

# Multi-Unit Dynamic PRA

D. Mandelli, C. Parisi, A. Alfonsi, D. Maljovec, R. Boring,  
S. Ewing, S. St Germain, C. Smith, C. Rabiti

*Idaho National Laboratory (INL), 2525 Fremont Ave, 83402 Idaho Falls (ID), USA*

M. Rasmussen

*Norwegian University of Science and Technology (NTNU) Social Research, Høgskoleringen  
1, 7491 Trondheim, Norway*

---

## Abstract

Dynamic Probabilistic Risk Analysis (PRA) methods couple stochastic methods (e.g., RAVEN) with safety analysis codes (e.g., RELAP5-3D) to determine risk associated to complex systems such as nuclear plants. Compared to classical PRA methods, which are based on static logic structures (e.g., Event-Trees, Fault-Trees), they can evaluate with higher resolution the safety impact of timing and sequencing of events on the accident progression. Recently, special attention has been given to nuclear plant sites which consist of multiple units and, in particular, on the safety impact of system dependencies, shared systems and common resources on core damage frequencies. In the literature, classical PRA methods have been employed to model multi-unit sites in a limited number of cases while Dynamic PRA methods have never been applied to analyze a full multi-unit model. This paper presents a PRA analysis of a multi-unit plant using Dynamic PRA methods. We employ RAVEN as stochastic tool coupled with RELAP5-3D. The site under consideration consists of three units (each unit is composed by a reactor and its associated spent fuel pool) while the considered initiating event is a seismic induced station blackout event. This paper describes in detail how the multi-unit site has been constructed and, in particular, how unit dependencies and shared resources are modeled from both a deterministic and stochastic point of view.

*Keywords:* Multi-unit, PRA, Dynamic PRA, Reduced Order Modeling

---

## 1. Introduction

Multi-unit Nuclear Power Plant (NPP) sites are defined as sites that are composed by more than one reactor. As of 2017, the U.S. fleet is composed by 99 operating reactors which are distributed as follows: 26 power plant sites have 1 reactor, 31 power plant sites have 2 reactors, 3 power plant sites have 3 reactors and 1 power plant site has 2 reactors and 2 additional reactors under construction. The situation is similar for other countries such as Canada, South

Korea and Japan were several NPP sites include a large number of reactors (6, 7 or even 8 reactors). Worldwide about 80 plants have more than 2 reactors and 32 power plants have more than 3 reactors.

Following the accident event that occurred in 2011 at the Fukushima Daiichi nuclear power plant [1] special attention has been given to multi-unit plant sites. This attention has focused on the safety aspects of nuclear reactors that cannot be considered as entities isolated from each other.

Historically, the first multi-unit Probabilistic Risk Analysis (PRA) at the industry level has been performed for the Seabrook power station [2] using classical Probabilistic Risk Assessment (PRA) methods based on Event-Tree (ET) and Fault-Trees (FT) tools. Canada studies have been recently published about CANDU multi-unit stations [3, 4]. Furthermore, the analysis of the safety aspects of multi-unit plants has been performed for few selected cases [5, 6, 7].

The objective of this paper is to propose the first analysis of a multi-unit power plant not by using classical ET/FT tools [8] but by employing a simulation-based (i.e., Dynamic PRA [9]) approach: the Risk Informed Safety Margin Characterization (RISMC) approach [10, 11]. The rationale behind this choice is that great modeling improvements can be achieved by employing system simulators instead of static Boolean structures like ETs/FTs. Accident dynamic is in fact not set a-priori by the analyst (like in an ET/FT structure) but it is entirely simulated given a set of initial and boundary conditions. Note that timing and sequencing of events are implicitly modeled in the analysis along with interactions between accident evolution and system dynamics.

Modeling limitations of ETs/FTs based methods are even more limiting when dealing with multiple complex systems that are coupled to each other. In fact, the presence of shared systems and structures among the units provides additional degrees of freedom in the accident progression temporal evolution. Furthermore, note that in a multi-unit site, multiple radiological sources are in fact present, both nuclear reactors and Spent Fuel Pools (SFPs).

The paper presents in Section 2 an overview of the RISMC approach and it describes the PRA elements of the analysis. Section 3 introduces the multi-unit modeling from a mathematical point of view while Section 4 describes the multi-unit site considered in the analysis. Section 5 describes in detail of the RISMC modeling of all elements of the chosen multi-unit site. Sections 6 and 7 describe how the data have been generated and analyzed. Finally, Section 8 presents the results obtained by the described RISMC analysis.

## 2. RISMC Approach to Dynamic PRA

The RISMC approach [10] employs both deterministic and stochastic methods in a single analysis framework. In the deterministic method set we include modeling of the thermal-hydraulic behavior of the plant [12, 13], external events such as flooding [14] and operators' responses to the accident scenario [15]. Note that deterministic modeling of plant or external events can be performed by employing specific simulator codes but also Reduced Order Models (ROMs) [16].

ROMs would be employed as substitute of the actual code in order to decrease the high computational costs of employed codes.

The modeling of the actual plant is performed by using system analysis codes which simulate the temporal evolution of the plant given the sampled values of timing/sequencing of events. Example of codes that can be employed are RELAP5-3D [17] and MELCOR [18]. Example of RISMC analysis performed using the RISMC approach can be found in [12, 13, 19]. The plant stochastic modeling is performed by employing stochastic analysis tools (e.g., RAVEN [20]) that are interfaced with the chosen system analysis codes. This interface is responsible to:

1. Perturb the input file of the system analysis code by inserting in it the values sampled by the stochastic analysis tools
2. Execute the simulation run given the input file generated in Step 1
3. Collect the output of the simulation run so that a link between sampled input values and simulation outcome is created.

### 3. Multi-Unit Modeling

From a mathematical point of view, a single simulator run can be represented as a single trajectory in the phase space. The evolution of such a trajectory in the phase space as function of time  $t$  can be described as  $\frac{\partial \boldsymbol{\theta}}{\partial t} = \boldsymbol{\Xi}(\boldsymbol{\theta}, \mathbf{p}, \mathbf{s}, t)$  where:

- $\boldsymbol{\theta} = \boldsymbol{\theta}(t)$  represents the temporal evolution of a simulated accident scenario, i.e.,  $\boldsymbol{\theta}(t)$  can represent temperature inside the reactor core, the pressure level inside a containment building, the radionuclide concentration at a specific point outside the plant, etc.
- $\boldsymbol{\Xi}$  is the actual simulator code that describes how  $\boldsymbol{\theta}$  evolves in time
- $\mathbf{s} = \mathbf{s}(t, \mathbf{p})$  represents the status of components and systems of the model (e.g., status of emergency core cooling system, AC system)

By using the RISMC approach, if Monte-Carlo sampling is chosen, the PRA analysis is performed by [12]:

1. Associating a probabilistic distribution function (pdf) to the set of uncertain parameters  $\mathbf{p}$  (e.g., timing of events)
2. Performing stochastic sampling of the pdfs defined in Step 1
3. Performing a simulation run given  $\mathbf{p}$  sampled in Step 2, i.e., solve  $\frac{\partial \boldsymbol{\theta}}{\partial t} = \boldsymbol{\Xi}(\boldsymbol{\theta}, \mathbf{p}, \mathbf{s}, t)$
4. Repeating Steps 2 and 3  $M$  times and evaluating user defined stochastic parameters such as Core Damage (CD) probability  $P_{CD} = \frac{M_{CD}}{M}$  where  $M_{CD}$  is the number of simulations that lead to CD.

In a multi-unit type of scenario, the dynamic behavior of each unit is not independent but it can actually interact with the other units (e.g., electrical cross-ties and shared plant resources such as portable AC generators). Since  $\frac{\partial \boldsymbol{\theta}}{\partial t} = \boldsymbol{\Xi}(\boldsymbol{\theta}, \mathbf{p}, \mathbf{s}, t)$  refers to a single unit plant site, if multiple units are considered then it is needed to track the temporal evolution of each unit, i.e., multiple  $\boldsymbol{\theta}$  needs to be evaluated (one for each unit). Assuming that a three-unit plant is considered,  $\frac{\partial \boldsymbol{\theta}}{\partial t}$  now becomes as follows:

$$\begin{cases} \frac{\partial \boldsymbol{\theta}_1}{\partial t} = \boldsymbol{\Xi}_1(\boldsymbol{\theta}_1, \mathbf{p}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, t) \\ \frac{\partial \boldsymbol{\theta}_2}{\partial t} = \boldsymbol{\Xi}_2(\boldsymbol{\theta}_2, \mathbf{p}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, t) \\ \frac{\partial \boldsymbol{\theta}_3}{\partial t} = \boldsymbol{\Xi}_3(\boldsymbol{\theta}_3, \mathbf{p}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, t) \end{cases} \quad (1)$$

Note that now the vector  $\mathbf{s}_i (i = 1, \dots, 3)$  of each unit is shared among other units. This feature captures shared resources and possible system cross-ties among units. In addition, intra-unit interactions (such as a sub-set of human actions in a unit) may be driven by the actual status of other unit (e.g., thermo-hydraulic limit and operational boundaries). Again, these actions may have cascade effects on the other units. This is particularly relevant for severe accident scenarios. Thus, now Eq. 1 becomes:

$$\begin{cases} \frac{\partial \boldsymbol{\theta}_1}{\partial t} = \boldsymbol{\Xi}_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \mathbf{p}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, t) \\ \frac{\partial \boldsymbol{\theta}_2}{\partial t} = \boldsymbol{\Xi}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \mathbf{p}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, t) \\ \frac{\partial \boldsymbol{\theta}_3}{\partial t} = \boldsymbol{\Xi}_3(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \mathbf{p}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, t) \end{cases} \quad (2)$$

From a modeling point of view, solving Eq. 1 or Eq. 2 poses different challenges. Equation 1 can in fact be solved by:

- Sampling the set of uncertain parameters  $\mathbf{p}$
- Determining the temporal profile of  $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$
- Run the simulator for each unit independently given  $\mathbf{p}, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$

On the other side, solving Eq. 2 requires a system simulator that allows running the simulation of each unit simultaneously and sharing the variables  $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3$  among them. This paper focuses on the case described by Eq. 1.

## 110 4. Test Case

### 4.1. Plant Description

For our analysis we have chosen a 3-unit plant site as shown in Fig. 1. The chosen layout is not representative of any existing plant but it is simply fictitious.

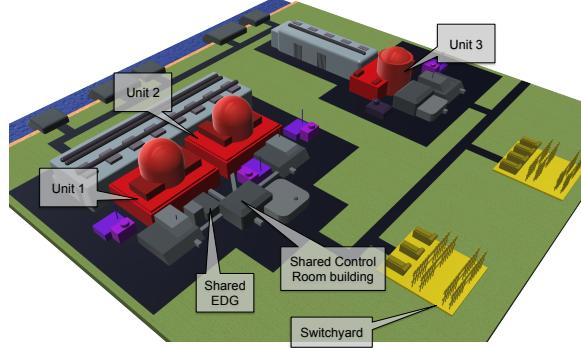


Figure 1: Overview of the multi-unit plant.

From a topographical perspective, a large body of water is located in proximity  
<sup>115</sup> of the NPP and it is employed as ultimate heat-sink for the plant.

All three units are composed by Pressurized Water Reactor (PWR) systems; the design of the PWR systems are identical for all the three units and it can be considered generic, i.e., it is not specific to an existing plant. For each PWR, the systems considered in the analysis are the following: High Pressure  
<sup>120</sup> Injection System (HPIS), Low Pressure Injection System (LPIS), Residual Heat Removal (RHR) system, Accumulators (ACCs), Auxiliary Feed-water (AFW) system, Charging pumps and Component Cooling Water (CCW) and Service Water (SW).

Special attention has been given to the design of the electrical and hydraulic  
<sup>125</sup> systems:

- The plant electrical system is shown in Fig. 2. Two electrical switch-yards can provide electrical power to all units. All units have a set of Emergency Diesel Generators (EDGs) and, in addition, a swing EDG (i.e., EDGS) can be employed to provide AC power to either Unit 1 or Unit 2. Note that also the 6.6 KV emergency buses of Unit 1 and Unit 2 can be cross-tied.  
<sup>130</sup>
- The AFW system of Unit 1 and Unit 3 can be cross-tied. Thus cooling to the secondary side can be provided from one unit to the other one.
- The Condensate Storage Tanks (CSTs) of Units 2 and Unit 3 can be cross-tied. Thus, the water source for the secondary side of either unit can be used as water source for the other one.  
<sup>135</sup>
- Plant recovery crew is a shared resource within the plant. Emergency Portable Equipments (EPEs) can be employed in order to restore water flow or AC power into the PWRs or SFPs. Each unit has its own set of EPEs but it is here assumed that a single EPE team is present within the plant boundaries.  
<sup>140</sup>

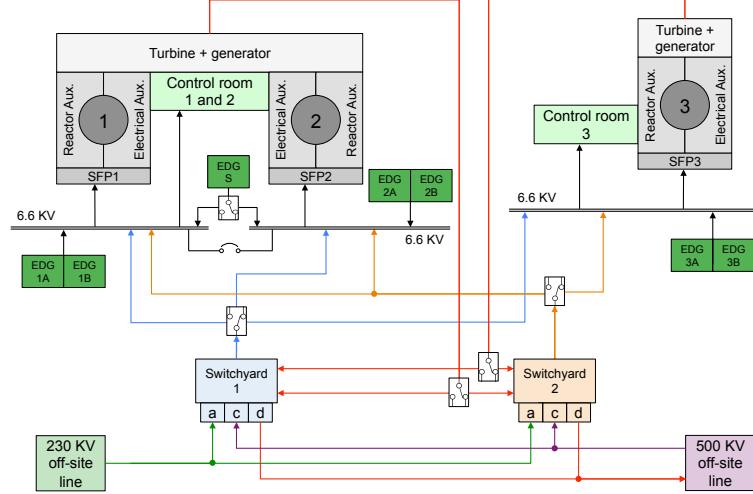


Figure 2: Plant electrical scheme.

#### 4.2. Initiating Event

The considered initiating event is a seismic event which causes the following events:

- Both switch-yards are disabled
- All EDGs are disabled except EDGS which is operating and it is initially aligned to Unit 2
- CSTs of Unit 2 and 3 have respectively lost 80% and 100% of their water inventory due to structural failure
- The seismic event might also rupture the SFPs. Thus a leak might be present during the accident scenario. Note that this could come from an aftershock or from a progressive structural degradation of the SFP

The proposed accident scenario resembles a Station Black Out (SBO) event at the plant level except for the fact that the EDGS is the only source of AC power available and it can be directed toward either Unit 1 or Unit 2. Prior to the seismic event, the initial conditions of three units are summarized in Table 1.

#### 4.3. Accident progression

Given the initiating event and the status of the plant, the analyzed accident focuses on the recovery strategy in order to place all PWRs and SFPs in a safe condition. For the scope of this analysis we consider a unit in safe state when EPEs are connected to the unit (i.e., both PWR and SFP). These EPEs are located between the plant site boundaries and once connected to a unit they can provide AC power and water injection.

Unit	Initial Conditions
1	The PWR of Unit 1 (i.e., PWR1) is at full power and it own SFP
2	The PWR of Unit 2 (i.e., PWR2) is in mid-loop operation (i.e., shut-down mode) and it own SFP. The mid-loop status is characterized by a primary coolant system drained to the hot leg centerline and the existence of openings which a further reduction of the mass inventory poses a serious risk, due to boil off and possible entrainment or spill over of liquid
3	The PWR of Unit 3 (i.e., PWR3) is at full power (nominal power level of PWR3 is 8% higher than PWR1 and PWR2) that restarted a few weeks earlier and its own SFP with a higher heat load since it contains used fuel recently moved from the reactor.

Table 1: Initial status of the three unit prior the accident scenario.

An EPE is available for each unit and we assume that the seismic event have not damaged the three EPEs available within the NPP. It is assumed that the EPE team can assist only one unit at a time, i.e. if three units need their own EPE, then the EPE team must first prioritize the units that require immediate assistance.

In our case, even though Unit 2 it is the one with AC power available, it is in the most vulnerable situation since it is in mid-loop condition: low water inventory in the primary system and Reactor Pressure Vessel (RPV) head removed. Hence, in case of CD condition, radioactive material will be directly released in the containment. Unit 3 is the unit in the most critical condition since it is in SBO condition and heat removal capabilities are limited due to the fact that only 20% of the CST inventory is available. Unit 1 is in SBO condition like Unit 3 but it has more time margin before reaching CD since CST water inventory of Unit 3 is lost.

*Given this assessment, three different strategies have been hypothesized as possible courses of action; these strategies are described in detail in Sections 4.3.1, 4.3.2 and 4.3.3. The set of recovery strategies have been designed given the status of the three units right after the initiating event through a prioritization scheme given the set of cross-ties (either electrical or hydraulic) that can be performed. This scheme includes not only the evaluation of the units in critical conditions (i.e., Unit 1 and 3) but also the unit with high potential of radioactive release (i.e., Unit 2). Units 3 and 1 are the most critical ones (both without AC power) from a level 1 PRA point of view while Unit 2 is considered the most critical one from a radioactive release point of view (since it is in mid-loop condition). The set of strategies has been formulated depending on which prioritization scheme is followed: either a level 1 PRA or radioactive release centric scheme. Strategy 1 and 2 are very similar and they follow a radioactive release centric prioritization scheme; thus the sequence is: Unit 2, Unit 3 and Unit 1. Strategy 3 follows a level 1 PRA prioritization scheme; thus the sequence is: Unit 3, Unit 1 and Unit 2.*

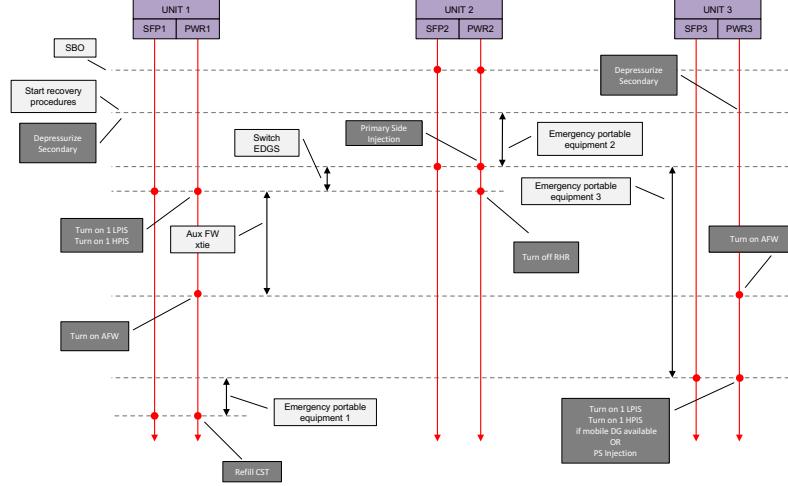


Figure 3: Sequence of events of recovery strategy 1.

#### 4.3.1. Strategy 1

This strategy prioritizes Unit 2 since it is in mid-loop condition. Hence the EPE team is initially directed toward this unit. Once this task has been accomplished, Unit 3 becomes now the next priority. In order to put this unit in a safe condition two parallel directions are followed:

- Move the EPE team toward Unit 3
- Align the EDGS from Unit 2 to Unit 1 so that also Unit 1 can be placed in a safe condition
- Provide cooling through AFW cross-tie between PWR1 and PWR3 (once completed). Note that the AFW cross-tie does not provide cooling to the SFP

Finally, once Unit 3 EPE has been connected, the EPE team move Unit 1. At this point Unit 1 should be already in a safe condition since EDGS has been aligned to Unit 1; this step has been added since the simulation (for both PWR and SFP) stops one hour after the EPE connected to the unit. A temporal scheme of the temporal evolution of this accident scenario is shown in Fig. 3.

#### 4.3.2. Strategy 2

This strategy, similarly to Strategy 1, prioritizes Unit 2 since it is in mid-loop condition; thus, the EPE team is initially directed toward this unit. Once this task has been accomplished, Unit 3 becomes again now the priority. In order to put this unit in a safe condition two parallel directions are followed:

- Move the EPE team toward Unit 3

- <sup>215</sup>
- Align the EDGS from Unit 2 to Unit 1 so that also Unit 1 can be placed in a safe condition
  - Provide CST inventory from PWR 2 and PWR 3 (once completed). Note that the CST cross-tie does not provide cooling to the SFP

<sup>220</sup> Finally, once Unit 3 EPE has been connected, the EPE team move Unit 1 as indicated also for Strategy 1.

#### *4.3.3. Strategy 3*

<sup>225</sup> This strategy prioritizes Unit 3; hence the EPE team is initially directed toward this unit. Once this task has been accomplished, Unit 1 becomes now the priority. In order to put this unit in a safe condition two parallel directions are followed:

- <sup>230</sup>
- Move the EPE team toward Unit 1
  - Perform an AC cross-tie so that AC power generated by EDGS can be employed to provide power to both Unit 1 and Unit 2. Note that it is here assumed that a correct AC management is implemented in order to avoid over-load of the EDGS.

Finally, once Unit 1 EPE has been connected, the EPE team move Unit 2. At this point Unit 2 should be already in a safe condition since EDGS is still aligned to Unit 2; again, this step has been added since the simulation run (for both PWR and SFP) stops when the EPE connected to the unit.

#### *4.3.4. EDGS Erroneous Alignment*

<sup>235</sup> As part of the simulation we have introduced a stochastic event: the erroneous alignment of the EDGS. This event can occur anytime during the simulation and it represents the erroneous alignment of the EDGS (initially aligned to Unit 2) to Unit 1. While this event can be considered an error of commission from an HRA perspective, from an operational point of view it does leave Unit <sup>240</sup> 2 in an unsafe situation but it provides a safe condition to Unit 1. Note that now this stochastic element adds an additional degree of freedom in the accident progression since the occurrence of this event changes the prioritization of the next unit to have the EPE connected.

#### *4.4. EPE Actions*

<sup>245</sup> Depending on the unit status, different actions can be employed when connecting EPEs to their respective unit. The EPE actions involve injection of water into the Primary Side (or directly into the CST) by employing mobile pumps or by providing AC power to the unit with auxiliary generators so that HPIS and LPIS can be employed.

## 5. RISMC Approach to Multi-Unit Modeling

The actual multi-unit model has been modeled using both RELAP5-3D and RAVEN. The RELAP5-3D models are employed to determine the temporal response of all PWRs and SFPs while the plant connections and dependencies have been coded by employing the RAVEN ensemble model capabilities. The multi-unit model is structured so that it receives in input the set of sampled values (refer to Section 5.5 for the complete list of the chosen stochastic parameters) and it provides in output the status of the six models (PWRs and SFPs). Internally, the multi-unit model consists of a plant model in series with the six RELAP5-3D models (which are arranged in a parallel configuration). The plant model determines timing and sequencing of events of the accident scenario and it provide these values to the RELAP5-3D models. The following sections describe in detail the RELAP5-3D models for the SFPs and the PWRs while Section 5.3 describes the plant model.

### 5.1. PWR1 and PWR3

RELAP5-3D PWR models are based on the so-called INL Generic PWR (IGPWR) model [21, 22]. The input deck is modeling a 2.5 GWth Westinghouse 3-loop PWR, including the RPV, the 3 loops and the primary and secondary sides of the Steam Generators (SGs). Four independent channels are used for representing the reactor core. Three channels model the active core and one channel models the core bypass. Different power values are assigned to the three core channels in order to take into account the radial power distribution. Passive and active heat structures simulate the heat transfer between the coolant and fuel, the structures and the secondary side of the IGPWR. The possible operator actions that can occur during a SBO event with the reactor at full power (Unit 1 and Unit 3) are implemented through the RELAP5-3D control logic. These actions include SG cool-down, feed and bleed, AFW flow control, primary/secondary side emergency injection, etc.

### 5.2. PWR2

Unit 2 is based on the same RELAP5-3D IGPWR model, consistently modified for simulating the mid-loop conditions [23]. The model represents Unit 2 during the refueling outage phase, with steam generator 3 manway opened for maintenance and pressurizer safety valves removed. Decay heat is removed by the Residual Heat Removal System (RHRS). Because the openings on the primary system (in the pressurizer and steam generator 3), the use of steam generators for reflux cooling in case of loss of RHR is prevented. However cooling by gravity drain from the RWST is possible.

The water level in the primary circuit is kept at the middle plane of the hot and cold legs by a recirculation circuit modeling the RHRS. The parts of the primary and the secondary systems with no water contain air at atmospheric condition.

### 5.2.1. SFPs

The three SFPs were also modeled [24] using RELAP5-3D. For the sake of simplicity, the two cooling loops were modeled just as boundary conditions (assigned water inlet/outlet mass flow rates and temperatures). The opening of a valve on the pool bottom could simulate the fuel pool break by seismic effects. The thermal-hydraulic scheme is based on 43 volumes and 49 junctions. Two Heat Structures are used for modeling  $15 \times 15$  Westinghouse FA. The two independent cooling systems are modeled as boundary conditions (i.e., mass flow and temperature inlet are imposed). A couple of valves are used to model medium and large breaks of the pool bottom. The cooling systems pumps trip if SFP liquid level lower than 0.1 m or if SFP temperature greater than 349 K. Emergency crew action (emergency water injection) is assumed when both the recirculation pumps trip.

### 5.3. Plant Model

The plant model has been coded in Python script and interfaced with RAVEN as an external model. Its main purpose is to determine timing and sequencing of events for all six system models (i.e., PWRs and SFPs) given the sampled values of the stochastic parameters.

### 5.4. Human Models

To consider the performance of field workers, a simple human reliability analysis was performed and incorporated into the event simulation. Although many human actions are required, the analysis considers two human events: time to start recovery procedures after initiating event and EDGS erroneous alignment.

Regarding the first, we employed a detailed subtask modeling using the HUNTER [25] framework. The details about the model can be found in [26]; for the scope of this report, this event is modeled as a stochastic variable with a pdf derived from [26].

Regarding the second human related event, in order to screen the impact of the possible error, the human error probability (HEP) was obtained from the Technique for Human Error Rate Prediction (THERP) method [27]. EDGS erroneous alignment corresponds well to THERP Table 20-13, Item 5: “Making an error of selection in changing or restoring locally operated valve when the valve to be manipulated is unclearly or ambiguously labeled, part of a group of two or more valves that are similar in all of the following, size and shape, state, and presence of tags.” This THERP item captures the nature of the task, particularly regarding complexity and ambiguity about pipe arrangements and corresponding valve operation.

THERP produces an HEP equal to  $1.0 \cdot 10^{-2}$ , with an uncertainty error factor of 3. THERP includes provision for considering additional degradation of performance due to lack of experience and situational stress. These factors were not deemed likely contributors to the event outcome. Event recovery was not modeled. Thus in the analysis, we have modeled erroneous alignment of EDGS with a Bernoulli distribution with value of  $p = 1.0 \cdot 10^{-2}$ .

### 5.5. Plant Stochastic Modeling

For the scope of this analysis, we have identified 23 stochastic parameters. We have grouped these parameters based on their area of interest.

Regarding the SFPs, we have have identified seismic induced rupture, i.e.,  
340 a SFP Loss Of Coolant Accident (LOCA), as element to include into the analysis. SFP LOCA has been represented by two stochastic parameters: time of occurrence and size of the SFP LOCA. In this work we have identified with locaTimeSFP1, locaTimeSFP2 and locaTimeSFP3 as the time of occurrence of the SFP LOCAs while locaSizeSFP1, locaSizeSFP2 and locaSizeSFP3 represents the actual size of the SFP LOCAs. *The choice of the distribution of the size was based on the design of the SFPs while distribution of the time was arbitrarily chosen in order to identify the impact of SFP LOCA timing on accident progression.*

Regarding the PWRs, we selected two elements: lifetime of the batteries  
350 and the LOCA associated to the seal of the Reactor Coolant Pumps (RCPs). Battery systems provides DC power to I&C systems of the PWRs such as the control of the Pilot Operated Relief Valves (PORVs). DC systems are considered for only Unit 1 and Unit 3; since Unit 2 is in mid-loop operation mode, its DC systems are not considered. *The distribution of the batteries lifetime and the RCP LOCA were derived from the SOARCA report [28].*

For the EDGS, we have identified the following parameters: the probability to erroneous align the EDGS from Unit 2 to Unit 1, the time of occurrence of EDGS erroneous alignment and time required to perform EDGS voluntary alignment. *Such distributions are described in Section 5.4.*

Each cross-tie, CST (between Unit 2 and Unit 3), AFW (between Unit 1  
360 and Unit 3) and AC (between Unit 1 and Unit 2), has been considered in the analysis and, in particular, they have been modeled by assigning to each of them the time required to perform such cross-tie. *Such distributions have been generated by taking into considerations operational data including the SOARCA report [28].*

Regarding the recovery of each unit through the EPEs, we have modeled them by representing them with a single stochastic parameter which represents the time to connect the EPE to its own unit. *Distributions associated to these events were arbitrarily chosen since the objective of the paper was to investigate the impact of timing of events on accident progression.*

Lastly, the recovery plan followed by the plant crew has been modeled using a single parameter: recoveryStrategy (see Section 4.3). *The distribution associated to this stochastic parameter was designed to give similar probability to each recovery strategy.*

In addition to recovery strategy, we have given an additional degree of freedom on the actual procedure associated to the EPE for Unit 3. As indicated in Section 4.4, depending on the recovery strategy, then the EPE connection on Unit 3 can be performed in different modes. *In this case, it has been chosen to provide higher likelihood to the AC recovery compared to PS injection.*

375 A summary of the chosen stochastic parameters are listed in Table 2 along with their description and probabilistic distribution function.

Parameter	Description	Unit	Distribution
AUXFWxtieTime	Time to perform AFW cross-tie	hour	Uniform (lower bound=.5, upper bound=1.5)
CSTxtieTime	Time to perform CST cross-tie	hour	Uniform (lower bound=.5, upper bound=1.5)
recoveryStrategy	Recovery strategy to be followed	-	Categorical(1,2,3) (p(1)=.3, p(2)=.3, p(3)=.4)
recovProcedTime	Time to start plant recovery procedure	hour	Truncated normal (mean=1., sigma=.2, lower bound=.5, upper bound=1.5)
EPETime1	Time to connect EPE to Unit 1	hour	Truncated normal (mean=2., sigma=.3, lower bound=1., upper bound=3.)
EPETime2	Time to connect EPE to Unit 2	hour	Truncated normal (mean=2., sigma=.3, lower bound=1., upper bound=3.)
EPETime3	Time to connect EPE to Unit 3	hour	Truncated normal (mean=2., sigma=.3, lower bound=1., upper bound=3.)
EDGSerrAlign	Probability of occurrence for EDGS erroneous alignment	-	Bernoulli (p=0.01)
EDGSerrAlignTime	Time of occurrence for EDGS erroneous alignment	-	Uniform (lower bound=.0, upper bound=1.)
batteryTime1	Battery life for Unit 1	hour	Triangular (lower bound=6., upper bound=8., peak=7.)
batteryTime3	Battery life for Unit 3	hour	Triangular (lower bound=6., upper bound=8., peak=7.)
EDGSSswitchTime	Time required to change EDGS alignment	hour	Uniform (lower bound=.25, upper bound=.75)

Table 2: Summary of the stochastic parameters chosen for the multi-unit analysis and their associated distribution.

Parameter	Description	Unit	Distribution
ACxTieUnit12	Time to perform AC cross-tie	hour	Uniform (lower bound=.5, upper bound=1.)
locaTimePWR1	Time of occurrence for PWR1 seal LOCA	hour	Uniform (lower bound=.1667, upper bound=.25)
locaTimePWR3	Time of occurrence for PWR3 seal LOCA	hour	Uniform (lower bound=.1667, upper bound=.25)
locaSizeSFP1	LOCA size for SFP1	gpm	Categorical(0.0004,0.0035,0.056) (p(0.0004)=.85, p(0.0035)=.1, p(0.056)=.05)
locaSizeSFP2	LOCA size for SFP1	gpm	Categorical(0.0004,0.0035,0.056) (p(0.0004)=.85, p(0.0035)=.1, p(0.056)=.05)
locaSizeSFP3	LOCA size for SFP1	gpm	Categorical(0.0004,0.0035,0.056) (p(0.0004)=.85, p(0.0035)=.1, p(0.056)=.05)
locaTimeSFP1	Time of occurrence for SFP1 LOCA	hour	Categorical(0.,.1667,.333,.5,24.) (p(.0)=.025, p(.1667)=.025, p(.333)=.025, p(.5)=.025, p(24.)=.9)
locaTimeSFP2	Time of occurrence for SFP2 LOCA	hour	Categorical(0.,.1667,.333,.5,24.) (p(.0)=.025, p(.1667)=.025, p(.333)=.025, p(.5)=.025, p(24.)=.9)
locaTimeSFP3	Time of occurrence for SFP3 LOCA	hour	Categorical(0.,.1667,.333,.5,24.) (p(.0)=.025, p(.1667)=.025, p(.333)=.025, p(.5)=.025, p(24.)=.9)
flex3Strategy13	Type of EPE connection for Unit 3 during recovery strategy 1 and 3	-	Categorical(1,2) (p(1)=.3, p(2)=.7)
flex3Strategy2	Type of EPE connection for Unit 3 during recovery strategy 2	-	Categorical(1,2) (p(1)=.4, p(2)=.6)

Table 2: Summary of the stochastic parameters chosen for the multi-unit analysis and their associated distribution (cont.ed).

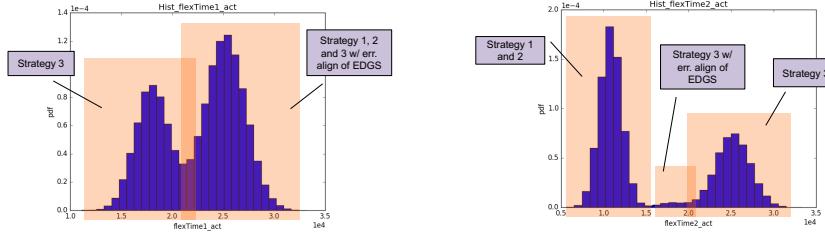


Figure 4: Histogram of EPE actuation time for Unit 1 (left), Unit 2 (right).

## 6. Multi-Unit PRA

An issue related to the model shown in Section 5 is that the overall computation time can be very high (about 10 hours) if the outcome of each of the six models (i.e., PWRs and SFPs) is determined using RELAP5-3D. Since the scope of our PRA analysis is to generate a large number number of scenarios the computational time of the RISMIC analysis would be too large even for modern high performance computing systems.

In order to decrease the computational time of the analysis, we have employed ROMs instead of running the RELAP5-3D models. The objective is to employ a ROM for each PWR and SFP instead the actual code.

The actual workflow that implements such approach is as follows:

1. Sample the response of each PWR and SFP using the RELAP5-3D code
2. Train a ROM using the data generated in Step 1
3. Validate the response of the ROMs against the response of the actual code
4. Insert the validated ROMs in the ensemble model in place of the RELAP5-3D code

For the six plant models considered in this analysis, we have chosen to use  $k$ -nearest neighbor classifiers as surrogates to predict the presence of core damage in each of the models. We leave in Appendix A the details about how the ROMs were generated from the RELAP-5 models along with their validation analysis.

Hence, given the model described in Sections 5 and 10 along with the set of stochastic parameters listed in Section 5.5 we have simulated  $10^6$  accident scenarios using RAVEN Monte-Carlo sampling capabilities. For each simulation, we have collected the binary output (OK or CD) from each model/ROM (i.e., all PWRs and SFPs). The use of ROMs instead of the actual codes allowed us to generate this very large database of data which helped us to visualize and understand timing and sequencing of events at the unit level.

Figure 4 gives an example of pdf of the actual EPE actuation time (i.e., in absolute time) for Unit 1 and Unit 2 . The presence of three distinct site recovery strategies and the possibility to erroneous align EDGS from Unit 2 to Unit 1 strongly affect the time convolution of such timing of events.

## 7. Data Analysis Methods

Historically the concept of CD probability has been associated to a single unit. At a plant level, a separate value of CD probability can be associated to all PWRs and SFPs. However, note that there is a high correlation among the six models of the plant site (PWRs and SFPs). Thus, a high correlation among CD probabilities of the six models is also expected.

Instead of defining a single CD probability value for each PWR and SFP we have defined a probability value to a Plant Damage State (PDS). A PDS is a 6-dimensional vector where each vector element describes the status of a plant model. Since two possible values to each element of the vector are permitted (OK or CD),  $2^6 = 64$  possible combinations are allowed. In order to analyze the data generated by RAVEN we have selected a three-step approach: 1) group simulation runs based on their own PDS, 2) evaluate probability associated to each PDS and rank PDSs based on their probability values and 3) identify commonalities that characterize each PDS.

### 7.1. Error Estimation

As indicated earlier, for each PDS a probability value needs to be determined. Part of this determination includes the evaluation of the statistical error associated to the probability value. Such evaluation has been performed using classical Bayesian inference.

The results for the PRA analysis are consistently provided as an occurrence of event divided by the total number of events to produce a percentage. Instinctually it is clear that this type of data corresponds to a beta-binomial distribution. Since it is not desired to project the rate of occurrence or probability for each event, a Jeffreys non-informative prior is implemented on the beta-binomial distribution [29]. The parameters of the beta-binomial distribution are then fed to a function that automatically produces the 5<sup>th</sup> and 95<sup>th</sup> percentiles of the distribution. This is reported rather than the variance, as variance in a beta-binomial distribution does not adhere to the classic normal distribution format.

## 8. Results

The  $10^6$  samples have been post-processed by partitioning the data set in 64 subsets, a subset for each PDS. For each subset/PDS a value of probability and error estimate associated to it have been determined (see Section 7).

Table 3 summarizes these findings and it ranks the PDS based on their probability. First of all, note that 14 out of 64 PDSs were actually generated; i.e., none of the  $10^6$  samples belong to 50 PDSs.

*Since the actual response of each model has been determined by using ROMs instead of the actual RELAP5 model, as indicated in Appendix B, such response is affected by errors. The evaluation of the prediction error of each of the six ROMs is summarized in Table 4. Even though the prediction error is below 1%, we have evaluated its impact on the probability values provided in Table 3. This*

ID	PDS						Probability		
	PWR1	PWR2	PWR3	SFP1	SFP2	SFP3	mean	5 <sup>th</sup>	95 <sup>th</sup>
8	OK	OK	CD	OK	OK	OK	0.8902	0.8897	0.8907
12	OK	OK	CD	CD	OK	OK	5.89E-2	5.85E-2	5.93E-2
10	OK	OK	CD	OK	CD	OK	3.39E-2	3.37e-2	3.43E-2
9	OK	OK	CD	OK	OK	CD	1.26E-2	1.24E-2	1.28E-2
24	OK	CD	CD	OK	OK	OK	2.10E-3	2.03E-3	2.18E-3
13	OK	OK	CD	CD	OK	CD	1.17E-3	1.11E-3	1.23E-3
14	OK	OK	CD	CD	CD	OK	5.81E-4	5.42E-4	6.21E-4
11	OK	OK	CD	OK	CD	CD	1.65E-4	1.44E-4	1.87E-4
26	OK	CD	CD	OK	CD	OK	1.56E-4	1.36E-4	1.77E-4
28	OK	CD	CD	CD	OK	OK	1.11E-4	9.43E-5	1.289E-4
25	OK	CD	CD	OK	OK	CD	1.10E-5	6.17E-6	1.70E-5
15	OK	OK	CD	CD	CD	CD	6.00E-06	2.61E-06	1.05E-05
30	OK	CD	CD	CD	CD	OK	5.00E-06	1.97E-06	9.15E-06
29	OK	CD	CD	CD	OK	CD	1.00E-06	5.13E-08	3.00E-06

Table 3: Multi-unit analysis results.

455 has been performed by employing a Markov model approach where each PDS is  
 represented by a state and transitions among states are dictated by evaluating the  
 impact of ROM prediction errors. As an example, transition from PDS 8 to PDS  
 12 is dictated by the erroneous prediction of the SFP1 ROM with a transition  
 probability equal to  $2.8E - 3$  (i.e.,  $1.0 - 0.9972$ ). We use this notation:  $P$  is  
 460 the state probability vector and  $A$  is the transition matrix. Accordingly to this  
 model, vector  $P^{ROM}$  contains PDS probability shown in Table 3 while vector  
 $P^{code}$  represents the PDS probabilities if the RELAP5 models were employed  
 instead of ROMs. From this Markov model we can write  $P^{ROM} = AP^{code}$  and,  
 thus, given  $A$  and  $P^{ROM}$  we were able to obtain  $P^{code}$ . The obtained  $P^{code}$   
 465 vector compared to  $P^{ROM}$  presents discrepancies in the  $1.E - 2$  to  $1.E - 5$   
 range.

From Table 3 note that none of the recovery strategies are able to recovery  
 PWR3: its condition at the beginning of the accident is the worst among the  
 three units (lost of CST inventory on top of SBO condition). From separate  
 470 calculations, PWR3 could be saved only if EPE3 would be connected withing  
 the first 50 minutes after SBO condition. Such condition, cannot be met given  
 the boundary conditions of the accident progression.

PWR1, on the other side, never reach CD condition: this is due to the fact  
 that CST inventory is intact (compare to PWR3) and, thus, the time required  
 475 to reach CD condition is much longer. In addition, PWR1 can be put in safe  
 condition through several ways (see Section 4).

PWR2 and the SFPS appear to reach both CD and OK condition. The  
 objective of the analysis is now to understand what are the driving factors  
 behind each PDS instead of focusing only on PDS probabilities.

480 In the next sections, each of the 24 data subsets (i.e, 24 PDs) is analyzed in

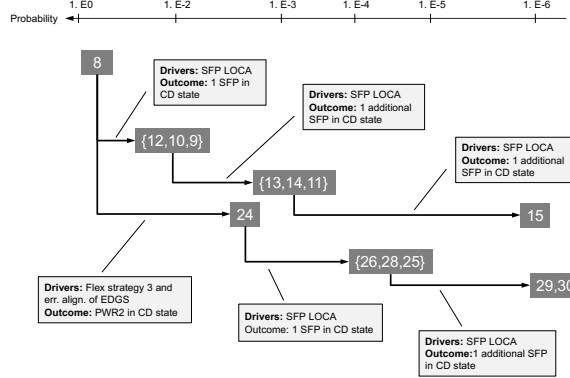


Figure 5: Summary of the relationships among PDSs.

order to discover the input drivers for each PDS. For the majority of them we report the histogram of the input variables for that particular PDS and compare with the full data set. The comparison is made visually by presenting:

- 485
- the histogram of a specific variable for the full data set in gray color, and,
  - the histogram of the same variable for the PDS subset in a brighter color (i.e., not gray)

A summary of the relationships among PDSs pictured in a hierarchical fashion is shown in Fig. 5; this figures also summarizes the most important drivers and the consequences related to the drivers.

490 *8.1. PDS 8*

This PDS contains the majority of the data generated (about 890,000 samples fall in this category). A first observation is about the SFPs: by looking at the histograms of the variable locaTime for each SFP we see a drift of the histogram toward the highest bin (see Fig. 6). The variable locaTimeSFP indicates 495 when the actual LOCA occurs: recall that locaTimeSFP=86400 implies LOCA does not occur. Figure 6 implies that despite the presence of a loss of fluid in a SFP, it is possible to put the SFP in safe condition if certain conditions are met.

The next step is to discover which are these conditions: this can be accomplished by considering the data samples in PDS8 that actually have a SFP LOCA (i.e., samples characterized by locaTimeSFP2 different than 86400) and creating a scatter plot in a 2-dimensional space where one dimension is the size of the SFP LOCA and the the second one is the absolute time (i.e., the exact time) when the SFP is put in a safe state. Depending on the unit, this safe state can be reached in different ways:

- 500
- Unit 1: when EDGS is aligned to Unit 1 or when EPE1 is connected to Unit 1

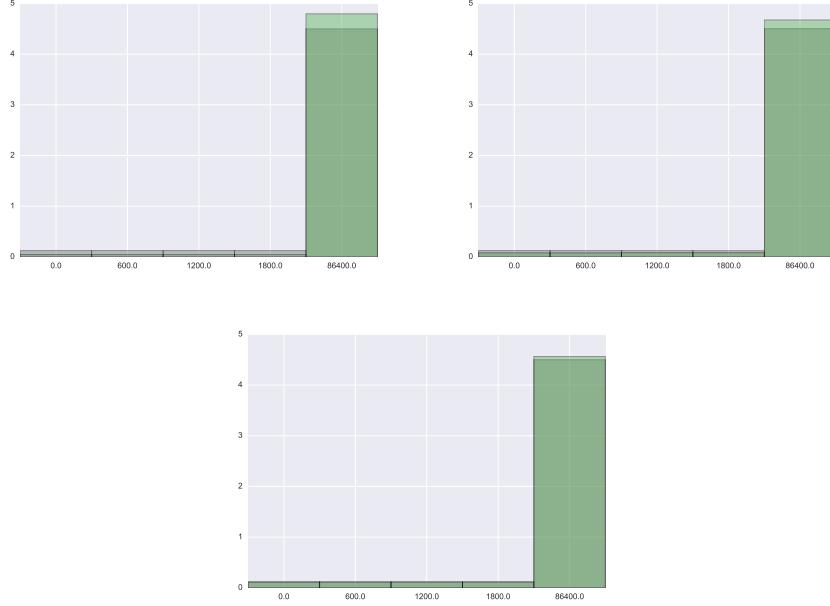


Figure 6: PDS8: histograms of the variables locaTimeSFP1 (top left), locaTimeSFP2 (top right) and locaTimeSFP3 (bottom).

- Unit 2: when EPE2 is connected to Unit 2; note that the erroneous alignment of EDGS plays a role here since its occurrence brings it in an critical condition (no cooling in SFP)
- 510
- Unit 3: when EPE3 is connected to Unit 3

Figure 7 shows four images; each image contains a scatter plot and an histogram along each dimension. Recall that all points in this scatter plots are characterized by SFP in an OK state. From Fig. 7 we can see the following:

- Unit 1 (top left image): this scatter plot shows locaSizeSFP1 vs. SFP1 recovery time (which is represented as the minimum value among EDGS is aligned to Unit 1) and EPE1 connected to Unit 1. Note that a SFP1 recovery time less than 25,000 seconds (points on the left of the scatter plot) can recover a small SFP LOCA (i.e., 5.E-4). Note also that a few points are clustered at around 12,000 seconds for SFP1 recovery time and a medium SFP LOCA (i.e., 3.5E-3). This small group of points are characterized by the following distinctive features: recovery strategy 3, no EDGS erroneous alignment and very early AC12 cross-tie (i.e., AC power of Unit 2 is provided to Unit 1 through a AC cross-tie). This feature implies that even a medium SFP LOCA can be recovered only if recovery strategy 3 is chosen and, the AC cross-tie between Unit 2 and Unit 1 is
- 515
- 520
- 525

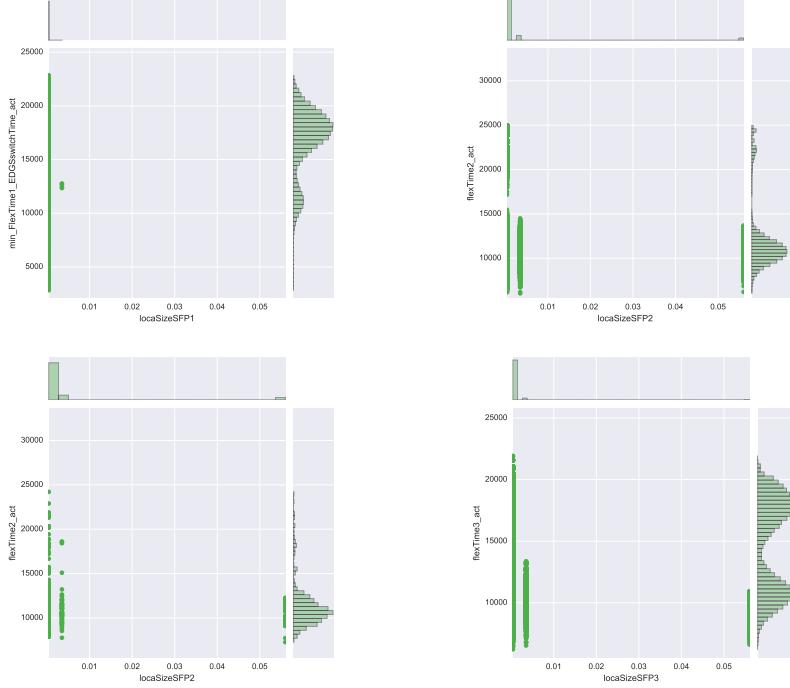


Figure 7: PDS8: scatter plot for SFP1 (top left), SFP2 with EDGS erroneous alignment (top right), SFP2 without EDGS erroneous alignment (bottom right) and SFP3 (bottom right).

completed before 12,700 s. From this plot it can be observed that a large SFP LOCA can not be recovered.

- Unit 2 (top right and bottom left images): these two scatter plots shows similar thresholds for Unit 2. Note two features: a large SFP LOCA can be recovered and the EDGS erroneous alignment impacts the threshold value small and medium SFP LOCA.
- Unit 3 (bottom right image): since SFP3 as a higher heat load it is expected that the thresholds decrease. This is confirmed by comparing the bottom right scatter plot with the upper left plot of Fig. 7.

### 8.2. PDSs 12, 10, 9

PDSs number 12, 10 and 9 are characterized by a single SFP in CD condition (on top of PWR3): SFP1, SFP2 and SFP3 respectively. The main driver is the loss of water inventory due to the seismic induced SFP LOCA. This conclusion might be obvious given the nature of the system; however, if we observe the histogram of the recovery strategy in each of these three PDSs (see Fig. 8) we observe a pattern. PDS12 and PDS9 are dominated mainly by samples that

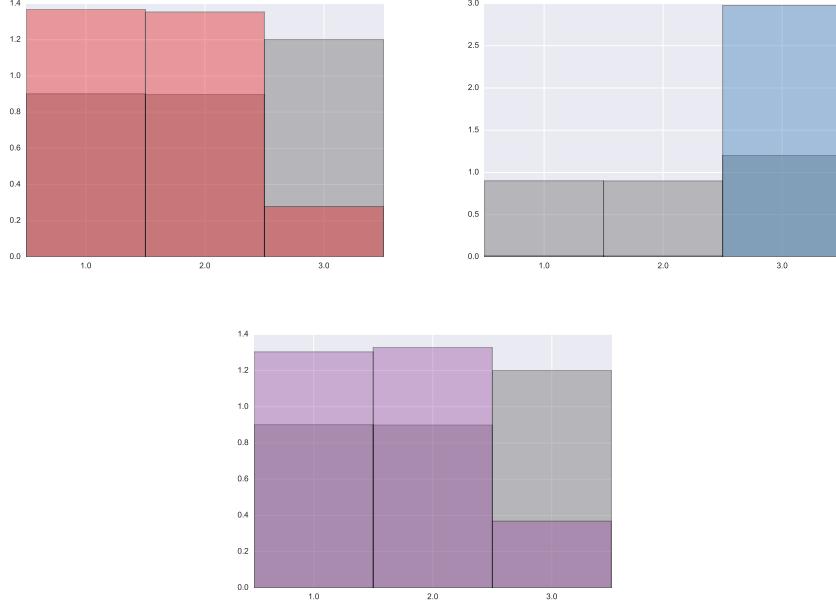


Figure 8: Histograms of the variable recovery strategy for PDS12 (top left), PDS10 (top right) and PDS9 (bottom)

follow Strategy 1 and 2 while PDS10 is exclusively characterized by simulations that followed Strategy 3. This is due to the fact that unit prioritization allows to recover only the first SFP through EPEs. Heating-up of the SFP is so fast that does not allow for two consecutive EPE timings to occur.

### 8.3. PDS 24

PDS24 is the first PDS that characterize an additional PWR to reach CD (on top of PWR3): PWR2. By looking at the histogram of the input parameters (see Fig. 9) that belong to this PDS we have identified that PWR2 reaches CD only if recovery strategy 3 is chosen. In addition, erroneous alignment of EDGS plays the major driver to reach PDS24. Interestingly, the time of such switch is also important: by looking at bottom histogram Fig. 9, the distribution of the variable EDGSerrAlignTime is characterized by two modes, an early mode and a late mode. This feature is due to the fact that, in strategy 3, EDGS erroneous alignment might run in parallel with EPE3 or EPE1. If this erroneous action occurs when EPE3 or EPE1 have just started, then PWR2 reaches CD almost certainly due to the PWR2 heat-up. If this erroneous action occurs when EPE3 or EPE1 are almost completed, then the EPE team has time to prioritize Unit 2 and quickly recover it. The two modes of the bottom histogram of Fig. 9 correspond to an EDGS erroneous action that occurs right after EPE operation for Unit 3 (early mode) and for Unit 1 (late mode) has started.

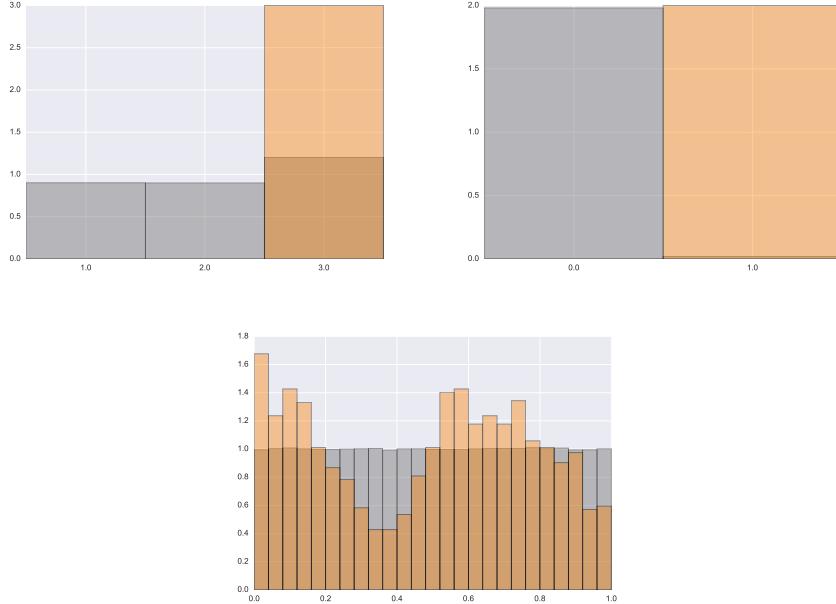


Figure 9: PDS24: histograms of the variables recovery (top left), EDGSerrAlign (top right) and EDGSerrAlignTime (bottom)

#### 8.4. PDSs 13, 14, 11

565 PDSs 13, 14 and 11 is a blend of PDS 12, 10 and 9: they contains 2 SFPs in CD condition (in addition to PWR3). These PDS can be simply characterized by the occurrence of 2 SFP LOCAs which are not correlated events; i.e., SFP LOCA have been modeled as independent events. Thus, the same conclusions derived from PDSs 9, 10 and 12, can be transposed for PDSs 13, 14 and 11.

#### 8.5. PDSs 26, 28, 25

570 PDSs 26, 28 and 25 are characterized by 1 SFP along with PWR2 and PWR3 in CD condition; thus it represents a mix of PDS 24 and PDS 12, 10 and 9. These PDS are in fact characterized by recovery strategy 3 and EDGS erroneous align along with a SFP LOCA. Similarly to what has been presented for PDS24, the interesting histogram of EDGS erroneous time for these three PDSs (see Fig. 10). Note these histograms follow the same pattern of Fig. 9 (bottom plot).

#### 8.6. PDS 15

580 PDS15 is characterized by having all SFPs in a CD state (along with PWR2). Similar to the considerations presented for PDSs 12, 10 and 9 (and also similar to PDSs 13, 14, 11), the main driver is a medium/large LOCA for all SFPs coupled with long EPE time.

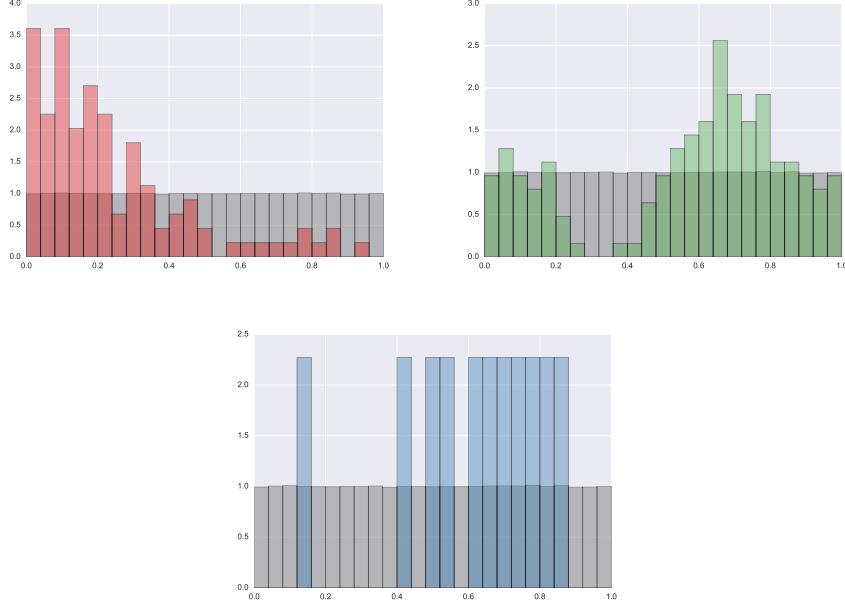


Figure 10: Histograms of the variable EDGSerrAlignTime for PDS28 (top left), PDS26 (top right) and PDS25 (bottom)

## 9. Conclusions

In this paper we have presented a first step toward a Dynamic PRA approach to analyze multi-unit sites. We have described in detail a method to perform the simulation-based PRA of a generic multi-unit site by employing RAVEN and RELAP5-3D codes.

The presented analysis exhaustively covered all major steps required to perform a RISMC analysis: plant deterministic and stochastic modeling, plant stochastic analysis and analysis of results.

Regarding Step 1, the considered plant site and accident scenario have been modeled in a great level of detail from both a deterministic and stochastic point of view. In particular, the RAVEN Ensemble Models allowed us to create the links among the six RELAP5-3D models and to model both system dependencies and timing/sequencing of events at the plant level.

An important feature of this step is that we have employed ROMs to predict the outcome of each unit model (both PWRs and SFPs). All ROMs have been trained by using large number of simulation runs and the ROM prediction has been properly validated. Regarding the stochastic modeling, note the presented analysis focused more on the NPP recovery actions while we have not introduced additional potential failures of system and components.

The plant stochastic analysis has been performed by using classical Monte-Carlo approach. This has been a natural choice since the computational time

605 was already decreased by employing ROMs instead of the actual codes. A very large number of simulation runs were calculated in order to significantly reduce the statistical error of the analysis.

610 We have presented a detailed analysis of the the generated data; this was accomplished by employing high performance computing systems due to high computational time of each simulation run and due to the high number of simulation runs requested. We have shown that more quantitative analysis details can be obtained from this kind of approach if compared to classical PRA methods that are based on ET/FT algorithms. We were able to identify how sequencing and timing of events affected the final outcome (i.e., the PDS).

615 *The presented approach to solve a multi-unit problem employed simulation tools instead of classical boolean structures such as ET and FT. These classical tools are static in nature, i.e., timing and sequencing of events are not implicitly modeled, and accident progression is set by the analyst. When modeling multi-unit systems, classical tools model system inter-dependencies by adding additional branches in each unit event-tree (e.g., CST cross tie for unit 3), adding additional basic events in fault-trees (erroneous alignment of EDGS for unit 2), solving the PRA model for each unit separately, and post-processing the newly obtained cut sets in order to eliminate event sequences that are not possible (basic event accounted for in Unit 1 and same event not accounted for in Unit 3).*

620 This work enhances state-of-art on multi-unit PRA modeling by implicitly considering timing and sequencing of events for all three units simultaneously in a single PRA framework, i.e., a PRA model is not solved separately for each unit and then post-processed to account for plant inter-dependencies. Each simulation directly evaluated timing and sequencing of events at the plant level (no need to post-processed the data to eliminate impossible sequences). Plant accident progression is also not set by the analysis prior the analysis but it is predicted by the simulation codes given the set of initial an boundary conditions.

625 Lastly, the statistical models for each event (e.g., CST cross-tie) do not require complex convolution integrals which need to be solved in classical PRA models since such events are tightly coupled. This event coupling is implicitly solved in the Monte-Carlo analysis.

## 10. Appendix A: Plant ROM Modeling

630 The  $k$ -nearest neighbor classifier is able to classify unobserved samples by employing a weighted voting system over the input space of the model. In this process, first a set of observed training samples,  $\mathcal{T}$ , are classified according to our ground truth model. In order to predict the class label of an unobserved query point,  $q_i$ , the  $k$  nearest samples in  $\mathcal{T}$ , which we will denote as  $p_{i,1}, \dots, p_{i,k}$ , are identified. Each of the  $k$  neighbors is assigned a weight and votes for its observed label. The weight of the vote for  $p_{i,j}$  is given by the inverse Euclidean distance between  $q_i$  and  $p_{i,j}$ . Similarly, we define “nearest” in terms of Euclidean or  $L^2$  distance on the input space of the model being trained.

Each surrogate is trained on a separate set of Monte Carlo samples, and cross-validated using 3-fold validation in order to select an optimal value for  $k$ . In 3-fold validation, the data is split into three roughly equal subsets and trained on two of the subsets where the third subset is used to evaluate the prediction accuracy. The  $k$ -value that fits the data with the highest accuracy is considered the “best” parameter setting. In order to achieve a stable optimum, we randomly select half of the training data and perform 3-fold validation over multiple iterations. The results are collected into a histogram and the most frequent “best” parameter is selected for  $k$ . The values of  $k$  considered are the multiple of 5 up to and including 50 (i.e., 5, 10, 15, 20, 25, 30, 35, 40, 45, 50).

*To further validate the choice of  $k$ , we performed paired  $t$ -tests between the collection of average accuracies of each  $k$  value. The resulting  $p$ -values were near or below machine precision, that is in each instance they were much smaller (several orders of magnitude) than even an aggressive  $p$  threshold such as 0.001. Thus, we reject the null hypothesis that the accuracies reported by each individual  $k$  value come from the same population. In other words, the choice of  $k$  for each model was based on statistically significant improvements in prediction accuracy.*

In order to report on the stability and convergence of each ROM, we evaluate the prediction accuracy of each ROM over increasing training sizes. To this end, we compute the mean and standard deviation of the prediction accuracy over 10 iterations of each training size. The prediction accuracy is always computed from the entire validation set. In each case, we randomly select a set of 100 training samples, train our surrogate using this subset of training data, and predict the accuracy of the entire validation set. We repeat this process ten times using a different set of randomly selected training samples and compute the mean and standard deviation for this training size. Next, we add 100 training samples and repeat the entire process. The process is repeated until the full training set is used. Note, in the case where the full training set is used, the standard deviation will be zero since every trial of randomly selected samples will be represent the full training set. This final accuracy value is the prediction accuracy reported in Table 4. An example of this convergence is shown in Fig. 11.

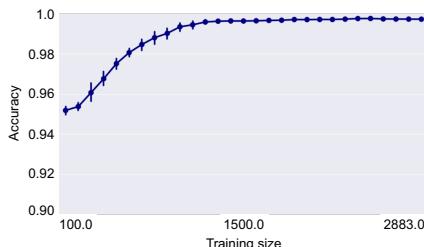


Figure 11: Convergence of the Prediction Accuracy with increasing sample sizes for the SFP1 model. Each data point is the average of ten trials worth of data with error bars representing one standard deviation. For this dataset the full training pool consists of 2883 points and accuracy is tested on an independent validation set of 2876 samples.

As we can see, the results are both highly stable and highly accurate. The stability and precision of the results are exhibited by the presence of small error bars and a steady, gradual improvement as the size increases. There are no large spikes or dips in the plot. In terms of accuracy, the model accuracy is never below 90% even when using only 100 samples. The final parameter settings and accuracies of each model trained on its full training size are given in Table 4.

Model	Parameter Count	Optimal $k$	Training size	Validation size	Prediction accuracy (%)
PWR1	7	5	4596	2500	100.0
PWR2	3	25	4951	2662	99.36
PWR3	10	50	12000	5988	100.0
SFP1	3	5	2883	2876	99.72
SFP2	5	5	4695	4714	99.02
SFP3	3	5	2807	2817	99.04

Table 4: Surrogate model settings and validation information.

We can also compare the relative distance between datasets used for training and validation and also the actual simulation. Figure 12 shows histograms reporting the minimum, unnormalized distances to a point in the training set for each point in the validation set (left column) or simulation set (right column). In all but one case, the farthest point in the simulation data used for actual analysis later in this paper is closer to the training data than the validation data sets. The exception, PWR1, has a farthest distance that is on par with the validation set. The remaining datasets are well within the bounds of the validation sets and all of the simulation results report an average minimum distance that is less than that of the validation set. Since the distance to the validation set is in the worst case larger and on average farther from the training data than the simulation data, we can reasonably expect similar or better accuracy than that reported in Table 4 for our  $k$  nearest neighbor classifier for the simulation results reported in Section 8.

## References

- [1] International Atomic Energy Agency, The Fukushima Daiichi Accident, Technical Report, International Atomic Energy Agency (IAEA) STI/PUB/1710, 2015.
- [2] Pickard, Lowe & Garrick Inc., Seabrook Station Probabilistic Safety Assessment, Technical Report, Public Service Company of New Hampshire and Yankee Atomic Electric Company, PLG-0300, 1963.
- [3] K. Dinnie, Considerations for future development of SAMG at multi-unit CANDU sites, in: Proceedings of 11th International Probabilistic Safety Assessment and Management Conference and the Annual European Safety and Reliability Conference (PSAM11 and ESREL 2012), 2012.

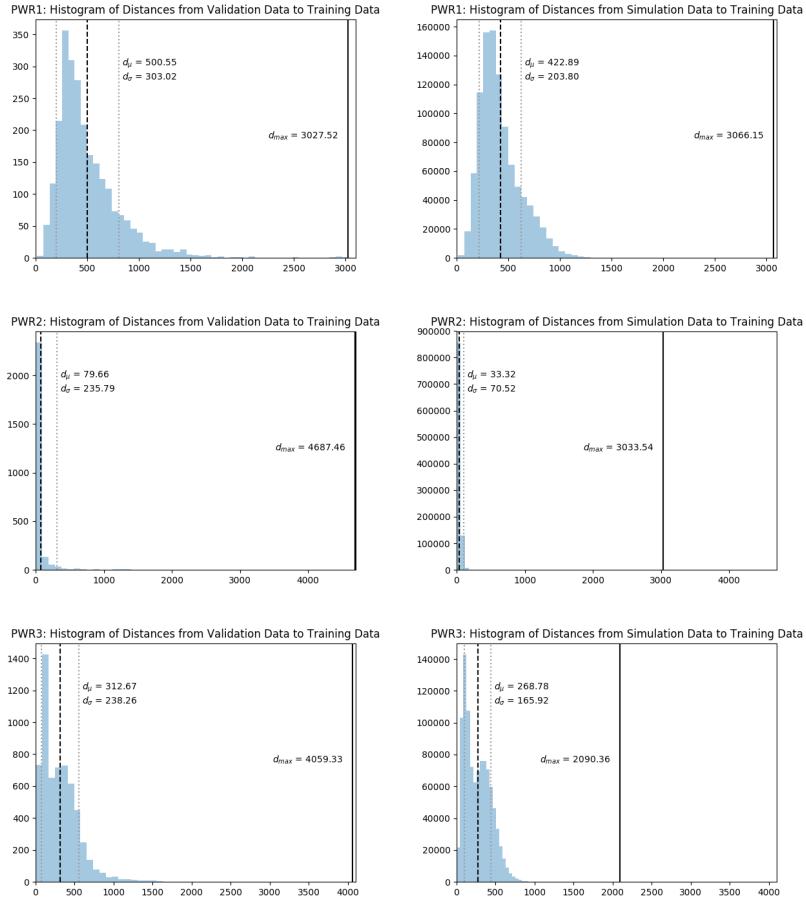


Figure 12: Histograms of the minimum distance between the training set and each of the validation set (left column) and the simulation set used for analysis later in this paper (right column) for each of the six models.

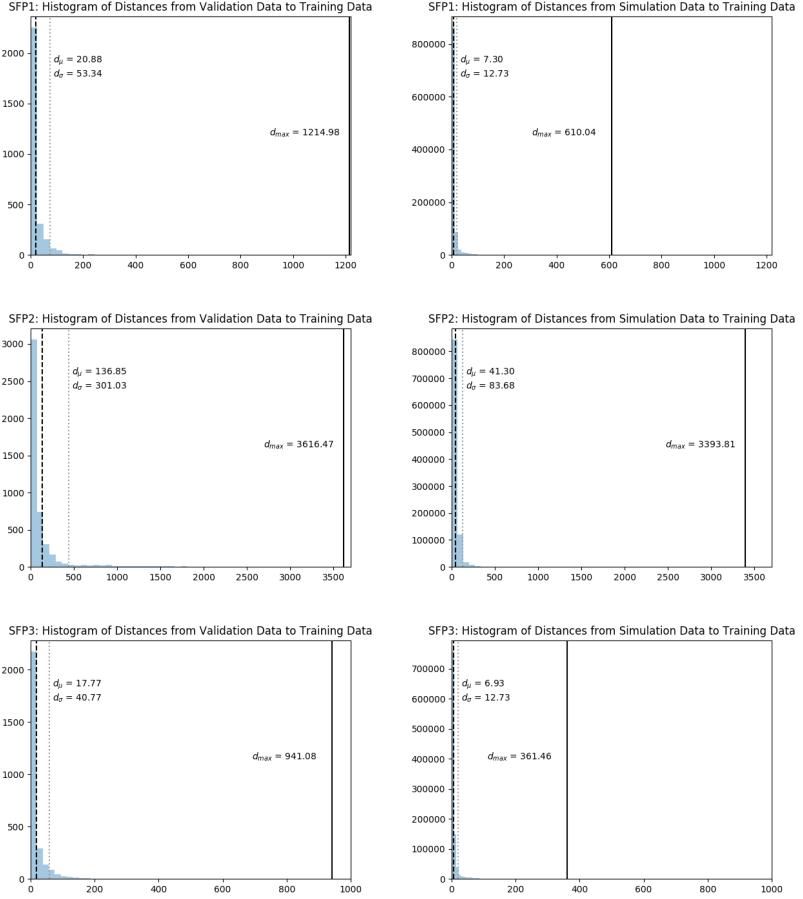


Figure 12: Histograms of the minimum distance between the training set and each of the validation set (left column) and the simulation set used for analysis later in this paper (right column) for each of the six models (cont.ed).

- 710 [4] Ontario Power Generation Inc., Darlington NGS Risk Assessment Summary Report NK38-REP-03611-10072, Technical Report, Ontario Power Generation Inc. Toronto, 2012.
- 715 [5] C. S. Kumara, V. Hassija, K. Velusamy, V. Balasubramaniyan, Integrated risk assessment for multi-unit NPP sites-a comparison, Nuclear Engineering And Design 293 (2015) 53—62.
- 720 [6] M. Modarres, T. Zhou, M. Massoud, Advances in multi-unit nuclear power plant probabilistic risk assessment, Reliability Engineering & System Safety 157 (2017) 87—100.
- 725 [7] S. Zhang, J. Tong, J. Zhao, An integrated modeling approach for event sequence development in multi-unit probabilistic risk assessment, Reliability Engineering & System Safety 155 (2016) 147—159.
- [8] U. S. NRC, Severe Accident Risks: An Assessment For Five U.S. Nuclear Power Plants Final Summary Report, Technical Report, U.S. Nuclear Regulatory Commission, Washington DC, 2005.
- 730 [9] E. Zio, M. Marseguerra, J. Devooght, P. Labeau, A concept paper on dynamic reliability via monte carlo simulation, Mathematics And Computers In Simulation 47 (1998) 371–382.
- [10] C. Smith, C. Rabiti, R. Martineau, Risk Informed Safety Margins Characterization (RISMIC) Pathway Technical Program Plan, Technical Report, Idaho National Laboratory Technical Report: INL/EXT-11-22977, 2011.
- 735 [11] D. Mandelli, C. Smith, C. Rabiti, A. Alfonsi, R. Youngblood, V. Pascucci, B. Wang, D. Maljovec, P.-T. Bremer, T. Aldemir, A. Yilmaz, D. Zamalieva, Dynamic PRA: An overview of new algorithms to generate, analyze and visualize data, in: Proceedings of American Nuclear Society (ANS) Winter Meeting, 2013.
- 740 [12] D. Mandelli, C. Smith, T. Riley, J. Nielsen, A. Alfonsi, J. Cogliati, C. Rabiti, J. Schroeder, BWR station blackout: A RISMIC analysis using RAVEN and RELAP5-3D, Nuclear Technology 193 (2016) 161–174.
- [13] D. Maljovec, S. Liu, B. Wang, D. Mandelli, P. T. Bremer, V. Pascucci, C. Smith, Analyzing simulation-based PRA data through traditional and topological clustering: A BWR station blackout case study, Reliability Engineering & System Safety 145 (2015) 262–276.
- 745 [14] D. Mandelli, S. Prescott, C. Smith, A. Alfonsi, C. Rabiti, J. Cogliati, R. Kinoshita, Modeling of a flooding induced station blackout for a pressurized water reactor using the RISMIC toolkit, in: Proceedings of PSA 2015 International Topical Meeting On Probabilistic Safety Assessment And Analysis, 2015.

- [15] R. L. Boring, R. B. Shirley, J. C. Joe, D. Mandelli, C. Smith, Simulation And Non-Simulation Based Human Reliability Analysis Approaches, Technical Report, Idaho National Laboratory Technical Report: INL/EXT-14-33903, 2014.
- [16] H. S. Abdel-Khalik, Y. Bang, J. M. Hite, C. B. Kennedy, C. Wang, Reduced order modeling for nonlinear multi-component models, International Journal On Uncertainty Quantification (2012) 341–36.
- [17] RELAP5-3D Code Development Team, RELAP5-3D Code Manual, Technical Report, Idaho National Laboratory Technical Report, 2005.
- [18] R. O. Gauntt, MELCOR Computer Code Manual, Version 1.8.5, Vol. 2, Rev. 2, Sandia National Laboratories, NUREG/CR-6119, 2000.
- [19] D. Mandelli, Z. Ma, C. Smith, Dynamic and classical PRA: a BWR SBO case comparison, in: Proceedings of PSA 2015 International Topical Meeting on Probabilistic Safety Assessment and Analysis Columbia, SC, on CD-ROM, American Nuclear Society, LaGrange Park, IL, 2015.
- [20] C. Rabiti, A. Alfonsi, J. Cogliati, D. Mandelli, R. Kinoshita, RAVEN, a new software for dynamic risk analysis, in: Proceedings of the Probabilistic Safety Assessment and Management (PSAM) 12, 2014.
- [21] C. Parisi, S. R. Prescott, R. H. Szilard, J. L. Coleman, R. E. Spears, A. Gupta, Demonstration of External Hazards Analysis, Technical Report, Idaho National Laboratory Technical Report: INL/EXT-16-39353, 2016.
- [22] R. H. Szilard et al., RISMC Toolkit and Methodology Research and Development Plan for External hazards Analysis, Technical Report, Idaho National Laboratory Technical Report: INL/EXT-16-38089, 2016.
- [23] J. Jo, et al, NUREG/CR-6144: Evaluation of Potential Severe Accidents During Low Power and Shutdown Operations at Surry, Unit 1. Evaluation of Severe Accident Risk During Mid-Loop Operations, Technical Report, U.S. Nuclear Regulatory Commission, Washington DC, 1995.
- [24] C. Parisi, et al, External events analysis for LWRS/RISMC project: Methodology development and early demonstration, in: Proceeding Of American Nuclear Society (ANS), New Orleans, 2016.
- [25] R. Boring, D. Mandelli, M. Rasmussen, S. Herberger, T. Ulrich, K. Groth, C. Smith, Human unimodel for nuclear technology to enhance reliability (HUNTER): A framework for computational-based human reliability analysis, in: Proceedings of 13th International Conference on Probabilistic Safety Assessment and Management (PSAM 13), Paper A-531, 2016.
- [26] R. Boring, D. Mandelli, M. Rasmussen, S. Herberger, T. Ulrich, K. Groth, C. Smith, Integration of Human Reliability Analysis Models into the

<sup>785</sup> Simulation-Based Framework for the Risk-Informed Safety Margin Characterization Toolkit, Technical Report, Idaho National Laboratory Technical Report: INL/EXT-16-39015, 2016.

- <sup>790</sup> [27] A. D. Swain, H. E. Guttmann, NUREG/CR-1278: Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications, Technical Report, U.S. Nuclear Regulatory Commission, Washington DC, 1983.
- [28] S. N. Laboratories, NUREG/CR-7110: State-of-the-Art Reactor Consequence Analyses Project Volume 2: Surry Integrated Analysis, Technical Report, U.S. Nuclear Regulatory Commission, Washington DC, 2012.
- <sup>795</sup> [29] N. L. Johnson, S. Kotz, N. Balakrishnan, Continuous univariate distributions, volume 2, Wiley & Sons, 1995.