Transactions of the Korean Nuclear Society Spring Meeting
Jeju, Korea, May 7-8, 2015

Development of a System Analysis Toolkit for Sensitivity Analysis, Uncertainty Propagation, and Estimation of Parameter Distribution

Jaeseok Heo * and Kyung Doo Kim
Korea Atomic Energy Research Institute
1045 Daedeok-daero, Yuseong-gu, Daejeon, 305-353, South Korea
**Corresponding author: jheo@kaeri.re.kr*

1. Introduction

Statistical approaches to uncertainty quantification and sensitivity analysis are very important in estimating the safety margins for an engineering design application. The methodologies have been widely used in many engineering fields for the last several years, but due to computational burden, they were not usually applicable to a multi-physics multi-scale system, e.g., nuclear power plants, that has multiple parameters and responses. For the complex system, it has been desired to develop system analysis methodologies with an efficient tool to conduct a cost-effective calculation.

This paper presents a system analysis and optimization toolkit developed by Korea Atomic Energy Research Institute (KAERI), which includes multiple packages of the sensitivity analysis and uncertainty quantification algorithms. In order to reduce the computing demand, multiple compute resources including multiprocessor computers and a network of workstations are simultaneously used. A Graphical User Interface (GUI) was also developed within the parallel computing framework for users to readily employ the toolkit for an engineering design and optimization problem.

2. Methods and Results

A general description is presented in this section by summarizing multiple packages of the system analysis methodologies with a GUI framework. Figure 1 shows the main page of this toolkit, which has deterministic and probabilistic approaches of data assimilation, forward uncertainty propagation methodology, Chi-square linearity test, sensitivity analysis, and Fast Fourier Transform Based Method (FFTBM).

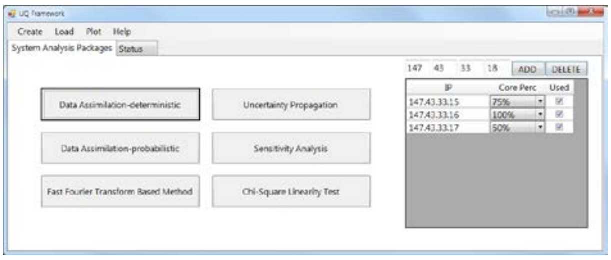

Figure 1. Main Page of the Toolkit

Once one of the packages is chosen and the IP addresses of the computers used for the simulations are entered, the toolkit directs users to the next step to define the parameters and import the input files of the simulation code for calculations. The defined parameters and simulation inputs will then be utilized by each package shown in the main page of the software. The computers, whose IP addresses are entered, are the computing resources for the parallel computation. The compute resources are typically a single computer with multiple processors/cores and an arbitrary number of such computers connected by a network. The following subsections present the system analysis methodologies implemented in the framework with some results obtained utilizing each package in the framework.

2.1 Data Assimilation – Deterministic

Inaccuracy in the prediction of physical phenomena can arise from multiple sources of uncertainties, including physical models, initial and boundary conditions, numerics, etc. Data assimilation procedure provides the means for integrating a "new" observed data to improve the model's prediction accuracy. It introduces a statistical approach for data adjustment indicating how prior knowledge can be updated by additional experimental data. During the data assimilation procedure, given measurements of the observables and the initial distributions of the model parameters, one calibrates the model by adjusting the parameter values to achieve better agreement between the measured and predicted values. The calibrated parameter distribution is called the *a posteriori* distribution of the parameters. The parallel computing framework includes data assimilation packages that solve both linear and nonlinear problems with the calibrated distribution of parameters derived based upon Bayes' theorem [1]. Sensitivity matrix is required for the deterministic approach, which can be calculated by perturbing each parameter and running the simulation code with the perturbed parameter. To accomplish this, users enter the parameter information on GUI including their distribution, initial mean value and standard deviation, etc. Parameters for the given engineering problem will then be defined and perturbed by clicking each parameter on the simulation input as shown in Figure 2.

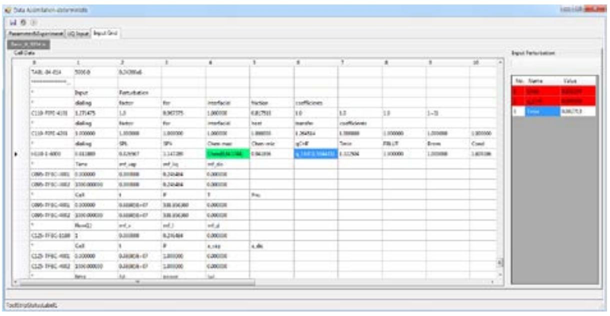

Figure 2. Input Perturbation Process

2.2 Data Assimilation – Probabilistic

In many cases of simulations, nonlinear and discontinuous behaviors can be observed due to the complexity of the system. The deterministic approach of data assimilation, based upon a first-order truncated Taylor series for the output, is inappropriate to treat this behavior owing to the nonlinear relationship between the simulation output and the parameters, hence the potential for a non-Gaussian nature of the calibrated parameter distributions. Thus the simulations that generate nonlinear system responses must be differentiated from those that behave relatively linearly. To address the nonlinear responses in data assimilation, a sampling approach such as MCMC [2], [3] can be employed by propagating the parameter uncertainties through the simulation model to predict the calibrated distributions of the parameters. The MCMC simulation is implemented in the toolkit with metropolis algorithm to efficiently compute the calibrated parameter distributions.

2.3 Uncertainty Propagation

When modeling and analyzing real physical systems with a computer code, one must consider not only nominal values of the calculation results but also their uncertainties. To calculate the distributions of the simulation results, one propagates the parameter uncertainties through the simulation model to predict uncertainties on the simulation output. If the parameter distributions are Gaussian and the calculation results respond linearly over the range of the parameter uncertainties, the covariance matrix of the simulation result can be readily calculated by the sandwich rule. Difficulty can arise from the physical process itself when the system is not linear. If the system is highly nonlinear, one can perturb multiple parameters simultaneously consistent with their uncertainties, and determine the simulation output through Monte Carlo simulation. In order to reduce the computing demand for the engineering system calculation, the Response Surface Methodology (RSM) is used to develop a surrogate for the complex system. The framework includes uncertainty propagation tools using both random sampling and the RSM.

2.4 Chi-Square Linearity Test

A linearity test is required to evaluate the degree of linearity of the target system. Chi-square linearity test is implemented in the framework for the purpose of determining the degree of nonlinearity of the simulation output with respect to the parameters. A linear system is a system that produces output equal to the linear combinations of the input variables. Thus if the parameter distributions are Gaussian and the system responds linearly over the range of the parameter values, then the calculation result distributions are Gaussian as well. In order to determine the linearity of the system, a random sampling was employed to develop distributions of the simulation results assuming Gaussian distributions for the parameters. For the Chi-Squared goodness of fit, the distribution of the simulation output is divided into *K* bins, each bin spanning a range of data values, and the test statistic is defined as:

$$\chi^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i} \tag{1}$$

where O_i and E_i are the observed and the expected frequencies for bin i, respectively. Using a Gaussian distribution to obtain the values for E_i and the code calculation to obtain the values of O_i, Chi-Square values can be obtained for the simulation output. If the degree of discrepancy between the simulation result and the Gaussian distribution is small, the Chi-Square value is also small, implying that the system is linear. If it is large, then the system is nonlinear.

2.5 Sensitivity Analysis

In many engineering design applications, sensitivity analysis is useful in identifying which of the design parameters have the most influence on the calculation results. This information is helpful prior to the system analysis as it can be used to remove model parameters that do not strongly influence the simulation results. For example, a ranked list of parameter influences can make forward uncertainty propagation more tractable over a reduced set of parameters. In an optimization problem, sensitivity information is useful in determining whether or not the response functions are robust with respect to small changes in the optimum design point. The sensitivity analysis package of this program supports this type of study through numerical finite differencing. The module was examined using Becker's post-dryout test [4], and the sensitivity coefficients for all simulation results with respect to the selected parameters are obtained.

2.6 Fast Fourier Transform Based Method

Recently the number of quantitative comparison between experimental data and calculation results in nuclear engineering has increased. The FFTBM is the tool widely used to quantify the accuracy of thermal hydraulic code calculations. It shows the discrepancies between measurements and predictions in the frequency domain. The FFTBM is included in the framework to validate the simulation results against the experimental data.

3. Conclusions

The goal of this work is to develop a GUI framework for engineering design and scientific analysis problems by implementing multiple packages of system analysis methods in the parallel computing toolkit. This was done by building an interface between an engineering simulation code and the system analysis software packages. The methods and strategies in the framework were designed to exploit parallel computing resources such as those found in a desktop multiprocessor workstation or a network of workstations. Available approaches in the framework include statistical and mathematical algorithms for use in science and engineering design problems. Currently the toolkit has 6 modules of the system analysis methodologies: deterministic and probabilistic approaches of data assimilation, uncertainty propagation, Chi-square linearity test, sensitivity analysis, and FFTBM. GUI was developed within the parallel computing framework for users to readily perform the system analysis.

REFERENCES

[1] Thomas Bayes, "An Essay Towards Solving a Problem in the Doctrine of Chance," *Philosophical Transactions of the Royal Society*, 53, 370-418, 1763.

[2] W. K. Hastings, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, **57**, No. 1, pp. 97-109, 1970.

[3] D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafeo and R. D. Ryne, "Combining Field Data and Computer Simulation for Calibration and Prediction," SIAM J. Sci. Comput. Vol. 26, No. 2, pp. 448-466, 2004.

[4] K. M. Becker, C. H. Ling, S. Hedberg and G. Strand, "An Experimental Investigation of Post Dryout Heat Transfer," Department of Nuclear Reactor Engineering, Royal Institute of Technology, KTH-NEL-33, 1983.