Nuclear Engineering and Technology 48 (2016) 684–701

Available online at ScienceDirect

Nuclear Engineering and Technology

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

Original Article

Implicit Treatment of Technical Specification and Thermal Hydraulic Parameter Uncertainties in Gaussian Process Model to Estimate Safety Margin

*Douglas A. Fynan and Kwang-Il Ahn**

Korea Atomic Energy Research Institute, 989-111 Daedeokdaero, Yuseong-gu, Daejeon, 305-353, South Korea

ARTICLE INFO

Article history:
Received 14 September 2015
Received in revised form
21 December 2015
Accepted 13 January 2016
Available online 9 February 2016

Keywords:
Gaussian Process Model
Large-Break Loss-of-Coolant Accident (LBLOCA)
Success Criteria
Safety Margin

ABSTRACT

The Gaussian process model (GPM) is a flexible surrogate model that can be used for nonparametric regression for multivariate problems. A unique feature of the GPM is that a prediction variance is automatically provided with the regression function. In this paper, we estimate the safety margin of a nuclear power plant by performing regression on the output of best-estimate simulations of a large-break loss-of-coolant accident with sampling of safety system configuration, sequence timing, technical specifications, and thermal hydraulic parameter uncertainties. The key aspect of our approach is that the GPM regression is only performed on the dominant input variables, the safety injection flow rate and the delay time for AC powered pumps to start representing sequence timing uncertainty, providing a predictive model for the peak clad temperature during a reflood phase. Other uncertainties are interpreted as contributors to the measurement noise of the code output and are implicitly treated in the GPM in the noise variance term, providing local uncertainty bounds for the peak clad temperature. We discuss the applicability of the foregoing method to reduce the use of conservative assumptions in best estimate plus uncertainty (BEPU) and Level 1 probabilistic safety assessment (PSA) success criteria definitions while dealing with a large number of uncertainties.

1. Introduction

Safety margin is an important concept for nuclear power plant (NPP) design and safe operation. Adequate safety margin ensures that the plant design can withstand transients and accidents without fuel damage and the release of radionuclides into the environment. Operationally, safety margin provides flexibility allowing for optimization of plant operations and maintenance, improving the safety, performance, and economics of the plant. Accurate characterization of safety margin has become increasingly important as many older NPPs seek power uprates changing the design basis.

* *Corresponding author.*
E-mail address: kiahn@kaeri.re.kr (K.-I. Ahn).
http://dx.doi.org/10.1016/j.net.2016.01.016

A detailed safety analysis must be performed to determine the safety margin of the NPP. The safety analysis can include operational and experimental data from scaled separate effects tests and integral experimental facilities, design basis accident (DBA) analysis, and probabilistic safety assessment (PSA). Deterministic simulation of transients using best-estimate thermal hydraulic computer codes is commonly used. However, accurate characterization of the safety margin for all plant states and configurations over the plant lifetime and all possible accident scenarios is extremely challenging, so practical approaches such as incorporating conservative and bounding assumptions must be implemented. Furthermore, if large changes to the plant design basis such as a power uprate or many small additive changes occur, a potentially large number of safety analyses including computer simulations must be redone. Accounting for the uncertainties of the computer models adds an additional layer of complexity. Best estimate plus uncertainty (BEPU) methodologies have been developed to address uncertainties in DBA analysis.

In this paper, we propose a methodology to make realistic estimates of the safety margin, reducing the need for excessive conservative and bounding assumptions in a safety analysis. The methodology uses best-estimate computer models to simulate a large (but manageable) number of transients that span a range of possible NPP safety system configurations and timing of safety system actuation to resolve a spectrum of plant responses during an accident. Simultaneously, many code input parameter uncertainties are sampled representing technical specifications, limiting conditions for operation, and thermal hydraulic model parameter uncertainties. The methodology adopts the Gaussian process model (GPM) to serve as a surrogate model used to characterize the safety margin. The GPM performs multivariate regression on the dominant input parameters, safety system configuration and sequence timing, for a predictive model of the safety parameter of interest, while modeling the other uncertainty contributors implicitly as measurement noise terms. The safety parameter probability distribution that can be used to quantify the safety margin is expressed as the GPM mean function and local uncertainty bounds defined by the GPM prediction variance. The unique features of the GPM as a nonparametric regression method with an automatic quantification of prediction model uncertainty are key aspects of the methodology.

This paper is organized as follows. Section 2 provides an overview of GPMs for nonparametric regression analysis. Specifically, the unique features of the GPM including the prediction variance, covariance function selection, and implementation issues are discussed. Section 3 presents the MARS code [8] model of the reference plant (Hanul Units 3&4, formerly known as Ulchin Units 3&4; one of the optimized power reactor, OPR1000, series) used to simulate the injection phase of a large-break loss-of-coolant accident (LBLOCA) serving as the demonstration application for the methodology. Section 4 presents the best-estimate simulation data of the LBLOCA and the training process of the GPM. Section 5 presents an analysis and discussion of the safety margin results derived from the GPM. Finally, limitations of the proposed methodology, future work and applications, and some concluding remarks are provided in Section 6.

2. Gaussian process model

In the context of thermal hydraulic simulations of NPP accidents, the best-estimate code or model can be interpreted as a general nonlinear function of the form

$$y = h(\mathbf{x}) \tag{1}$$

with a vector of inputs $\mathbf{x} = [x_1, x_2, \ldots x_p]^T$ and a limiting safety parameter of interest as the output variable y. The actual code output from a simulation contains time histories of many plant parameters from which any number of limiting safety parameters can be obtained. However, for clarity we will consider only a single output. A regression analysis can be performed on a dataset from simulations $\{X = [\mathbf{x}_1, \ldots \mathbf{x}_n], \mathbf{y}\}$ to estimate the functional relationship between the input variables and output variable. The dataset used in regression is called the training set. The regression function becomes a surrogate model to the best-estimate code and can be evaluated many times with minimal computational cost to obtain large samples used for uncertainty quantification, design optimization, safety margin characterization, etc.

GPMs are a popular class of surrogate models that can be used for multivariate regression. Rasmussen and Williams [1] and their associated GPML code package [2] are prominent resources on GPMs. Chapter 4 of Yurko [3] provides a nice summary of and practical implementation recommendations for GPMs. For consistency, we will generally follow the notation in Rasmussen and Williams [1] to present the mathematical formulation of the GPM and describe the unique features in the context of regression and characterizing large datasets from computer simulations.

2.1. GPM mean function and prediction variance

The GPM is unique among regression methods because it defines a *predictive distribution* of the dependent variable y at any input test location x_*. The GPM is fully defined by the mean function and prediction variance. The predictive distribution is assumed to be Gaussian parameterized by the mean function and prediction variance. The mean function and prediction variance are

$$\bar{y} = \bar{f}(\mathbf{x}_*) = \mathbf{k}_*^T \left(K + \sigma_n^2 I\right)^{-1} \mathbf{y} \tag{2}$$

$$V[f(\mathbf{x}_*)] = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^T \left(K + \sigma_n^2 I\right)^{-1} \mathbf{k}_* \tag{3}$$

The predictive distribution for y is

$$y \left| \mathbf{x}_* \sim N\left(\bar{f}(\mathbf{x}_*), V[f(\mathbf{x}_*)] + \sigma_n^2\right)\right. \tag{4}$$

From the perspective of conventional regression analysis, the mean function of Eq. (2) can be interpreted as the regression function approximating Eq. (1). The prediction variance of Eq. (3) is an empirical estimate of the GPM prediction uncertainty derived from the data measurement noise variance σ_n^2, density of the training data set, and the complexity of the inputs/output relationship estimated by the GPM. Although it is not necessarily required, the GPM usually assumes a zero mean so the vector of data output **y** has been shifted by its

sample mean. The sample mean is added back in after the evaluation of Eq. (2).

The fundamental building block of Eqs. (2) and (3) is the covariance function defining the covariance between pairs of function evaluations from a user-defined basis function or kernel of the inputs

$$\mathrm{cov}\left[h(\mathbf{x}_q), h(\mathbf{x}_r)\right] = k(\mathbf{x}_q, \mathbf{x}_r) \tag{5}$$

$$k(\mathbf{x}_q, \mathbf{x}_r) = \sigma_f^2 \exp\left[-\frac{1}{2}(\mathbf{x}_q - \mathbf{x}_r)^T \Lambda^{-1}(\mathbf{x}_q - \mathbf{x}_r)\right] + \sigma_c^2 \tag{6}$$

$$\Lambda = \mathrm{diag}\left(r_1^2, \ldots, r_p^2\right) \tag{7}$$

For this study, we adopt the squared exponential covariance function, which assigns covariance using a simple distance-based metric.

Although a very large number of covariance functions including exotic options exist [1,2], the squared exponential covariance function is probably the most widely used in GPMs, and from our experience it is flexible and performs well. Perhaps the most advantageous feature of the squared exponential function is the length scale parameter r_i of Eq. (7) for each input dimension that can naturally lead to automatic relevance determination [1,4]. The length scale parameters, when inferred from the training set, are similar to sensitivity coefficients. The scaling factor σ_f^2 is the signal variance. A small constant term $\sigma_c^2 \ll \sigma_f^2$ is added for numerical stability issues when performing the matrix inverse. One percent of σ_f^2 is an appropriate value for σ_c^2.

The covariance matrix K of the training set defines the covariance between all n training data points. The entries of K are

$$K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j); \{i, j = 1, \ldots, n\} \tag{8}$$

The covariance between the training data and a test point is the vector

$$\mathbf{k}_* = [k(\mathbf{x}_*, \mathbf{x}_1);\ k(\mathbf{x}_*, \mathbf{x}_2); \ldots;\ k(\mathbf{x}_*, \mathbf{x}_n)] \tag{9}$$

Evaluating the mean function and prediction variance involves matrix and vector operations with Eqs. (8) and (9). First the inversion of the $n \times n$ matrix $(K + \sigma_n^2 I)$ must be performed requiring $O(n^3)$ floating point operations. Note that this operation only needs to be performed once and the matrix inverse is stored. A practical upper limit for modern (desktop) computers on n is $O(10^4)$ for matrix inversion, so the number of training points must be less than 10,000. In reactor safety analysis using best-estimate codes, a few hundred to a maximum of 1,000 simulations and corresponding data points would be generated for a realistic study of any particular transient, so the numerical algebra requirements should not be an issue. Only for specific benchmarking purposes would more than a thousand simulations of a particular transient with a best-estimate code be justified. However, some experimental databases, such as for heat transfer or cross-section measurements, may contain tens of thousands of data points. In these types of cases, the training set size could pose challenges. Rasmussen and Williams [1] discuss approximation methods for GPM regression on large datasets. For the mean function, the matrix-vector product $O(n^2)$ involving the training outputs $\mathbf{y}$ only needs to be performed once. To make a prediction at the test point, the vector $\mathbf{k}_*$ must be constructed and vector product $O(n)$ with the stored $(K + \sigma_n^2 I)^{-1}\mathbf{y}$ vector performed. The prediction variance at a test point requires $O(n^2)$ because every $\mathbf{k}_*$ is uniquely determined based on the distances of the test point to every training point. The GPM retains all of the training data within the model and uses the data when making predictions, whereas the training data are discarded after the initial regression is performed and the parameter values of the regression function are obtained in conventional regression analysis.

2.1.1. *Asymptotic analysis of GPM mean and predictive variance*

Following the discussion of Eqs. (2) and (3) as vector and matrix operations, Eq. (2) can be recast as

$$\bar{f}(\mathbf{x}_*) = \sum_{i=1}^{n} \alpha_i \mathbf{y}_i \tag{10}$$

where α_i is the i^{th} entry of the row vector $\mathbf{k}_*^T (K + \sigma_n^2 I)^{-1}$. Eq. (10) shows that the GPM mean is simply a weighted average or data smooth of the training data. The weights are local due to the distance-based squared exponential covariance function. The GPM mean function is a form of locally weighted nonparametric regression [5].

An asymptotic analysis can provide useful insights on the local weights α_i and behavior of the mean function as a data smoother. For a noiseless problem with $\sigma_n^2 \to 0$, in the limit a test point is at an infinite distance from all training points, $|\mathbf{x}_* - \mathbf{x}_i| \to \infty$ for all $\{i = 1,\ldots,n\}$, Eq. (9) becomes the zero vector and all $\alpha_i \to 0$. This result just says if two points are far apart, the outputs are uncorrelated (covariance is zero). The best estimate of the output value at the distant test location in the absence of any additional information is the sample mean of the training data. In the limit as the test point approaches the i^{th} training point $|\mathbf{x}_* - \mathbf{x}_i| \to 0$, $\alpha_i \to 1$ and $\alpha_{j \neq i} \to 0$, and the prediction is equal to the training point y_i. The transpose of Eq. (9) is equal to the i^{th} row of K if $\mathbf{x}_* = \mathbf{x}_i$. From the properties of the matrix inverse, $KK^{-1} = I$ so $\alpha_i = 1$. When a test point lies in the vicinity of training points, the largest weight is assigned to the closest y_i and the weights decay exponentially for distant training points.

In the limit the data noise becomes large, $\sigma_n^2 \to \infty$ or $\sigma_n^2 \gg \sigma_f^2$, the GPM mean will return the sample mean for all predictions. With the data noise variance added to the main diagonal of the covariance matrix, the leading term of the determinant used in the calculation of $(K + \sigma_n^2 I)^{-1}$ is $(\sigma_n^2)^n$. The inverse of a matrix A is defined as $A^{-1} = (1/|A|)C$ so $(\sigma_n^2)^n$ is in the denominator. Even if $\mathbf{x}_* = \mathbf{x}_i$, the product of $\mathbf{k}_*^T$ with the i^{th} column of $(K + \sigma_n^2 I)^{-1}$ is $<< 1$. In this case, the measurement noise has the effect of masking any covariance between data points. When the ratio of the signal variance to noise σ_f^2/σ_n^2 is not at an extreme, the weights are determined by the distance between the test point and training points and the dampening effect of the noise variance. When noise or random error is present in a dataset, the ratio can be adjusted to prevent over fitting, a benefit of the GPM as a data smoother. In the mean function, the signal variance and measurement noise

variance are tuning parameters of the data smoother and do not have a physical meaning with respect to the output parameter y. The signal variance favors local smoothing of data for resolving nonlinear behavior by interpolating between adjacent points in contrast to the measurement noise variance favoring global averaging of the dataset. The signal-to-noise ratio reflects the balance between local interpolation and global smoothing performed by the GPM.

The leading term of the prediction variance in Eq. (3) is always positive and evaluates to the signal variance σ_f^2. For all cases, if $|\mathbf{x}_* - \mathbf{x}_i| \to \infty$, the second term in Eq. (3), $\mathbf{k}_*^T(K + \sigma_n^2 I)^{-1}\mathbf{k}_*$, vanishes because Eq. (9) is the zero vector. The prediction variance is maximized when the test point is far from training points. This result also gives the physical interpretation of the signal variance as a global measure of the expected variation of y in the input space. Signal variance reflects the range of $h(x)$ over $\Delta\mathbf{x}$ when $\Delta\mathbf{x}$ is on the order of the characteristic length scales r_i. If Eq. (1) is slowly varying, the prediction variance will be small because the training set data will only show small deviations from the sample mean. Conversely, the prediction variance is minimized and evaluates to zero when $\mathbf{x}_* = \mathbf{x}_i$ and the measurement noise is zero. If an exact data point already exists at an input location, all information is known about the output; there is no uncertainty about the output. In the limit that the measurement noise is large, prediction variance is also maximized because the training data give very little information about the function. With uncertain data, the predictive capability of the model should also be uncertain.

2.2. *Learning hyperparameter values from training data*

The GPM parameters $\{\Lambda, \sigma_f^2, \sigma_n^2\}$ are referred to as hyperparameters to distinguish the GPM from conventional regression models that parameterize the input/output relationship with an assumed functional form. The GPM hyperparameters parameterize the local weights used in the data smooth. The hyperparameter values should be learned from the training data using Bayesian techniques. Much of the GPM research is focused on Bayesian inference methods for hyperparameter learning such as Markov chain Monte Carlo and gradient-based optimization; see Ch. 5 of Rasmussen and Williams [1]. For this study, we adopt the leave-one-out method implemented in the GPML code [2].

Learning the optimal hyperparameter values from training data is a powerful feature of the GPM as a nonparametric multivariate regression method. The inverse of the length scale parameters are semi-quantitative sensitivity coefficients so the sensitivity information is automatically calculated during GPM training. Automatic relevance determination provided by the GPM eliminates the need to perform a separate formal sensitivity analysis. However, Bayesian inference and the associated algorithms are mathematically complicated and cumbersome, and the results obtained do not always make physical sense. From our experience with GPM regression on datasets from thermal hydraulic simulations, overfitting and non-informative models can be issues due to gradient-based methods optimizing to local maxima and minima in the multi-dimensional space. Any GPM and inferred hyperparameter values *must be properly verified through additional statistical analysis and engineering judgment*. We always perform additional simulations for an independent cross-validation dataset and verify whether the magnitudes of the hyperparameters are consistent with the transient and thermal hydraulic phenomena being studied. Cross-validation procedures and a simple dimensional analysis for verifying the length scale parameters will be demonstrated in Section 4.

2.3. *Proposed methodology for implicit treatment of uncertainties in GPM*

The measurement noise variance term of the GPM allows measurement uncertainty present in experimental data to be explicitly incorporated into the regression and prediction variance. Computer experiments are generally deterministic and repeatable. The code will always yield the same output for a set of inputs up to machine precision so the output data are exact with no uncertainty. When surrogate models are developed to emulate computer experiments, one figure of merit used to measure the performance of the surrogate is the regression error and value is often placed on models that can reproduce the training data with high accuracy. For sophisticated computer codes with many inputs, increasingly complex surrogate model structures can be selected to achieve agreement with the training data. However, the pitfall of such an approach is that one blackbox function, the computer code, is analyzed with a second blackbox function, the complex surrogate, which may suffer from overfitting to the training data resulting in poor predictive capabilities.

For the current study, we take a slightly different approach to regression analysis and expectations of the information we want to obtain from the surrogate. First, we recognize that, while computer results are exact, the model is only an approximation or best estimate of the underlying physics and system being studied. As a result, the surrogate is at least two levels of approximation from the truth. We expect a reasonable level of accuracy from the surrogate so that engineering decisions can be made about the system with the information, but the surrogate should not be expected to reproduce code results with precision. If a precise quantity is needed, for example the sensitivity of a system parameter to an empirical coefficient in a specific sub-model of the code, then the computer code should be used directly.

Secondly, there are many input uncertainties to account for in a computer simulation, including nodalization and numerical error of finite differencing methods, material property data, coefficient values of empirical sub-models, decay heat data, etc. The BEPU methodologies provide a systematic way of quantifying these types of uncertainties in the context of DBA analysis with thermal hydraulic codes. The NPP being studied is not a static system so the boundary conditions and plant states used to initialize a simulation are also uncertain. The parameters defining a transient and system response can span a range of values such as break size of a LOCA, availability or configuration of safety systems, and timing of safety system actuation or failure. These latter types of parameter uncertainties are represented in Level 1 PSA event trees (ETs). ET branches represent different safety system configurations, and deterministic simulations are performed to determine if the sequences lead to core damage, or alternatively, the

success criteria are defined for the fault tree (FT). In the conventional ET/FT approach, accounting for sequence timing uncertainty across the many ET branches is difficult so conservative or bounding approaches are used. PSA can benefit from more rigorous approaches to handle uncertainties as they relate to thermal hydraulic simulations that support the probabilistic framework of ET/FT analysis.

We propose a new methodology to simultaneously consider the code input parameter uncertainties addressed in BEPU and the technical specifications, safety system configuration, and sequence timing uncertainties present in Level 1 PSA success criteria and ET branching. The methodology uses best-estimate computer models to simulate a large but manageable number of transients with all uncertainty contributors sampled for each run. The GPM will be the data synthesis tool to process the information contained in the simulation data. The GPM performs multivariate regression on the dominant input parameters, safety system configuration and sequence timing, for a predictive model of the system parameter of interest. The regression function only explicitly considers a subset of the input parameters sampled for the code simulations. The other inputs, such as the thermal hydraulic parameters and minor technical specification ranges, are not completely eliminated from the regression but rather their contribution to the output is lumped into the measurement noise variance term of the GPM. These inputs are implicitly represented in the GPM as measurement noise. In this manner, we are interpreting the code output as an experimental measurement with measurement error. The measurement error is quantified by the noise variance value learned during the GPM training. The resulting GPM is a predictive model of the system parameter through the mean function with local uncertainty bounds defined by the prediction variance.

3. Description of LBLOCA model

The LBLOCA is an extensively studied DBA. In the present work, we study the LBLOCA in the context of both BEPU and Level 1 PSA where thermal hydraulic system codes are used to simulate the transient. The methodology combines sampling of input parameter uncertainties representative of BEPU with many configurations of the safety systems which are considered in the Level 1 PSA ET. The GPM is the data synthesis tool of the simulation data and will be used to characterize the safety margin. This section describes the relevant safety systems of the reference NPP to apply the aforementioned method, the MARS code model to simulate the LBLOCA, and the important plant parameters and thermal hydraulic phenomena.

3.1. *Overview of the reference plant safety systems*

Hanul Units 3&4 are representative optimized power reactors 1000 MW_e (OPR1000), formerly known as the Korean Standard Nuclear Power Plant. The OPR1000 is a two-loop pressurized water reactor design with two cold legs (CLs) and one hot leg per loop and a rated power of 2815 MW_{th}. The safety injection (SI) systems designed to mitigate the LBLOCA are one safety injection tank (SIT) per CL, a high-pressure safety injection system (HPSI) with two pumps, and a low-pressure safety injection system (LPSI) with two pumps. The SITs are predominantly passive components and automatically inject water when the reactor coolant system (RCS) pressure falls below set points. Assuming offsite power is lost, the HPSI and LPSI pumps actuate following the start and loading of the emergency diesel generators (EDG) after the safety injection actuation signal (SIAS) is received on low pressurizer pressure. The flow from each HPSI pump is split to all four CLs safety injection headers while flow from each LPSI pump is split to only two CLs. From their initial configurations, the active SI systems are designed to operate in injection mode until the refueling water tank (RWT) is depleted and recirculation mode is initiated.

The Hanul 3&4 success criteria for the injection phase of the cold-leg LBLOCA are to inject water through at least two of three intact CLs through two of three SITs and inject RWT

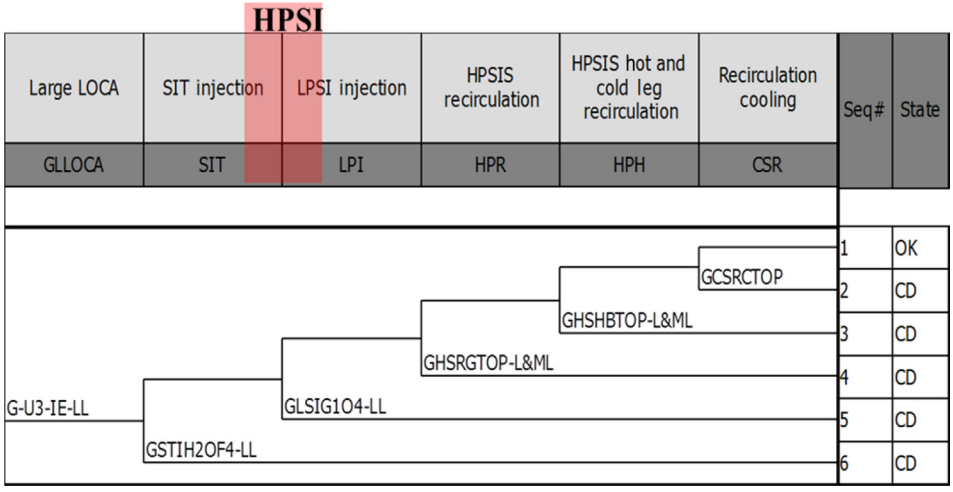

Fig. 1 – Hanul 3&4 LBLOCA event tree with potential HPSI branch. HPSI, high-pressure safety injection system; LBLOCA, large-break loss-of-coolant accident; LPSI, low-pressure safety injection system; SIT, safety injection tank.

water through at least one of three intact CLs using one of two LPSI pumps [6]. The success criteria do not credit any flow to the broken loop. The LPSI success criterion represents the operating pump that is aligned to the broken loop with one intact CL. Fig. 1 shows the LBLOCA ET. HPSI during the injection phase is not included in the ET although the HPSI system is designed to automatically actuate and the HPSI system must be available for recirculation cooling. HPSI availability can lead to injection phase success even if the LPSI system is completely failed for some LBLOCAs so HPSI should be considered in realistic studies [7].

3.2. MARS code model for Hanul 3&4

The MARS code version KS1.3 [8] is used to simulate the LBLOCA with the Hanul 3&4 input model [9] as a representative OPR1000. MARS is a thermal hydraulic system code based largely on RELAP5/MOD3.3 and has been developed by the Korea Atomic Energy Research Institute for best-estimate analysis of light water reactor transients. The reactor coolant system is modeled by 250 hydrodynamic volumes, 280 hydrodynamic junctions, and 259 heat structures. The reactor core is modeled with two coolant channels representing the hot rod channel and a core-average rod channel. Each channel has 12 axial nodes and the power distribution is set as the top-skewed cosine shape specified for LBLOCA analysis in the final safety analysis report (FSAR) [10]. For the hot rod, the linear heat generation rate in Node 8, the power peak location, is set to the limiting condition for operation of 13.9 kW/ft.

Safety systems are modeled as a series of time-dependent volumes and junctions controlled by trip logic interfacing with the RCS at the injection headers on each CL. The HPSI and LPSI pumps are time-dependent junctions with the flow rate calculated from a lookup table as a function of the injection header pressure. The lookup tables can be sampled to interpolate between the maximum and minimum rated flow rate curves for each pump type. The RWT is modeled as time-dependent volumes for each pump junction and the water temperature is user defined. The SITs are modeled as accumulator components. Trip logic calculates the SIAS and initiates SIT and HPSI/LPSI at the appropriate set points and time delays. Fig. 2 shows the SI configuration of the MARS model at a cold leg. Volume 396 is the SIT and junctions 385 and 387 are the HPSI and LPSI pumps, respectively. Unique trip numbers control each component so many configurations of pump and valve failures can be easily modeled representing ET branches.

A double-ended guillotine break is modeled to occur in CL 1A between the SI header and reactor vessel inlet. The containment volume is not explicitly modeled but time-dependent volumes serve as sinks for the mass and energy flowing from the break junctions. The time-dependent containment pressure response is input as a lookup table. Any SI flow to cold leg 1A spills into the containment through the break.

3.3. LBLOCA reference cases

Three LBLOCA reference cases are simulated with the MARS model representing configurations of HPSI and LPSI injection. The break is modeled in Loop 1A and the LPSI is assumed to inject into Loop 1B, the intact leg of the broken loop. Case 1 is two of two HPSI pumps injecting to three of three intact cold legs and no LPSI injection. The total injection flow rate to the RCS is approximately 82 kg/s, representing the minimum rated flow from the pumps. Case 2 is the current success criteria of one of two LPSI pumps injecting to one of three intact CLs at the minimum rated flow of approximately 138 kg/s. Case 3 is one of two SI trains consisting of one LPSI pump injecting to one of three intact CLs and one HPSI pump injecting to three of three intact CLs at the minimum rated flows for a total injection flow rate of approximately 179 kg/s. For all cases, two of three SITs are assumed available to inject to intact CLs of Loop 1B and Loop 2A.

The technical specification for the EDG is to start on either loss of offsite power or SIAS and warm-up in 10 s followed by the loading sequence. The HPSI pumps load at 5 s followed by the LPSI pumps at 20 s. The pumps are designed to accelerate to full speed in 5 s. An uncertainty range of 15 s to 30 s for pump start delay time is usually assumed for best-estimate LBLOCA analysis. For the reference cases, all pumps are assumed to start with a 30-s delay after the receipt of the SIAS.

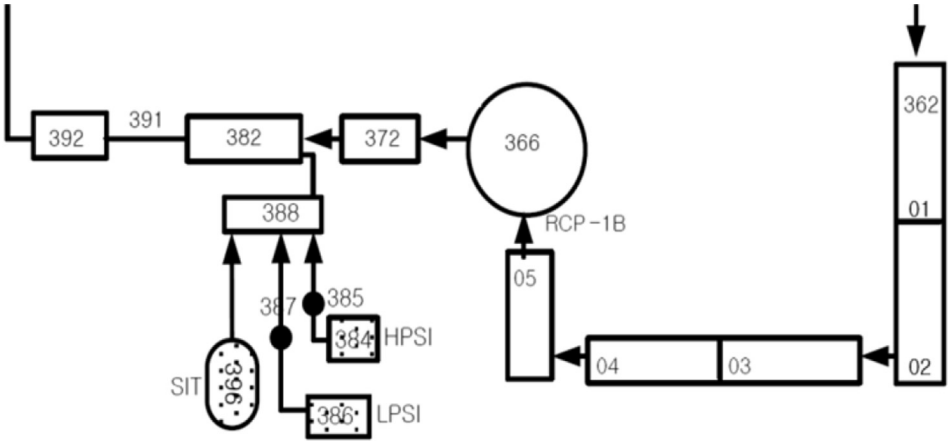

Fig. 2 – Nodalization of cold leg and SI systems in MARS model.

3.3.1. *Thermal hydraulic phenomena*

Fig. 3 shows the time-dependent axial clad temperature profiles for the hot pin. The active core is subdivided into 12 axial nodes. In ascending order, Node 1 is the bottom of the active core and Node 12 is the top of the active core. During the injection phase of the LBLOCA, the thermal hydraulic conditions in the RCS progress through three major phenomenological regimes: the blowdown, refill, and reflood. The blowdown occurs immediately following the break at the start of the transient with the RCS rapidly depressurizing when the core inventory flashes to steam and exits the break. The core fission power rapidly drops from the large negative void coefficient of the moderator. Due to the large amount of stored energy in the fuel and the decrease in heat transfer to the coolant, the fuel and cladding heat up, as shown in Fig. 3, by the large increase in clad temperature from 0.0 s to 20 s.

The refill is indicated by the clad temperature peak occurring between 20 s and 100 s in Fig. 3. The SITs inject very large mass flows of cold water into the RCS from approximately 16 s to 85 s. The SIT inventory refills the largely voided volumes of the CL piping, downcomer, core lower plenum, and the active core. The decrease in clad temperature by the end of the refill is the result of declining decay power and the increased flow rate of coolant through the core. The SIAS is sent at 6.9 s so the active SI systems begin to inject during the refill phase at 36.9 s assuming a 30-s delay. The flow rates delivered by the HPSI and LPSI pumps are only a small fraction of the SIT flow rates so the SITs are the dominant safety system during refill.

Figs. 4 and 5 show the collapsed water level in the downcomer and core for each reference case. During blowdown, both levels drop rapidly as most of the RCS voids and then the levels are restored during refill. After the SITs deplete at 85 s, the levels drop again at rates approximately inversely proportional to the effective flow rate of the active SI systems. Over time, the continuous injection of cold water from the RWT restores downcomer and core water levels. During reflood, the upper one-third of the active core can experience two-phase flow and less efficient heat-transfer regimes, such as transition boiling, film boiling, and vapor heating, leading to the heatup of the fuel and cladding. Fig. 3 shows Case 1 with only two of two HPSI pumps available experiencing heatup

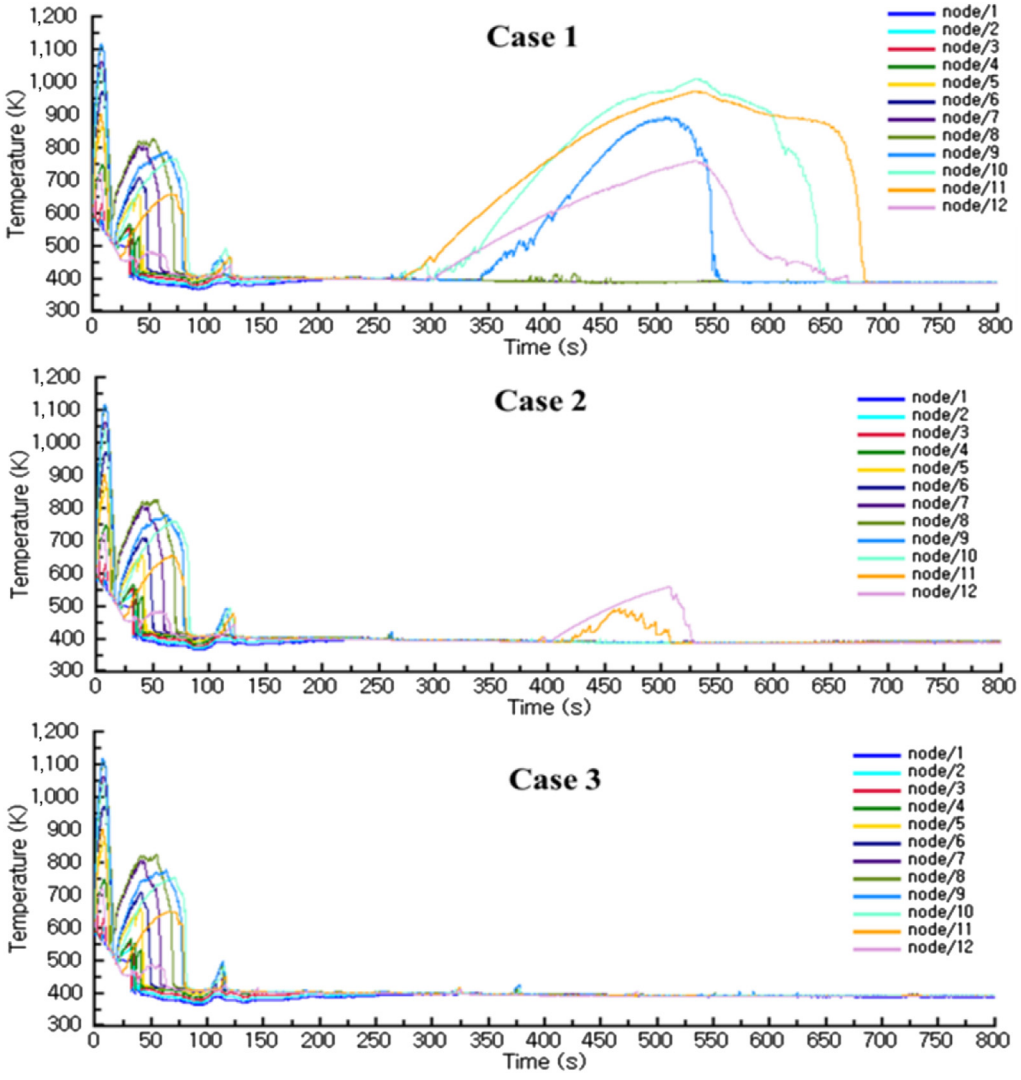

Fig. 3 – Axial clad temperature profiles of hot pin during injection phase for different safety injection system configurations.

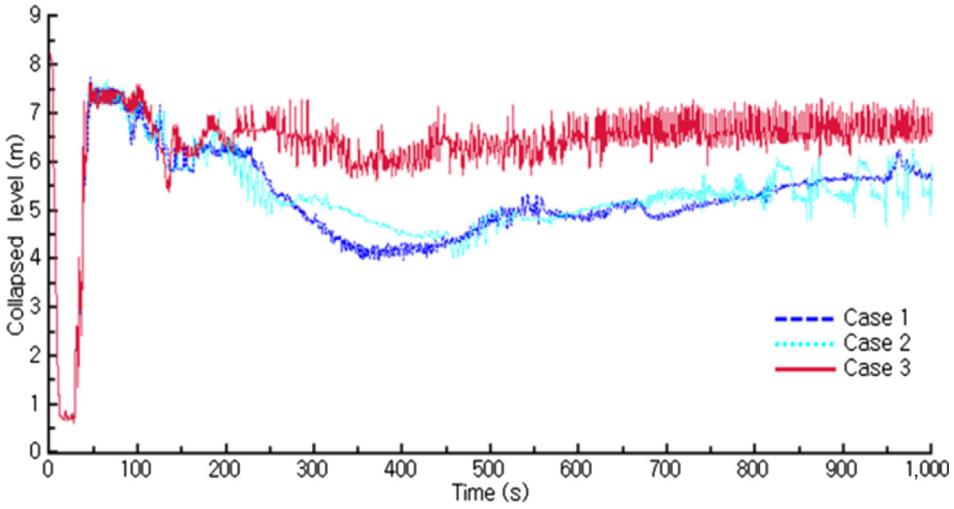

Fig. 4 – Downcomer collapsed water levels for reference cases.

during reflood and to a lesser extent heatup in Case 2 with only one of two LPSI pumps available.

3.3.2. *MARS calculation of heat-transfer modes*

The reference case results show that the thermal hydraulic behavior of the core is sensitive to the SIT injection and HPSI/LPSI with the SIT injection most important during refill and active SI systems driving the reflood. Various system parameters, including the clad temperature, downcomer level, and core level, can serve as figures of merit for system success. For example, if the peak clad temperature (PCT) does not exceed a regulatory limit the system is in a success state. Comparison of Figs. 3 to 5 shows a correlation between clad temperature, downcomer level, and core level. The water levels as a function of safety injection are easy to understand as the SI systems simply deliver mass flows of water to the finite volumes of the downcomer and core. However, the clad temperature is calculated by the MARS code through the interaction of many sub-models, specifically the heat conduction model for the fuel pin and the hydrodynamic model for the two-phase flow in the fuel channel. To the first order, the flow in the channel is correlated to the downcomer and core water levels and the decay heat and material properties of the fuel pin determine heat conduction.

The heat conduction and hydrodynamic model exchange energy through wall-to-fluid heat transfer using empirical correlations, flow regime modes, and heat-transfer modes to calculate the heat-transfer coefficient. Fig. 6 shows the heat-transfer modes activated by the MARS code during the reference case simulations. The calculated heat-transfer coefficients are shown in Fig. 7. The heat-transfer modes and coefficients are for Axial Node 11 of the hot pin near the top of the active core. The temperature excursions in the Node 11 cladding during the reflood for Cases 1 and 2 directly correspond to the code switching from nucleate boiling heat transfer (mode 43/44) to transition boiling (mode 45/46), film

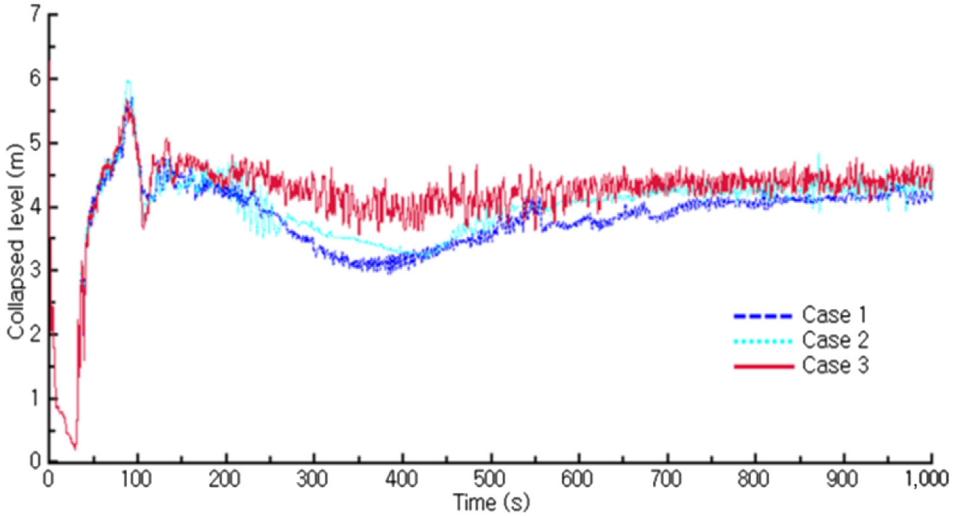

Fig. 5 – Core collapsed water levels for reference cases.

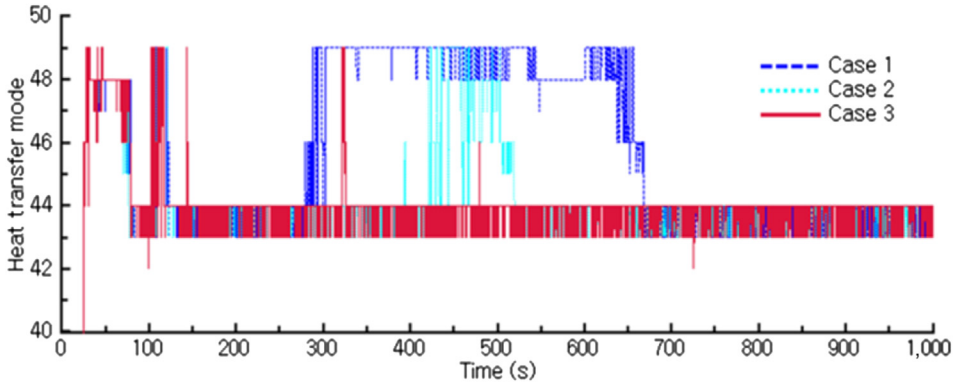

Fig. 6 – Heat-transfer modes activated by MARS code for Axial Node 11 of hot pin. Mode 43/44, nucleate boiling; Mode 45/46, transition boiling; Mode 47/48, film boiling; Mode 49, vapor heating.

boiling (mode 47/48), and vapor heating (mode 49). The latter heat-transfer modes are much less efficient, as indicated by the very small heat-transfer coefficients in Fig. 7 when the modes are active. During these periods, more energy is deposited in the fuel pin from decay heat than energy is transferred to the coolant so the fuel and cladding must increase in temperature. The fuel pin experiences near adiabatic heating conditions.

The heat-transfer modes and empirical correlations used to calculate heat-transfer coefficients were developed to model the boiling and heat-transfer phenomena observed during thousands of heat-transfer experiments. The MARS code uses finite differencing schemes to solve differential equations for a time-dependent solution of the flow and temperatures throughout the plant model. The NPP and components are also discretized by the plant nodalization. MARS uses Boolean logic to select the specific sub-models used to calculate parameters, such as heat-transfer coefficients, needed to advance the solution from time step to time step and from cell to cell. Step changes and discontinuities in parameters that would be continuous in nature are inherent to Boolean logic and finite differencing as evidenced by the step changes in heat-transfer modes in Fig. 6.

A potential consequence of the solution methods of deterministic computer codes are bifurcations in solution results for two simulations with inputs that are almost identical. For example, a nominal change in the delay time, flow rate of the LPSI pump, time step size, or nodalization in Case B may yield a clad temperature similar to Case C with no reheating of the clad during reflood as a consequence of the local flow conditions approaching but not exceeding a threshold in the Boolean logic used to switch the heat-transfer mode from nucleate boiling to transition boiling or the flow types from the normal to hot wall flow regime maps. In this context, the precise numerical value of a result such as PCT from a deterministic code is not exact but is subject to some randomness arising from the solution structure of the model. Using the results in Figs. 3 and 6 to quantify this statement, the rate of clad heatup due to adiabatic heating of the fuel pin is on the order of 3 K/s to 5 K/s when heat transfer is vapor heating. An uncertainty of the timing of the transition to vapor heating of on the order of 10 s could result in a change in PCT of approximately 30 K to 50 K. Due to extensive verification and validation of best-estimate codes with experimental results, we have confidence in the code's ability to predict the overall thermal hydraulic behavior and trends in a realistic manner. However, a single specific numerical result quoted with many significant figures may not be itself adequate for decision making in a safety analysis. Computer code results are not exact. The methodology proposed in Section 2.3 is an

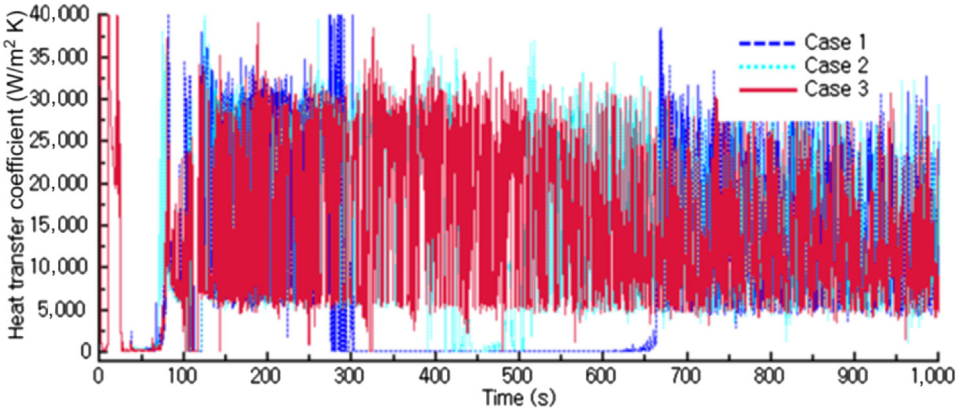

Fig. 7 – Heat-transfer coefficients calculated by MARS code for Axial Node 11 of hot pin.

attempt to present the valuable information in the numerical results of computer codes while addressing some of the uncertainties and limitations of the data.

The discussion of heat-transfer modes and two-phase flow regimes highlights that the analyst cannot properly interpret code results in blackbox mode using only the input and output data streams. The analyst must be aware of the solution techniques of the code and how the output parameter of interest might be sensitive to the interacting submodels of the code, not just the sensitivity of the observed change in the output parameter due to input parameter variation. The MARS code theory manual [8] discusses in detail the two-phase flow heat-transfer phenomena and the implemented solution methods for reactor vessel reflood that must resolve the entire boiling curve and many flow regimes. The high-frequency oscillations of the variables plotted in Figs. 4–7 are primarily physical artifacts of the axial and time heterogeneity of the complex two-phase flow patterns and the model structure of flow and heat transfer. These are not artifacts of numeric stability issues of the finite differencing methods used to solve the model equations with regard to time step size and Courant limits.

4. GPM for LBLOCA injection phase

The reference cases simulated three configurations of the SIT system, HPSI system, and LPSI system representative of branches of the LBLOCA ET model. In a conventional safety analysis these simulations could support success criteria definitions for the ET/FT model. Success criteria definitions only identify configurations of safety systems leading to success sequences in the ET. Now we want to extend the analysis to establish a more general statement of the safety margin of configurations accounting for additional uncertainties of the plant systems, modeling assumptions, and thermal hydraulic code parameters using the GPM to process and interpret the information from MARS simulations. The PCT, minimum downcomer-collapsed water level, and minimum core-collapsed water level during reflood will be figures of merit for the safety margin.

4.1. *Selection of input parameters and distributions*

Table 1 lists the input parameters and uncertainty distributions selected for LBLOCA application of the GPM methodology. The list is representative of sequence timing, safety system function and configuration, technical specification ranges for safety system components, and thermal hydraulic code parameters. The list of parameters is not comprehensive of all uncertainties and was not generated from a formal phenomena identification and ranking table (PIRT). The research scope of the paper is to demonstrate the GPM methodology so the parameters and distributions were selected using engineering judgment and are realistic for Hanul 3&4 and the MARS model in the context of our application.

The delay time of the EDG start, warm-up, and loading of the SI pumps is the primary sequence timing parameter of the injection phase of the LBLOCA. For this study we extend the uncertainty range from the loading sequence technical specification of 15 s to 30 s out to 600 s in order to analyze the sensitivity of the safety margin to the demanding requirements on the EDG. The delay time appears in the MARS input deck as a time trip that turns true after the SIAS trip is activated and the sampled delay time passes. The flow rate uncertainty range for each configuration of SI pumps represents the minimum-to maximum-rated flows of each pump type given in the FSAR. The GPM regression will be explicitly performed on the delay time and SI flow rate because these are the two dominant parameters affecting the clad temperature and the downcomer and core collapsed water levels during reflood.

Table 1 – Input parameter uncertainties for GPM regression.

Input uncertainty	Distribution	Range or $\mu \pm \sigma$
Sequence timing		
EDG delay time (s)	Uniform	15–600
Safety injection flow rate		
Configuration 1: 2/2 HPSI to 3/3 CL (kg/s)	Uniform	80–115
Configuration 2: 1/2 LPSI to 1/3 CL (kg/s)	Uniform	135–170
Configuration 3: 1/2 HPSI to 3/3 CL and 1/2 LPSI to 1/3 CL (kg/s)	Uniform	190–215
RWT water temperature (K)	Uniform	276–323
SIT technical specification ranges		
Initial water volume (m^3)	Uniform	51–54.2
Gas pressure (MPa)	Uniform	3.992–4.42
Water temperature (K)	Uniform	280–320
T/H code parameters		
Heat transfer coefficient multipliers:		
Transition boiling	Normal	1.0 ± 5%
Film boiling	Normal	1.0 ± 5%
D-B vapor HT	Normal	1.0 ± 5%
Decay heat:		
Fission product yield factor	Normal	1.02 ± 0.03

1/2, One of two; 1/3, one of three; 2/2, two of two; 3/3, three of three; CL, cold leg; D-B, Dittus-Boelter; EDG, emergency diesel generators; GPM, Gaussian process model; HPSI, high-pressure safety injection system; HT, heat transfer; LPSI, low-pressure safety injection system; RWT, refueling water tank; SIT, safety injection tank; T/H, thermal hydraulic.

The uncertainty ranges for RWT water temperature and SIT parameter uncertainties, initial water inventory, water temperature, and gas pressure, are the technical specification ranges for the components in the FSAR. During the operation of the plant, these parameters may fluctuate with the ambient environment and the plant procedures require personal to periodically verify measured values do not exceed the specifications. These variables will be treated implicitly in the noise term of the GPM because they are related to the key safety systems, SITs and HPSI/LPSI, but only introduce minor variations in mass and energy balance during simulations with respect to nominal conditions.

The MARS code allows uncertainty multipliers to be applied to heat-transfer coefficients calculated by the heat-transfer modes. We assume a normal distribution with 5%

standard deviation for the transition boiling, film boiling, and vapor heating multipliers. Clad heatup occurs during reflood when these three heat-transfer modes are active. A figure of merit, the reflood PCT, is sensitive to the heat-transfer modeling. The decay heat is calculated using the ANS79-1 standard and the nominal rated power of 2815 MW_{th}. The decay heat model uncertainty is modeled by assuming a normal distribution with a mean of 1.02 and 3% standard deviation for the fission product yield factor. These input parameters will also be treated implicitly in the noise term of the GPM to demonstrate how to incorporate code parameter uncertainties into the analysis.

4.2. *Sampling of inputs and MARS simulations for training data generation*

The training set for surrogate construction should be a space-filling design to provide adequate coverage of the input parameter space. We sample in a two-stage approach because the EDG delay time and SI flow rate are the explicit regression variables while the RWT water temperature, SIT parameters, and MARS code parameters are modeled implicitly as the noise term in the GPM. For the two-dimensional input space of EDG delay time and SI flow rate, the unscented transform with random orthogonal matrix sampling algorithm [11] is used to generate 100 data points for each configuration of SI pumps sampled between the minimum- and maximum-rated flows and the EDG delay time uncertainty range for a total training set size of 300 data points. The unscented transform with random orthogonal matrix experiment design uses Eigen decomposition of a randomly generated 50 × 50 matrix to obtain a random orthonormal basis whose projection onto the two-dimensional input space when properly scaled provides excellent space-filling properties and randomness similar to Latin hypercube designs. The second stage of sampling is Monte Carlo sampling of the remaining eight input variables for every point in the two-dimensional input space. By Monte Carlo sampling, the measurement noise of each training point is a random contribution from the eight probability density functions (PDFs). For each configuration of SI pumps, an additional 20 simulations were performed with all 10 input PDFs randomly sampled. These samples are used independently of the training set for cross validation to assess the predictive accuracy of the GPM and learned hyperparameter values.

The input samples were processed by the MOSAIQUE software [12] to automate the generation of input decks and scheduling and execution of the simulations. The simulations were performed in batches of 100 on a single modern desktop computer with a quad-core processor. Each batch of simulations required approximately 3 hours to 5 hours of run time, so computational expense was not a factor during the study, with most simulations performed overnight. The output for a few key system variables was manually checked for all simulations to verify the MARS code results were physically reasonable. One simulation crashed during the blowdown phase from either numerical failure of the code or a power outage to the computer. The exact cause of failure was not investigated and the failed run was discarded from the training set since all other simulations were successful.

Data smoothers using locally weighted regression, including the GPM, perform best as interpolating functions and can be prone to instability and extrapolation error near the boundary of the input space. Data smoothers will also perform better on noisy data sets if more data points are available. To provide an easy fix for numerical issues with

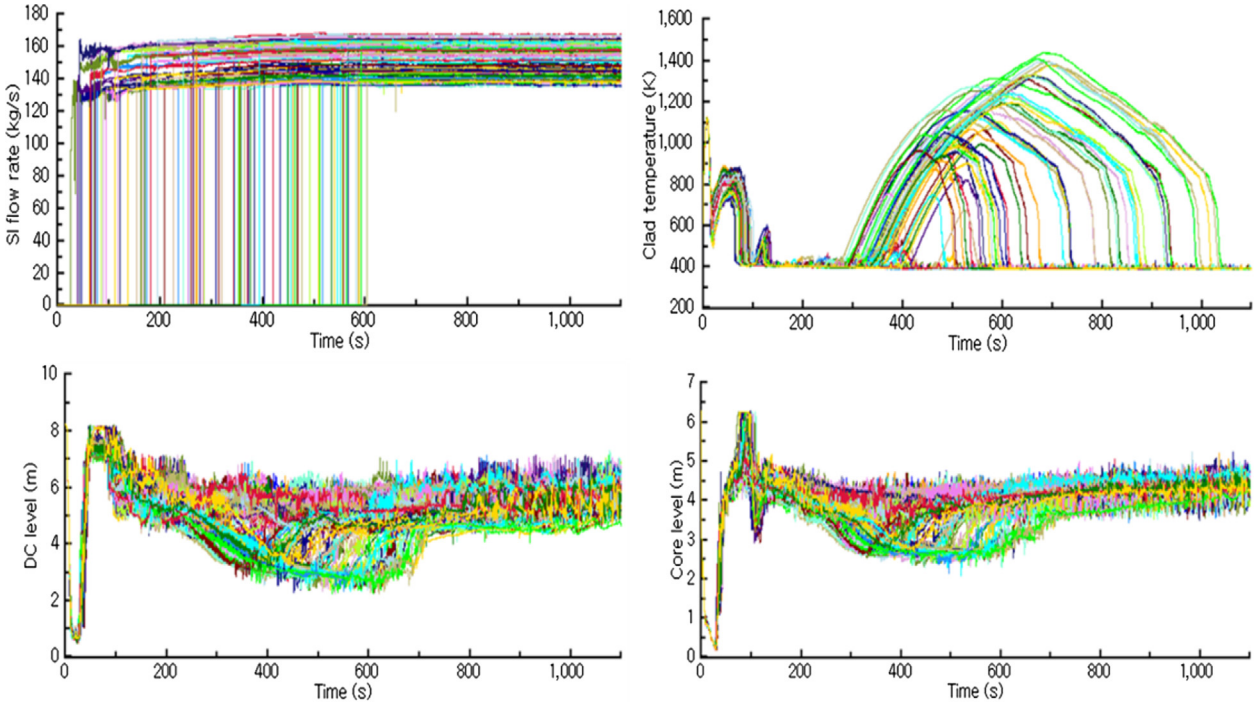

Fig. 8 – MARS simulation data for one of two LPSI pumps injecting to cold leg 1B for LBLOCA. Clad temperatures are from axial Node 9 of the hot pin. DC; LBLOCA, large-break loss-of-coolant accident; SI, safety injection.

locally weighted regression near boundaries and regions with larger noise levels, we added additional data points to the training set. The alternative is to implement special edge treatment procedures in each specific data smoothing algorithm. Eleven additional MARS simulations were performed with an EDG delay time of 15 s for the minimum and maximum flows for the two of two HPSI pump configuration. The safety parameter data appear to be more noisy in the lowest flow rate region of the two of two HPSI pump configuration. For EDG delay times between 600 s and 630 s, 11 artificial data points were added to the model with hot pin PCT equal to 1500 K, and downcomer- and core-collapsed water levels equal to 2 m. Beyond 600 s, the only fluid in the reactor vessel is in the lower plenum and most of the clad has heated to the point where core damage is imminent. The artificial data points allow the GPM using the zero mean to converge to the core damage condition at the right-hand side time boundary rather than reverting to the sample mean of the training set.

Every MARS simulation gives a time history of the clad temperatures and the downcomer- and core-collapsed water levels. Fig. 8 shows the time histories for the 100 simulations for Configuration 2 with one of two LPSI pumps injecting. Also shown are the sampled SI flow rates and delay times. From the time histories, the maximum clad temperature and the minimum downcomer- and core-collapsed water levels during reflood are extracted from the time histories becoming the output y values in the GPMs.

4.3. GPM training and model verification

Using the 299 data point training set, GPM hyperparameters were learned using the leave-one-out method with Gaussian priors implemented in the GPML code [2]. Table 2 lists the learned hyperparameter values for GPM surrogates predicting the PCT of the hot pin, minimum downcomer collapsed water level, and minimum core collapsed level during reflood. A surrogate for the PCT of the core averaged fuel channel is also provided. The maximum linear heat generation rate for the core averaged channel is 8.25 kW/ft compared to 13.9 kW/ft for the hot pin.

4.3.1. Cross validation

Fig. 9 shows the surface plots for the GPM mean functions compared to the training data. The mean function surfaces can be interpreted as response surfaces in the context of conventional regression analysis. By visual inspection, the nonlinear surfaces appear to be smooth interpolants between the training points and overfitting does not appear to be an issue. Using the 60 additional data points from the MARS simulations that were left out of the training set, a cross validation was performed. Fig. 10 shows the MARS cross validation data compared to the 95% probability intervals obtained from the GPM mean functions and associated prediction variances, which define normal distributions at prediction points. If the GPM is the truth probabilistic model of the expected MARS results, 95% of random samples drawn from the MARS model should, on average, fall within $\pm 2\sigma$ of the GPM mean represented by the error bars. The eight implicit input parameter PDFs were randomly sampled for the cross validation points simulated by the MARS code so approximately 3 of 60 points in the cross validation set should lie outside the GPM 95% probability intervals. Fig. 10 shows no statistically significant results that would indicate the GPMs are inappropriate from overfitting or wrong model assumptions.

4.3.2. Dimensional analysis for hyperparameter verification

The hyperparameter values in Table 2 have associated units of measurement and each hyperparameter has a physical interpretation in relation to the thermal hydraulic behavior of the NPP. The signal standard deviation is a measure of the order of magnitude of Δy given Δx; how much the output can vary through a change in the input space. The length scale parameters of the input define the magnitude of Δx. We performed a simple dimensional analysis to verify the hyperparameter values in Table 2. Conservation of mass can be used to validate the minimum downcomer- and core-collapsed water level surrogates. The minimum downcomer- and core-collapsed water levels represent a net change in the fluid mass inventory in the downcomer and core volumes from the end of the refill at 100 s when the volumes are filled with single-phase fluid to the times during the reflood when the levels reach the minimum values. If the boundary conditions change, the SI flow rate and timing of injection start assumed in the MARS model, the net change in fluid mass inventory will show a deviation from a nominal reference case. From the downcomer- and core-level surrogate hyperparameters presented in Table 2, the signal standard deviation quantifies the deviation of the net change in fluid mass inventory relative to the changes in timing and flow rate quantified by r_1 and r_2, respectively. SI systems inject mass or volumetric flow rates into the finite volume of the CL piping, downcomer, lower plenum, and active core. A SI flow rate of 1 kg/s is approximately 10^{-3} m^3/s. The free volumes of the downcomer and active core are approximately 21 m^3 and 17 m^3, respectively. The signal standard deviations for the minimum downcomer level and core level surrogates are

Table 2 – GPM hyperparameter values for LBLOCA reflood safety parameters.

	PCT core avg.	PCT hot pin	Minimum DC level	Minimum core level
Delay time scale, r_1 (s)	60.33	59.81	110.52	85.87
Flow rate scale, r_2 (kg/s)	80.38	30.21	26.01	30.93
Signal standard, σ_f	156.41 K	162.62 K	0.793 m	0.245 m
Noise standard, σ_n	110.65 K	114.82 K	0.074 m	0.223 m

DC, downcomer; GPM, Gaussian process model; LBLOCA, large-break loss-of-coolant accident; PCT, peak clad temperature.

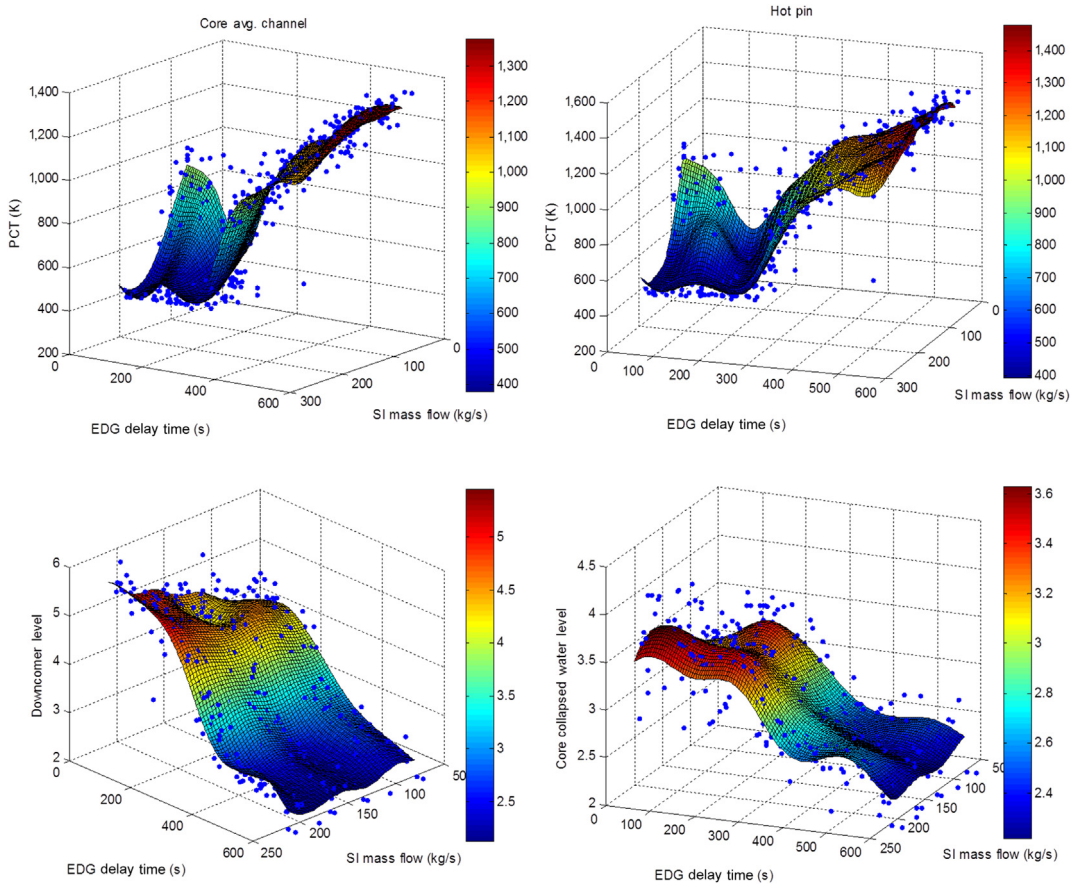

Fig. 9 – GPM mean function surface plots for LBLOCA reflood safety parameters. EDG, emergency diesel generators; GPM, Gaussian process model; LBLOCA, large-break loss-of-coolant accident; PCT, peak clad temperature; SI, safety injection.

0.793 m and 0.245 m, corresponding to volumes of approximately 2.75 m^3 and 1 m^3, respectively. We define a dimensionless parameter relating the delay time length scale (r_1), flow rate length scale (r_2), the signal standard deviation expressed as a volume (V_{σ_f}), and the density of the SI water (ρ_{H_2O})

$$\pi_{\text{level}} = r_1 r_2 \Big/ V_{\sigma_f} \rho_{H_2O} \tag{11}$$

Eq. (11) evaluates to approximately 1.05 for the downcomer level surrogate and 2.65 for the core level surrogate. The π_{level} values are approximately unity, indicating the GPMs and hyperparameter values inferred from the training data are consistent with conservation of mass and volume. A $\pi_{\text{level}} \gg 1$ or $\pi_{\text{level}} \ll 1$ would signify an unphysical significant scale distortion.

The core volume hydrodynamics are coupled to the downcomer, so the difference in Eq. (11) between the two sets of hyperparameters, although not extreme, should be investigated. First, we recognize the GPMs also assume the noise term when relating Δy and Δx, and the signal to noise ratio $\sigma_f^2/\sigma_n^2 = 1.2$ for the core-level surrogate. The measurement noise standard deviation is significant and is an independent and additive contribution to the deviation of the net change in fluid mass inventory. Eq. (11) is modified to include the contribution to volume change from the measurement noise standard deviation expressed as a volume (V_{σ_n})

$$\pi_{\text{level}} = r_1 r_2 \Big/ \sqrt{V_{\sigma_f}^2 + V_{\sigma_n}^2}\,\rho_{H_2O} \tag{12}$$

The adjusted π_{level} for the downcomer level and core level are 1.04 and 1.96, respectively. The factor of two difference in adjusted π_{level} is not surprising because the flow rate of coolant entering the core is determined by the hydraulic head of the fluid level in the downcomer and the flow resistance through the hot legs. The flow regimes and heat transfer in the core are much more complex than the thermal hydraulics in the downcomer so there are other factors present affecting the core level. These factors are partially accounted for in the large measurement noise variance. The larger adjusted π_{level} for the core level also means the length scale parameters are larger when compared to the downcomer level length scales on a normalized basis relative to volume or mass change. With respect to data smoothing, longer length scales promote better smoothing of noisy data by widening the weighting window of the covariance function. Here the dimensional analysis provides insight into the data smoothing properties of the GPM.

A similar dimensional analysis is performed for the PCT surrogates based on conservation of energy. During reflood,

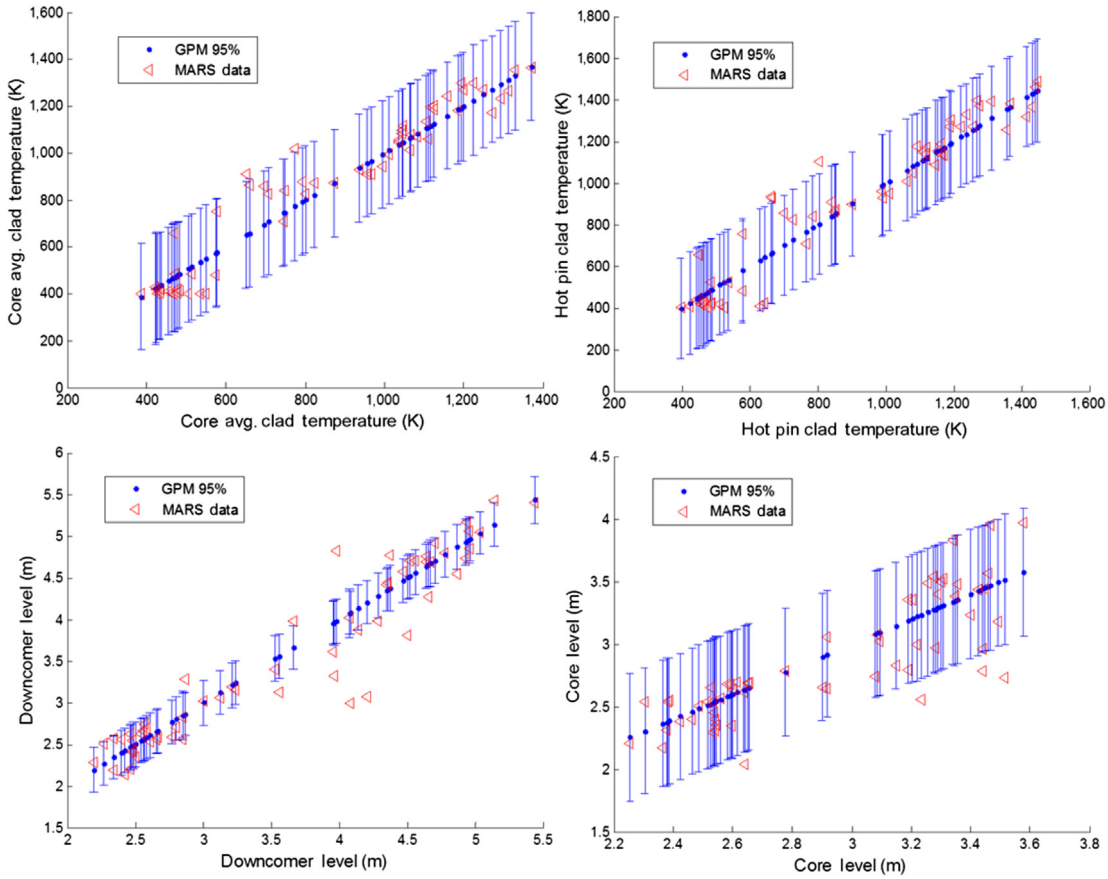

Fig. 10 – Comparison of cross validation data points to 95% probability intervals predicted by GPMs. GPM, Gaussian process model.

the decay power decreases from approximately 105 MW_{th} at 100 s to 65 MW_{th} at 1000 s. During this time, the decay heat will primarily boil the coolant or be stored in the fuel pellets and cladding of the fuel pins. The coolant and fuel pins are two competing heat sinks for decay energy. The SI flow rate provides new liquid coolant that is available to boil with a latent heat of vaporization (H_{vap}) of approximately 2.257 MJ/kg. The volumetric heat capacities of uranium dioxide fuel (C_{p,UO_2}) at 600 K and Zircaloy cladding ($C_{p,Zr}$) at 573 K are approximately 3.043 MJ/m^3 K and 2.079 MJ/m^3 K, respectively. The total volumes of fuel (V_{UO_2}) and cladding (V_{Zr}) in the core are approximately 9.15 m^3 and 1.61 m^3, respectively. A dimensionless number relating the SI flow rate (r_2) to decay power (P_d) is

$$\pi_{SI} = H_{vap} r_2 / P_d \tag{13}$$

For the core averaged PCT GPM, π_{SI} will range from 1.7 at 100 s to 2.8 at 1000 s. A $\pi_{SI} > 1$ means a flow rate through the core equal to r_2 has enough thermal capacity to remove the generated decay heat and has excess capacity to remove some stored energy from the fuel pins if all of the fluid boils. The ability of constant SI flow to remove stored energy from the fuel pins increases with time as decay power decreases.

A dimensionless number representing adiabatic heating of the fuel pins over the delay time length scale (r_1) is

$$\pi_{\Delta T} = \sqrt{\sigma_f^2 + \sigma_n^2}\left(V_{UO_2} C_{p,UO_2} + V_{Zr} C_{p,Zr}\right)/r_1 P_d \tag{14}$$

For the core averaged PCT GPM, $\pi_{\Delta T}$ is approximately 0.94 at 100 s and 1.5 at 1000 s using the signal standard deviation (σ_f) and noise standard deviation (σ_n) from Table 2. The ratio $\pi_{\Delta T}/\pi_{SI}$ is close to unity during reflood, indicating that the GPM allocates *equal weighting* to the importance of the fuel pin and the coolant as heat sinks. Eq. (13) overestimates the heat removal by the coolant because some fluid is entrained in the vapor in the two-phase flow boiling regimes with clad temperatures above the surface rewetting temperature [8]. One final observation is that in the absence of coolant boiling, when the heat-transfer mode is vapor heating and the heat-transfer coefficient is very small, most of the decay energy is deposited in the fuel rod and the fuel and cladding must heat up adiabatically. The slopes of the clad heatup during reflood shown in Fig. 8 are approximately 3 K/s to 5 K/s and the slope defined by the ratio $(\sigma_f + \sigma_n)/r_1$ is approximately 4.4 K/s. Although the GPM only predicts the PCT, the limiting value of a safety parameter during the transient, the GPM hyperparameters are consistent with the dynamics or time-dependent behavior during the transient.

The hyperparameters for the core averaged PCT and the hot pin PCT are almost identical except that the flow rate

length scale for the hot pin PCT is approximately 60% less than the core averaged PCT, indicating the hot pin clad heatup is more sensitive to changes in the SI flow rate. This scale distortion reflects the difference between the relative pin powers of the hot channel and the core averaged channel. The maximum linear heat generation rate for the hot pin is approximately 60% greater than the core average fuel pin. Power distribution, power peaking factors, maximum linear heat generation rates, etc., are common uncertainties that are considered in safety analysis. Based on the comparison of the core averaged PCT to hot pin PCT surrogates, a power distribution parameter could be introduced as a third dimension of the input space used for explicit regression. A power distribution parameter that changes the pin power cannot be treated implicitly in the noise term regression variable because the GPM training and dimensional analysis have identified significant scale effects due to pin power and flow rates.

Eqs. (11–14) provide useful insights into the first-order behavior of the thermal hydraulic phenomena of the reflood phase based on simplified expressions representing conservation of mass and energy. The dimensional analysis validates the hyperparameter values for the explicit regression variables and can also be used to assess the other input parameters in Table 1 that are implicitly represented in the regression in the GPM measurement noise term. The fission product yield factors affects the decay power so a variation of P_d on the order of 3% in Eqs. (13) and (14) has a negligible impact on the calculated π_{SI} and $\pi_{\Delta T}$ values. The SIT technical specifications and RWT water temperature mainly affect the subcooling and specific volume of the SI water sources. The subcooling enthalpy of the SI water is on the order of 10% of the heat of vaporization so subcooling effects can be neglected in Eq. (13) and in the GPM regression. The 5% expansion of the fluid from subcooled conditions to saturation would be a minor volume effect in Eqs. (11) and (12). The MARS code has detailed models and closure relations to accurately model these minor effects but in the GPM surrogates, reduced-order models, explicit representation of these minor variables is not needed and the use of the measurement noise term is justified. Dimension reduction of the regression problem is a practical benefit and the contribution of the implicit variables is not discarded but rather quantified by the measurement noise variance.

5. Estimating safety margin with Gaussian process model

From the cross validation study and dimensional analysis of the GPM hyperparameters, we are confident using the GPM to make inferences about the safety margin of the NPP during the LBLOCA without directly using the MARS code. Figs. 11–13

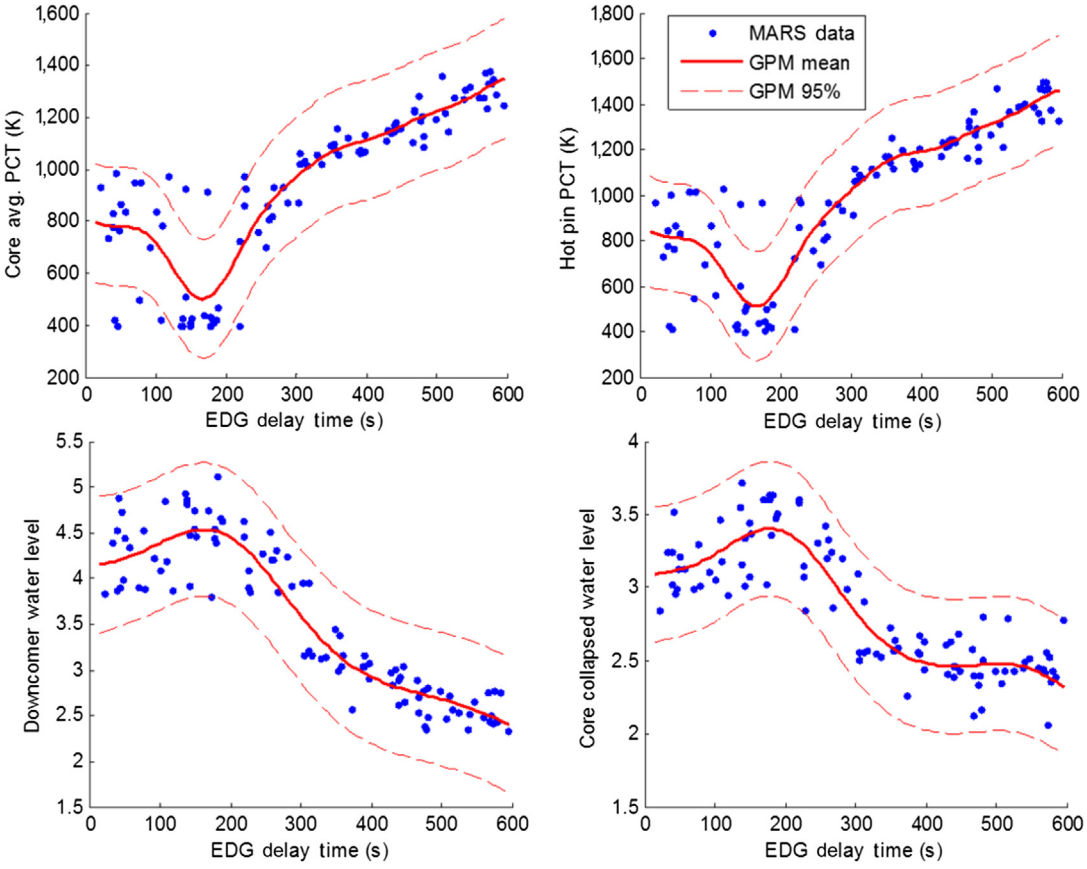

Fig. 11 – GPM probability intervals for two of two HPSI pumps injecting to three of three CLs. EDG, emergency diesel generators; GPM, Gaussian process model; HPSI, high-pressure safety injection system; PCT, peak clad temperature.

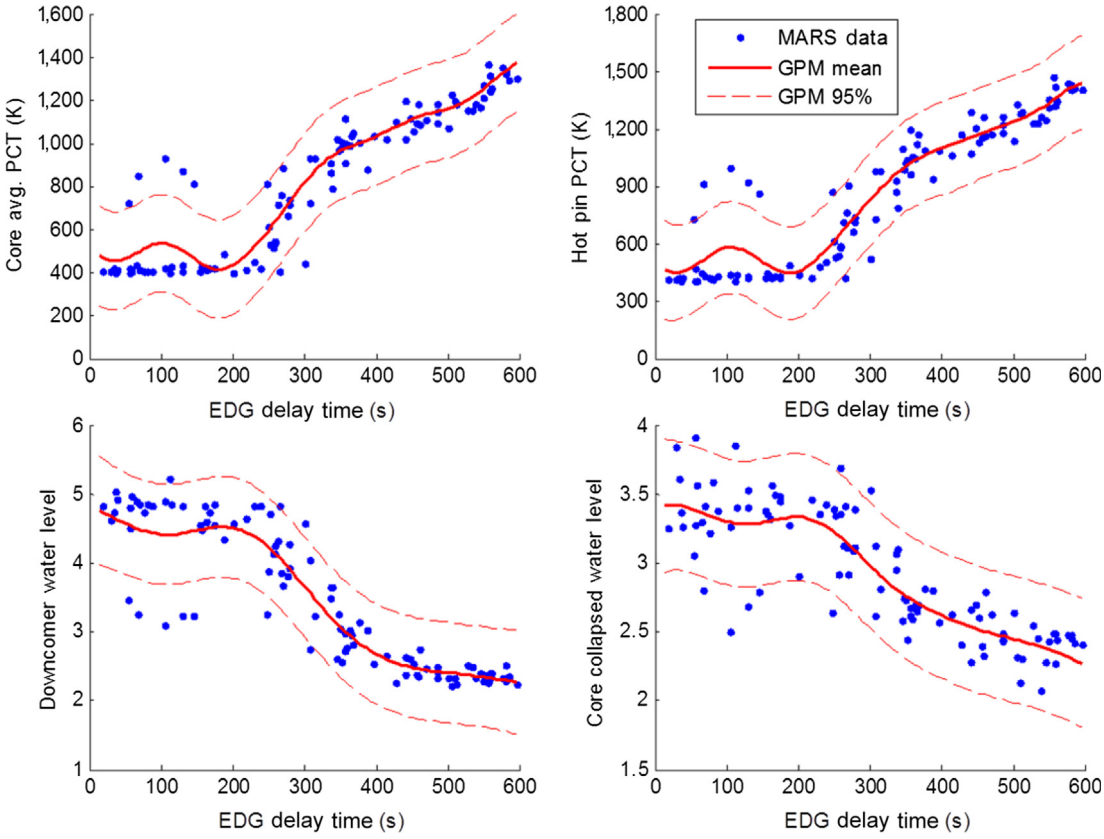

Fig. 12 – GPM probability intervals for one of two LPSI pumps injecting to one of three CLs. CL, cold leg; EDG, emergency diesel generators; GPM, Gaussian process model; LPSI, low-pressure safety injection system; PCT, peak clad temperature.

show the 95% probability intervals predicted by the GPMs for the three configurations of SI pumps as a function of EDG delay time. The SI flow rates for each configuration are held at the mean value between the minimum and maximum rated flows to allow for easier visualization. The training set data for each configuration are overlain for reference.

Fig. 12 shows several limiting system parameters during the LBLOCA with two of three SITs injecting to two of three intact cold legs and one of two LPSI pumps injecting to one of three intact CLs. This configuration represents the current success criteria for CL LBLOCAs in the Hanul 3&4 PSA model. Using the 1477 K PCT acceptance criteria from 10CFR50.46 as the safety limit to measure the safety margin and assuming the EDG loading sequence technical specification of a 30-s to 35-s delay time for the start of the LPSI pump, the probability of exceeding the acceptance criteria can be estimated using the normal distributions for the safety parameters predicted by the GPMs. The GPMs for the core average PCT and hot pin PCT predict 50/97.5 percentiles of 452 K/680 K and 448 K/692 K, respectively, for a delay time of 35 s. The 1477 limit is over eight standard deviations from the mean and the probability of exceeding the acceptance criteria is infinitesimal. Configuration 1, shown in Fig. 11 with two of two HPSI pumps injecting, is the most limiting configuration studied and the 50/97.5 PCT percentiles are 789 K/1015 K and 832 K/1072 K. The probability of exceeding the acceptance criteria in Configuration 1 is less than 10^{-7}. Of course, the true distributions for the reflood PCTs subject to all input and plant uncertainties and their assumed PDFs are not exactly the normal distributions assumed by the GPMs; however, using the GPMs as first-order approximations, we can conclude that a large safety margin exists for these configurations of safety systems operating under the conditions of the plant design basis.

What if the plant design basis changes? Training the surrogate on a large number of simulations representing a wide range of plant configurations and modeling assumptions allows changes in the safety margin to be quickly assessed if a plant modification or technical specification amendment is proposed. For example, a conventional safety analysis supporting conservative success criteria definitions may depend on a few simulations employing bounding assumptions. If a proposed change exceeds a bounding assumption used in the previous analysis, it may be difficult to make an informed decision about the safety margin by extrapolating the former results. The safety analysis must be redone, often requiring the reconstitution of legacy input models and knowledge transfer to the new analyst.

To demonstrate how the GPMs can be used to assess a design basis change in the context of PSA, we consider a hypothetical change to the loading sequence technical specification of the EDG. Cold start, short warm-up time, and rapid loading of EDG during regular testing and unplanned starts is known to cause irregular wear and premature aging on engine components reducing the reliability of the EDG

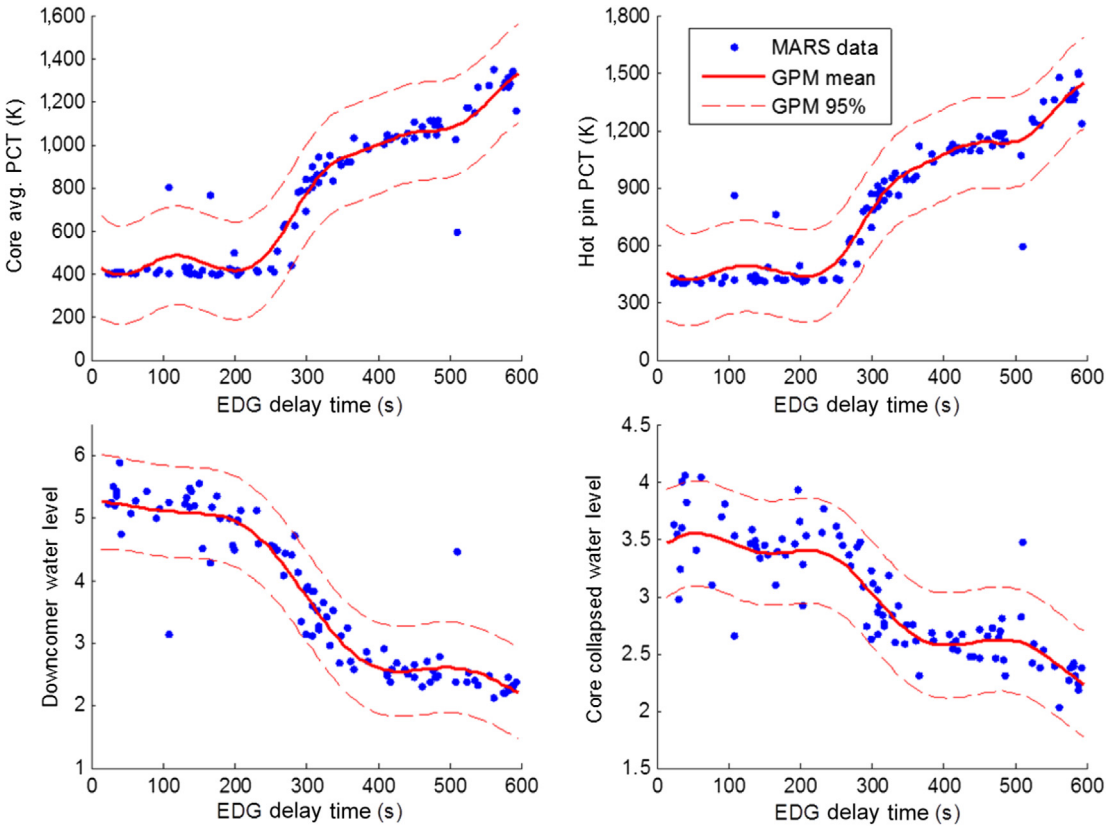

Fig. 13 – GPM probability intervals for one of two LPSI pumps injecting to one of three CLs and one of two HPSI pumps injecting to three of three CLs. CL, cold leg; EDG, emergency diesel generators; GPM, Gaussian process model; HPSI, high-pressure safety injection system; LPSI, low-pressure safety injection system; PCT, peak clad temperature.

and has been a long standing safety concern in the nuclear industry [13]. The LBLOCA is the most demanding event that requires prompt SI so the stringent loading sequence technical specification and testing programs are directly related to LBLOCA mitigation but come at the cost of decreased EDG reliability increasing the risk of other events such as loss-of-offsite power (LOOP) and station blackout. Figs. 11–13 show that the PCT and uncertainty bounds remain relatively constant for EDG delay times from 15 s to 200 s. The safety margin is not sensitive to a change in loading sequence of loading HPSI and LPSI pumps up to 200 s and a very large safety margin greater than 400 K exists. Beyond 200 s, the PCT increases and the probability of exceeding the safety limit will also increase, reducing the margin. The minimum downcomer and core collapsed levels are also relatively constant for delay times of up to 200 s which provides supporting evidence that the safety margin is large and constant. The downcomer and core levels are the driving force of reflood cooling which determines the clad temperature behavior from which the PCT is used as the safety parameter to measure margin. The plant operators could optimize the loading sequence to minimize engine wear using a mission time of three minutes for warm-up and loading. The probability of the core experiencing single-phase vapor heating and subsequent heatup during reflood is very low based on the best-estimate code results and uncertainty analysis through GPM regression for a mission time of 3 minutes.

6. Summary and conclusion

A new methodology to estimate the safety margin of an NPP has been proposed and demonstrated for best-estimate simulation of an LBLOCA in support of Level 1 PSA success criteria definitions. The methodology simultaneously considers sequence timing, safety system configuration, technical specifications, and code model parameter uncertainties. A key aspect of the methodology is the partitioning of the input parameter space into two subsets of inputs: explicit regression variables consisting of the dominant input uncertainties that are the fundamental drivers of thermal hydraulic behavior of the transient and implicit noise variables. A Gaussian process model performs regression on the explicit regression variables, and output uncertainty is quantified by a measurement noise term representing the contribution of the implicit input noise variables to local variation or uncertainty of the safety parameter. This approach retains high-fidelity

treatment of all input uncertainties during best-estimate simulation of the transient, but allows the analyst to focus the regression analysis on the most important application specific parameters thereby overcoming the curse of dimensionality inherent to the analysis of complex systems.

The demonstration of the methodology for the LBLOCA success criteria definitions used a two-dimensional explicit variable input space for regression. The GPM is well suited for regression in higher dimensions on the order of 20 to 30 variables so the LBLOCA GPM could be expanded to include more inputs relevant to pressurized water reactors such as power distribution uncertainty (linear heat generation rates) and power uprate conditions. For other transients studied in PSA such as the small-break LOCA, parameters such as break size could be represented as an explicit regression variable. Future applications of the methodology could include safety analysis of severe accident phenomena and mitigation strategies where the timing and consequences of uncertain phenomena (e.g., timing and magnitude of pump seal leakage during station blackout) need to be studied in conjunction with mitigation strategies (e.g., injection into RCS and steam generator secondary side with portable equipment) to determine the effectiveness of the strategies and key mission times. The implicit treatment of many input uncertainties as measurement noise is a novel approach to managing surrogate complexity when performing regression on large data sets. The GPM provides a framework for data synthesis where excessive conservatisms and bounding assumptions used in safety analysis can be reduced while considering many types of uncertainties.

Dimensional analysis has a long-proven track record in engineering analysis and the design of experiments. A simple dimensional analysis was performed to validate the surrogate model derived from machine learning techniques. The dimensional analysis also provided justification that the implicit treatment of some input parameters would not introduce significant scale distortions into the analysis. With the expanding use of complex computer models and complex reduced-order modeling methodologies to interpret large data streams, the analyst should be remain vigilant and should not rely on the blackbox approach. We believe dimensional analysis should be regularly used in computer experiments as an independent and sound basis for data validation and model assessment. Dimensional analysis is especially suited to support safety analyses using best-estimate thermal hydraulic codes to simulate nuclear power plant transients and should be performed early on by the analyst to provide a clear benchmark for comparison against computed results.

Conflicts of interest

All authors have no conflicts of interest to declare.

Acknowledgments

This work was supported by the Nuclear Research & Development Program of the National Research Foundation of Korea (NRF) grant, funded by the Korean government, Ministry of Science, ICT & future Planning (MSIP) (No. NRF-2012 M2A8A4025989).

REFERENCES

[1] C.E. Rasmussen, K.I. Williams, Gaussian Processes for Machine Learning, MIT Press, Cambridge, 2006.
[2] C.E. Rasmussen, H. Nickisch, Gaussian process for machine learning (GPML) toolbox, J. Machine Learning Res. 11 (2010) 3011.
[3] J.P. Yurko, Uncertainty quantification in safety codes using a Bayesian approach with data from separate and integral effect tests, PhD Dissertation, Massachusetts Institute of Technology, Cambridge, 2014.
[4] R.M. Neal, Bayesian learning for neural networks, No. 118, in: Lecture Notes in Statistics, Springer, New York, 1996.
[5] W.S. Cleveland, Robust locally weighted regression and smoothing scatterplots, J. Am. Stat. Assoc. 74 (1979) 929.
[6] KOPEC, Probabilistic safety assessment for Ulchin Units 3&4, Korea Power Engineering Company, Yongin (Korea), 2004.
[7] D.H. Lee, H.G. Lim, H.Y. Yoon, J.J. Jeong, Improvement of the LOCA PSA model using a best-estimate thermal-hydraulic analysis, Nucl. Eng. Technol. 46 (2014) 541.
[8] KAERI, Code structure, system models, and solution methods, KAERI/TR-2812/2004, in: MARS Code Manual, Volume I, Korea Atomic Energy Research Institute, Daejeon (Korea), 2009.
[9] J.J. Jeong, K.D. Kim, S.W. Lee, Y.J. Lee, B.D. Chung, M. Hwang, Development of the MARS input model for Ulchin 3/4 transient analyzer, KAERI/TR-2620/2003, Korea Atomic Energy Research Institute, Daejeon (Korea), 2003.
[10] KEPCO, Ulchin Units 3&4 final safety analysis report, Final Safety Analysis Report, Korea Electric Power Corporation, Naju (Korea), 1996.
[11] D.A. Fynan, K.I. Ahn, J.C. Lee, PCT uncertainty analysis using unscented transform with random orthogonal matrix, Trans. Korean Nucl. Soc., Spring Meeting, Jeju (Korea), (2015).
[12] H.G. Lim, S.H. Han, J.J. Jeong, MOSAIQUE—A network based software for probabilistic uncertainty analysis of computerized simulation models, Nucl. Eng. Design 21 (2011) 1776.
[13] K.R. Hoopingarner, F.R. Zaloudek, Aging mitigation and improved programs for nuclear service diesel generators, NUREG/CR-5057, U.S. Nuclear Regulatory Commission, Washington D.C. (WA), 1989.