a submission queue on the cluster to allow simultaneous submission of multiple code executions on multiple processors. Coarse-grained parallelization is a direct consequence of the use of randomized techniques to identify the active subspaces, wherein each code run the model parameters are selected randomly, i.e., independently of all other code executions. This is one of the primary advantages of the ROM process algorithms proposed in this framework, as it allows the analyst to reach the theoretical maximum efficiency of the parallel computing environment. The availability of a computing cluster and the associated submission queue may be considered hard requirements for the UCF. Absent these requirements, the computational cost for UC will be practically infeasible.

With regard to the second component, the computational algorithms require first the aggregation of all the output results followed by performing statistical analysis and ROM techniques on the results collected from all processors. These computations require fine-grained parallelization, implying that the algorithms must be written using parallel environment instructions such as OpenMP and MPI multi-threading or combination of the two. Given that the bulk of UC/ROM process algorithms are mainly inner product and matrix-vector products operations, their parallelization is seamless resulting in a very high efficiency that is close to the theoretical maximum value achieved coarse-grained parallelization. This requirement can be easily realized as most of the UC/ROM process algorithms available in the public domain have already been parallelized to maximize their performance.

6.3.2 Data Management Server

A data management server must be developed to facilitate the exchange of data between the various UC/ROM process algorithms and the CANDU reactor analysis codes, expected to vary considerably in their associated data types and structures. Data management is therefore an integral component to the UCF application to CANDU cores. This is because CANDU reactor analysis codes must be executed repeatedly to generate realizations that can be used to properly represent the wide ranges and types of the uncertainties inherent in the simulation. Given the complexity of CANDU models, the sizes of the data streams associated with these realizations are expected to be very large, typically numbering in the millions. A brute force data management server that relies on sequential text files to process, temporarily store, and recall the data as needed by the various process algorithms would result in poor performance and hence must be avoided.

Data management servers also have the advantage of standardizing the formats of the data being processed. This is an important characteristic as the UC/ROM

process algorithms are expected to be agnostic to the data being processed, and hence standardization of data formats is essential.

Therefore, a modern data management server must be either developed or adopted to ensure flexibility and efficiency in handling the CANDU numerically-intensive data structures in order to minimize the time required for their temporary storage and recall as required by the proposed framework. Given that this is not a unique problem to CANDU simulation, we recognize that many data management servers have been proposed and successfully employed over the past two decades in different scientific disciplines. One notable data management approach is the HDF5 data format, which we believe would be ideal for the UCF application.

The HDF5 format provides a versatile data structure that can represent very complex data objects and a wide variety of data, with efficient recall and temporary data storage. Moreover, enough experience exists in the nuclear engineering community as evident by recent applications by fuel vendors and government laboratories. For example, the RAVEN [Alfonsi, 2013] environment has demonstrated the ability to store the mapping between the variations of the input files and the resulting values for the responses of interest in the output files in the form of databases constructed using HDF5 formats.

6.3.3   Job Management Software

The job management software is needed to handle and coordinate the execution of the two fundamental components of the UCF described in section 6.3.1. In the first component, the management software must generate the needed input files for the various models, dispatch the input files to the parallel environment, execute the models on the various processors, and aggregate the resulting output files from all processors. In the second component, the software needs to execute UC/ROM process algorithms to determine active subspaces, which are employed by the first component to determine the samples used to construct the needed input files. The integration between these two components is therefore crucial for the successful and efficient implementation of the UCF.

The RAVEN environment, developed at the Idaho National Laboratory, is less than two years old. The main developer of RAVEN was invited to give a talk to the contractor about its recent status. A preliminary evaluation of RAVEN shows that it is carefully designed to suit the implementation of UC/ROM process algorithms. In particular, RAVEN allows the user to specify the number of jobs it is allowed to run simultaneously, and the number of MPI environment per job, and number of MPI multithreads. RAVEN will interfaces with the queuing system on the cluster and request a total number of cores equal to the product of the above three values.

Once obtained the needed numbers of cores RAVEN will use the MPI implementation (a version of MPI needs to be installed on the cluster) to distribute the job over the available resources and ensuring that, if computational resources are freed by the conclusion of one of the submitted jobs, new jobs are submitted as long as new samples are needed. RAVEN can use the queuing system but could be run on single nodes, in this case it will use MPI to allocate the resources but it will not need to interact with the queuing software. When used on a single machine, RAVEN will simply submit as many job as the size of the batch and relay on the OS to distribute them.

6.4 Reactor analysis software requirements

This subsection discusses the requirements of all the software tools employed in typical CANDU reactor analysis calculations. These requirements include:

6.4.1 Accessibility to input and output files formats

Because the UC algorithms need to perturb the parameters to the various computer codes employed in the overall analysis, access to the input file formats is necessary. If the input files are only available in binary format, the code developers or the programmer manuals must be consulted to identify these formats. The output files formats are also needed, because as explained earlier, ROM approaches will be designed to devise reduction at the interfaces of the various codes, e.g., the few-group cross-sections at the interface between cell lattice and core calculations.

6.4.2 Variance reduction capability

To characterize the uncertainties resulting from modeling assumptions, access to Monte Carlo continuous cross-section models is a must. The challenge here is that one must execute the Monte Carlo model in the order of several hundred times, sometimes for a standard cell lattice, sometimes for a super lattice, and other times for the whole reactor core. This is computationally expensive, and is unlikely to be possible with some considerable acceleration of Monte Carlo simulation.

Given the nature of the repeated simulations, each representing a perturbation off of a nominal case, variance reduction techniques provide an excellent tool to speed up Monte Carlo convergence to render the process computationally tractable. If the variance reduction capability is not available, one must increase the computational resources discussed in section 6.3.1 by two orders of magnitude, which is the approximate computational savings expected with variance reduction techniques.

6.4.3 Capability to calculate few-group cross-sections with Monte Carlo models

In support of cell lattice calculations, the UCF intends to calculate the modeling uncertainties resulting from the use of deterministic multi-group cross-sections in cell lattice calculations, and their impact on the few-group cross-sections used in downstream core-wide calculations. To achieve this, one must be able to generate the few-group cross-sections using a Monte Carlo continuous cross-section model. For some Monte Carlo codes, e.g., SERPENT [SERPENT, 2011] and KENO [SCALE, 2011], this is a standard feature, and for others, such as MCNP [MCNP Website], some minor work is needed to define tallies that homogenize the cross-sections in the form of few-groups.

6.4.4 Availability of deterministic adjoint transport solver

The most significant reduction in reactor analysis calculations occurs at the cell lattice level, therefore one must devise the best ROM algorithm to render the maximum reduction possible. As discussed in section 2.0, the best ROM algorithm is one that combines both gradient-based and gradient-free reduction. The gradient-based reduction requires access to the derivatives of the model responses with respect to its input parameters. Therefore, ideally, one would want to use a cell lattice physics code that has an adjoint solver. However, if this capability is not available, ROM provides an alternative approach wherein a different but representative transport code could be used in lieu of the original lattice physics code to calculate the adjoints. The adjoints are employed to calculate the active subspace which is then fed to the original lattice physics code to propagate and prioritize uncertainties. The only requirement of this alternative approach is that the representative code must employ the same multi-group cross-section energy structure to ensure consistency of the active subspace between the two codes. If the two codes do not employ the same energy structure, one could in principle transfer the active subspace, but this process is expected to introduce an additional source of uncertainties, which could be easily avoided.

In our experience [Bang, 2012], [Wang, 2013], the most significant reduction in the size of the uncertainty space occurs in the third stage of CANDU computational sequence, i.e., cell lattice calculations, cf. section 3.0. The implication is that in order to get the best reduction, one needs the adjoint solver or a representative code with an adjoint capability, as discussed in section 4.2.2, in order to realize the maximum reduction possible using the gradient-based ROM approach. For the other three stages, i.e., ENDF library generation, multi-group cross-section generation, and core-wide calculations, the gradient-free ROM approach is sufficient to render adequate reduction which relaxes the need for adjoint solvers.

7.0 <u>OVERVIEW OF EXISTING UC TOOLKITS</u>

Given the importance of UC to the wide engineering and scientific community, our development and implementation philosophy will focus on utilizing as much as possible any of the existing, i.e., open-source, tools and process algorithms for UC and ROM. A number of US and other foreign institutions have already expended considerable resources to develop UC toolkits for a number of scientific disciplines, including the DAKOTA toolkit from Sandia National Laboratory [DAKOTA Website], the PSUADE toolset from Lawrence Livermore National Laboratory [PSUADE, 2011], the RAVEN environment from Idaho National Laboratories [Alfonsi, 2014], and the SCALE's SAMPLER super-sequence from Oak Ridge National Laboratory [Williams, 2013], and the SCALE's CRANE super-sequence from Purdue University [Mertyurek, 2014], and the TMC from the Netherlands NRG group [Rochman, 2014]. These tools contain the basic process algorithms required for any UCF such as sampling algorithms, range finding algorithms, filtering techniques, surrogate construction techniques, etc. They also contain sophisticated strategies for submitting and distributing the numerous jobs in a parallel computing environment, which is one of the pre-requisites of a successful UCF, given the large computational overhead required for typical problems.

An informed development plan should employ select components from these toolkits as a starting point for developing the UCF for CANDU reactor analysis application. This section provides an overview of these toolkits with our intent being to familiarize the readers with the advantages and disadvantages of each toolkit, and the needed developments to render them useful for the UCF application to CANDU problems.

Before listing the various tools in section 7.3, we reiterate a short overview of UC methods which are classified into two categories, sampling-based and ROM-based methods, described respectively in the next two sections of 7.1 and 7.2.

7.1 Sampling-based Methods

Sampling-based methods, as the name implies, involve the execution of the model many times by sampling the various sources of uncertainties, often in the form of parameters with known priori PDFs. The result is a recorded set of response variations which are statistically manipulated to calculate quantities such as the mean, variance, tail probabilities. The methods belonging to this category differ in the way the samples are selected [Helton, 2006] . Some of the sampling techniques employed include, pure random sampling, stratified sampling, Poisson disk sampling, etc. These sampling techniques are designed to minimize the number of samples required to calculate the quantity of interest, such tail probability.

While many studies have compared the performance characteristics of the various sampling techniques, the initial implementation of the proposed framework is not expected to be sensitive to the choice of the sampling method. This is because the size of the uncertainty

space in CANDU reactor analysis is large enough to overwhelm any of the existing sampling techniques. This follows because it has been reported that when the number of uncertain parameters is approximately above 50 (much lower than the size of the CANDU uncertainty space), the basic random sampling strategy proves to be the most efficient approach for uncertainty characterization [Cousins, 2012, Gerstner, 1998]. Therefore, we will not discuss here the specifics of the various sampling techniques and refer the interested reader to the literature for more details [Helton, 2006].

7.2 ROM-based Methods

ROM-based methods achieve computational efficiency in one of two approaches, either by reducing the dimensionality of the uncertainty space, or by building a surrogate model that can be used in lieu of the original model. Appendix D provides details on the differences between these two approaches.

7.2.1 Dimensionality Reduction

DR operates on the assumption that a large part of the uncertainty space is discarded completely from the analysis, as it is deemed to be non-influential with respect to the quantities of interest. To determine what is non-influential, DR employs randomized model executions using both the gradient-free and gradient-based ROM techniques discussed in section 4.2.2. The advantage of the DR approach is that one can rigorously determine an upper-bound on the errors resulting from the reduction.

7.2.2 Surrogate Model Construction

The second approach operates on the assumption that model response variations can be approximated parametrically by a response surface or polynomial splines with a number of undetermined features, such as the coefficients of expansion, or splines' knots placement [Box, 1987]. To determine the unknown features, one must execute the model a number of times and use an optimization technique to determine the features that minimize the discrepancies between the original model and surrogate model predictions. Once the surrogate is constructed, one can apply UC algorithms efficiently, because the surrogate form is selected to be computationally inexpensive to evaluate. The prominent cost is in the evaluation of the surrogate coefficients, since they require the execution of the original model. This introduces two challenges. First, one must be able to sample the entire space adequately to ensure that the surrogate is representative of the real model, which is computationally exhaustive in real applications. Second, it is very difficult to assess the quality of the predictions at points not included in the construction of the surrogate. To achieve that, many additional model executions would be required,

which is typically bypassed given the computational burden. Instead one has to rely on expert judgment to determine whether the surrogate predictions are adequate.

The surrogate construction methods differ primarily in the form of the surrogate model employed, and the sampling techniques employed to generate its training points. Examples include, polynomial fitting with regular intervals for the training points, adaptive stochastic collocation methods, which adapts the intervals based on the degree of model response variations (see [DAKOTA Website] for a great number of response surfaces implementation and their associated theory). A great deal of research goes into the selection of the model forms that render the optimization search computationally stable, i.e., not ill-conditioned, especially when the number of uncertain parameters is more than a few.

7.2.3   *Sensitivity Analysis*

It is worth mentioning here that sensitivity analysis is closely related to the area of surrogate modeling. Appendix A provides an overview of different SA methods. When local SA methods are employed to calculate first and/or higher order derivatives, one could use the results to construct a polynomial expansion, which serves as a surrogate model. When global SA methods are employed however, they serve to provide an initial screening of which parameters or parameters interactions are important, to be included in a subsequent surrogate model construction exercise.

7.3 Overview of Open-Source Tools

This section provides a short overview of existing open-sources tools that are proposed to provide an initial starting point for the proposed UCF. This discussion is intended to provide an idea about the various UC and ROM process algorithms available in open-source format. It is however not intended to imply that any of these toolkits can be used directly to perform UC for CANDU reactor physics calculations. This is due to two reasons. First, all the tools described below focus on propagating parameter uncertainties only. This is because parameter uncertainties can be treated in a generic form that treats all models as black boxes. Modeling uncertainties however require intimate knowledge of the mechanics of the models and basic understanding of the modeling decisions, approximations, and assumptions that give rise to these errors. Based on this knowledge, the analyst can devise special strategies for propagating modeling errors, examples include the proposed UQ^m and UQ^m mapping algorithms presented in 2.2. The second reasons is that significant level of reduction must be done a priori before propagating uncertainties for CANDU reactor physics problems. This follows because the standard black box approach employed by most of the tools described below assume that the user can execute the model enough number of times to fully characterize uncertainties. This is not possible for typical CANDU problems.

7.3.1 DAKOTA

The DAKOTA toolkit is developed by the US Sandia National Laboratory researchers as a general-purpose environment for conducting many engineering-oriented analyses such as uncertainty quantification (UQ), surrogate model construction, design optimization, parameter estimations, and sensitivity analysis (SA), etc. Given that all these analyses require the ability to execute the model repeatedly with variable input parameters, DAKOTA has been developed using an object-oriented design which provides a flexible and extensible environment for repeated model executions on high performance computing environment. Another important feature of DAKOTA is that each of the above engineering analyses can be conducted using a wide range of methods, all available in a single toolkit. This allows the user to experiment with different methods to assess their relative performance for the application of interest.

The primary handicap of DAKOTA is that none of its UQ and SA methods can be used directly for CANDU reactor physics applications; this is due to the high dimensionality of the uncertainty space. It is worth mentioning here that the DAKOTA code has been initially considered as a toolkit for completing UQ and SA analyses for LWR reactors under the US DOE CASL hub initiative, but it was soon recognized that the associated computational cost is unrealistically large. Instead, ROM techniques have proven to be the right tool for rendering initial reduction of dimensionality prior to the application of UQ and SA analyses, see [Abdel-Khalik, 2014] for a demonstration of the use of ROM techniques to complete UQ and SA analyses using an advanced neutronics solver. This is now widely accepted as the most reasonable approach for characterizing uncertainties for reactor physics calculations.

DAKOTA also comes equipped with many surrogate construction techniques which can be used, once constructed, to complete UC. These methods are inadequate for the present CANDU application because on the one hand, the cost associated with the construction of the surrogate model for typical CANDU problems is as expensive as the cost of characterizing uncertainties. On the other hand, DAKOTA does not provide rigorous upper-bounds on the errors resulting from the surrogate model predictions, which introduces an additional source of uncertainty that must be captured to render the results of uncertainty propagation reliable.

7.3.2 XSUSA

The XSUSA methodology represents one of the basic methodologies under the DAKOTA toolkit [XSUSA Website]. XSUSA is designed to complete forward-based UQ only by generating randomized samples for the input parameters based

on a prior covariance matrix, and dispatches the random samples to the various processors, and finally aggregates the responses from all the model runs to calculate statistical quantities such as means and standard deviation, and correlation structure between multiple response. Although the method is pretty basic, it has pioneered the UQ work based on forward sampling in the nuclear engineering community, and has been the basis for the SAMPLER super-sequence used by the US Nuclear Regulatory Commission for criticality safety problems [Williams, 2013].

The primary handicap for XSUSA is the inability to extract the sensitivities of the responses of interest with respect to all input parameters. When applied to small-scale models with few parameters, analysis of variance type measures, similar to Sobol indices [Saltelli, 2000], can be employed to determine key dominant parameters.

7.3.3 *GMPSA*

GPMSA is a toolkit developed at the US Los Alamos National Laboratory with primary focus on parameter estimation and constructing response surfaces (i.e., surrogate model) relating the input parameters and the responses of interest. Unlike the other toolkits discussed in this section, GPMSA is the only toolkit capable of estimating modeling uncertainties. The approach employed is based on a seminal paper by [Kennedy, and O'Hagan, 2001], where modeling errors are represented by additional terms in the response surface model with additional parameters that are calibrated via fitting against measurements.

$$ y = f\left(x, \alpha\right) + \mu\left(x, \beta\right) + \varepsilon \ , $$

where the *f* term represents the response surface model and α are its associated parameters. The μ is an additional term which represents the modeling error, with functional form selected by the user. The β are additional set of parameters for characterizing the modeling errors. Based on the available experimental value for the response *y*, both α and β parameters are simultaneously calibrated. Once determined, the response surface can be used for UQ, and for mapping modeling errors to other operating conditions.

Although GPMSA provides a new capability to characterize modeling errors, its methodology is currently being debated by mathematicians and statisticians, and continues to receive many updates and improvements by many researchers from different scientific backgrounds, see [Oliver, 2014] for an excellent overview on the current challenges of the Kennedy/O'Hagan methodology. To practitioners the method remains questionable as it continues to rely on extrapolation of the response surface predictions and calibration against a limited set of experimental data which

may not necessarily capture the wide range of variations over the expected range of operation.

GPMSA allows one to perform dimensionality reduction for the responses of interest via projection onto a user-defined active subspace. The advantage of this approach is that one can build a surrogate model directly in terms of the reduced dimensions. The disadvantage is that the user has to provide the basis for the active subspace, rather than letting the toolkit calculate it automatically. For most realistic applications, the optimum active subspace is not known a priori. Moreover, with the subspace being user-defined, it is difficult to upper-bound the errors resulting from the reduction, similar to the situation described earlier with the DAKOTA toolkit.

7.3.4 PSUADE

PSUADE has been developed by the US Lawrence Livermore National Laboratory as an efficient response surface construction toolkit for expensive computational models. PSUADE is the only toolkit that has sophisticated ROM techniques used directly to reduce the computational cost required for the construction of the surrogate models. In particular, it has both gradient-free and gradient-based algorithms for rendering reduction at both the response and parameters levels, respectively. Its capabilities provide an excellent starting point for a comprehensive approach to construct ROM models and capture their associated errors.

Given its comprehensiveness, the PSUADE toolkit has been selected as the basis for the FOQUS framework, developed by the Carbon Capture Simulation Initiative [CCSI Website]. CCSI is a US DOE-funded partnership between US national laboratories and university to develop modeling and simulation tools that can accelerate the commercialization of carbon capture systems. FOQUS represents one of the important CCSI products; it is designed as an integrated framework for optimization and quantification of uncertainty and sensitivity for the CFD models associated with the CCSI project. FOQUS has the capability to develop reduced models that can be used as surrogate models to allow studying a wide range of carbon capture systems, and to allow rigorous propagation of uncertainties.

7.3.5 SAMPLER

SAMPLER is developed by the US Oak Ridge National Laboratory as a super-sequence capable of propagating nuclear data and fuel composition uncertainties throughout any of the codes under the SCALE environment [Williams, 2013]. The main UC engine for SAMPLER is the XSUSA methodology. In addition to that, SAMPLER provides a comprehensive job launching software and queueing system

developed in C++. Because of its versatility and flexibility, it allows one to seamlessly implement ROM techniques as done by CRANE.

The primary deficiencies of SAMPLER is that it is not capable of calculating the sensitivity coefficients of model responses with respect to the parameters. This makes it difficult to determine dominant sources of uncertainties. Moreover, SAMPLER is not capable of generating a surrogate model that can be used in lieu of the original SCALE models.

7.3.6 CRANE

CRANE leverages the SAMPLER super-sequence to render reduced order modeling for the various codes that can be executed via SAMPLER, e.g., KENO, NEWT, TSUNAMI, etc. In particular, CRANE provides three functionalities, a) dimensionality reduction of the input space (representing the cross-sections and number density) and the output space (representing the flux solution); b) construction of surrogate model that can be used in lieu of the radiation transport code employed; and c) calculations of sensitivity coefficients for all model responses.

With regard to the first functionality, both gradient-free (which requires only the forward transport solver) and gradient-based (which requires the adjoint solver) ROM techniques are implemented to help reduce the dimensionality of the parameter space. For typical cell lattice calculations, the gradient-free approach can help reduce the dimensionality of the few-group cross-sections that are typically functionalized in terms of large number of core conditions. The gradient-based approach helps reduce the dimensionality of the multi-group cross-sections, also expected to be very large.

With regard to the construction of surrogate models, CRANE employs a novel physics-based approach which allows the analyst construct upper-bounds on the surrogate model predictions, as compared to the original physics model. Recall that the state-of-the-art surrogate modeling techniques are essentially function approximation techniques which employ a predetermined response surface to approximate (via fitting) the behavior of the physics model over the expected range of parameter variations. CRANE however employs the physics model to directly determine the form of the response surface in a manner that precludes the need for fitting. This is possible via the use of exact-to-precision generalized perturbation theory, or EₚGPT for short.

EₚGPT is an ROM rendition of conventional generalized perturbation theory (GPT), which has been receiving continuous developments in the nuclear community since the 1940s. GPT is premised on the fact that via the solution of an

adjoint function, one can estimate response variations for any general parameter variations without having to re-execute the forward model. The primary challenge is that this premise is only applicable for small linear variations. When the size of the perturbation is large to excite the nonlinear modes of the solution, the application of GPT becomes computationally taxing as now higher order adjoint functions must be calculated which quickly becomes even more expensive than executing the forward model.

By noticing that radiation transport models are inherently reducible, E$_P$GPT employs ROM techniques to recast GPT into a form that is amenable for routine execution with both linear and nonlinear variations. The CRANE modules represents an automation of the E$_P$GPT methodology in the SCALE code package. Appendix E may be consulted for details on the E$_P$GPT theory and recent application in CRANE.

7.3.7 RAVEN

RAVEN is a software tool developed at the US Idaho National Laboratory (INL). Its primary function is to act as the control logic for RELAP-7 thermal-hydraulic code, also currently being developed at INL. Although the specific models RAVEN employs are not of direct interest to us here, RAVEN has been designed as a multi-purpose software framework that allows one dispatch the models for a wide range of functionalities. For example, it contains comprehensive tools for the processing of input/output data streams, and the automated application of dimensionality reduction techniques. Its general pluggable design allows one to work easily with different solvers/models in order to allow for a seamless experimentation with different solvers/models for multi-physics applications. Another advantage of the RAVEN tool is that all its process algorithms for UC and ROM are available via open-source licenses. Given its versatility and flexibility, RAVEN provides an excellent starting point for the proposed UCF.

7.4 Observations

Following this review, we would like to summarize the following key points. The construction of the UCF will require acquiring or developing basic UC/ROM algorithms, customizing the algorithms for CANDU-specific applications, constructing a computational environment for the handling of the model-to-model interfaces, the dispatching of the jobs on parallel environments, and the manipulation/storage of the data streams from the various models.

After reviewing the available toolkits, it is safe to say that most of the UCF components have been demonstrated and are already available in the public domain, leaving the bulk of the remaining work to focus on the customization of the algorithms for CANDU-specific

applications. For example, both the FOQUS and RAVEN environment contain comprehensive job management and launching software, adapted to parallel computing environments, which may be directly leveraged for the UCF project. They also both contain the basic UC/ROM algorithms that can be easily acquired, and further developed. Efficient ways to manipulate and store the voluminous UC data already exist, and have been widely practiced inside and outside the nuclear engineering community.

8.0 UCF PILOT STUDY

In support of investigating the UCF feasibility, a pilot study is conducted here to demonstrate the standard application of UC algorithms, i.e., UQ^p and UQ^m, and provide an initial assessment of the relative importance of the various sources of uncertainties. The study will be limited to cell lattice and core wide calculations only, focusing on the propagation of representative parameter of modeling uncertainty sources.

The model employed represents a transient simulation of large break loss of coolant accident (LBLOCA), based on a recent study conducted by CNSC, which is fully reported in [Serghiuta, 2014]. The core simulator model for this accident is based on a two-group cross-section nodal diffusion theory model of the core employing the NESTLE-C code [NESTLE, 2003]. The BE (i.e., reference) cross-sections used in NESTLE-C have been previously generated using a CANDU version of the HELIOS code [Casal, 1992].

The core analyzed is a standard CANDU core containing 380 channels, each channel consists of 12 bundles, with each bundle containing 37 fuel elements. The bundle model configurations are shown in Fig. 8.1 and Table 8.1. The layout in Fig. 8.1 is based on a NEWT model of the ORNL's SCALE package [SCALE, 2011]. Our objective is to evaluate the effect of the two-group cross-section uncertainties on key core attributes, including peak core reactivity, peak bundle enthalpy, peak bundle power, and peak bundle energy.
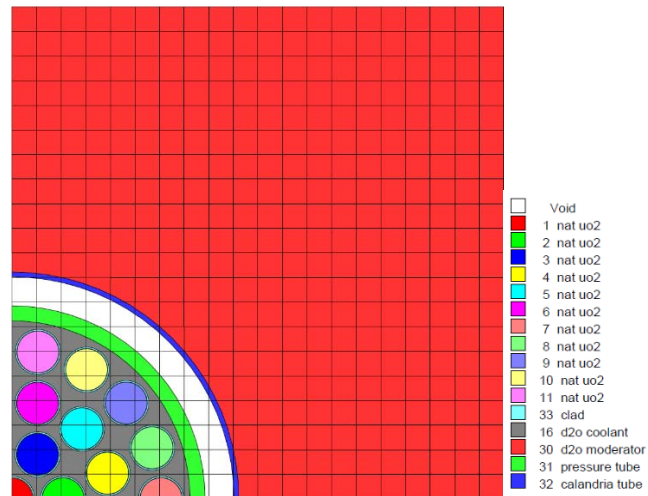
Fig. 8.1 NEWT Model of ¼ CANDU BUNDLE

For the propagation of uncertainties, different neutronics codes are employed for the cell lattice calculations in order to study the impact of different modeling assumptions such as the use of multi-group cross-sections and deterministic transport models. The codes employed are depicted in Fig. 8.2. The starting point for the UC analysis is the 238-group cross-sections uncertainties, available in the form of a covariance matrix that is pre-generated by the SCALE code package. The covariance matrix contains uncertainties for all cross-sections of relevance

to CANDU calculations, including fission, absorption and scattering cross-sections for coolant, clad, and fuel materials.

Table 8.1 Lattice Configurations

	Composition	Inner Radius (cm)	Outer Radius (cm)
Fuel	Natural UO_2: 10.39475 g/cc	0.605295	
Gap	Void		0.610295
Cladding	Zirc-II: 6.3918 g/cc		0.652183
Coolant	Heavy water 98.39 wt%, 0.80623 g/cc		5.1689
Pressure tube	Zr-Nb Alloy: 6.5041 g/cc		5.6032
Calandria tube	Zirc-II: 6.4003 g/cc	6.4478	6.5875
Moderator	Heavy water 99.935 wt%, 1.08579 g/cc	Cuboid length:	14.2875

NEWT is a 2D deterministic multi-group S_N transport solver, and KENO is a 3D Monte Carlo model capable of using both continuous and multi-group cross-sections. By comparing the predictions of these codes for the few-group cross-sections, it is possible to provide an initial assessment of the various sources of modeling and parameter uncertainties. In particular, we investigate the impact of two sources of uncertainties, parameters and modeling uncertainties, described in the following two subsections.
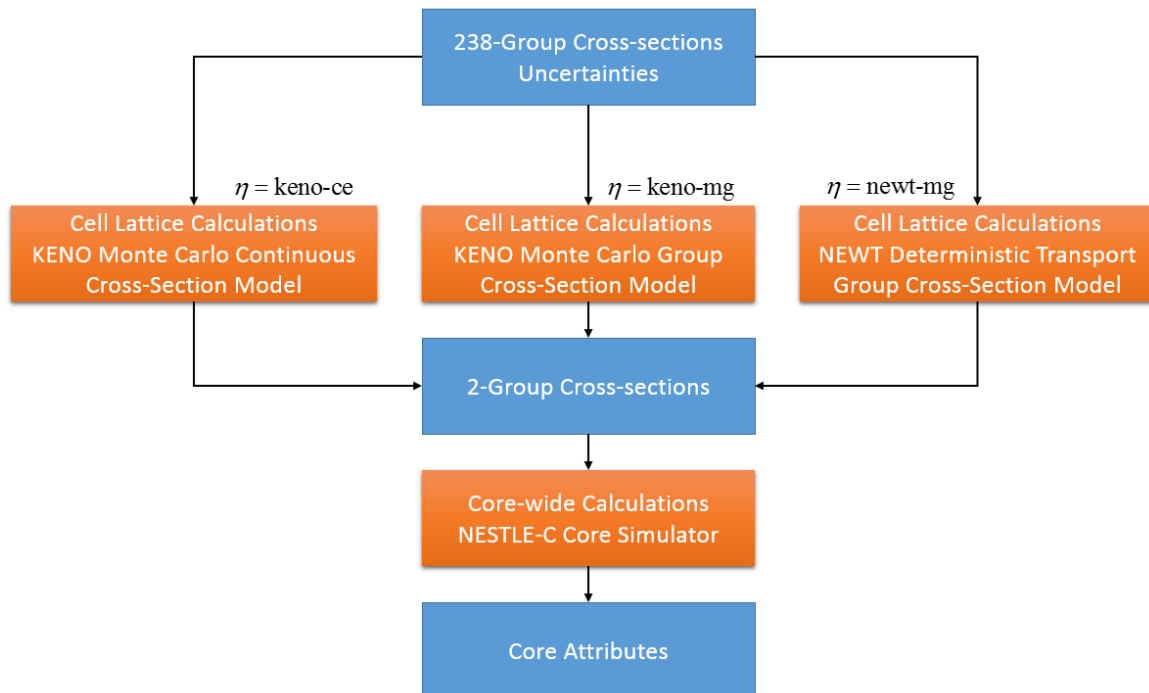
Fig. 8.2 Codes Layout for UC Pilot Study

8.1 Propagation of Cell Lattice Modeling Uncertainties

Recalling from section 3.3, there are multiple sources of modeling uncertainties in cell lattice calculations; we focus here on the uncertainties resulting from the deterministic transport solver, and the multi-group cross-sections approximations only. As discussed in section 4.0, one must have access to a high fidelity model that can explicitly characterize the modeling sources of uncertainties using the UQ^m approach, which is demonstrated below. We employ three different solvers:

a. 2D NEWT S_N multi-group transport solver, denoted by decision variable $\eta = \text{newt-mg}$
b. 3D KENO continuous cross-section Monte Carlo Model, denoted by $\eta = \text{keno-ce}$
c. 3D KENO multi-group Monte Carlo Model, denoted by $\eta = \text{keno-mg}$

The UQ^m algorithm is applied to the following combinations of decision variables:

a. $\eta = \text{newt}$ vs. $\eta = \text{keno-mg}$: This allows one to quantify the errors resulting from the transport solution only, i.e., not taking into account the effect of multi-group cross-section approximations. The higher fidelity model here is assumed to be the $\eta = \text{keno-mg}$ since it models the neutron transport using Monte Carlo directly with enough number of particle histories to ensure acceptable convergence.
b. $\eta = \text{keno-ce}$ vs. $\eta = \text{keno-mg}$: This allows one to quantify the uncertainties resulting from the use of multi-group cross-sections. The higher fidelity model here is $\eta = \text{keno-ce}$.

Each of these applications allows one to generate a bias that measures the errors resulting from the respective modeling assumptions. To turn these biases into PDF distributions, one must investigate the impact of other control parameters, such as fuel composition, fuel temperature, coolant density, reactivity devices insertion, etc. In this pilot study, we focus on the impact of fuel composition only. To get representative compositions over the life of the fuel in the core, an initial depletion of the cell lattice is completed to an average end-of-life burnup, taken to be 7.0 GWD/MTU, and the composition is recorded at 11 points during depletion.

The UQ^m algorithm is implemented as follows:

1. Initial depletion of CANDU cell lattice (Fig. 8.1) to 7.0 GWD/MTU with burnup step of 0.7 GWD/MTU.
2. DO $i = 1, \ldots, K$, where K=11 looping over the 11 different compositions.
3. Set up a steady state CANDU cell lattice model with composition # i
4. Calculate the two-group cross-sections using the three modeling decisions, $\eta = \text{keno-mg}$, $\eta = \text{keno-ce}$, and $\eta = \text{newt}$. This is achieved by running KENO twice, (in multi-group and continuous energy modes), and NEWT once.

5. Calculate the bias for the following two cases
   - $\eta = $ newt vs. $\eta = $ keno-mg *"transport model errors"*
   - $\eta = $ keno-ce vs. $\eta = $ keno-mg *"multi-group errors"*
6. END DO
7. Plot the biases for the two cases over the range of compositions.

The results of this study are shown in Fig. 8.2 and Fig. 8.3 for the two-group cross-sections of U^{235} and U^{238}, respectively. The following notations are used: fast-ab235 denotes the fast absorption cross-section of U235, and thermal-f238 denotes the thermal fission cross-section of U^{238}. In all graphs, the blue graph denotes the discrepancies in the two-group cross-sections calculated as follows:

$$bias(\text{blue}) = \frac{\sigma(\eta = \text{newt-mg}) - \sigma(\eta = \text{keno-mg})}{\sigma(\eta = \text{keno-mg})}$$

which describes the modeling errors resulting from the transport solver. The orange graph is:

$$bias(\text{orange}) = \frac{\sigma(\eta = \text{keno-ce}) - \sigma(\eta = \text{keno-mg})}{\sigma(\eta = \text{keno-mg})}$$

which describes the modeling errors from the multi-group cross-sections approximation.

Detailed investigation of these results can provide great insight into the sources of uncertainties and their relative impact on the quantities of interest. While this is beyond the scope of the current study, we make few initial observations to complete the presentation:

- The modeling errors are in the order of few percent, which is the same order of magnitude of the cross-section uncertainties (shown in the next section).
- The absolute magnitude of modeling errors resulting from the use of multi-group cross-sections are consistently higher than the errors resulting from the transport solution for the thermal cross-sections. This behavior is reversed for the fast cross-sections.
- The magnitude of the errors for the fast cross-sections are two to three times higher than the errors in the thermal cross-sections.
- Some errors show possible correlations with fuel composition, while others do not.

These results indicate that to render a comprehensive evaluation of all uncertainty sources, one must include the effects of modeling assumptions and approximations, such as the use of multi-group cross-sections, and the use of deterministic transport solver. The initial results indicate that their magnitude could be in the same ball park as the uncertainties resulting from nuclear data.
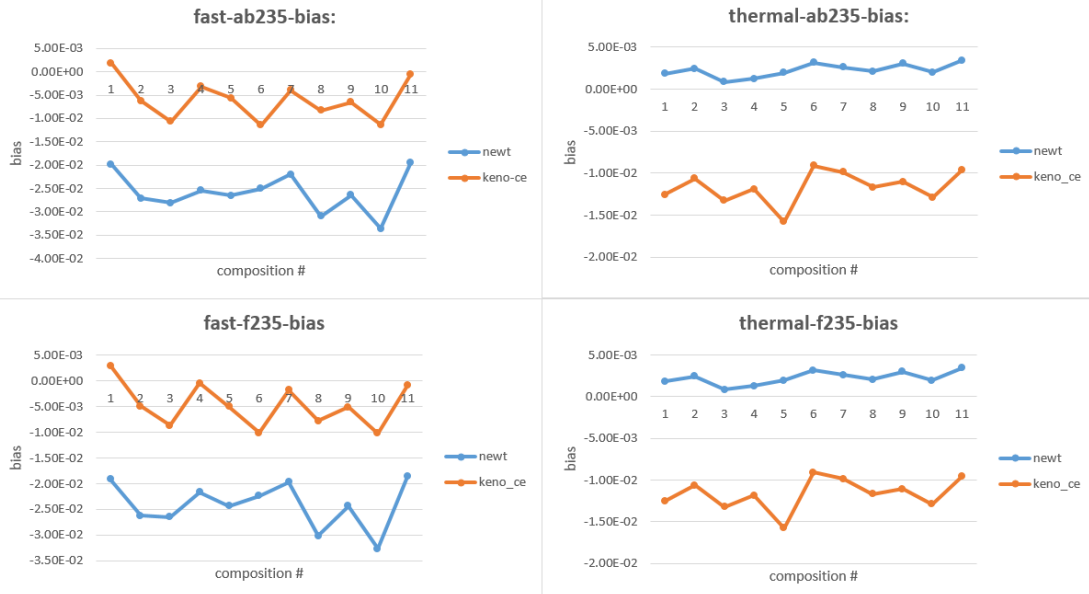
Fig. 8.3 U^{235} Thermal and Fast Absorption and Fission Cross-sections
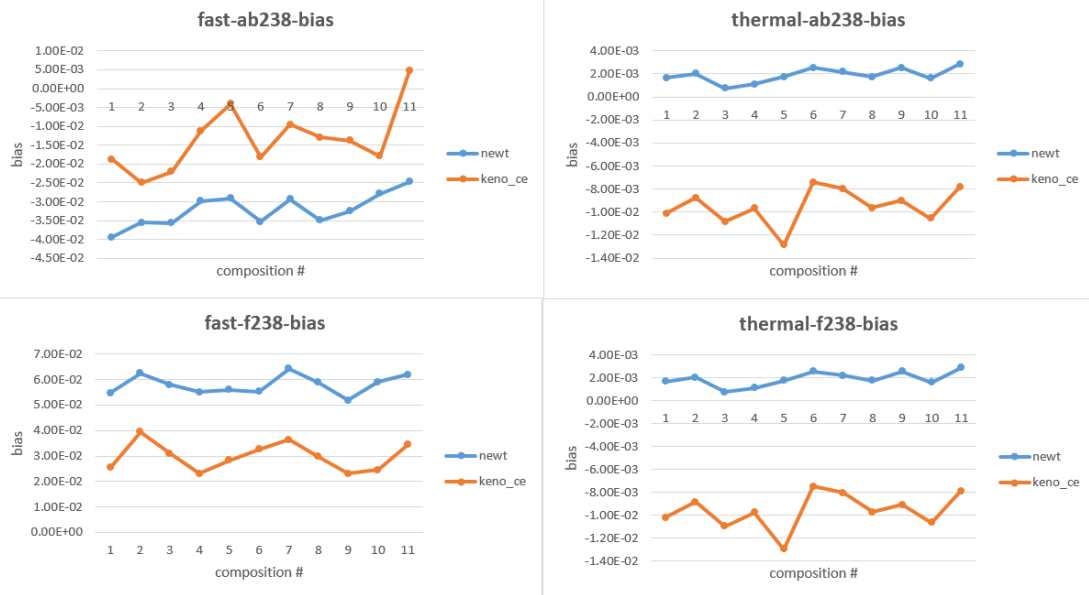
Fig. 8.4 U^{238} Thermal and Fast Absorption and Fission Cross-sections

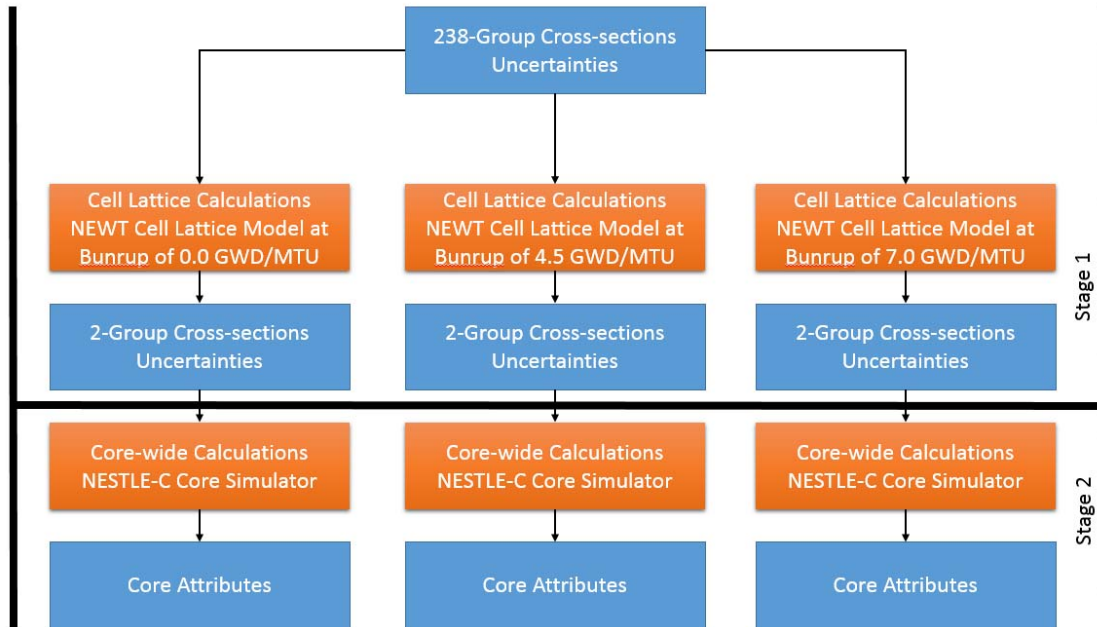## 8.2 Propagation of multi-group cross-section uncertainties

The second part of the pilot study deals solely with uncertainties resulting from the multi-group cross-section uncertainties. As discussed in section 3.2, the basic nuclear data uncertainties, resulting from differential cross-section measurements and fitting to nuclear reaction models, are propagated to a multi-group format which represents the basis for cell lattice calculations. Our goal is to propagate the multi-group cross-section uncertainties to core attributes of interest, such as peak core reactivity, bundle enthalpy rise, etc. This is achieved over two stages as shown in Fig. 8.5. The first stage propagates the multi-group cross-sections uncertainties through cell lattice calculations to the few-group cross-sections, and the second stage further propagates the few-group uncertainties through core-wide calculations to the core attributes of interest.

In the first stage, the CRANE super-sequence (cf., section 7.3.6 and Appendix E) is employed to propagate the 238 group cross-section uncertainties to the two-group cross-sections uncertainties using a representative CANDU cell lattice as shown in Fig. 8.1. CRANE is based on the SAMPLER super-sequence of SCALE with an imbedded ROM capability to reduce the size of the uncertainty space. By proper selection of the size of the active subspace, cf. section 4.2, one can render agreement between SAMPLER and CRANE results within machine precision. Preliminary verification results on the CRANE's predictions against the SAMPLER code are reported elsewhere [Wang, 2013]. The result of CRANE application is the covariance matrix of the two-group cross-sections. The two-group cross-sections considered in this pilot study are the transport, fission, and absorption cross-sections.

In the second stage, the two-group cross-section uncertainties are propagated through the NESTLE-C code using a sampling-based UQ^p approach. In this stage, some simplifications were introduced to limit the scope of the study. In the full UCF application, the two-group cross-sections uncertainties should be functionalized in a similar form to the reference cross-sections, i.e., in terms of core conditions such as fuel temperature, coolant density, soluble boron content, etc. In this limited-scope study however, we will assume that the uncertainties are independent of all core conditions. They are applied as multipliers to the NESTLE-C's cross-sections after they have been interpolated. This implies a single multiplier per cross-section type over the wide range of conditions in the core.

In reality, one must consider the impact of various core conditions on this multiplier. To relax this assumption a bit, we consider the fuel depletion impact on the propagated two-group uncertainties. This is done by repeating the UC study with different two-group uncertainties, evaluated using cell lattice calculations at three different fuel compositions. The selected compositions represent different burnup points over the life of the fuel bundle in the core, which is assumed to accrue an average discharge burnup of 7.0 GWD/MTU. The multi-group uncertainties are propagated using three cell lattice models, one

representing fresh fuel, i.e., 0.0 GWD/MTU, an intermediate burnup of 4.5 GWD/MTU, and discharged fuel at 7.0 GWD/MTU. The two-group uncertainties generated in each case are subsequently propagated through NESTLE-C code. This provides an initial estimate off the impact of burnup on the propagated uncertainties. Representative results of the two-group cross-section uncertainties generated with CRANE are shown in Table 8.2, and the corresponding NESTLE-C core attributes uncertainties are in Table 8.3. The uncertainties are described in relative units of standard deviation divided by the mean value (which is the best estimate value used in reference calculations).

Fig. 8.5 Propagation of Multi-group Cross-Sections Uncertainties

Table 8.2 Two-Group Cross-Section Uncertainties

Cross-section	Zero Burnup σ/μ (%)	Intermediate Burnup σ/μ (%)	Discharge Burnup σ/μ (%)
Burnup (GWD/MTU) at which cross-sections uncertainties are generated	0.0	4.5	7.0
Fast Absorption	0.70	0.64	0.61
Thermal Absorption	0.51	0.47	0.49
Nu-Fast Fission	2.64	2.85	2.71
Nu-Thermal Fission	0.61	0.91	0.99
Kappa-Fast-Fission	2.32	2.49	2.37
Kappa-Thermal-Fission	0.55	0.64	0.69
Fast Transport	0.98	0.98	0.99
Thermal Transport	1.37	1.46	1.37

Table 8.3 Core Attributes Uncertainties

Response	Zero Bunrup	Intermediate Bunrup	Discharge Burnup
Burnup (GWD/MTU) at which cross-sections uncertainties are generated	0.0	4.5	7.0
Peak Core Reactivity (*mk*)*	0.09	0.52	0.07
Peak Bundle Energy (%)	3.40	2.73	4.37
Peak Bundle Power (%)	0.93	1.24	0.82
Peak Reactor Power (%)	0.93	1.24	0.82

Peak Reactivity Reference Value = 3.2 mk

In Table 8.2, kappa is the energy released per fission. In the two-group macroscopic cross-section model employed by NESTLE, the product Kappa-Thermal-Fission defines an effective value for the kappa times the fission cross-section over the entire inventory of nuclides present in the fuel, reflecting the fact that each isotope has a slightly different value for fission energy release and significantly different fission cross-section.

The results in Table 8.3 may be interpreted as follows. The 'Intermediate Bunrup' column implies that all fuel bundles in the core are assumed to have the same uncertainties (given by the 'Intermediate Bunrup' column in Table 8.2) that are generated using a cell lattice model with an intermediate burnup of 4.5 GWD/MTU. Clearly, this assumption is not valid since the fuel bundles in the core have a wide range of bunrups; however by repeating the study with three different burnups, one can get a preliminary idea about whether or not the burnup dependence is important.

9.0 UCF DELIVERABLES/MILESTONES AND SCHEDULE

The milestones/deliverables and proposed schedule are shown below, to be carried out over 5 years by 4 junior researchers (e.g., PhD students), 1 senior researcher, and the project director.

#	Milestone/Deliverable	FY1		FY2		FY3		FY4		FY5	
1	**Initial UCF Setup**	G	G								
1.1	Computational environment specifications	Y									
1.2	Code specifications	Y									
1.3	Buffer codes design for various selected codes	Y									
1.4	Initial bank of CANDU cell lattice and core models	Y	Y								
1.5	Enable parallelization of computational environment	Y	Y								
2	Implementation of UC Algorithms	G	G	G							
2.1	Setup of data server for data storage/manipulation		Y								
2.2	Automation & initial testing of UQ^p algorithm	Y	Y								
2.3	Automation & initial testing of UQ^m algorithm		Y	Y							
2.4	Automation & initial testing of ROM algorithms		Y								
2.5	Overall UCF testing with simplified models		Y								
3	**ENDF Pointwise Cross-sections Generation**	G	G	G	G						
3.1	Propagation of uncertainties from differential measurements	Y	Y								
3.2	Propagation of uncertainties from nuclear reaction models		Y								
3.3	Propagation of overall uncertainties			Y							
3.4	Verification of propagated uncertainties			Y							
4	**Multi-group Cross-section Generation**	G	G	G	G	G	G				
4.1	Propagation of continuous cross-sections uncertainties		Y								
4.2	Propagation of uncertainties from assumed flux shaped		Y	Y							
4.3	Propagation of overall uncertainties			Y							
4.4	Identification and ranking of dominant uncertainties			Y	Y						
4.5	Verification of propagated uncertainties				Y						
5	**Cell Lattice Calculations**		G	G	G	G	G				
5.1	Propagation of multi-group cross-sections uncertainties		Y								
5.2	Propagation of uncertainties from multi-group transport solver		Y	Y							
5.3	Propagation of uncertainties from boundary conditions			Y	Y						
5.4	Propagation of overall uncertainties				Y						
5.5	Variance reduction techniques		Y	Y							
5.6	Identification and raking of dominant uncertainties				Y						
5.7	Verification of propagated uncertainties				Y						
6	**Downstream Core-Wide Calculations**					G	G	G	G		
6.1	Propagation of few-group cross-sections					Y	Y				
6.2	Propagation of uncertainties from diffusion solver					Y	Y				
6.3	Propagation of overall uncertainties for steady state conditions						Y				
6.4	Propagation of overall uncertainties for transient conditions						Y	Y			
6.5	Identification and ranking of dominant uncertainties							Y			
6.6	Verification of propagated uncertainties							Y			
7	**Delivery of the Framework**								G	G	G
7.1	Full UC demonstration from ENDF to core attributes								Y	Y	
7.2	Documentation: programmer and user manuals									Y	
7.3	Beta users testing and feedback									Y	Y
7.4	Feedback from CNSC and project finalization										Y

10.0 APPENDIX A: UC DEFINITIONS

Whenever one makes a prediction about the behavior of an engineering system, one must provide a quantitative measure of his/her own confidence about such prediction, referred to as the uncertainty of the prediction. This is because all predictions are based on some approximate physics models, i.e., mathematically-driven hypotheses that describe how one believes the system behaves under influence of given initial and boundary conditions. All models contain errors and/or uncertainties, resulting from simplifying assumptions and lack of full knowledge about the behavior of the system. Understanding and quantifying these uncertainties is necessary before the model predictions can be used to make decisions about the system design, operation, and safety. This is especially relevant when dealing with high consequence systems such as nuclear reactors.

10.1   BASIC DEFINITION OF PROBABILITY

Let the model response be described by a variable y. The uncertainty in y is described by a probability density function (PDF) $p(y)$ such as:

$$P\left(y_1 \leq y \leq y_2\right) = \int_{y_1}^{y_2} p(y)dy \qquad (A.\ 1)$$

is the probability of y lying between two fixed values y_1 and y_2. Understanding this definition is very important as it determines the way in which the uncertainties are processed. The term 'probability' as used here has *two* meanings that are fundamentally different from one another. The *first definition* employs the probability to estimate the frequency of finding y in a specified interval. The implication here is that y is a stochastic variable, i.e., it is expected to attain a wide range of values due to the stochastic nature of the physical phenomena. The uncertainty here is a declaration of the fact that the actual value for y at a given instant/condition is never known, only the frequency of occurrence in a given interval can be estimated. The width of the PDF, as measured in units of standard deviation, provides a quantitative measure of the spread of the values that y can attain.

The associated model that predicts y as a stochastic variable is often referred to as a stochastic or probabilistic model. This is as opposed to a deterministic model which generates a single value for the response. One can argue that a deterministic model is the limit of a probabilistic model when the standard deviation goes to zero, i.e., the PDF becomes a delta function.

The neutron interaction with matter provides an excellent example to this situation (Fig. A.1). When a given neutron bombards a specific nucleus, it is not known deterministically whether an interaction will occur. However when considering a high flux of neutrons impinging on a macroscopic number of nuclei, one can determine the average number of interactions to very high accuracy, albeit the process is stochastic in nature. This implies

that the spread of the PDF describing the total number of interactions is very small compared to the average number of interactions.
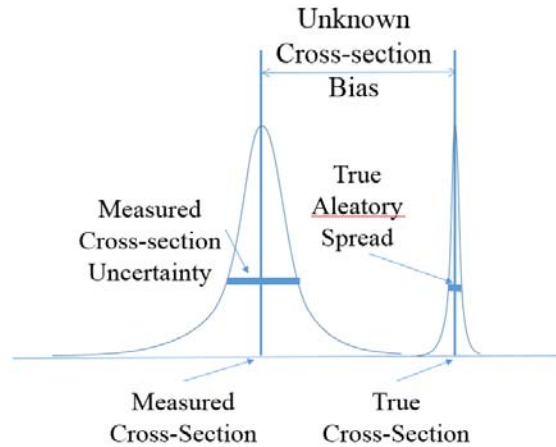
Fig. A.1 Cross-sections Aleatory vs. Epistemic Description

The *second definition* of probability is concerned with deterministic modeling when the true value of the parameter is single-valued but unknown. In this case, the probability as defined in Eq. (A.1) provides a quantitative measure of one's confidence about the true value for *y* lying in the given interval.

One's confidence? What does that really mean? To understand the significance of this definition of probability, consider a simple example, where a physicist asks *two* experimentalists to determine the value of a given physical parameter, say *v* of U^{235}. How can the physicist assess/use the values returned by the two experimentalists, expected to be different? If each experimentalist simply returns a single value for *v*, more information will be needed to determine which value is more credible. To obtain such information, the physicist will have to ask each experimentalist how their experiments were conducted. Qualitative comparison of the experiments is of course possible, however in most realistic situations, a quantitative measure would be desirable especially when the details of the experiments overwhelm simple qualitative reasoning. Therefore, some measure of confidence should be available. A convenient measure is a numerical value between 0 and 1, where 0 means no confidence and 1 means absolute confidence. This measure is called probability as well. The common terminology sometimes makes it difficult to distinguish between the two definitions for probability. More interesting, the same probability can be treated in two different manners depending on the outcome of the analysis. For example, cross-sections uncertainty can be viewed as aleatory and epistemic. As described earlier, the nature of neutrons interaction with matter is purely stochastic implying that the cross-section is a stochastic variable, however as noted earlier its spread is very small as depicted in Fig. A.1. When measuring cross-sections, one is attempting to measure this true distribution. Due to imprecise measurements, another distribution is measured, one with a

mean value that is far from the true mean value of the true distribution, and a much wider standard deviation as compared to the standard deviation (i.e., spread) of the true distribution. The spread of the measured distribution is an indication of the measurement uncertainty. Because of this, cross-sections are typically treated as epistemic uncertainties by cross-section adjustment techniques (i.e., model calibration) which attempt to improve knowledge about the true cross-section values using integral experiments measurements, such as critical eigenvalue, and flux measurements [Cacuci, 2003], [Kennedy, 2001].

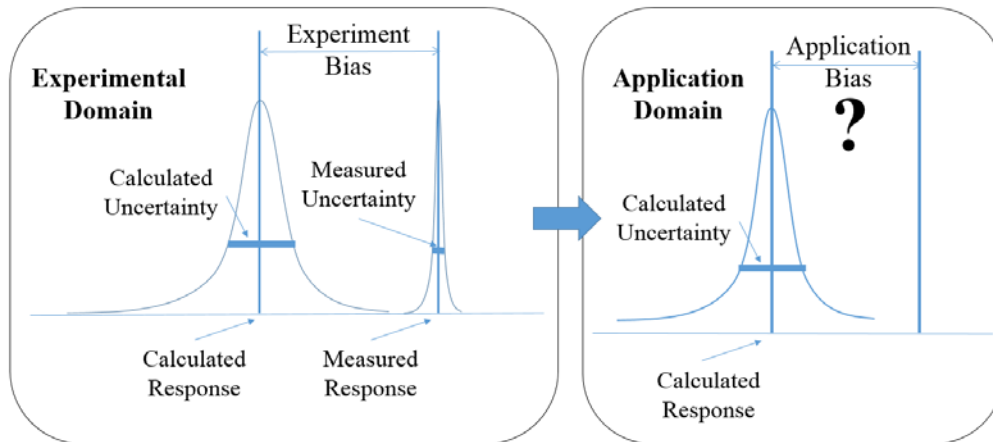## 10.2   UNCERTAINTY CHARACTERIZATION

To describe the role of uncertainty characterization (UC), one must first describe model validation. Model validation refers to all activities conducted to establish, based on a scientific basis, the degree to which the model represents the real physical phenomena being simulation over the range of expected operating conditions. Notice that the range of applicability is an important component of model validation, because no model is perfect enough to capture everything for all possible conditions. It is therefore important to recognize that model validation only validates a model over a given range of applicability. Naturally, as the model becomes more accurate and more comprehensive, its range of applicability increases, which is the premise of the various advanced modeling and simulation initiatives currently underway at many government laboratories and academic institutions.

UC's role in model validation is to provide the scientific quantitative metric by which the validation domain, i.e., domain of model applicability, can be mapped. In particular, UC deals with the following situation: given a computer model that is intended to simulate the behavior of the reactor system, and given a number of small-scale experiments, conducted to support the validation of the computer model, how can one answer the following questions?

(a) What are the key sources of uncertainties in the simulation?
(b) How can one propagate and prioritize these sources of uncertainties to determine the uncertainties of the responses of interest, i.e., model responses used to judge performance of the reactor system, and identify the key contributors to their propagated uncertainties?
(c) How can the biases (discrepancies between measurement and predictions from the small-scale experiments, denoted by the experimental domain) be scaled to the full-scale reactor conditions (denoted by the application domain) expected during normal and off-normal operation? And what are their uncertainties?

To answer these questions, consider a simple example in which one is interested in validating a two-group diffusion theory-based model for the calculation of pin-powers during steady state conditions. Assume that the maximum pin power in the core is designated as the response of interest. To help validate this model, one must construct an experiment to check whether the computer model provides adequate prediction of reality. Assume that the reactor employs a single fuel enrichment of 3.0 w/o in U^{235}. The small-scale experiment comprises a 3x3 array of fuel pins with the same enrichment and geometry details, i.e., fuel to moderator ratio, of the

reactor core fuel. The experiment is repeated several times, and the bias between the measured and calculated value is calculated "denoted as the experiment bias in Fig. A.2" on the average at steady state conditions with a measurement uncertainty "denoted as the measured uncertainty in Fig. A.2". The results of the experiments are typically described using a PDF as shown in Fig. A.2.

Fig. A.2 Uncertainty Characterization for Experimental and Application Domain

Regarding the first UC question, the sources of uncertainties in the simulation originate from the numerical iterative strategy employed, the modeling assumptions, and the parameters input to the model. Numerical uncertainties originate due to the conversion of the continuous mathematical model into discretized equations than can be digitized on a computer. The modeling uncertainties originate from the use of diffusion theory and multi-group approximations instead of a more accurate representation such as Monte Carlo simulation with continuous cross-section representation. Parameter uncertainties originate from the basic nuclear data, such as the diffusion coefficient, absorption and fission cross-sections, etc., because all cross-sections are measured using differential experiments with their own uncertainties. Other sources of uncertainties also include technological parameters such as geometry and composition information, all subject to manufacturing tolerances. If this question is not answered thoroughly, discrepancies between measurements and predictions may either remain unexplained, or more often they are compensated for using model calibration techniques, where the model is fitted against measurements.

Regarding the second question, one must be able to propagate all sources of uncertainties throughout the simulation "denoted by calculated uncertainty in Fig. A.2" and following that employ a quantitative measure that assigns a numerical value that describes the importance of each source of uncertainty. Based on these numerical values, one can order the various sources of uncertainties from high to low. This will provide valuable information to the analyst and the experimentalists, because they can now refine their analysis/experiments to reduce these key sources of uncertainties. In our example, answering this question implies the estimation of

maximum pin power uncertainty resulting from the various sources of identified uncertainties. The uncertainty can be described using a PDF as shown in Fig. A.2. The mean value represents the BE for the response. The standard deviation (std) is used as a measure of the uncertainty for the BE value. Clearly, a small std implies high confidence in the BE value, and vice versa, a high std implies little confidence therein. In our example the std for the maximum pin power is indicated in the figure as by the thick blue line labeled as calculated uncertainty".

Regarding the third UC question, can one determine what the bias and its uncertainty should be for the reactor conditions during steady state and transient scenarios, referred to as "the application"? Will the bias and bias uncertainty simply be the same values determined from the small-scale experiments, referred to as "the experimental domain", or should they be scaled somehow? What if the bias uncertainty is found to be higher than the uncertainties propagated through the model describing the reactor application? Should one discard the bias information or the propagated uncertainties? Typically the calculated uncertainty for the application is found to be very large, which would negatively impact the economy of the reactor. To help reduce this uncertainty, the experimental biases are used to determine an application bias. This is one of the key requirements for a successful validation exercise.

The challenges facing UC methods may be described as follows:

- Computational challenges: propagating all sources of uncertainties and prioritizing their key sources is a computationally daunting task using complex reactor analysis models. To address this challenge, algorithmic development of more efficient techniques is needed to help realize the benefits of UC methods.
- Methodology challenges: scaling the biases between the experimental and application domain is extremely difficult in the general case. There currently exists no unified theory on how to do that. Highly customized strategies driven by expert judgment are employed in the nuclear industry to justify the scaling of biases and the identification of the validation domain. Basic research in this area is currently being pursued by many researchers from a wide range of scientific backgrounds.

11.0   APPENDIX B: SENSITIVITY ANALYSIS

Sensitivity analysis (SA) is concerned with the following question: how the responses variations could be apportioned to the input parameters perturbations [Saltelli, 2000]. The results of SA have many important utilities. For example, results could be used to render reduction in the parameters space which could be used to construct an ROM model. Also, the results could be used to determine the most influential parameters on responses of interest which proves useful when designing or altering the design to reach a certain objective. Additionally, SA results could be used to propagate parameters uncertainties and identify the key parameters contributing to responses uncertainties. Also, SA may be used to perform data assimilation (also known as data adjustment of model calibration [Kennedy, 2001]) in order to improve model's predictions.

SA methods may be classified in two different manners: according to their range of application or the algorithmic approach employed (see [Bang, 2012] and the references within). First, according to the range of applicability, they may be classified in terms of global and local methods. Global methods are concerned with finding an approximate relationship between the responses and input parameters that is globally accurate over all possible parameters variations. Local methods however only focus on a small region around nominal parameters values. Global methods attempt to find all unique features such as maxima and minima, turning points, etc. Local methods however calculate local derivatives around some nominal parameters values up to an order specified by the user.

According to algorithmic approaches, SA methods may be classified into forward (sometimes denoted sampling) and adjoint (also known as variational) methods. Forward (sampling) methods employ the forward model only which allows for non-intrusive implementation; whereas adjoint methods require the construction of the adjoint model which represents the dual of the forward model. In contrast to forward methods, adjoint methods allow one to directly calculate the response variations resulting from parameters perturbations without explicitly calculating state (i.e., flux) variations. This however requires intrusive access to the model's equations to allow for the construction of the adjoint model. In general, the forward method is more suitable for nonlinear models with few input parameters and many responses, while the adjoint method suits linear or quasi-linear models with many input parameters and few responses. The reasons are described in the following discussion.

Global sampling-based methods employ the forward model to randomly sample all input parameters from their prior probability density functions (PDF). After each run, responses deviations from reference values are recorded and the procedure is repeated with different random samples until a reliable estimate of responses PDFs is obtained. This method is advantageous because of its simplicity and ability to obtain detailed (i.e., all moments) PDFs for all responses. This is primarily important for a general nonlinear model and general input parameters PDFs where the responses PDFs are expected to deviate considerably from the

Gaussian shape (often employed to describe PDFs when only first and second moments are available). The disadvantage is that sensitivity information is more difficult to infer (often sought via a response surface or an analysis of variance approach), and the number of model executions can be too large to render the approach practical for high dimensional models. When the number of input parameters is sufficiently small, one could exhaustively sample the input space to identify important sensitivities, especially when the models are highly nonlinear. With the number of input parameters numbering is large (i.e., greater than 10), the search algorithm becomes computationally prohibitive. Computational scientists have devised numerous strategies to limit the number of samples with considerable success [Adams, 2009, Ghanem, 1991, Xiu, 2010]. However, the state-of-the-art techniques are still limited to models with small number of input parameters, i.e. $n \sim 10$, for complex multi-physics models.

Global variational-based methods employ a rigorous mathematical framework to determine all the unique features of the model in the combined phase space formed by the model parameters, state dependent variables, and adjoint variables [Cacuci, 2003]. The identified critical points are then further analyzed using local sensitivity analysis methods. The framework employs a functional analysis approach to forming a Lagrange functional of system response of interest, the set of linear or nonlinear equations relating model's input parameters to system dependent variables, and the set of equality and/or inequality constraints used to delimit parameters value. The Lagrange functional is then minimized using Kuhn-Tucker necessary conditions to identify the model's critical points. While being genuinely effective in identifying all unique features of the model, the computational requirements to solve the minimization problem is expensive; thereby limiting its applicability to models with few parameters. Although, the software requirements of these methods are extremely taxing since they require intensive code modifications for their implementation.

Alternatively, the local variational methods trade obtaining detailed responses PDFs for achieving computational efficiency and obtaining detailed sensitivity information. In their most practical implementation, the model is linearized and only first order derivatives of responses with respect to input parameters are determined which requires a single adjoint model execution per response. With this information only, the responses PDFs are approximated as linear combinations of the input parameters PDFs. If the input parameters uncertainties are described by Gaussian PDFs, the responses will also have Gaussian PDFs. If the model is nonlinear however, responses PDFs are expected to deviate from the Gaussian shape. The advantage is that sensitivity information can help identify influential input parameters responsible for propagated uncertainties. The disadvantage is that if the model deviates considerably from being linear, this approach will poorly approximate the responses PDFs, especially towards the tail of the distribution which are typically associated with failure events.

To identify higher order effects, the mathematical framework for variational methods has been extended to nonlinear models [Gandini, 1981]. For example, to characterize second order

effects, an extra execution of the adjoint model is required for each input parameter in order to characterize its second order derivatives which capture its interaction with other parameters [Greenspan, 1978]. Clearly, for a high dimensional model, this approach becomes computationally prohibitive. Moreover, because the computational cost is proportional to the number of responses, unlike forward methods, the adjoint approach becomes less effective for models with many responses.

The ROM algorithms proposed to support the UCF may be viewed as hybrid algorithms that combine the benefits of both local and global SA methods while circumventing some of their deficiencies. In principle, the ROM techniques provide the means to render simultaneous reduction for both the input parameters space and responses space. This is achieved over two steps. In the first step, local methods are employed to identify a subspace (denoted by the active subspace) that captures the dominant parameters and their cross-interactions in order to account for all high orders effects inherent in the original nonlinear model. This is done based on a numerical tolerance specified by the user which sets an upper bound on the maximum allowed discrepancy between the original and reduced model predictions (see Appendix C for how the tolerance is set). This implies that all parameters cross-interactions that are below the specified tolerance will be considered non-influential and hence not included in the active subspace. In the second step, global methods are used to build a surrogate which restricts the samples to the active subspace only, thereby reducing the effective dimensionality of the input parameters space, and rendering the construction of the surrogate model computationally feasible.

12.0 APPENDIX C: RANGE FINDING ALGORITHMS

Range finding algorithms (RFAs) have been primarily developed in the linear algebra community and the machine learning community [Halko, 2011]. They represent the cornerstone of the ROM techniques. For the sake of this preliminary introduction, RFAs are described here for the simple case of linear operators. The reader is referred to the literature for the extension of RFA to nonlinear operators.

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a general matrix operator whose elements cannot be accessed directly, however the following operation is possible for a given user-defined vector p_i:

$$w_i = \mathbf{A}p_i \qquad (C.1)$$

This equation emulates the execution of a model with randomized input parameter perturbations for the general nonlinear case. In the gradient-free ROM approach, the w_i vectors represent the response variations corresponding to different randomized parameter perturbations, whereas in the gradient-based ROM approach, the w_i vectors represent the variations in the gradient, which is used to find an active subspace for the parameters space.

The objective of RFA is to find an effective range for $\mathbf{A}$ defined by a matrix $\mathbf{Q}$ such that:

$$\left\| \left(\mathbf{I} - \mathbf{Q}\mathbf{Q}^T \right) \mathbf{A} \right\| \leq \varepsilon_{user} \qquad (C.2)$$

where ε_{user} is user-defined; $\mathbf{Q} \in \mathbb{R}^{m \times r}$; and r is denoted the effective rank of the matrix $\mathbf{A}$, or the size of the effective range of the matrix $\mathbf{A}$. If r is much less than m, the implication is that the components of the vector w_i are highly correlated, and the associated variations can be constrained to a subspace (range of $\mathbf{Q}$) at a small but acceptable loss in accuracy, which can be made as small as the numerical precision of the computational model predictions, e.g., double precisions.

It has been shown recently by applied mathematicians that the condition in Eq. C.2 could be met with a high probability $p = 1 - 10^{-s}$ by employing only $r + s$ matrix-vector products of the form in Eq. C.1 [Halko, 2011]. This may be achieved as follows:

1. Let s be a small integer (see discussion below).
2. Pick s random vectors: $\omega_1, \dots, \omega_s$
3. Calculate: $\varpi_i = \mathbf{A}\omega_i$, $i = 1, \dots, s$
4. Let t_0 represent the current estimate of the effective rank of the matrix $\mathbf{A}$
5. Pick t additional random vectors: $p_1, \dots, p_t$
6. Calculate: $w_i = \mathbf{A}p_i$, $i = 1, \dots, t$
7. Find an orthonormal set such that: $\text{span}\{w_1, \dots, w_{t_0 + t}\} = \text{span}\{q_1, \dots, q_{t_0 + t}\}$

8. Let $\mathbf{Q} = \begin{bmatrix} q_1 & \cdots & q_{t_0+t} \end{bmatrix} \in \mathbb{R}^{m \times (t_0+t)}$

9. Calculate: $z_i = \left(\mathbf{I} - \mathbf{Q}\,\mathbf{Q}^T \right) \varpi_i$, $i = 1, \ldots, s$

10. Let $t_0 = t_0 + t$

11. If $\varepsilon_{user} < 10 \sqrt{\dfrac{2}{\pi}} \max_{i=1,..,s} \| z_i \|$, go back to step 5.

12. Let $r = t_0$

13. With probability $1 - 10^{-s}$, the following statement is true: $\left\| \left(\mathbf{I} - \mathbf{QQ}^T \right) \mathbf{A} \right\| \leq \varepsilon_{user}$

This algorithm is based on the observation that when a matrix is multiplied by random vectors, the resulting vectors are expected to be independent with a very high probability. If the matrix has a low rank representation that satisfies the condition in Eq. C.2, then one can find this subspace with at most r matrix-vector products. The extra s oversamples are employed to ensure the condition in Eq. C.2 is met with a high probability.

This analysis could be extended to nonlinear operators, where now the active subspace is captured by sampling the gradient of the responses with respect to the input model parameters. One can establish rigorous error bounds similar to Eq. C.2, which upper-bounds the errors resulting from discarding all parameter components that are orthogonal to the active subspace. Appendix D provides intuitive understanding of how these errors are quantified. For the interested reader, we recommend consulting the following references [Bang, 2012], and [Abdo, 2015].

13.0   APPENDIX D: ROM DEFINITIONS AND APPROACHES

Reduced order modeling, dimensionality reduction, surrogate modeling, and fitting techniques are typical analysis tools that are required by any uncertainty characterization (UC) approach. This is because in its most basic form, any UC approach requires repeated model execution to help quantify and prioritize the various sources of uncertainties. In most realistic situations, repeated model execution is computationally infeasible, especially when the model describes in details the behavior of the system, in terms of its various scales and physics feedback. For example, a brute force application of basic UC approaches to a nuclear reactor core analysis model could easily require one million model executions, which is not practically possible. This represents the top challenge for any UC practitioner,

How to reduce the number of required model runs and/or how to make them affordable?

13.1   REDUCED ORDER MODELING

ROM implies any attempt to reduce the complexity (i.e., order) of the analysis models. If the complexity is reduced, the number of uncertainty sources could be reduced, which will reduce the number of required model runs. Also, when the complexity is reduced, the cost of the calculations will decrease which could render repeated model execution to be affordable. Examples of ROM approaches include:

- Simplifying the physics employed to reduce computational cost.
- Simplifying the physics coupling to reduce the cost of the iterative solvers.
- Reducing the model dimensions by using coarser mesh and/or time steps.
- Discarding sources of uncertainties that are expected to be negligible.
- Employing 1D or 2D models instead of 3D models.
- Constraining the solution to be a polynomial of given order.
- Expanding the solution in terms of truncated Fourier expansion.
- Reducing the effective degrees of freedom of the input parameters.
- Reducing the effective degrees of freedom of the state solution.
- Replacing the model with a function (or a model) with undetermined coefficients.

The result of these applications is a model with reduced complexity that can be used in lieu of the original model, hence the common terminology of 'surrogate' model, i.e., it can be used to replace/represent the original model for the sake of completing UC analysis.

We will focus here on two ROM approaches, dimensionality reduction (DR) and function approximation techniques. Let the original complex model be defined by two equations:

$$\text{Constraints: } \Pi\left(\phi, x\right) = 0 \tag{D.1}$$

$$\text{Response: } y = f\left(x, \phi\right) \tag{D.2}$$

where $x \in \mathbb{R}^n$ is a vector representing n input model parameters, and $y \in \mathbb{R}^m$ are m output model responses, and $\phi \in \mathbb{R}^k$ the state function. The first equation determines the state for the given input parameters and the second equation calculates the responses of interest as function of the input parameters and the determined state. In static neutronics calculations, the constraint equation is represented by the Boltzmann equation, the input parameters are the nuclear cross-sections, the state is the neutron angular flux, and the responses include the count rates of a number of detectors placed throughout the core, the core's critical eigenvalue, pin powers, etc..

Since the state is an intermediate variable, the Eq. (D.1) and Eq. (D.2) could be compressed into a black-box form as follows:

$$y = \theta(x) \tag{D.3}$$

13.2   FUNCTION APPROXIMATION (FITTING)

Function approximation (or response surface methodology [Box, 1987]) techniques attempt to replace the original model in Eq. (D.3) by an approximate function ϑ, referred to as surrogate model such that:

$$y^{(approx)} = \vartheta(x, a) \tag{D.4}$$

where $a \in \mathbb{R}^l$ are a set of l undetermined coefficients. The rational here is that the modeler has enough experience with the model behavior such that he/she can hypothesize that the model's trend for a range of x values can be well-approximated by a pre-determined function ϑ. This approach is referred to as function approximation. The approximate function is selected such that it offers a certain level of flexibility, introduced via a set of undetermined parameters a, to allow the analyst match or at best minimize the discrepancies between the original model and the approximate function predictions at a number of pre-selected points in the x domain, referred to as training points or samples. Mathematically, the parameters a are defined such that:

$$\min_a \left\{ \theta(x_i) - \vartheta(x_i, a) \right\}_{i=1}^M \tag{D.5}$$

This expression implies the ability to minimize the differences in predictions, i.e., discrepancies, at M points in the x domain. Finding values for a that simultaneously minimize all the M discrepancies is not generally possible, because each term is likely to have a minima at different values of a. To overcome this problem, the M discrepancies are combined together into a single function to be minimized. The most common approach uses the Euclidean norm defined by:

$$\min_a \sum_{i=1}^{M} \left\{ \theta(x_i) - \mathcal{G}(x_i, a) \right\}^2 \tag{D.6}$$

The most notable challenge here is to ensure that the surrogate model thus generated is reliable which is only possible if one can constraint the differences between the surrogate and original model predictions as follows:

$$\left| \theta(x) - \mathcal{G}(x, a) \right| \leq \varepsilon \text{ for all } x \in S \tag{D.7}$$

This expression implies that the errors at any point x in a given domain S can be upper-bounded by a limit ε. This is unfortunately not possible with any function approximation technique, simply because the assumed function form $\mathcal{G}$ does not know how the original model behaves at every point in the x domain. In practice, some choice of the function $\mathcal{G}$ could actually result in an unphysical oscillatory behavior while still matching the original model predictions at the training points. Therefore, it is reasonable to expect the function approximation to be successful only when enough experience with model execution has been accumulated to justify the use of a given function approximation, and it is safe to say that the experience is what provides credence to the approximation rather than the quality of the fit as determined by the minimizer of Eq. (D.6).

13.3   DIMENSIONALITY REDUCTION

In dimensionality reduction (DR), one attempts to reduce the effective dimensionality of the model interfaces without changing the form of the functions or the constraint equation. If described in compressed form as in Eq. (D.3), DR attempts to find reduced variables $x^{(DR)} \in \mathbb{R}^n$ such that:

$$y^{(DR)} = \theta\left(x^{(DR)}\right), \text{ and } x^{(DR)} = u(x) \tag{D.8}$$

where the reduced variables are related to the original variables x by a function u that is to be determined by the DR approach.

Note that both the original input variables and the reduced variables have the same dimensionality n, so where does the reduction come from? Instead of throwing around some mathematical jargon, we will first give a simple example to explain how the reduction is rendered: let x live in a three dimensional space implying that x has three degrees of freedom; this is because each of the components can change freely, and not because x has three nominal components. Now, what if one places some restrictions on the movement of the components, say for example, all components are now constrained to an oblique plane that passes through the origin. In this case, although all three components are changing as one travels from one point to the next on the plane, the effective degrees of freedom are only two. Said differently, although all three components are changing as we move over the plane, perfect correlations must exist between these components variations.

This restriction/correlations can be mathematically described using the concept of projection. Let's start with a one-dimensional projection for simplicity, wherein the three components are constrained to a line that passes through the origin. This line may be described by a unit vector q. To generate a random point that lives on that line, one can simply multiply the unit vector q by a random scalar, i.e., $x^{(DR)} = \alpha q$. The superscript (DR) implies that the randomly generated points are constrained in some way.

One may generate a random point using a more complicated two-step approach, where first a random point in the three-dimensional space is generated, expected to be outside the line, and next x is projected onto the line as depicted in Fig. D.1. This may be described mathematically as follows:

$$x^{(DR)} = q\left(q^T x\right) = \left(q q^T\right) x \tag{D.9}$$

The $q^T x$ is simply the component of x along q, which is a scalar quantity. Multiplying this component by the unit vector q converts the scalar component into a vector component. The brackets in Eq. (D.9) denote the order of the operations. In the first scenario just described, the component of x along q is calculated first, then all the components of the q vector are scaled by that component. In this scenario, $q^T x$ is called the inner product of the two vectors q and x, and it produces a single scalar quantity, typically referred to as the component (or the scalar component) of x along q.
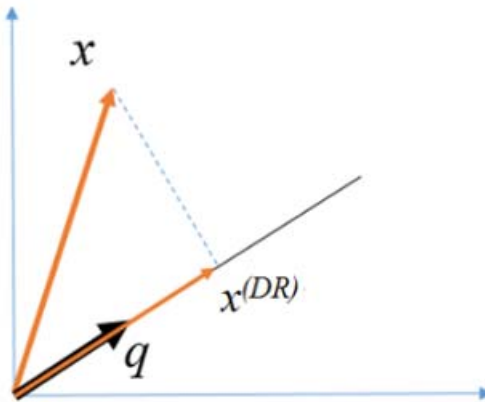
Fig. D.1 Reduction via Projection

One can also place the brackets around $q q^T$ which implies one has to form $q q^T$ first, then multiply it by x to produce $x^{(DR)}$. In this scenario, $q q^T$ must be a matrix, and is referred to as the outer product of the two vectors. Outer products always generate matrices, whereas inner products generate scalars. Inner products require that the two vectors have the same length, whereas outer products do not require that. The inner product between two vectors a and b of same length n is given by:

$$a^T b = \sum_{i=1}^{n} a_i b_i$$

whereas the outer product between two vectors a and c of length n and l, respectively is given by:

$$\left[ac^T \right]_{ij} = a_i c_j \quad \text{and} \quad ac^T \in \mathbb{R}^{n \times l}$$

where $\left[\mathbf{Z} \right]_{ij}$ denotes the element at the intersection of the ith row and j^{th} column of the matrix $\mathbf{Z}$.

Fig. D.2 depicts the difference between inner and outer products (assume that both a and c are unit vectors for simplicity, denoted in black). The inner product operation $c^T x$ finds the scalar component of x along c. The outer product ac^T applied to x takes the scalar component $c^T x$ and mounts it onto the vector a with the result being a vector of length $c^T x$ pointing in the direction of a. Outer products therefore involve two operations, a projection operation followed by a rotation operation. When either the vectors a or c are not of unit length, an additional scaling operation is also involved.
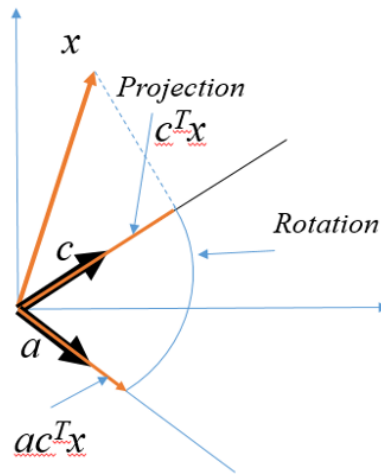
Fig. D.2 Inner vs. Output Product

The power of the outer products stems from the fact that any matrix operator can be written as a summation of mutually orthogonal outer products, this represents the basic premise of singular value decomposition of a general matrix $\mathbf{Z}$ given by:

$$\mathbf{Z} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{D.10}$$

where both $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices, and $\mathbf{S}$ is a diagonal matrix. This decomposition may also be re-written as:

$$\mathbf{Z} = \sum_{i=1}^{r} s_i u_i v_i^T \qquad (D.11)$$

where $\{u_i\}_{i=1}^{r}$ are the columns of the matrix $\mathbf{U}$ and $\{v_i\}_{i=1}^{r}$ are the columns of the matrix $\mathbf{V}$. The s_i are scalars. Because $\mathbf{U}$ and $\mathbf{V}$ orthonormal, all their columns are unit length and orthogonal to each other. Notice that each term of the form $u_i v_i^T$ is an outer product which when applied to a vector x, identifies the component of x along v_i, then rotates it to the direction u_i. Because of the scalar s_i, the rotation is also accompanied by dilation of the component. Note also that this expression involves a summation of r outer products, implying that when the matrix $\mathbf{Z}$ multiplies a vector x, only r degrees of freedom of the vector x will be projected and rotated by the operator $\mathbf{Z}$. Since x lives in an n dimensional space, the implication is that n-r degrees of freedom will be lost upon multiplying x by $\mathbf{Z}$. This represents the basic idea of DR.

Now we show how outer products can be used to constraint the movement of a vector to a plane, as described earlier. If one wishes to constraint x movements to a plane, one needs two unit vectors that describe the plane, say q_1 and q_2 such that $q_1 \perp q_2$, assume these vectors are also unit length for simplicity. Eq. (D.9) now reduces to:

$$x^{(DR)} = q_1\left(q_1^T x\right) + q_2\left(q_2^T x\right) = \left(q_1 q_1^T\right)x + \left(q_2 q_2^T\right)x = \begin{bmatrix} q_1 & q_2 \end{bmatrix} \begin{bmatrix} q_1^T \\ q_2^T \end{bmatrix} x = \mathbf{Q}\mathbf{Q}^T x \quad (D.12)$$

This expression describes how a general vector x that lives in a 3D space is projected onto a plane described by two vectors q_1 and q_2. First, the component of x along q_1 is formed; next the component along q_2 is formed; and finally the two components are added back to form one vector that lives in the plane. This process is referred to as orthogonal projection. Note that this is still an outer product operation but with a zero degree rotation since the two vectors forming the outer products are the same.

The columns of the matrix $\mathbf{Q}$ are said to form a basis for the said plane. The number of columns of the matrix $\mathbf{Q}$ refers to the number of DOFs available for the movements of the points x in the plane. This idea can be generalized to an n dimensional space, and a plane with r_x DOFs.

$$x^{(DR)} = \mathbf{Q}\mathbf{Q}^T x \text{ and } \mathbf{Q} = \begin{bmatrix} q_1 & q_2 & .. & q_{r_x} \end{bmatrix} \in \mathbb{R}^{n \times r_x} \qquad (D.13)$$

Returning now to Eq. (D.8), one can write:

$$u(x) = \mathbf{Q}\mathbf{Q}^T x \qquad (D.14)$$

This type of reduction, based on projection operators, is referred to as linear dimensionality reduction (LDR), because the relationship between the reduced variables and the original

variables is linear. Nonlinear DR techniques are also possible when the relationship between the reduced and original variables become nonlinear. We will restrict our discussion here to LDR since its payoff is extremely high in neutronics calculations, representing the focus of the UCF.

The objective of LDR is to pick the matrix $\mathbf{Q}$ that minimizes the function:

$$\min_{\mathbf{Q}} \sum_{i=1}^{M} \left\{ \theta(x_i) - \theta\left(\mathbf{Q}\mathbf{Q}^T x_i\right) \right\}^2 \tag{D.15}$$

This expression evaluates the errors in the responses of interest resulting from discarding components in the x space. We reiterate here that in this type of reduction, the function θ remains unchanged, implying that the errors result only from the input variables reduction. Because of that, it is possible to upper-bound the associated errors.

To demonstrate that, we employ a simple single-valued function: let $y = \theta(x)$

$$y = \frac{(a_1 x_1 + a_2 x_2 + a_3 x_3)}{1 + (a_1 x_1 + a_2 x_2 + a_3 x_3)^3} \exp\left\{-(a_1 x_1 + a_2 x_2 + a_3 x_3)^2\right\} \tag{D.16}$$

This may be rewritten as: $y = \theta(z)$ and $z = a^T x$. This is a function of three variables, however one can eyeball that it has only a single DOF such that:

$$y = \frac{z}{1+z^3} \exp\left\{-z^2\right\} \text{ and } z = a_1 x_1 + a_2 x_2 + a_3 x_3 = a^T x \tag{D.17}$$

An LDR here implies finding a direction in the space that can be used to approximate the function well. By solving the minimization problem in Eq. (D.15), one can easily find that the single needed direction is given by:

$$q = \frac{a}{\|a\|} \tag{D.18}$$

This follows because the function only varies along this direction. The division by the norm of a is only to ensure that q is a unit vector as defined earlier. Now assume that one does not have know the optimum direction q, instead some other approximate direction w is assumed. Our job is to decide whether the errors resulting from this reduction can be quantified and potentially upper-bounded. The reduction error in this case is given by:

$$\left| f(z) - f(\tilde{z}) \right|$$

where $\tilde{z} = a^T w w^T x$. Note that when $w = q \Rightarrow \tilde{z} = z$. If the difference $z - \tilde{z}$ is sufficiently small, one can approximate the above error by:

$$\left| f\left(z\right) - f\left(\tilde{z}\right)\right| = \frac{df}{dz}\left(z - \tilde{z}\right) + \ldots.$$

Taking the upper limit of both sides using Triangle inequality:

$$\left| f\left(z\right) - f\left(\tilde{z}\right)\right| \leq \left| \frac{df}{dz}\right| \left| z - \tilde{z}\right|$$

This equation implies that if one can upper-bound the term $z - \tilde{z}$ and the function's derivative is also upper-bounded over the range of domain of interest, the resulting reduction error in the estimated function value must also be upper-bounded. Examples of functions with upper-bounded derivatives and non-upper-bounded derivatives in Fig. D.3.
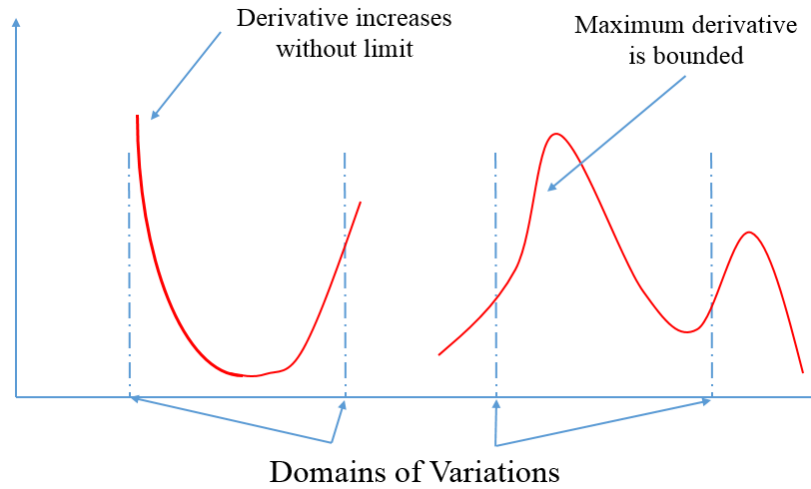
Fig. D.3 Bounded vs. Unbounded Variations

Finally, estimating $z - \tilde{z}$ implies the need to estimate $\left\| a^{T}\left(I - ww^{T}\right)\right\|$, since one can write:

$$z - \tilde{z} = a^{T}\left(I - ww^{T}\right) = \mathbf{A}x \rightarrow \mathbf{A} = a^{T}\left(I - ww^{T}\right)$$

Since the range of x variations is known a priori, as determined by the analysis model, one only needs to estimate the norm of the matrix $\mathbf{A}$. In 1983, Dixon [Dixon, 1983] proved a theory that allows one to estimate an upper-bound on the norm of any general matrix using simple random matrix-vector products.

$$P\left\{ \left\| \mathbf{A}\right\| < 10\sqrt{\frac{2}{\pi}}\max_{i=1,..,s}\left\| \mathbf{A}x_{i}\right\| \right\} = 1 - 10^{-s}$$

This theorem states that if one can calculate s random matrix vector products of the form $\mathbf{A}x$, where x is of unit norm, and randomly selected from Gaussian distribution, the norm on the matrix $\mathbf{A}$ may be upper-bounded with high probability.

14.0   APPENDIX E: CRANE AND E$_P$GPT

This appendix provides an overview of a new functionality, the exact-to-precision generalized perturbation theory (E$_P$GPT) [Wang, 2013] within the SCALE super-sequence CRANE [Mertyurek, 2014]. CRANE (**C**omplexity **R**eduction **A**lgorithms in **N**uclear **E**ngineering calculations) is a new 'supper-sequence' in SAMPLER implemented recently under the SCALE [SCALE, 2011]. CRANE recognizes the need for ROM techniques to complete UQ and SA analyses for neutronics calculations. E$_P$GPT represents one of the fundamental ROM applications, wherein a physics-based surrogate model is constructed to replace the high fidelity neutron transport code, e.g. SCALE's NEWT. Because of its reliance on ROM techniques, E$_P$GPT is capable of calculating rigorous upper-bounds on its prediction errors for all quantities of interest, which is necessary to render its general use credible. This short discussion demonstrates the use of E$_P$GPT to replace the NEWT model for performing standard cell lattice calculations.

14.1   BACKGROUND ON E$_P$GPT

E$_P$GPT represents a rendition of two fundamental ROM approaches combined together to formulate a surrogate model that can be used in lieu of expensive neutron transport calculations. Unlike existing GPT methods [Williams, 1986], [Cacuci, 2003], E$_P$GPT calculates all higher order variations with efficiency taking advantage of the DR of the flux space. E$_P$GPT is designed to achieve two important goals, computational efficiency and accuracy. With regard to efficiency, E$_P$GPT employs DR techniques described in Appendix D to reduce the number of forward and adjoint model executions required. To achieve efficiency, it employs a physics-driven approach to select the surrogate form. Recall from Appendix D that the selection of the surrogate form is instrumental in determining the reliability of its predictions. E$_P$GPT employs the adjoint functions derived from the solution of the adjoint transport equation, both in eigenvalue mode and inhomogeneous mode, to construct the surrogate model form.

This Appendix intends to provide qualitative discussion of EPGPT and its implementation into the SCALE code package. This is intended to provide a sense to the reader of the cost required to implement it for other neutron transport codes in support of CANDU analysis. For quantitative details on the theory and implementation, the reader may consult past publications and the references within [Wang, 2013].

E$_P$GPT requires five major steps, as depicted in the Fig. E.1, which are automated by CRANE. The *first* step involves the execution of the forward transport model few hundred times, each with randomized input parameters, i.e., randomized cross-sections and number density. The *second* step employs the flux variations aggregated from the first step to identify the so-called pseudo responses for the flux space. Pseudo responses are linear combination of the flux values everywhere in the phase space. They can be used to reconstruct flux variations for any given input parameter perturbation with quantified

accuracy. This represents of the core advantage of E$_P$GPT, as it allows one to calculate an upper-bound for the flux in all future model executions. The ***third*** step employs conventional GPT to calculate generalized adjoints, each corresponding to one of the pseudo responses. In our experience with reactor calculations, the number of pseudo responses is typically in the order of few tens to few hundreds, depending on the goal of the analysis. Each adjoint is employed to calculate the first order derivatives of a pseudo response with respect to all input model parameters. The ***fourth*** step employs ROM techniques to identify an active subspace for the parameters using the derivatives aggregated from the previous step. The ***fifth*** step employs the active subspaces for the parameters and responses to construct a number of matrices which form the functional form of the surrogate model. With these matrices calculated once, one can may call CRANE-E$_P$GPT as a separate module with the standard SCALE input file to calculate the flux for the user-defined perturbations.
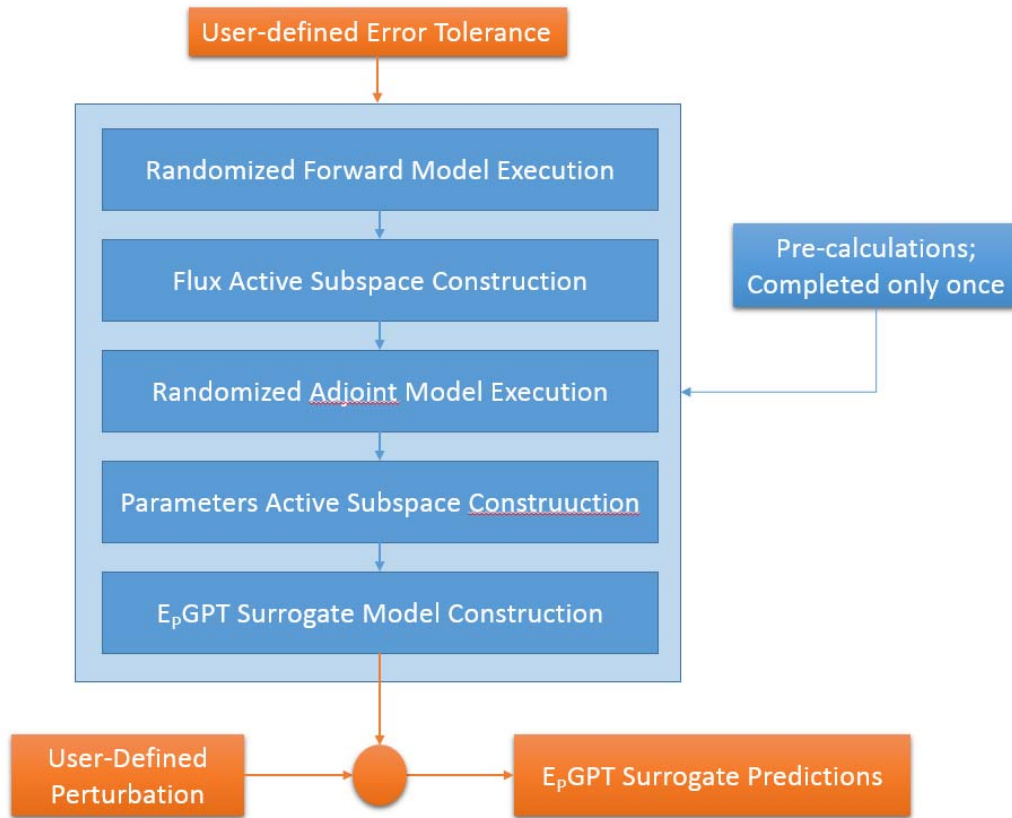

Fig. E.1. CRANE-E$_P$GPT Flowchart

14.2   Cross Section Perturbation

To generate the randomized forward runs, one must develop a buffer code to perturb the model's input parameters. For this application, we employed the ClarolPlus code developed under the SAMPLER super-sequence. Currently, ClarolPlus is automated to

read the specific perturbations employed by SAMPLER to propagate cross-section uncertainties based on the 44G covariance library. Since E$_P$GPT requires randomized perturbation, ClarolPlus has been modified to read user-defined perturbations. The perturbations are defined as follows:

$$f_{x,g} = 1 + \frac{\Delta\sigma_{x,g}}{\sigma_{x,g}} \tag{E.1}$$

where $\Delta\sigma_{x,g} / \sigma_{x,g}$ are the relative cross section variations, the subscript x denotes the nuclide-reaction type and g denotes the energy group number.

14.3   Active Subspace Construction

E$_P$GPT assumes that the flux variations for a general input parameter perturbation may be written as follows:

$$\Delta\phi = \sum_{i=1}^{r_\phi} \alpha_i q_i \tag{E.2}$$

where the r_ϕ vectors q_i represent a basis for active subspace of the flux, and they are calculated based on the randomized forward model executions. In addition, E$_P$GPT assumes a general input parameter perturbation may be written as follows:

$$\Delta\sigma = \sum_{i=1}^{r_\sigma} \beta_i u_i \tag{E.3}$$

where the r_α vectors u_i represent a basis for the active subspace of the input parameter, and they are calculated based on the randomized GPT model executions.

The r_ϕ and r_α denote the size of the active subspaces, i.e., the number of independent vectors used to span their ranges. With the basis defined, one can calculate an upper-error bound on the flux given by:

$$\left\| \phi^{exact} - \phi^{E_P GPT} \right\| \leq \varepsilon \tag{E.4}$$

where ϕ^{exact} is the exact flux value obtained by running the original transport code, and the other value denotes the one calculated by the E$_P$GPT surrogate. If the epsilon value is not satisfactory, one needs to execute the code additional number of times, until the desired error tolerance is achieved. This error bound can be applied to the entire flux vector or on an individual component basis without being affected by the amplitude of the perturbation in the prediction step with respect to the range of perturbation used in the active subspace construction (see Ref. [Bang, 2012] for the detailed discussion).

14.4   Generalized Adjoint Calculations

If r_ϕ is much smaller than m (the dimension of the flux), one can recast GPT in terms of a set of r_ϕ pseudo responses, implying that the flux does not show appreciable changes over the remaining $m - r_\phi$ directions. The pseudo responses are defined by:

$$R_{pseudo} = \langle q_i, \phi \rangle, \text{for } i = 1, \cdots, r_\phi \tag{E.5}$$

Note that the evaluation of the generalized adjoints corresponding to the pseudo responses requires a straightforward application of conventional GPT [Williams, 1986]. The only difference here is that one needs to evaluate r_ϕ pseudo adjoints only to represent all possible model responses. This is because any model response is a function of the flux which is fully described by the pseudo responses as ensured by the error bound described earlier.

The relevance of the pseudo responses must be noted when considering the number of responses typically calculated using cell lattice physics models. For example in CANDU cell lattice calculations, one must generate the few-group cross-sections in terms of a wide range of conditions, such as coolant temperature, fuel temperature, soluble boron content, etc., over the entire range of depletion. Brute force application of GPT to this problem would result in an extremely high number of adjoints that must be evaluated. This cost can be significantly reduced using the concept of pseudo response as implemented in E$_P$GPT.

14.5   Surrogate for Transport Calculations

One can rigorously show that the E$_P$GPT surrogate model form is given by [Wang, 2013]:

$$\Delta\phi = -\mathbf{Q}(\mathbf{I} + \mathbf{C})^{-1}\varpi \tag{E.6}$$

where the column space of $\mathbf{Q}$ represents the active flux subspace, $\mathbf{I}$ is the identify matrix, $\mathbf{C} \in R^{r_\phi \times r_\phi}$ and $\varpi \in R^{r_\phi}$. Each element of $\mathbf{C}$ and ϖ is an inner product of the form:

$$\langle \Gamma^*, \Delta\mathbf{P}\Psi \rangle \tag{E.7}$$

where Γ^* represents the generalized adjoint solutions associated with the pseudo responses, Ψ represents the basis for the flux active subspace, and $\Delta\mathbf{P}$ denotes the variations in the neutron transport operators due to a given perturbation. The matrix $\mathbf{C}$ and vector ϖ in Eq. (E.6) can be rewritten as:

$$\mathbf{C}(\Delta\mathbf{P}) = \mathbf{C}\left(\sum_{i=1}^{r_\alpha}\beta_i\Delta\mathbf{P}_i\right) = \sum_{i=1}^{r_\alpha}\beta_i\mathbf{C}(\Delta\mathbf{P}_i) = \sum_{i=1}^{r_\alpha}\beta_i\mathbf{C}_i \tag{E.8}$$

$$\varpi(\Delta\mathbf{P}) = \varpi\left(\sum_{i=1}^{r_\alpha} \beta_i \Delta\mathbf{P}_i\right) = \sum_{i=1}^{r_\alpha} \beta_i \varpi(\Delta\mathbf{P}_i) = \sum_{i=1}^{r_\alpha} \beta_i \varpi_i \qquad \text{(E.9)}$$

where $\mathbf{C}_i \in R^{r_\phi \times r_\phi}$ and $\varpi_i \in R^{r_\phi}$ denotes $\mathbf{C}(\Delta\mathbf{P}_i)$ and $\varpi(\Delta\mathbf{P}_i)$, respectively. With these matrices $\{\mathbf{C}_i\}_{i=1}^{r_\alpha}$ and vectors $\{\varpi_i\}_{i=1}^{r_\alpha}$ calculated once, one can call CRANE-EPGPT as a separate module with standard SCALE input file to calculate the flux for the given conditions of interest as follows:

$$\Delta\phi = -\left(\mathbf{I} + \sum_{i=1}^{r_\alpha} \beta_i \mathbf{C}_i\right)^{-1}\left(\sum_{i=1}^{r_\alpha} \beta_i \varpi_i\right) \qquad \text{(E.10)}$$

Previous work has demonstrated that EPGPT can be used to evaluate the variations in neutron flux and responses of interest due to various perturbations, such as cross sections, fuel enrichment, fuel densities, temperatures, etc., in lattice design calculations [Wang, 2013]. We will not duplicate these results here, but will demonstrate some representative results for the sake of a complete discussion.

14.6 Numerical Results

A PWR assembly model, a UO$_2$ Gösgen (ARIANE) sample identified as GU1 [Radulescu, 2010], is selected to demonstrate the new EPGPT capability of CRANE. The modeling layout is provided in Fig. E.2 with 3.5 wt% ^{235}U initial enrichment, and the original design characteristics are given in Ref. [Radulescu, 2010].
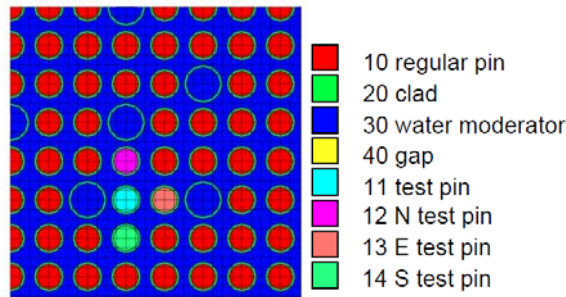
Fig. E.2. GU1 Assembly model layout

The standard SCALE 56-group cross section library (xn56v7 generated from ENDF/B-VII.0 64-group neutron library) is employed. The dimension of the angular flux is 3,929,856 (i.e. angular quadrature $S_N=6$, 56 energy groups and 2924 cells). The flux predicted by the CRANE-EPGPT surrogate, using a subspace with $r = 60$, are compared to the exact fluxes predicted by NEWT.

We will refer to $\left\|\phi_{Exact} - \phi_{E_PGPT}\right\|_p / \left\|\phi_{Exact}\right\|_p$ as the relative error in the flux. Large number of randomized cross-section perturbations, giving rise to 1% change in reactivity, were used to calculate to assess the quality of E$_P$GPT surrogate predictions. Specifically, using $r = 60$, the upper bounds on the angular flux errors were calculated in several norms as follows:

$$\frac{\left\|\phi_{Exact} - \phi_{E_PGPT}\right\|_1}{\left\|\phi_{Exact}\right\|_1} : 3.98E-3$$

$$\frac{\left\|\phi_{Exact} - \phi_{E_PGPT}\right\|_2}{\left\|\phi_{Exact}\right\|_2} : 4.57E-3$$

$$\frac{\left\|\phi_{Exact} - \phi_{E_PGPT}\right\|_\infty}{\left\|\phi_{Exact}\right\|_\infty} : 6.41E-3$$

The corresponding upper-bound on the critical eigenvalue is found to be 0.037%, or approximately 0.4 mk. In practice, one could employ higher r to improve the accuracy. A small value is used here to demonstrate that reasonable accuracy can be obtained with very small active subspace as compared to the original flux space.

Fig. E.3 shows the flux spectrums for fuel in four types of pins, the 11 test pins, the 12 N test pin, the 13 E test pin, and the 14 S test pin as shown in Fig. 2.
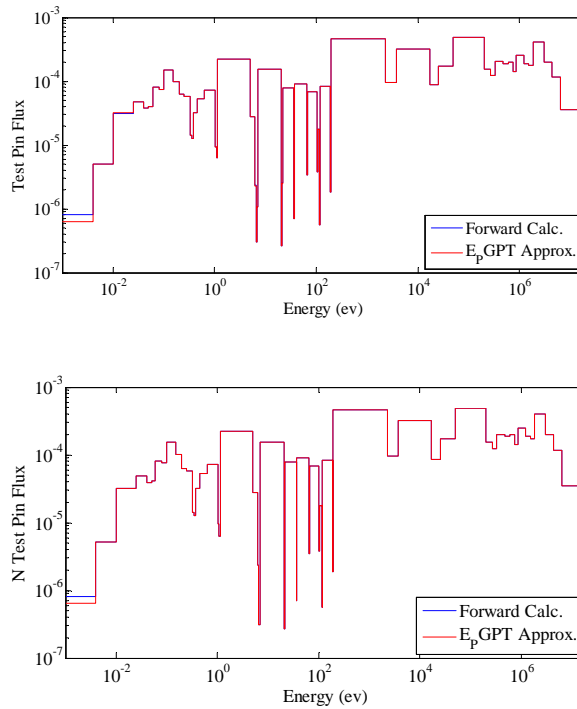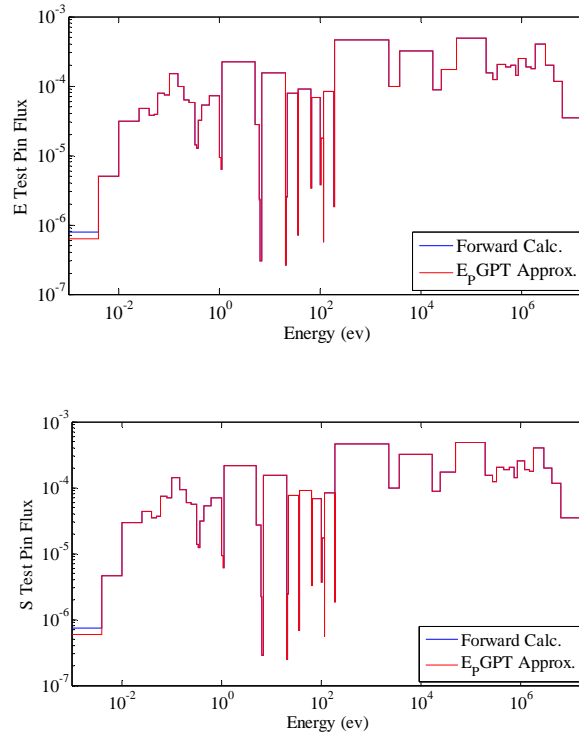
Fig. E.3. Flux spectrum for the test pins

The CRANE-E$_P$GPT tool is planned for release this year to beta users. A working copy has been transferred to ORNL for evaluation [Mertyurek, 2015]. We encourage interested parties to contact us for obtaining beta version for testing, evaluation, and feedback. Future work will primarily focus on improving the user-interface to match the versatility offered by the current SCALE sequence. We are currently working on breaking up the various steps into separate scripts to allow the users evaluate the separate steps which will help provide insight into the mechanics of the various parts of the algorithm.

15.0 **REFERENCES**

1. Abdel-Khalik, H. S., et. al., 2008, "Efficient Subspace Methods-Based Algorithms for Performing Sensitivity, Uncertainty, and Adaptive Simulation of Large-Scale Computational Models," Nuclear Science and Engineering, 159 (3), 256-272, 2008.

2. Abdel-Khalik, H., et. al., 2013, "Overview of Hybrid Subspace Methods for Uncertainty Quantification, Sensitivity Analysis," Annals of Nuclear Energy, 52, Pp. 28-46, 2013.

3. Abdel-Khalik, H. S., et. al., 2014, "Initial UQ of Insilico," CASL Milestone Rerport # L3:VUQ.V&V.P8.03, March 2014.

4. Abdo, M. G., et. al., 2015, "Multi-Level Reduced Order Modeling," Transactions of American Nuclear Society, June 2015.

5. Abdo, M. G., et. al., 2015, "Probabilistic Error Bounds for Reduced Order Modeling," Proceedings of International Conference on Mathematics and Computations in Nuclear Science and Engineering, Nashville, TN, April, 2015

6. Alfonsi, A., et. al., 2013, "Raven as a Tool for Dynamic Probabilistic Risk Assessment: Software Overview," International Conference on Mathematics and Computational Methods Applied to Nuclear Science & Engineering, Sun Valley, Idaho, USA, May, 2013.

7. Antoulas A. C., et. al., 2001, "A survey of model reduction methods for large scale systems", Journal of Contemporary Mathematics, 280, (2001).

8. Arenas, C., et. al., 2013, "Uncertainty Analysis of Light Water Reactor Fuel Lattices," Science and Technology of Nuclear Installations, Volume 2013, Article ID 437409, 2013.

9. Bang, Y., et. al., 2012, "Hybrid Reduced Order Modeling Applied To Nonlinear Models," International Journal For Numerical Methods In Engineering, 91, Pp. 929-949, 2012.

10. Bang, Y. "Hybrid Reduced Order Modeling Algorithms for Reactor Physics Calculations," PhD Dissertation, North Carolina State University, Raleigh, North Carolina, 2012.

11. Box, G. E., et. al., 1987, "Empirical Model-Building and Response Surfaces," (1987).

12. Cacuci, D.G., 2003, *Sensitivity and Uncertainty Analysis*, Chapman & Hall, CRC Press, 2003.

13. Casal, J. et. al., 1992 "HELIOS: Geometric Capabilities of a New Fuel-Assembly Program," Proc. Int. Top. Mtg. Adv. Math. Comp. Reac. Phys., Pittsburg, PA, USA, Vol. II, Sect. 10.2.1, 1-13.

14. CCSI Website: https://www.acceleratecarboncapture.org/

15. Cousins, 2012, private communication, "Monkeys vs Gauss: Integration in High Dimensions," Presentation.

16. DAKOTA Website: http://dakota.sandia.gov/index.html

17. Dixon, J. D., 1983, "Estimating extremal eigenvalues and condition numbers of matrices," *SIAM*, **20**(2): 812–814, 1983.

18. Gandini, A., "Generalized Perturbation Theory for Nonlinear Systems for the Importance Conservation Principle", Nucl. Sci. Eng., 77, 316 (1981).

19. Gerstner T., et. al., 1998, "Numerical integration using sparse grids," Numerical Algorithms, 18:209-232, 1998.

20. Ghanem, R.G., et. al., *Stochastic Finite Elements: A Spectral Approach*, Courier Dover Publications, (1991).

21. GPMSA Website: http://www.stat.lanl.gov/staff/DHigdon/Shortcourse.html

22. Greenspan, E., et. al., "Second-Order Generalized Perturbation Theory for Source-Driven Systems," Nuclear Science and Engineering, **68**, pp. 1-9 (1978).

23. Halko, N., et. al., 2011, "Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions," *SIAM Rev.*, Survey and Review Section, June 2011.

24. Helton, J.C., et. al., 2006, "Survey of sampling based methods for uncertainty and sensitivity analysis," *Reliability Engineering and System Safety*, **91**, pp. 1175–1209, 2006

25. Holmes, P., et. al., 1996, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge Monographs on Mechanics, Cambridge University Press, (1996).

26. Hubbard, D., 2010, *How to Measure Anything: Finding the Value of Intangibles in Business*, 2nd ed. John Wiley & Sons, 2010.

27. Jaynes, E., 2009, *Probability Theory: The Logic of Science*, 2009

28. Jessee, M., et. al., 2011, "Many-Group Cross-Section Adjustment Techniques for BWR Adaptive Simulation," accepted for publication, Nuclear Science and Engineering, **169**,1 2011.

29. Khuwaileh, B. A., et. al., 2013, "Exploratory Development of Multi-Physics Reduced Order Modeling II", *Transactions of American Nuclear Society*, 109, 757, (2013).

30. Kramer, O., 2011, "Unsupervised K-Nearest Neighbor Regression,"arXiv:1107.3600v2, 2011

31. Kennedy, M. C., et. al., 2001, "Bayesian Calibration of Computer Models," Journal of the Royal Statistical Society, Series B (Statistical Methodology), Vol 63, No. 3, pp. 425-464, 2001.

32. Klein, M., et. al., 2011, "Influence of Nuclear Data Uncertainties on Reactor Core Calculations, " Kerntechnick, **76**, 3, 2011

33. Larson, N. M., 1984, "Updated Users' Guide for SAMMY: Multilevel R-Matrix Fits to Neutron Data Using Bayes' Equations," ORNL/TM-9179, 1984.

34. MCNP Website: https://mcnp.lanl.gov/

35. Mertyurek, U., 2014, "CRANE: A Prototypic SCALE Module for Reduced Order Modeling," Transactions of the American Nuclear Society, Reno NV, 2014.

36. Mertyureek, U., 2015, private communication.

37. Meyer, C., 2001, *Matrix Analysis and Applied Linear Algebra*, SIAM, 2001

38. NASA 2012, UQTOOLS, NASA/TM–2012-217561, 2012

39. NESTLE, 2003: Code System to Solve the Few Group Neutron Diffusion Equation Utilizing the Nodal Expansion Method (NEM) for Eigenvalue, Adjoint, and Fixed-Source Steady-State and Transient Problems, Version 5.2.1, June 2003. RSICC CCC-641

40. Oliver, T., et al., 2014, "Validating Predictions of Unobserved Quantities," Comp. Meth. Appl. Mech. Eng., 283, 215, 2014

41. PSUADE, 2011, The PSUADE Uncertainty Quantification Project. Lawrence Livermore National Laboratory, 2011

42. Radulescu, G., et. al., 2010, "SCALE 5.1 Predictions of PWR Spent Nuclear Fuel Isotopic Compositions," Nuclear Science and Technology Division, ORNL/TM-2010/44 (2010).

43. Rochman, D., et. al., 2014, "Efficient Use of Monte Carlo: Uncertainty Propagation," Nuclear Science and Engineering, 177, 337–349 (2014).
44. Saltelli, A., et. al., 2000, *Sensitivity Analysis*, John Wiley (2000).
45. SCALE, 2011, A Comprehensive Modeling and Simulation Suit for Safety Analysis and Design, ORNL/TM- 2005/39, Version 6.1, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 2011.
46. Serghiuta, D., et. al., 2014, "A Stochastic-Deterministic Approach For Evaluation Of Uncertainty In The Predicted Maximum Fuel Bundle Enthalpy In A Candu Postulated LBLOCA Event," The 19th Pacific Basin Nuclear Conference (PBNC 2014) Hyatt Regency Hotel, Vancouver, British Columbia, Canada, August 24-28, 2014.
47. SERPENT, 2011. PSG2/Serpent Monte Carlo Reactor Physics Burnup Calculation Code, http://montecarlo.vtt.fi.
48. Silverman, B., 1986, Density Estimation for Statistics and Data Analysis, 1986.
49. Sirovich L., 1987, "Turbulence and The Dynamics of Coherent Structure. Part 1: Coherent Structures", Quarterly of Applied Mathematics, **45** (3), (1987).
50. Wang, C., et. al., 2013, "Exact-To-Precision Generalized Perturbation Theory for Eigenvalue Problems", Nuclear Engineering and Design, 256, 130-140, 2013.
51. Wang, C., et. al., 2013, "Exact-to-Precision Generalized Perturbation Theory for Neutron Transport Calculation," Proceedings of the 2013 International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering, Sun Valley, Idaho, May 5-9, 2013.
52. Wang, C., et al., 2014, "Intersection Subspace Method for Uncertainty Quantification," Transactions of American Nuclear Society, 111, 2014.
53. Williams, M. L., et al., 2013, "A Statistical Sampling Method for Uncertainty Analysis with SCALE and XSUSA," *Nuclear Technology*, **183**, 515 (2013)
54. Williams, M. L., 1986, *Perturbation Theory for Reactor Analysis*, CRC Handbook of Nuclear Reactor Calculations, CRC Press (1986).
55. Wiarda, D., et. al., 2006, PUFF-IV: A Code for Processing ENDF Uncertainty Data into Multigroup Covariance Matrices, ORNL/TM-2006/147
56. Xiu, D., 2010, *Numerical Methods for Stochastic Computations: A Spectral Method Approach*, Princeton University Press, (2010).
57. XSUSA Website: http://www.grs.de/en/simulation-codes/xsusa
58. Zhang, Q., et. al., 2014, "Global Variance Reduction for Monte Carlo Reactor Physics Calculations," Nuclear Engineering and Design, **280**, 76-85, 2014.

16.0   REVIEW COMMENTS AND DISPOSITIONS

Industry Comments	
Comment	**Disposition**
The presumption that the analysis uncertainties are dominated by the nuclear data uncertainties is, we believe, incorrect. With the current nuclear data libraries, modelling uncertainties, both with respect to the uncertainty in the "control parameters" but especially uncertainties introduced by modelling choices and methodologies are almost certainly more important than the uncertainties due to nuclear data.	This assertion assumes that reactor calculations are carefully customized to the reactor type of interest in order to maximize their predictability against the body of experimental data available from reactor operation and integral experiments. This results in reducing the contribution of modeling uncertainties against nuclear data.
The very significant correlations in the uncertainties for the multitude of different parameters at different conditions required at various stages in the calculation of the response may be difficult to identify and implement. Because of the presence of highly correlated uncertainties, the analysis will need to be very carefully controlled to ensure these correlations are not lost. We suspect the data manipulation requirements will be very expensive (and prone to error).	The correlations are identified using well-established algorithms designed to ensure that the randomized samples used by the UCF preserve the physics of the associated models.
The objective that using such a UCF should be possible as a routine exercise is also unrealistic. The complexity of all the components involved will definitely not allow this to be a routine process for many years. The streamlining and supporting documentation of the various processes will require several years of learning and refinement before this can become routine.	Past work which demonstrated application and the computational efficiency of UCF methods to realistic core reactor calculations, including LWRs and SFRs. A well planned effort, as guided by CANDU and UCF experts, will optimize the UCF implementation and its supporting documentations to ensure proper knowledge transfer and routine execution.
For a group of researchers already versed in the underlying UC methodologies, we expect the main obstacles will be the integration of the various codes into relatively seamless wholes with appropriate access to uncertainty information from the upstream codes (although there may also be some very real problems with the dimensionality reduction for realistic core configurations). This coordination of existing codes with added access to uncertainty information may be very	Expert CANDU researchers are expected to be involved in the planning and implementation stages of the UCF.

demanding, especially for a group of researchers primarily familiar with UC but with limited familiarity with CANDU physics (on-line refuelling, zone control, adjuster rods, fuel channel ageing effects, various gravity effects, etc.).	
P.11: "Specifically, we will assume that the Monte Carlo pointwise cross-section models provide perfect representation of the physics of neutron transport inside the CANDU core." This seems to relate to MCNP or KENO continuous-energy calculations. What about the multi-group Monte Carlo simulations performed by KENO? More precisely, how are we to assess how well the multi-group sensitivity profiles generated by KENO apply to MCNP or KENO-CE?	The uncertainties resulting from the use of multi-group cross-sections instead of continuous cross-sections must be quantified. Section 5.2.2 discusses preliminary analysis to determine the impact of the assumed flux shape on the lattice cell parameters. Specific strategies to assess the modeling errors in the sensitivity profiles will be developed as part of the UCF implementation.
It seems from the discussion in Section 3.1 that the nuclear data uncertainties included in the SCALE6 library covariance library include uncertainties for the resonance region. If so, how are these resonance uncertainties applied to specific self-shielding calculations (involving specific isotopic densities and geometry) at specific temperatures?	The reader may consult the SCALE's SAMPLER module for specifics on how the uncertainties in self-shielding calculations are propagated.
What about uncertainty in the fission product yields? Is there some method proposed to handle this, such as using uncertainty in some single or few representative fission products as surrogates for the more realistic multitude of fission products?	The standard UQ^p sampling approach, aided by ROM algorithms, is proposed here to propagate and prioritize the important contributors to this source of uncertainty.
P.14: If the rank r < dimension n, the implication is that not all inputs are independent (i.e., they end up being correlated by some constraint, for example, imposed by the neutron flux solution) and some relationship exists between them that would, in principle, allow one to write one or more of the inputs in terms of the others. If these relationships are non-linear, is the transformation to independent variables still useful? If so, does this mean that the original covariance matrix was not a good representation of the uncertainties in the various original parameters?	Details on the methodology that address these questions have been published elsewhere. The reader may consult Bang, et. al., 2012.

P.20: Here, the BE value seems to be determined as the average response from the sampled inputs. This is somewhat different from our usual definition of BE response, which corresponds to the response from the BE of each input parameter.	Agree. Correction is made to eliminate possible confusion with the standard NE use of BE term.
P.22: The need for "mapping of uncertainties" implies that the different operational conditions are not part of the original input uncertainties. Why is that?	Disagree. The operational conditions uncertainties are included in the uncertainty analysis.
P.24, mapping example: This is not at all clear. Step 1: What is $p(y, \alpha_o)$? Is this the uncertainty propagation with the low- or high-fidelity model? Or is this the propagation of the uncertainty due to the difference between the low- and high-fidelity models? Or is this even the propagation of the uncertainty due to the difference between the low- and high-fidelity models at fixed model input parameters, i.e., replacing the UQ^p sampling with an α_o "sampling"?	$p(y, \alpha_o)$ is defined on P.24 as the uncertainty in y evaluated with the high fidelity model at the conditions α_o.
Step 2: I do not understand this step either. Presumably the identification of the various input uncertainties and their PDFs would have been required for Step 1. What is the "Joint PDF"? I have difficulty visualizing a scatter plot depicting the distribution of say the cell Σ_a at a range α_1 of burnups plotted on the y-axis versus the distribution of Σ_a at a more limited set of burnups α_o on the x-axis. Is the idea that you would have such a scatter plot for each combination of α_o and α_1? Where do the high-fidelity calculations fit in? The response values plotted on the x-axis?	This discussion is intended to demonstrate the basic principles of mapping uncertainties. Specific details on the methodology as applied to CANDU analysis is expected to be part of the implementation plan for the UCF.
Would the representation of the mapping really require more of a 3D picture, with the x-axis representing α_o, the y-axis being the response (say, Σ_a) and PDF profiles being the third dimension? The mapping would then correspond to a parameterization of the PDFs as functions of α_o for extension to the more complete set of conditions α_1. It is still not clear how the high-fidelity calculations would fit in, unless the PDFs are specifically for the difference between the responses using the low- and high-fidelity models.	Details on the mapping algorithms, i.e., parametric vs. non-parametric approaches, are to be included in the UCF implementation plan.

P.28: "the basic assumption in most neutronic UQ studies is that nuclear reaction model parameters represent the main source of uncertainty for all downstream calculations." This is probably not the case for reactivity coefficients calculations, where the uncertainty impacts nuclear data are strongly correlated between different conditions.	Disagree. Nuclear reaction models and their associated parameters are the starting point for all downstream neutronic calculations.
PP.35-36: we have considerable skepticism about the generality of this Dimensional Reduction. It is true that for a given fuel design, the isotopics of the fuel will be strongly correlated between burnup points and that correlation depends primarily on only a few parameters. Also, the neutron spectrum will be strongly correlated with the local fuel burnup and a few other parameters such as moderator and coolant purity, moderator and coolant temperature, coolant density, fuel temperature, presence of fission products, local PT creep, etc. However, that still leaves a huge number of independent and potentially relevant degrees of freedom, particularly if spatial dependence of these various parameters is taken into account. A flux-squared or adjoint weighting of these various parameters would presume that a perturbation formalism (linearity) would be applicable and an unperturbed flux can be defined for the flux-squared weighting. It is unclear how Dimensional Reduction can capture all of the non-linear effects of self-shielding of multiple resonances from multiple nuclides at different burnups, without imposing some significant constraints on how the isotopics are related at different burnups, i.e., without imposing significant constraints on the response at different operating points. Does this effectively mean that a huge number of resonances can be combined into a few relevant degrees of freedom independent of the numerous other parameters that set the neutron flux environment? Is this realized because the nuclear data uncertainty is essentially largely irrelevant, i.e., the nuclear data are so good that the associated uncertainties are not important	The reader may consult with the numerous publications on ROM's application to nuclear reactor calculations, where similar issues in LWR reactor cores were identified and addressed by the ROM methodology.

to the calculation response or at least to the calculated response at different conditions? If so, are you really capturing all of the uncertainties or just some approximation to them?	
This reduction in complexity for the whole range of parameters consistent with reactor operation without loss of accuracy does stretch credibility. If Dimensional Reduction were truly reliable, the only need for all of our sophisticated and detailed models would be to use them to find the few hundred relevant degrees of freedom and the reduced order model. This is hard to absorb. If one of the responses is the bundle power distribution, it is difficult to imagine how you have fewer degrees of freedom than you have average channel burnups. It is true that bundle burnups are generally strongly correlated within a channel but burnups in individual channels are not nearly as strongly correlated. Are we to give up on channel refuelling times as being degrees of freedom and only analyse a specific core burnup distribution such as a time-average-equivalent burnup distribution?	Channel refueling times, and other operator-controlled parameters, are expected to be included as part of the analysis when determining the overall true dimensionality of the core simulator's model.
Is such a Dimensional Reduction valid only in the neighbourhood of a single operating point and to be repeated at multiple operating conditions? How would you determine the range (neighbourhood) of the DR applicability?	These specific issues were discussed in earlier publications, see for example Abdo, et. al., 2015, and Bang, et. al., 2012.
Is this dimensional reduction too ambitious? Dimensional reduction applicable to say, lattice modelling but not to the whole core?	The application of ROM methodology to both lattice cell and core wide calculations has been demonstrated in past work.
Figure 4.1 and discussion of Figure 4-3: The use of the single phrase "active subspace" to denote both the active degrees of freedom in the inputs space and their mapped responses in the response space is confusing when it comes to the discussion of Figure 4-3. Why do you need the gradients to determine the active subspace of inputs? Wouldn't a forward sampling also give you that? Or does the model need to be linear for the forward sampling to give the active subspace of inputs? This is not clear.	These specific details were discussed in previous publications. See for example Bang, et. al., 2012, and Khuwaileh, et. al., 2013.

Considerable dimensional reduction is already done for the full-core CATHENA modelling: 380 channels are typically represented by 28 channels (7 channels per core coolant pass). For the neutronics modelling, some similar dimensional reduction is achieved in the full core model by using only bundle burnups to discretize the burnup distribution as opposed to a burnup distribution based on the diffusion model mesh structure or a smooth functional form. The choice of two-energy groups is also a form of dimensional reduction. Does dimensional reduction primarily correspond to a means of transferring parameter uncertainty contributions to modelling uncertainties and limiting the range of states represented by the model?	These modes of reduction are mostly based on expert intuition of the physics models, which renders it difficult to quantify their associated errors. Dimensionality reduction however, as explained in this study, means identifying the states that are produced by the physics models due to the specific application set by the designer/operator using rigorous mathematical techniques that allows proper quantification and bounding of reduction errors.
Section 4.2.2, p.43: "Generate N samples of the parameters according to their expected ranges of variations" Are these samples intended to span the entire possible ranges of these input parameters or is it sufficient to be consistent with their PDFs?	The notion "expected ranges of variation" is employed here to cover a multitude of applications. For example, if one is interested in propagating uncertainties only, then the ranges of variation are consistent with the prior uncertainties determined by the prior PDFs. If however one is interested in building ROM that is applicable for the wide range of reactor states then the range of variation would be consistent with the expected range for the parameters values.
Section 5.1: The preparation of the ENDF library data is outside any scope of uncertainty characterization that would realistically be performed by analysts modelling the CANDU reactor. It is unrealistic to consider this as part of any feasible UCF for CANDU analysis.	This source of uncertainty needs to be performed only once just like the ENDF libraries are prepared only once, and hence they are not expected to be done routinely by CANDU analysts.
Section 5.2: With respect to CANDU analysis, this section corresponds to uncertainties introduced by the NJOY processing of various library data to determine, for example, a multi-group, multi-temperature library for use with WIMS-AECL or a continuous-energy, multi-temperature library for use with MCNP. A similar process is also used to prepare the multi-group library for use with KENO to support TSUNAMI analysis of CANDU. This is done at infrequent intervals and only a few	To streamline the UCF, these experts will need to be involved during the initial implementation to provide guidance on the matrix of conditions and the steps required to generate the reference multi-group cross-sections. These same conditions/steps will be used to propagate uncertainties using the UCF. This process will need to be done at the same frequency at which the reference cross-sections are generated by the cross-section experts at CANDU.

individuals within the CANDU industry have the training and experience required to do this.	
Section 5.2.1: This NJOY processing is also the analysis stage where molecular binding effects on the thermal scattering kernel and a temperature dependence to the nuclear data are introduced. It is unclear to me in what form the resonance information is present in the multigroup libraries. However, the propagation of resonance data uncertainties at different temperatures should be considered carefully.	The details on the inclusion of resonance uncertainties are model dependent. For an example on how resonance uncertainties are included in the SCALE code, consult Williams, et. al., 2013. As mentioned earlier, the UCF algorithms employ well-established sampling methods to ensure that the physics correlations are preserved, e.g., evaluations at different temperatures.
Section 5.2.2: The assumed flux shape will not only affect the multi-group cross section uncertainties, but also the multi-group cross sections themselves. That is, it could introduce a bias and the total variation in the multi-group cross sections with different assumed fluxes and simultaneous nuclear data variations would need to be examined.	That's the point made in sections 5.2.1, 5.2.2, and 5.2.3. Section 5.2.1 deals with propagating parameter uncertainties only, i.e., those propagated from the nuclear reaction model parameters. Section 5.2.2 deals with the bias generated by the assumed flux shape, which is treated as a source of modeling uncertainty. Section 5.2.3 attempts to estimate both sources of uncertainties simultaneously, taking into account that their correlation.
Section 5.2.2: It is not clear that the same multi-group cross section uncertainty will apply to the data at different temperatures for all isotopes. It is therefore difficult to see how appropriate lattice cell sensitivity profiles can be determined at different temperatures and then integrated with common-temperature multi-group uncertainties to propagate the uncertainties to reactivity coefficients. we suppose interpolation of the temperature-specific uncertainties is available as an option to provide an approximate evaluation.	The details of these algorithms will be developed during the initial implementation plan of the UCF.
Section 5.3: Lattice cell calculations do have control parameter uncertainties: solid material dimensions, compositions, and densities, including fresh fuel isotopics and other material impurities (and, perhaps, power density), and ageing effects on fuel channel geometry and tube materials. In fact, one would expect some of these control parameter uncertainties to be some of the most important uncertainties in the neutronic calculations.	Agree. All these uncertainties are considered part of the "control parameters" uncertainties. Their inclusion in the analysis is straightforward.

Section 5.3.1: Another source of uncertainty here will be the fission product yields as the fuel is irradiated.	Fission product yields and delayed neutron fractions are considered part of nuclear data for which uncertainties are to be propagated.
Section 5.3.1: Is the forward SA approach combined with the use of sensitivity profiles not practical here because of the many points to be investigated? Namely, many values of cell burnup, fuel temperature, coolant conditions, moderator temperature, heavy water purities, poison concentration, etc. Or is the forward SA approach impractical because one of the objectives will be to identify the main sources of uncertainty? In general, the number of cases to be studied for uncertainty quantification does not depend on the number of input parameters. The number of input parameters is only of importance if the main sources of the uncertainties are to be isolated.	Sensitivity analysis is needed only if one is interested in determining the contribution of the various sources of uncertainties to the total propagated response uncertainties. When the number of uncertainty sources become too high, the forward SA becomes impractical. When sensitivity profiles are combined with the forward SA approach however, one can reduce the number of required model executions using ROM techniques. If however one is only interested in estimating total response uncertainty without estimating the contributions of the various sources, then SA need not be performed, and one can rely on the standard forward UQ^p approach.
At this point in time, WIMS-AECL does not provide an adjoint flux solution. However, DRAGON does and would be suitable for this task. The author suggests the use of SCALE's NEWT code. The question would remain as to whether or not the somewhat different calculational methods used by DRAGON or NEWT would produce results with the same uncertainties as WIMS-AECL. However, if these alternative codes are only used to restrict the dimensionality of the parameters to be sampled, this difference may be irrelevant.	The different code is only used to restrict the degrees of freedom, however the code which is used as part of the standard calculational sequence, is the one used to propagate uncertainties. Abdo, et. al., 2015 has shown that errors resulting from the restriction of degrees of freedom can be reliably upper-bounded with overwhelmingly high probability. Therefore, one can use DRAGON or any other representative code with adjoint capability to reduce the dimensionality of the multi-group cross-section space. NEWT is proposed given its wide use and the recent introduction of the CRANE module which automates the process of active subspace generation.
Section 5.3.2: Various comparisons between results from the multi-group WIMS-AECL calculations and continuous energy MCNP calculations have been made in the past for different types of fuel lattices, primarily with respect to reaction rates, flux distributions and power distributions as opposed to the two-group cell parameters. The analysis proposed here would also evaluate the spatial homogenization approximation and the energy-group condensation.	Agree. The idea is to capture the discrepancies in the few-group cross-sections, and obtain a prior distribution for their variations that can be used in downstream core-wide calculations to account not only for cross-section uncertainties resulting from differential cross-section measurements but also from the modeling approximations.

Between Section 5.3 and Section 5.4: There is a missing section on uncertainty of few-group incremental cross sections, typically calculated with DRAGON. These may not be present in PWR and BWR calculations but they do form an important part of full-core CANDU simulations. They will introduce their own sources of uncertainty: assumed moderator conditions, fuel and fuel channel geometry modelling, assumed fuel channel conditions (including differences in burnup in fuel on one side of the device and on the other side of the device), aging assumptions (fuel channel and device itself), etc. These uncertainties will likely be highly be correlated with the lattice few-group cross section uncertainties.	The discussion assumes that the specific details of how the incremental cross-sections are calculated are considered part of lattice cell calculations. If two different codes are employed, i.e., one for the reference few-group cross-section generation, and another code for the incremental values, then one must employ the same train of random samples for the cross-sections and any control parameters that are common to the two codes, e.g. absorber rod configuration. This is a standard approach in any UQ study designed to ensure that the physics-introduced correlations between the two codes are preserved.
Section 5.4: Among the control parameters, there is also the need to consider the fuel burnup distribution, and various absorber rod configurations and LZC fill fractions. While the absorber rod configurations and LZC fill fractions may introduce only small uncertainties directly, they will indirectly introduce the uncertainty contributions from the incremental cross sections in different situations, correlated with the underlying lattice cross section uncertainties. What about the effects of an incompletely specified operating history?	Without availability of reliable prior uncertainties on the sources of uncertainties, it becomes difficult to assess the quality of the propagated uncertainties. To address this, one often uses conservative values when operating history information is not complete. Another approach is to employ sensitivity analysis to determine the key control parameters that contribute the most to the variability of the responses of interest. With this knowledge, one can refine ones estimate of the prior estimates for the control parameters that dominate the uncertainties of the responses of interest. This sensitivity analysis can be seamlessly done using the proposed UCF's ROM techniques. In regard to the correlations, they are preserved by the UCF sampling algorithm as discussed earlier.
Section 5.4.1: Again, as a result of the online refuelling in the CANDU reactor and the complicated and varied burnup distributions possible in the CANDU core, it is difficult to imagine that the number of active degrees of freedom can be reduced to the order of 100 or so unless one restricts the evaluation to specific core configurations such as a time-average-equivalent core.	Based on the experience with LWRs over multiple cycles of operation, the active degrees of freedom are consistently found to be in the order of 100, despite the large number of fuel lattice designs employed (e.g., BWRs) and the resulting wide range of isotopic composition at discharge. This behavior is a result of the careful design of the reactor and the fueling strategy, designed to flatten power distribution and reduce peaking. These design constraints result in reducing the degrees of freedom of the

	spatial distribution of the flux spectra and hence the formation and evolution of isotopic concentrations over the life of the fuel in the core. This behavior introduces strong correlation between the isotopic concentrations that persists over the life of the fuel in the core, and very insensitive to the initial fuel lattice design, because of the short mean free path of the neutrons.
Section 6.4.1: The WIMS-AECL nuclear data library is a binary file. This will pose a problem for the UC of current CANDU analyses based on current versions of WIMS-AECL.	This was identified as one of the challenges in the report, see section 6.2.3 for more details.
Section 6.4.2: What is meant by "two folds"? A factor of 2 or a factor of 100?	Yes, it is meant to be two orders of magnitude. It was changed in the report to clear confusion.
Figure 8.1: The CANDU 37-element bundle does not have 4-fold symmetry. For realistic calculations, I would hope that this figure does not represent the entire lattice model.	The development of a CANDU lattice model is outside the scope of this study. We have been provided with a fully detailed model for MCNP, but decided to employ a model by Oak Ridge and University of Tennessee for NEWT (Ref. available upon request) for similar scoping analyses. The results could be repeated seamlessly using the CRANE module in SCALE (currently available to the contractor and available upon request), which automates the process of ROM and UQ^p for SCALE neutronic models, for a detailed CANDU lattice model.
Table 8.2 (p.77): What are Kappa-fast-fission and Kappa-nu-fission? From another document found on the internet, it seems that in a similar notation, kappa represents the energy per fission (should be defined here).	Yes, that's correct. The definition is included in the document following Table 8.2
Table 8.3 (p.78): The uncertainty in the peak reactivity during a LBLOCA purely from 2-group parameter uncertainties resulting from the propagation of nuclear data uncertainties listed in this table at 4.5 GWD/MTU is difficult to believe. The uncertainty in the peak reactivity using uncertainties corresponding to mid-burnup fuel is quite different from the peak reactivity uncertainty based on uncertainties corresponding to fresh fuel or near-exit fuel. Furthermore, a significantly larger uncertainty in peak reactivity would be	The goal of this pilot study is to demonstrate the basic mechanics of the UC algorithms, highlight the relative importance of modeling and parameter uncertainties, and identify the possible challenges when applied to CANDU reactor analysis. To achieve that within the constraints of this study, numerous simplifying assumptions were made to limit the scope. Detailed analysis of these assumptions is therefore not attempted.

correlated with a significantly larger uncertainty in peak bundle power. That is not the case in this Table. Also, why are the values for the uncertainty in the peak bundle power identical to the uncertainties in the peak reactor power? Secondly, if the results from TSUNAMI are to be believed, the nuclear data contribution to the uncertainty in the full-core CVR in CANDU is only roughly 0.4 mk. The peak reactivity in the LBLOCA is achieved with the core only partially voided. What other feedback effects are important here? From the text, it does not seem that this uncertainty also includes the uncertainty in the delayed neutron fraction. Even if it did, one would not expect the variation in the peak reactivity just due to nuclear data uncertainties to be anywhere near 0.5 mk.	
Was this result determined by repeatedly performing the transient simulation or were certain reference transient results, such as the trip time and coolant conditions as a function of time, used in all cases? There are mitigating (non-linear) effects in a realistic analysis that make this peak reactivity relatively insensitive to the nuclear data. For example, if the nuclear data uncertainty results in a higher CVR or a smaller delayed neutron fraction, the transient will progress more quickly and result in a more rapid shutdown, such that the peak reactivity is reached sooner (at a smaller degree of voiding). The result is that, in a realistic transient simulation, the relative change in the peak reactivity is smaller than the relative change in the CVR.	Not applicable since thermal-hydraulic feedback was not considered in this study. Neutronic calculations are evaluated at fixed thermal-hydraulics conditions.
The uncertainty in the CVR cannot be determined from the uncertainties in the 2-group parameters at fixed conditions. The uncertainty in the CVR is determined by the uncertainty in the change in the 2-group parameters due to coolant voiding. The uncertainties in the 2-group parameters are highly correlated between the cooled and voided states. This correlation is absolutely critical to getting the correct result. For	Not applicable since the study does not calculate the CVR, it only calculates the core eigenvalue during the transient at fixed thermal-hydraulics conditions.

example, the *a priori* k-value uncertainties from nuclear data uncertainties reported by TSUNAMI for the cooled and voided time-average-equivalent CANDU core are both roughly 9.5 mk. However, the *a priori* uncertainty in CVR from these sources is only roughly 0.4 mk. The result shown in Table 8.3 leaves considerable doubt about how this correlation was taken into account.	
Section 12.0 (p.88): In the description of the Range-Finding Algorithm, some things are not clear. First of all, it is not always clear what vector space the various vectors correspond to, the input parameter vector space or the response vector space.	The discussion here is generic and aims only to explain the mechanics of randomization. If **A** represents a forward model, then p would represent the parameters, and w the responses (or the flux). If **A** represents the adjoint model, then the p represents the responses, and w the parameters. Although the discussion in this section implies **A** is a linear matrix operator, the analysis has been extended to include nonlinear operators. Consult Bang, et. al., 2012 for more details.
Secondly, in Step 5, "Pick t additional random vectors…" There is no indication what value t would initially have. (In Step 10, t_o is incremented by t.). The distinction between t and to+t seems to be rather loose in Steps 6 to 8.	This is one rendition of the algorithm. Many other variations are possible. In this example, t_0 and t are an initial guess for the rank and the additional number of code runs added if the rank estimate is too low.
There is no real definition of the phrase "Pick s random vectors" or "Pick t additional random vectors". Is the intent to span the whole range of input parameters or to sample the input parameters over there likely range. If the input parameters are not fully sampled in this range-finding algorithm, how can you claim that the probability of the reduced-order model being incorrect by more than ε is very small? This may not be a real problem in a linear problem such as is described here, but could be much more important in a non-linear problem.	The randomization here assumes that each parameter is sampled using its prior PDF, or its intended range of application, depending on the type of the parameter and the application of the reduced order model. If one is interested in propagating uncertainties of say the multi-group cross-sections, then the cross-sections are sampled from their prior PDFs. If however, one is interested in building a reduced order model that is representative of any fuel burnup, then the number densities for the various isotopes must be sampled over their expected range during operation. These ranges can be determined using a number of representative forward runs. This process has been demonstrated in recent work, and allows one to build an ROM for a general multi-physics model in a computationally efficient manner. See Bang, et. al., 2012, Khuwaileh, et. al., 2013, Abdo, et. al., 2015.

Canadian Nuclear Laboratories Comments	
Comment	**Disposition**
Regarding the Randomized Algorithm, which forms the core of the UCF. The reviewer comments: "It is completely reasonable. The algorithm he describes in a version of batch orthogonal matching pursuit, and it is provably convergence with specific probability for particular types of matrices. This was a very popular topic of research starting mid 2000s."	Agree.
When he says "[epsilon_user] can be made as small as the numerical precision of the computational model", this is true, but there is no free lunch. Asking for better precision will increase the dimension "r", and computer requirements will grow accordingly.	Agree. However for realistic LWR calculations, one finds that numerical precision of the calculations can be achieved with very small rank.
The algorithm Hany gives in Appendix C appears in Halko page 25, "Adaptive Randomized Range Finder". I have no reason to doubt the algorithm is correct.	
One minor variance: Halko page 25 states a failure probability of "1-m*10^-s" instead of "1-10^s" given by Hany step 13. It is possible that "the Contractor" is correct -- Halko notes the value he gives is a simplistic "pessimistic" estimate. In either case, it would be easy to take a couple additional samples to get the same result.	Agree. Halko requires an "m" in the failure probability expression because the randomized vectors employed are assumed to be normalized to one. However, in the UCF implementation, this assumption is relaxed, and all randomized vectors are sampled directly from their prior distributions. This allows one to get rid of the "m" dependence. See Abdo, et. al., 2015 and the references within for additional details.