




OPEN

DATA DESCRIPTOR

# An open time-series simulated dataset covering various accidents for nuclear power plants

Ben Qi<sup>1</sup>, Xingyu Xiao<sup>1</sup>, Jingang Liang<sup>1</sup> <sup>✉</sup>, Li-chi Cliff Po<sup>2</sup>, Liguang Zhang<sup>1</sup> & Jiejuan Tong<sup>1</sup>

Nuclear energy plays an important role in global energy supply, especially as a key low-carbon source of power. However, safe operation is very critical in nuclear power plants (NPPs). Given the significant impact of human-caused errors on three serious nuclear accidents in history, artificial intelligence (AI) has increasingly been used in assisting operators with regard to making various decisions. In particular, data-driven AI algorithms have been used to identify the presence of accidents and their root causes. However, there is a lack of an open NPP accident dataset for measuring the performance of various algorithms, which is very challenging. This paper presents a first-of-its-kind open dataset created using PCTTRAN, a pre-developed and widely used simulator for NPPs. The dataset, namely nuclear power plant accident data (NPPAD), basically covers the common types of accidents in typical pressurised water reactor NPPs, and it contains time-series data on the status or actions of various subsystems, accident types, and severity information. Moreover, the dataset incorporates other simulation data (e.g., radionuclide data) for conducting research beyond accident diagnosis.

## Background & Summary

Nuclear energy has been a great discovery in human history. After more than 100 years of development since humans discovered nuclear radiation in the late 19th century, nuclear energy is now closely linked to peoples' lives and jobs<sup>1</sup>. Nuclear power is one of the main forms through which human beings use nuclear energy to promote economic development and benefit society. Since the first commercial nuclear power plant (NPP) was built in the former Soviet Union in the late 1950s<sup>2</sup>, more than 450 nuclear power units have been in operation worldwide<sup>3</sup>. During this time, in more than 60 years, nuclear power technology has undergone iterative upgrades: from Generation 1 prototype reactors to Generation 2 commercial reactors and then to Generation 3 advanced high-power nuclear reactors. In addition, nowadays, Generation 4 nuclear power systems, which are safer and more economical than the above-mentioned systems, are being explored and experimented<sup>4</sup>. Under the current wave of global industrial intelligence, numerous countries have conducted research on the integration of the technologies of nuclear power and artificial intelligence, especially through the development of digital instrumentation and control (I&C) systems, which can collect large amounts of operational data<sup>5</sup>. However, the value of such large data amounts has not been fully explored, and intelligent nuclear power is still a technique to be developed in the future.

The experience of the previous three serious nuclear accidents in history shows that relying only on nuclear power plant operators to perform early accident diagnosis can result in serious consequences due to human errors<sup>6</sup>. Artificial intelligence (AI) refers to the technology of expressing human-like intelligence through computational models<sup>7</sup>. AI systems can manage complex situations and efficiently process multi-source information, making them suitable for the task of rapidly and accurately diagnosing accidents in nuclear power plants. Therefore, many studies have been performed to develop artificial intelligence-based accident diagnosis technologies for nuclear power plants. Typical AI applications, such as face recognition<sup>8</sup> and autonomous driving<sup>9</sup>, have been developed by training kernel algorithms (i.e., artificial neural network<sup>10</sup>, support vector machine<sup>11</sup>, decision tree<sup>12</sup>) with massive data with the help of high-performance computing. Algorithms, data, and computing are the three core elements of AI. Among them, the use of datasets has always been a fundamental factor that directly affects the final performance of AI models in real-world scenarios. A high-quality dataset can be a good starting point for validating or building better algorithms. However, regarding nuclear power plants, it is

<sup>1</sup>Institute of Nuclear and New Energy Technology, Tsinghua University, Beijing, 100084, China. <sup>2</sup>Micro-Simulation Technology, Montville, New Jersey, 07045, United States of America. ✉e-mail: [jingang@tsinghua.edu.cn](mailto:jingang@tsinghua.edu.cn)

difficult to obtain real accident data, as accidents rarely occur in practice. Moreover, the high safety requirements of NPPs make it impractical to experimentally obtain data from commercial nuclear power plants. Thus, nuclear power plant simulators are often used to obtain large amounts of data.

Nuclear power plant simulation is a technique for simulating system characteristics using mathematical and theoretical models, and it has become an important tool for nuclear power plant design and characterisation. To the best of our knowledge, there are no open datasets for nuclear power plant accident diagnosis. Most studies use non-open nuclear power plant simulators to construct datasets and then train diagnostic models to verify the performance of new algorithms. For example, Yao *et al.* obtained five nuclear power plant accident data using a RELAP5-HD simulator, and they compared the performance of five artificial intelligence algorithms for nuclear power plant accident diagnosis<sup>13</sup>. Qi *et al.* used the simulator of three-loop pressurised water reactor to validate hybrid AI algorithms driven by both knowledge and data<sup>14</sup>. Wang *et al.* used an online pressurised water reactor (PWR) simulator to obtain seven types of nuclear power plant accident data and verify the accident diagnosis performance of a hybrid AI algorithm<sup>15</sup>. Lee *et al.* used another simulator (3KEYMASTER) to obtain ten types of nuclear power plant accident data so as to verify the feasibility of convolutional neural networks for nuclear power plant accident diagnosis<sup>16</sup>. Yang *et al.* studied the capability of various artificial intelligence algorithms for reactor transient analysis based on computational fluid dynamics (CFD) data<sup>17–19</sup>. Moreover, Wang *et al.* developed a simulator called Nuclear Steam Supply System and validated the feasibility of Long Short-term Memory neural networks with regard to small PWR accident diagnosis<sup>20</sup>. In summary, previous studies were usually performed based on private datasets to develop and optimise accident diagnosis algorithms. However, there are several issues with non-open datasets. First, it is difficult to compare the performance of different algorithms due to the lack of a common benchmark dataset. Second, the data quality of the used simulators in some studies may not have been verified with sufficient reliability. Third, the constructed datasets in some studies mostly cannot cover a comprehensive range of accident types. In addition, it is a repetitive effort for researchers to build their own datasets when developing and optimising algorithms.

To address the above-mentioned issues, we built an open dataset, Nuclear Power Plant Accident Data (NPPAD), with massive data that covers various accidents that can occur in nuclear power plants to help with the development and optimisation of artificial intelligence algorithms and other applications. The dataset was constructed based on a nuclear power plant simulation software, PCTRAN, which is one of the most widely used desktop simulators for nearly all types of nuclear reactors. PCTRAN was specifically designed for different light water plant types, such as PWR<sup>21</sup> and boiling water reactor (BWR)<sup>22</sup>. Since 1998, it has been used by the International Atomic Energy Agency's (IAEA) annual Advanced NPP Simulation Workshop as a sample model<sup>23</sup>. PCTRAN-based plant-specific models have been installed in nuclear power plants and institutions all over the world for practical applications in training, analysis, probabilistic safety assessment and emergency exercises.

In this study, an open dataset of the most common accidents of PWR nuclear power plants was constructed. The constructed dataset can be used by multiple domains. For example, AI experts can learn about nuclear power plant domain datasets to develop adapted algorithms, while nuclear power experts can use it as a benchmark dataset to compare the performances of various algorithms in NPP accident diagnosis. Notably, in the emerging research area, Gong *et al.* proposed a digital twin technique for nuclear reactor operations, which also presents an urgent need for nuclear power plant accident datasets<sup>24–26</sup>.

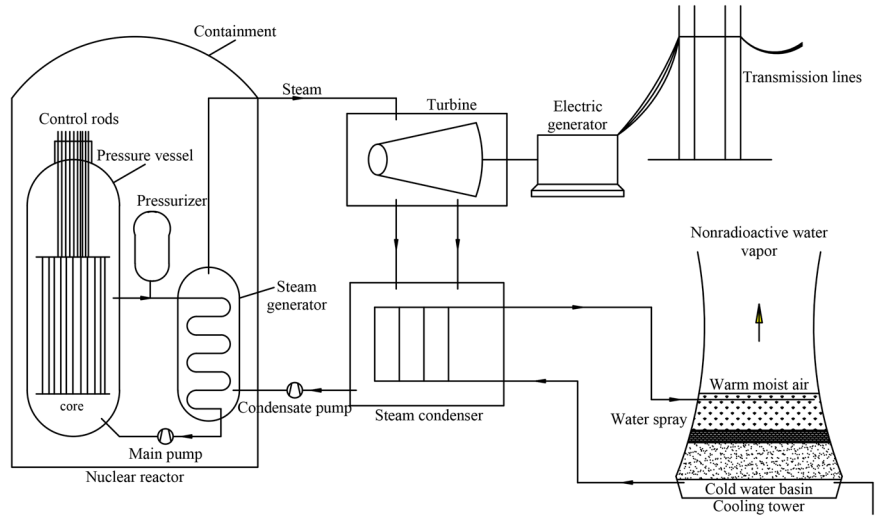
In the rest of the paper, we introduce the main NPP structure, theoretical models of PCTRAN, the methods used to generate the proposed dataset, the data records structure, and multiple aspects of technical validations.

## Methods

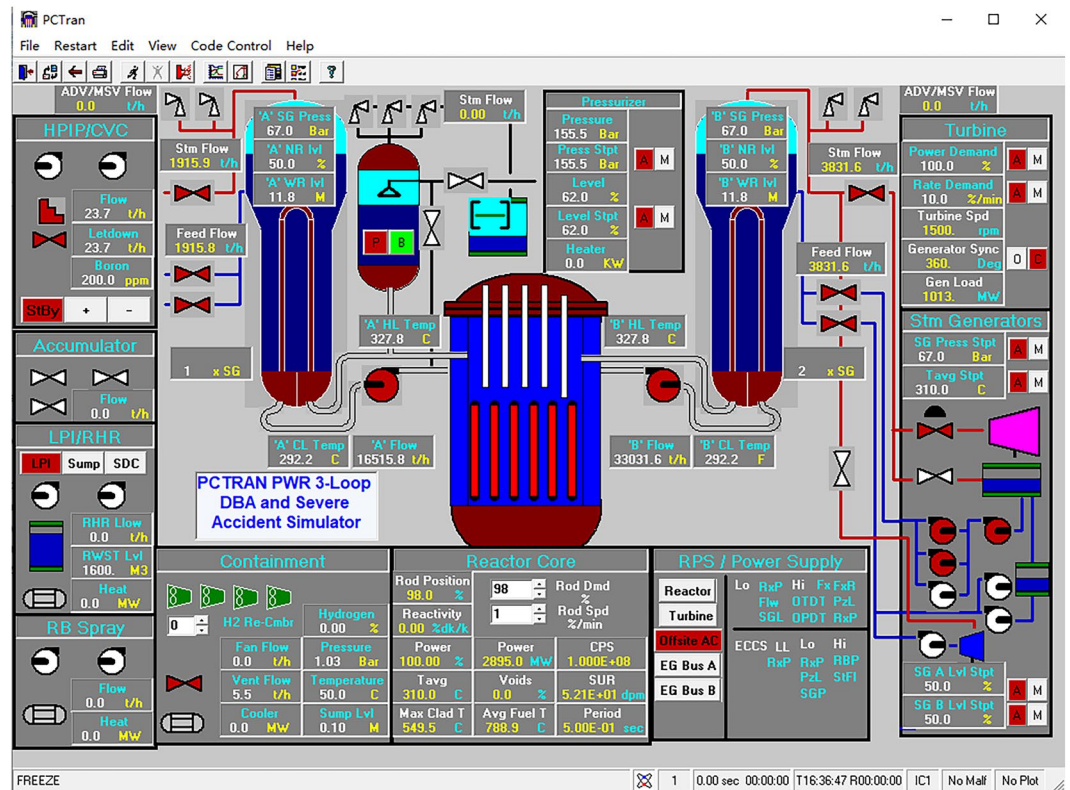
In this section, we describe the methods used to create NPPAD<sup>27</sup>, as well as a description of nuclear power plants, theoretical models of PCTRAN, and an overview of the processing workflow.

**Description of nuclear power plants.** The nuclear power plants currently in operation have numerous reactor types, including PWRs, BWRs and fast reactors, and two-thirds of these reactors are PWRs. Therefore, the constructed dataset in this study is based on PCTRAN. As shown in Fig. 1, the overall structure of a nuclear power plant consists of three main loops. The first loop is the nuclear reactor loop, which consists of a reactor pressure vessel, a pressurizer, a main pump, a steam generator (first loop side), and other components, all of which are located within the containment. The second loop consists of a steam generator, a condensate pump, a turbine, a steam condenser, and other components. The third loop consists of auxiliary system equipment, including electric generators, cooling towers, and other auxiliary equipment. The water in the first loop is heated by the generated heat through nuclear fission and is then transferred to the steam generator, in which it is turned into steam in the second loop. The generated steam rotates the turbine and is then condensed into water, which is returned by a pump to be heated and then turned into steam again. The turbine drives the electric generator, which produces electricity.

**Theoretical models of PCTRAN.** PCTRAN is a reactor transient and accident simulation software that is operated based on a personal computer, and it has a Graphical User Interface (GUI) adhering to the specifications of the Microsoft Windows environment. The data input and output are in MS Office's Access database format (MDB format). PCTRAN provides two control interfaces: the main control interface (Fig. 2) and the radiation dose simulation interface (Fig. 3). In the main control interface, users can control various types of pumps, valves, control rods, and other equipment and can also visualize data on real-time changes in pressure, temperature, flow, and other operating conditions. PCTRAN displays real-time changes in the radiation dose values and cumulative values for each area in the radiation dose simulation interface.



**Fig. 1** Main structure of a PWR nuclear power plant.



**Fig. 2** Main control interface of PCTRAN.

PCTRAN consists of several key simulation modules, including a reactor dynamics module, a reactor coolant system module, and a steam generator module. A brief description of the theoretical models of each module is given below.

**Reactor dynamics module.** This module simulates nuclear reactor cores by describing the variation of neutron densities and related quantities during transients. It includes neutron dynamics<sup>28</sup>, fuel dynamics<sup>29</sup>, steam generator dynamics<sup>30</sup>, and their associated feedback models. As shown in Eqs. 1 and 2, PCTRAN uses a classical set of slow-emitting neutron point reactor models<sup>31</sup>.

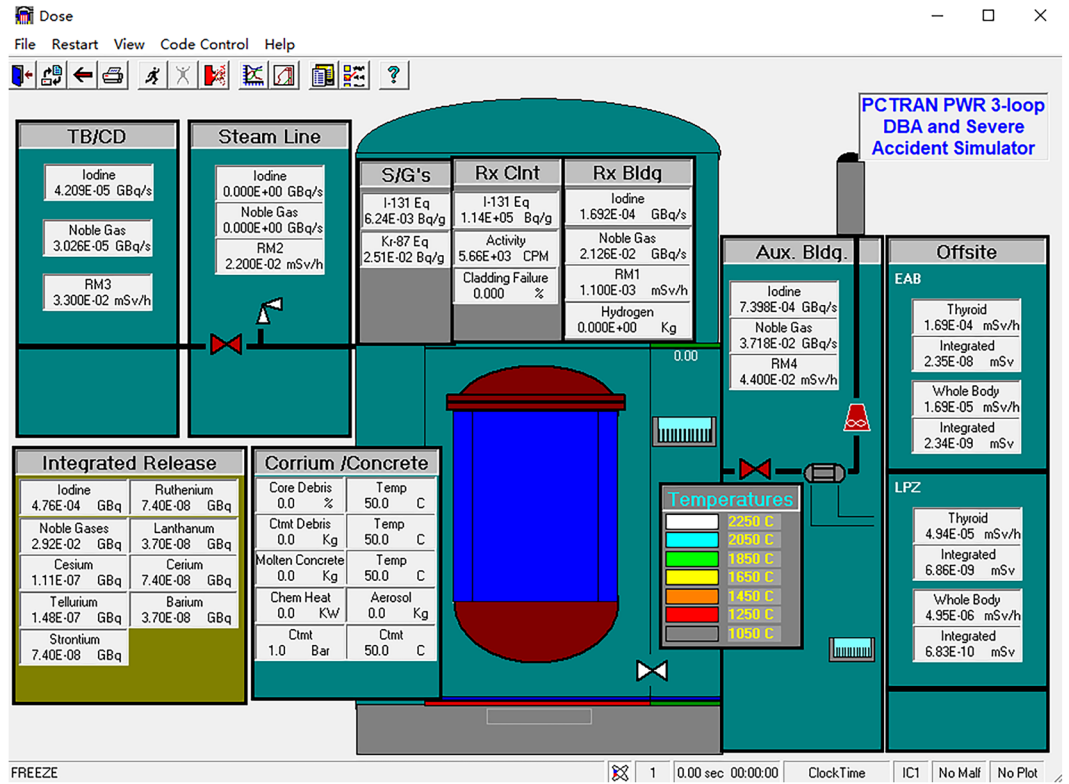


Fig. 3 Dose simulation interface of PCSTRAN.

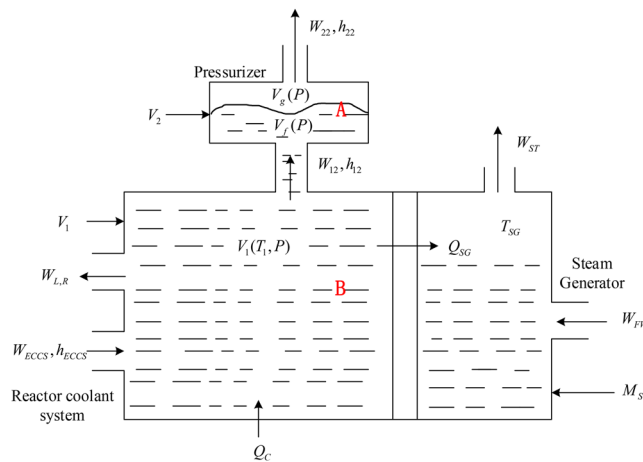


Fig. 4 Model of the reactor's coolant system module and steam generator module in PCSTRAN.

$$\frac{dn}{dt} = \frac{\rho - \beta}{l} n + \lambda C \tag{1}$$

$$\frac{dC}{dt} = \frac{\beta}{l} n - \lambda C \tag{2}$$

where  $n$  is the neutron density,  $\rho$  is the reactivity,  $\beta$  is the delayed neutron fraction,  $l$  is the neutron lifetime,  $\lambda$  is the decay constant, and  $c$  is the precursor concentration.

**Reactor coolant system module.** This module simulates the reactor coolant system and pressurizer of nuclear reactor loop, and the basic mathematical models are based on the first principles of mass and energy balance, thus ensuring credible and realistic simulations. As shown in Fig. 4, a fluid boundary that separates the saturated

two-phase fluid (A) from the subcooled liquid (B) was introduced. The saturated two-phase flow is the fluid within the pressurizer, while the subcooled fluid is the rest of the reactor coolant system's fluid. In transient operating, the boundary is allowed to move upwards and downwards. The upper two-phase fluid with a total volume  $V_2$  consists of a vapour space, which occupies a fraction  $\alpha$  of  $V_2$  and a saturated fluid space. The total volume of the lower subcooled fluid is  $V_1$ . The A and B fluids are treated separately.

First is the saturated two-phase fluid (A) model<sup>32</sup>. The specific enthalpies and volumes of liquid and vapour are denoted as  $h_f, h_g$  and  $v_f, v_g$ , respectively. The quality  $x$  and average mixture enthalpy  $h_m$  are related by the following equation<sup>33</sup>:

$$x = \frac{\alpha/v_g}{\alpha/v_g + (1 - \alpha)/v_f} \quad (3)$$

$$h_m = x \cdot h_g + (1 - x)h_f \quad (4)$$

The flow discharge leaving the two-phase volume is denoted by the flow rate  $W_{22}$  and enthalpy  $h_{22}$ .  $W_{12}$  and  $h_{12}$  in Fig. 4 correspondingly express the inter-connecting flow. Then, according to the conservation of mass,

$$\frac{dM_2}{dt} = W_{12} - W_{22} \quad (5)$$

According to the conservation of energy, the nuclear core heat is generated in this volume.

$$\frac{dU}{dt} = W_{12} \cdot h_{12} - W_{22} \cdot h_{22} \quad (6)$$

where  $U$  is the total internal energy in this volume and is expressed as follows:

$$U = M_2(h_m - P \cdot v_m) \quad (7)$$

where  $M_2$  is the total mass,  $P$  is the system pressure and  $v_m$  is the average specific volume.

According to the equation of the system's state,

$$V_2 = M_2[x \cdot v_g + (1 - x)v_f] = V_2(x, P, M_2) = \text{constant} \quad (8)$$

The solution to the system of equations can be obtained by eventually combining the above equations.

Second is the subcooled fluid (B) model<sup>34</sup>. Assuming that the emergency core cooling system (ECCS)'s injection flow is the only flow in this region and that the LOCA break flow  $W_{LR}$  is the net loss, the conservation of mass and energy balance equations in the subcooled region result in

$$M_1 \frac{dh_s}{dt} = W_{EC}(h_{EC} - h_s) - W_{LR}(h_{LR} - h_s) - W_{12}(h_{12} - h_s) + V_1 \left( \frac{dP}{dt} \right) + Q_C - Q_{SG} \quad (9)$$

where  $h_s$ , which is the specific enthalpy of the subcooled liquid, is a function of the system pressure  $P$  and liquid temperature  $T$ . Similarly, the state equation of the subcooled fluid is expressed as follows:

$$V_1 = M_1 v_s(P, T) = \text{constant} \quad (10)$$

According to the conservation of mass,

$$\frac{dM_1}{dt} = W_{EC} - W_{12} - W_{LR} \quad (11)$$

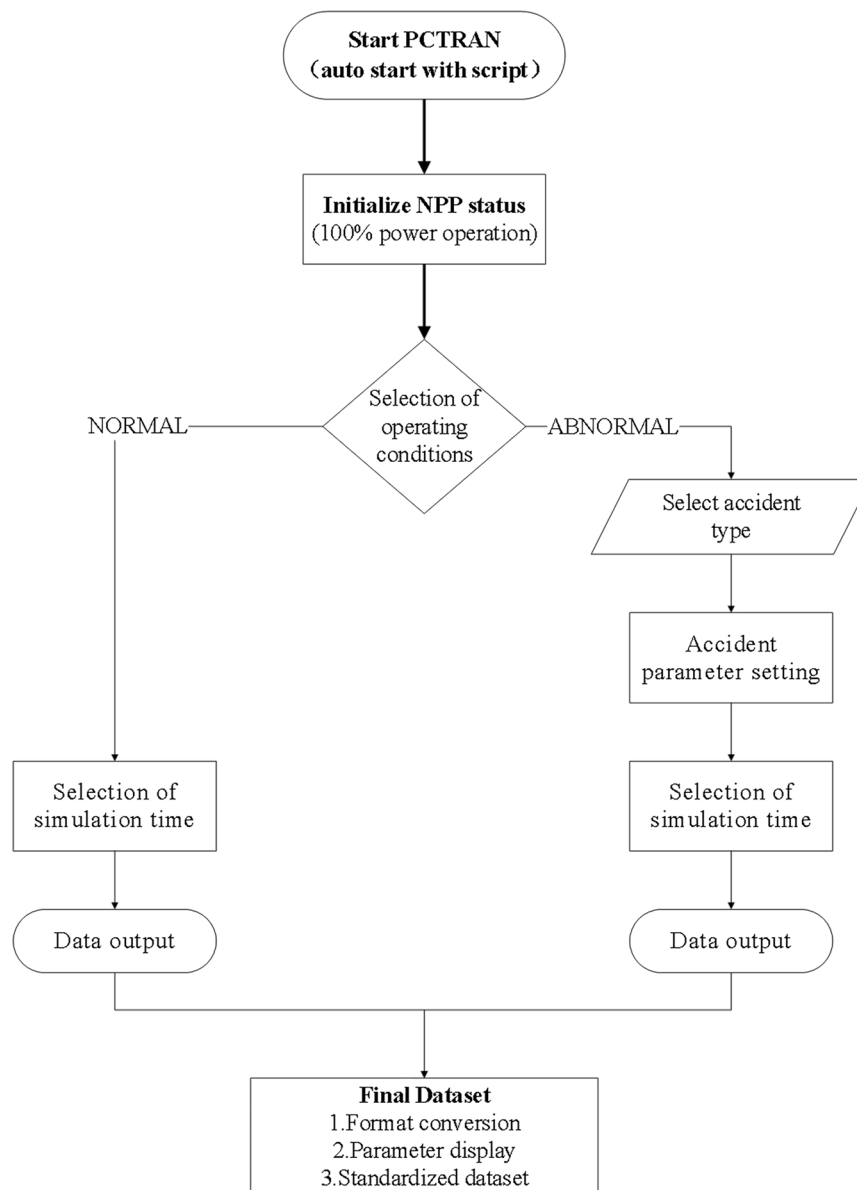
The solution to the equation system can be obtained by eventually combining the above equations.

**Steam generator module.** This module includes the heat flux transfer model, water level dynamic control model and pressure, and steam valve control model<sup>35</sup>. The water level dynamic control model is based on the principle of conservation of mass and energy. According to the feed water flow, steam flows to maintain the dynamic stability of the water level. The heat flux transfer model, i.e., the heat exchange equation between the second loop and first loop, is as follows:

$$Q = u_w \cdot A_w (T_{avg} - T_{SG}) \quad (12)$$

where  $u_w$  is the heat transfer coefficient,  $T_{avg}$  is the average temperature of the reactor's primary coolant,  $T_{sg}$  is the secondary saturated temperature of the steam generator, and  $A_w$  is the wet tube's surface area.

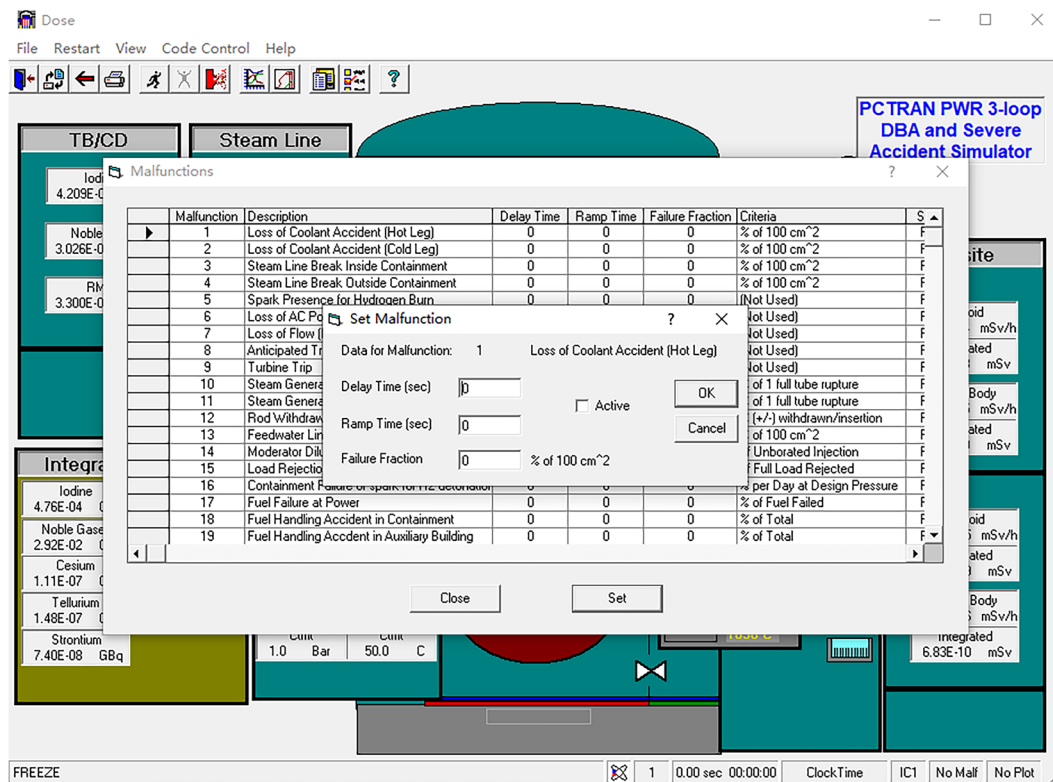
In addition to the above-mentioned core modules, PCTTRAN includes a nuclear fuel model, a containment model, and a radiation dose module<sup>36</sup>, whose radiation dose leakage calculation module is a unique function of PCTTRAN.



**Fig. 5** Overall workflow of the simulation data generation.

**Workflow overview.** PCTTRAN was used in this work to produce accident data on nuclear power plants. However, simulating each accident routinely requires manual key/mouse operations at various steps, such as initialising the operating conditions, selecting the accident type, setting the accident parameters, and ending the accident simulation. Specifically, each accident was inserted at the 20 s moment of full power operation of the nuclear power plant, the time step of data sampling was 10 s and the average simulation time for each accident was ~4000 s. When dealing with a large number of accident scenarios, complex manual operations become very cumbersome and inefficient. In this work, we tried to develop scripts to automate the process and manipulate PCTTRAN to generate large amounts of accident data quickly and easily. Using the automation script, a total of 1,217 samples (normal and abnormal operation conditions) were simulated to generate the whole dataset, and the operation took more than 1,350 hours in total with a common desktop computer. Such operations are almost impossible to perform using manual key/mouse operations.

The overall workflow implemented in the script to generate the nuclear power plant accident dataset is shown in Fig. 5. First, the PCTTRAN software is started by an automation script that replaces the manual key/mouse operations. Once the software is launched, the nuclear power plant (operating at 100% power) is initialised. Then, different operating conditions are selected. If the normal operating condition is treated, the simulator runs for the selected time to get the data output. For abnormal operating conditions (i.e., accidents), as shown in Fig. 6, various parameters, including the accident type, accident parameters, and simulation time, are configured. Then, the simulation data is output. Table 1 lists the accidents covered in this work, where almost all the possible nuclear accidents are simulated. Each of these accidents has the potential to cause reactor core damage,



**Fig. 6** Accident type selection and parameter setting.

Accident	Description	Type	Severity
NORM	Normal operating	—	—
LOCA	Loss of Coolant Accident (Hot Leg)	Severity	% of 100 cm <sup>2</sup>
LOCAC	Loss of Coolant Accident (Cold Leg)	Severity	% of 100 cm <sup>2</sup>
SLBIC	Steam Line Break Inside Containment	Severity	% of 100 cm <sup>2</sup>
SLBOC	Steam Line Break Outside Containment	Severity	% of 100 cm <sup>2</sup>
SP	Spark Presence for Hydrogen Burn	Other	—
LACP	Loss of AC Power	Other	—
LOF	Loss of Flow (Locked Rotor)	Other	—
ATWS	Anticipated Transient Without Scram	Other	—
TT	Turbine Trip	Other	—
SGATR	Steam Generator A Tube Rupture	Severity	% of 1 full tube rupture
SGBTR	Steam Generator B Tube Rupture	Severity	% of 1 full tube rupture
RW	Rod Withdrawal	Severity	% (+/−) withdrawn
RI	Rod Insertion	Severity	% (+/−) insertion
FLB	Feedwater Line Break	Severity	% of 100 cm <sup>2</sup>
MD	Moderator Dilution	Severity	% of unborated injection
LR	Load Rejection	Severity	% of full load rejected
LLB	Letdown Line Break in auxiliary buildings	Severity	% of nominal letdown flow

**Table 1.** Accident sets covered by NPPAD.

and whether or not they ultimately result in core damage depends on the successful response of the nuclear plant's accident mitigation system. The dataset in this work does not include cases in which mitigation system failures are superimposed on nuclear plant accidents, as such superimposed cases are too numerous to cover. The detailed process of accident simulation, which is executed by configuring a set of input parameters, is shown in Box 1. If an accident involves different levels of severity, such as the size of the first loop break, it is defined as “severity type,” which needs to be set as a severity parameter. Finally, we obtained the dataset *NPPAD.rar* with normal and abnormal conditions.

```

Box 1 The accident simulation process in PCTTRAN

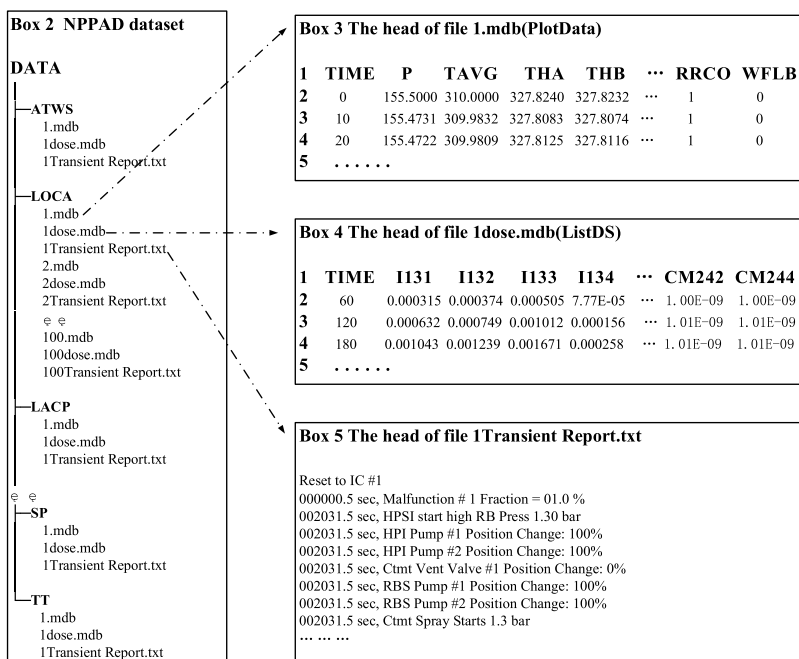
Input: accident_set, accident_type, severity, simulation_time, output_format

For accident_set do
  If accident_type = 'severity type' then
    for severity do
      accident_simulation(accident,severity, simulation_time);
      when time = simulation_time do
        data_output
      end
    end
  else
    accident_simulation(accident, simulation_time);
    when time = simulation_time do
      data_output
    end
  end
End

Output: NPPAD.zip
  
```

### Data Records

The dataset is available at Figshare<sup>27</sup>. Box 2 illustrates the general structure of the data records in NPPAD, in which accidents are stored in separate directories. The initial version of the dataset contains 18 types of operating conditions that are possible under the full power operation of a three-loop pressurised water reactor NPP. Each operating condition sample contains three files: two in the MDB format and the other in the plain text format. The MDB files can be opened directly using Microsoft Access. For example, as shown in Box 3, the *1.mdb* (*PlotData*) represents the time series of the status parameters with a 1% of 100 cm<sup>2</sup> break of LOCA. Moreover, *PlotData* represents the sub-table in the *1.mdb* file. As shown in Box 6, another useful sub-table is *ListPlotVariables*, as it describes the parameters corresponding to the abbreviations in *PlotData*. As shown in Box 4, *1Dose.mdb* represents the time series of the radionuclide in the nuclear power plant. In addition to the MDB format, we also provided a CSV format in the folders *Operation\_csv\_data* and *Dose\_csv\_data*. As shown in Box 5, *1Transient Report.txt* describes the actions in the subsystems of the nuclear plant over the simulation time for each accident, which can help users understand changes in the plant status. The numbers in front of the files in the other operating conditions (e.g., *1.mdb*, *2.mdb*) correspond to the severity of the accident, and the exact meaning can be determined by the column 'severity' of Table 1. The above-mentioned datasets are also stored on the GitHub website (<https://github.com/thu-inet/NuclearPowerPlantAccidentData>).



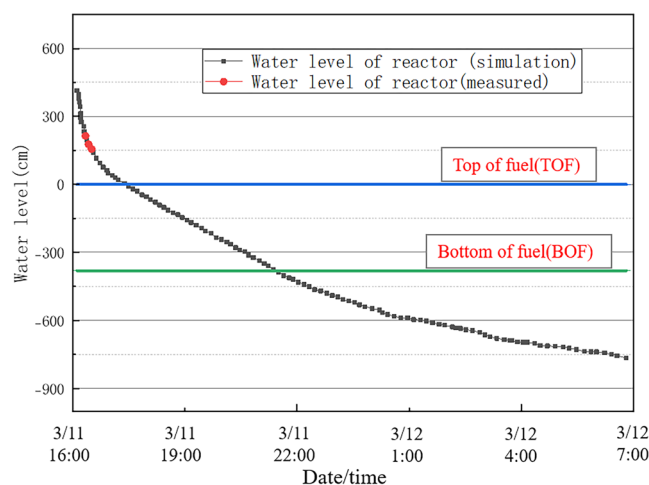
**Box 6 The head of abbreviated names of variables**

ID	Variables	Units	Names
1	Time	sec	TIME
2	Temp RCS average	°C	TAVG
3	Temp Hot leg A	°C	THA
4	...	...	...



Date	Time	Event
2011/3/11	14:46	Earthquake: reactor was automatically shut down. Decay heat was continuously generated
		Loss of off-site power: diesel generators were automatically started. Therefore, AC and DC power was available in this period
	14:52–15:34	IC cooling: reactor was cooled by Isolation Condenser (IC) with start-stop operation so that RPV cooling down rate did not exceed 55 °C/h. Unit 1 was operated to achieve a cold shutdown
	15:37	Tsunami hit: AC and DC were lost. IC was not in operation at this time
	After tsunami	Reactor Pressure Vessel (RPV) water inventory decreases due to no water injection
	After 20:00	(PCV) pressure increased
		RPV bottom damage: Corium (melted fuel) slumping to PCV pedestal
2011/3/12	14:30	Regarding the containment vessel vent, the operation of AO valve of suppression chamber side was implemented at 10:17 am, and a pressure decrease was confirmed at 2:30 pm
	15:36	Reactor building explosions

**Table 2.** Chronological accident description for Unit 1 of the Fukushima Daiichi NPP.



**Fig. 7** Reactor water level for Unit 1 of the Fukushima Daiichi NPP.

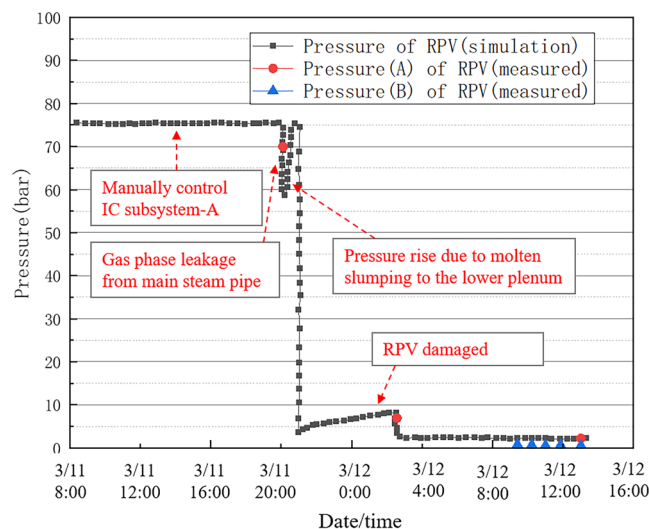
### Technical Validation

Since its introduction in 1985, PCTRAN has been constantly upgraded and expanded. The current software's scope covers numerous types of PWR and BWR plant designs, including both Generation II and Generation III plants. PCTRAN models have generally gone through detailed benchmarking and verifications. For example, in the International Atomic Energy Agency (IAEA)'s handbook<sup>37</sup>, PCTRAN was used to simulate the nuclear accident of Three Mile Island (TMI), which happened at 4 am on March 28, 1979, when the reactor was operating at 97% power. The accident occurred due to a relatively minor malfunction in the secondary cooling circuit, which caused the primary coolant's temperature to increase. In turn, the reactor was automatically shut down, which took approximately one second. At this point, a relief valve failed to close; however, instrumentation did not reveal this fact. A large amount of the primary coolant was then drained so that the residual decay heat in the reactor core was not removed. Thus, the core was severely damaged. The TMI accident simulation was analysed for up to 6,000 seconds, where the changes in key parameters, such as the water level of the steam generator, reactor coolant temperature and pressure, and fuel temperature, were accurately presented.

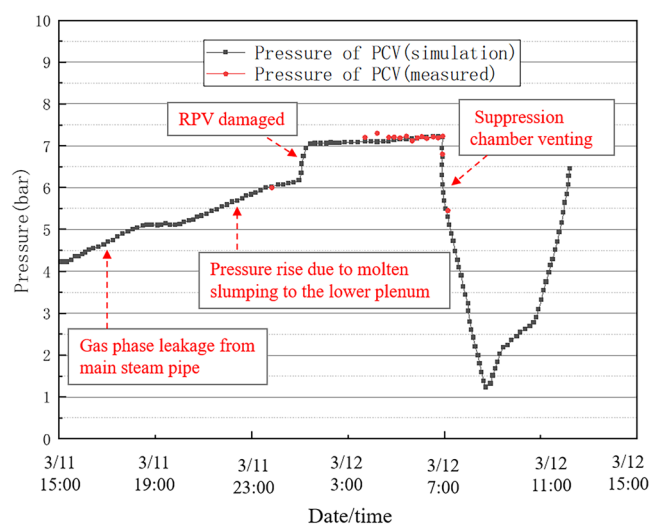
In this section, two simulations were conducted to demonstrate the accuracy and reliability of the PCTRAN simulation software with regard to nuclear power plant operations. The first simulation involved simulating the evolution of the Fukushima nuclear accident, and the simulation results were compared with the measured results (accident report data). The second simulation involved simulating two nuclear power plant operating conditions (load rejection and Large LOCA), and an analysis was performed to see whether the simulation results conformed to the expected physical phenomena.

**Validation using the Fukushima nuclear accident.** The PCTRAN simulation was validated against the accident progression of the Fukushima Daiichi nuclear power plant Unit 1, as shown in Table 2<sup>38,39</sup>. As a result of the tsunami, all the cooling capability was lost in Unit 1, which fell into a severe condition within 3 or 4 hours after the earthquake. It was not until the next morning (March 12) that Tokyo Electric Power Company (TEPCO) could inject water into RPV. Then, Primary Containment Vessel (PCV) venting was conducted at 14:30 on March 12. Afterwards, a hydrogen explosion occurred.

The simulation of the accident was performed using PCTRAN with assumed boundary conditions starting with the loss of off-site power, which was induced by the earthquake and tsunami, followed by the venting of



**Fig. 8** RPV pressure changes for Unit 1 of the Fukushima Daiichi NPP.

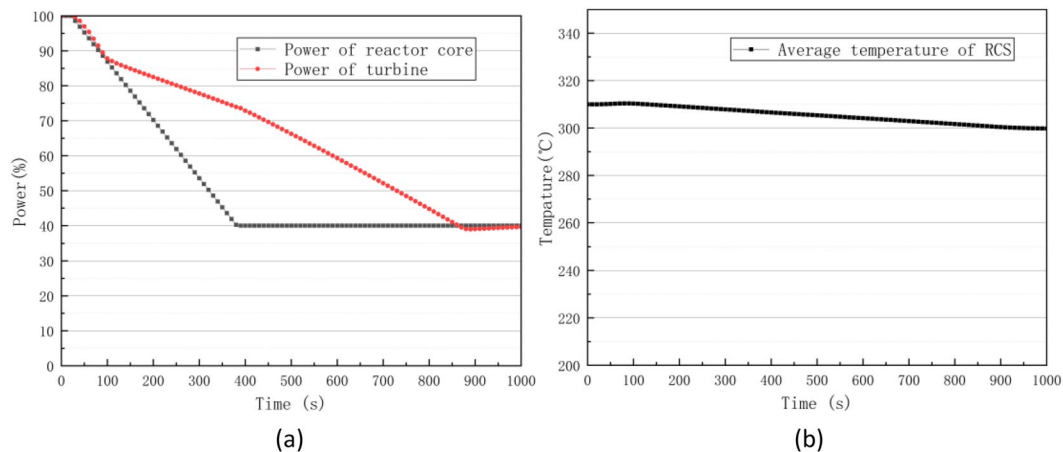


**Fig. 9** PCV pressure changes for Unit 1 of the Fukushima Daiichi NPP.

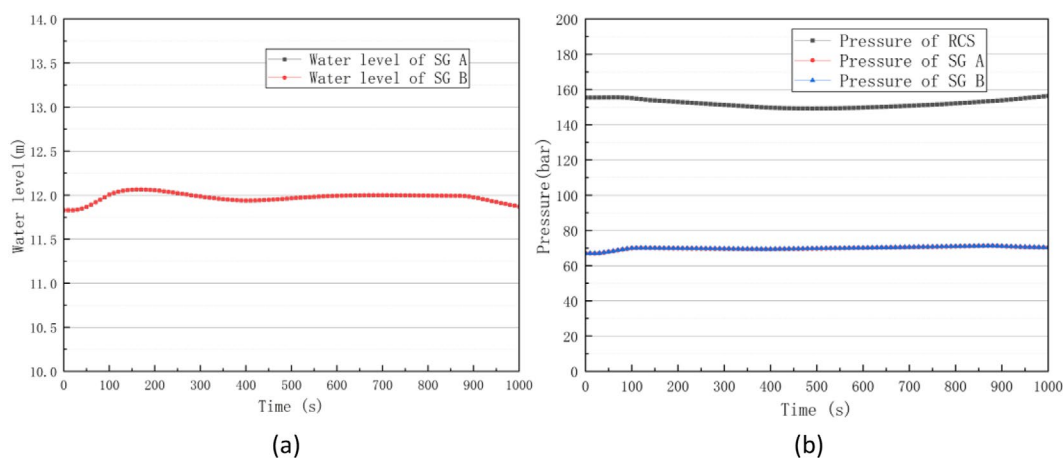
the over-pressurised containment unit and the later injection of seawater<sup>40</sup>. The results are shown in Figs. 7–9. Figure 7 shows the change in the reactor water level, while Figs. 8, 9 show the changes in the pressures of the reactor pressure vessel (RPV) and primary containment vessel (PCV). The real accident progressions are marked in Figs. 7–9. It can be noted that only limited measurements could be used due to the damage to the monitor from the accident<sup>41</sup>. In this study, for convenience purposes, we divided the accident process into four intervals, and the simulation results were compared with the measured data in each interval.

*From the Earthquake to tsunami arrival.* The reactor pressure increased due to the earthquake-caused shut down, and two isolation condenser (IC) systems were automatically activated. Afterwards, the two IC systems were manually shut down, and then an IC Subsystem-A was activated. As shown in Fig. 8, the reactor pressure was controlled by manually starting up and shutting down the IC subsystem-A to keep the pressure at a certain level.

*From the tsunami arrival to reactor water level decrease.* All the cooling capabilities, including the steam-driven cooling system and motor-operated pump, were lost due to the loss of control power. The water in the reactor continued to boil and evaporate, causing the reactor water level to continuously decrease (Fig. 7). From approximately 16:42 to 17:00 on March 11, 2011, the reactor water level could be measured for some time due to the temporary recovery of some DC power. As observed at 16:56, the water level was at the top of fuel (TOF) + 2,13 cm and had not yet decreased to TOF.



**Fig. 10** Power curves of reactor core and turbine (a) and average temperature curve of RCS (b) in a load rejection accident.



**Fig. 11** Pressure curves of the RCS and steam generator (a) and water level curve of steam generator (b) in a load rejection accident.

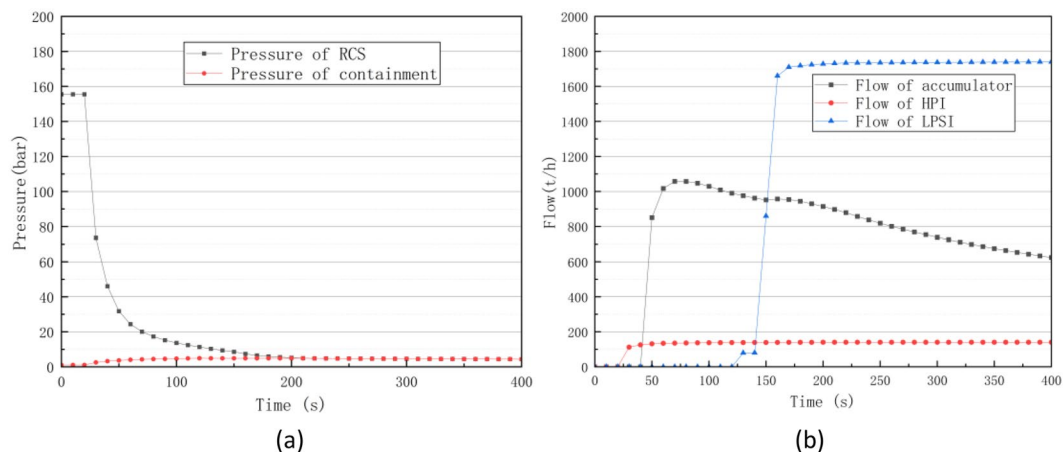
*From the reactor water level decrease to pcv pressure increase.* The RPV pressure was measured as 70 bar at 20:07 on March 11 and as 9 bar at 02:45 on March 12 (Fig. 8). The PCV pressure was measured as 6 bar at ~23:50 on March 11 (Fig. 9). It was observed that at a certain time after 20:00 on March 11, the RPV pressure decreased despite, and the PCV pressure showed a sharp increase, which is considered to be due to gas leakage from the main steam pipe.

*From the containment vessel pressure increase to containment venting operation.* On March 11, at ~23:50, the PCV pressure was 6 bar. However, it then increased and remained near 7–8 bar until the suppression chamber was successfully vented (Fig. 9). The reason was that the molten fuel descended to the bottom of the reactor vessel. Then, it further descended to the bottom of the PCV, thus further increasing the PCV pressure. When the molten fuel could not be sufficiently cooled, the concrete of the PCV floor was heated above its melting point, and a core–concrete reaction started, producing non-condensable gases, such as hydrogen and carbon monoxide, which have dramatic effects on the containment pressure.

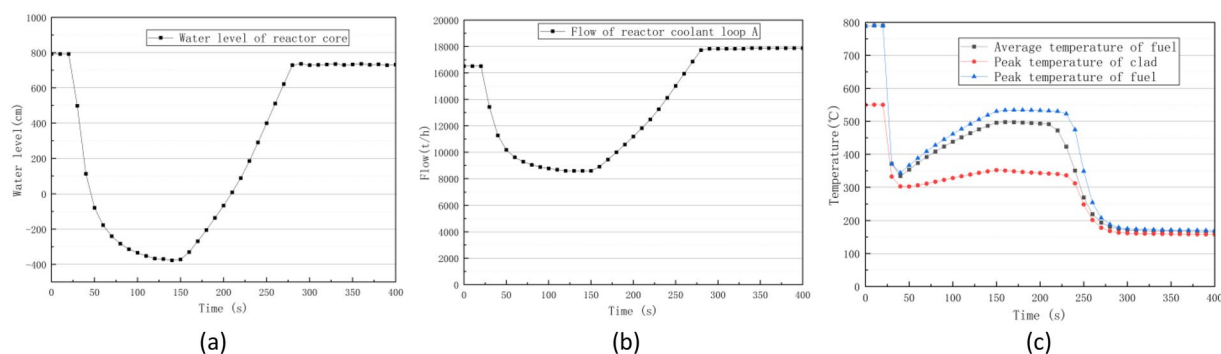
For all the intervals, the simulation results correctly showed the control logic and transient processes of Unit 1 of Fukushima Daiichi NPP. Partial measured data, which was acquired for using the temporary recovery of the power supply, was compared with the simulation results. Figures 7–9 show that the simulation results match the measured data. Thus, the reliability and quality of the used PCTAN data were validated.

**Validation using two operating conditions.** Two representative accident conditions of nuclear power plants were selected for the further technical validation of PCTAN.

*Normal operation with load rejection.* Failures in the speed control system of turbines or mislosures of steam piping valves can cause load rejection, and such conditions are among the common accident in nuclear power plants. Through PCTAN simulations, the processes of reactor core power and turbine power (load) changed



**Fig. 12** Pressure curves of the RCS and containment (a) and flow curve of auxiliary system (b) in a large LOCA accident.



**Fig. 13** Water level curve of the reactor core(a), flow curve of RCS(b) and temperature curves of reactor core in a large LOCA accident.

after inserting load rejection accidents at 20 s, as shown in Fig. 10a, where the turbine power first changed. Then, the turbine power reduction caused a heat imbalance in the second loop, in turn causing an increase in the average temperature of the first loop. As shown in Fig. 10b, the negative feedback effect of the temperature caused the core power to start decreasing.

The first loop pressure increased with the increase in temperature. Then, both of them began to decrease when the spray system and control rod drive system started working. As shown in Fig. 11a, the turbine power decrease also caused a brief pressure increase in the second loop. The pressure gradually decreased as steam was released to the outside through the atmospheric bypass valve. As shown in Fig. 11b, the water level of the steam generator (SG) temporarily increased due to the steam flow reduction. Afterwards, it gradually returned to normal through the control system. During this time, an oscillatory behaviour was caused by the pressure relief of the atmospheric bypass valve.

The above simulation results accurately demonstrate the control logic and transient processes of the load rejection accident.

**Normal operation with a large LOCA.** Coolant losses due to coolant pipe breaks are also among the common accidents in nuclear power plants. Under full power operation, a large break with an area of 2,300 cm<sup>2</sup> was assumed. As shown in Fig. 12a, a large coolant amount was discharged, and the reactor's cooling system pressure rapidly decreased. The containment pressure rapidly increased after the accident and gradually decreased as the containment spray system was operated.

As shown in Fig. 12b, the high-pressure injection (HPI) system, accumulator, and low-pressure safety injection (LPSI) system were successively put into operation as the pressure of the first loop was decreased to each threshold value. Figure 13a,b show that the core went through uncovered and re-flooded phases, with the core water level first decreasing and then gradually increasing. As shown in Fig. 13c, the fuel clad and peak fuel temperatures significantly increased during the core uncover phase and then decreased again as the core was re-flooded.

The above simulation results correctly show the control logic and transient response of a large LOCA accident.

## Usage Notes

The dataset is in the original MDB format and has a total size of ~15.1 GB. A more detailed description of the dataset and Python scripts for exploring the dataset are available on the GitHub page of the dataset (<https://github.com/thu-inet/NuclearPowerPlantAccidentData>). Users can reproduce the dataset using the PCTTRAN software and automation scripts described in this work. However, it is recommended to directly use the pre-built datasets as the building process is quite time-consuming. The *Data Processing.py* file provides the python code for converting the dataset's MDB format to Excel. Users can also use our code to generate datasets needed for AI models (e.g., training and test sets). In addition, the code can be used to plot time series graphs.

## Code availability

These simulations were conducted using PCTTRAN-PWR3LP (<https://github.com/thu-inet/NuclearPowerPlantAccidentData/tree/main/Simulator/>). The data processing step was performed using scripts written in the Python 3.10 programming language. More about this dataset can be found on the dataset's GitHub page (<https://github.com/thu-inet/NuclearPowerPlantAccidentData>).

Received: 25 October 2022; Accepted: 30 November 2022;

Published online: 13 December 2022

## References

- Murray, R. & Holbert, K. E. *Nuclear Energy: An Introduction to The Concepts, Systems, and Applications of Nuclear Processes* (Elsevier Press, 2014).
- Gu, Z. History review of nuclear reactor safety[J]. *Annals of Nuclear Energy* **120**, 682–690 (2018).
- Bogue, R. Robots in the nuclear industry: a review of technologies and applications. *Industrial Robot: An International Journal* **38**(2), 113–118 (2011).
- Marques, J. G. Evolution of nuclear fission reactors: Third generation and beyond. *Energy Conversion and Management* **51**(9), 1774–1780 (2010).
- Allen, G. & Chan, T. *Artificial Intelligence and National Security* (Belter Centre for Science and International Affairs Press, 2017).
- Friedman, S. M. Three Mile Island, Chernobyl, and Fukushima: An analysis of traditional and new media coverage of nuclear accidents and radiation. *Bulletin of the atomic scientists* **67**(5), 55–65 (2011).
- Ertel, W. *Introduction to Artificial Intelligence* (Springer Press, 2018).
- Brunelli, R. & Poggio, T. Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15**(10), 1042–1052 (1993).
- Levinson, J. *et al.* Towards fully autonomous driving: Systems and algorithms. *IEEE Intelligent Vehicles Symposium (IV)* **2011**, 163–168 (2011).
- Wu, Y. & Feng, J. Development and application of artificial neural network. *Wireless Personal Communications* **102**(2), 1645–1656 (2018).
- Suthaharan, S. *Support vector machine//Machine learning models and algorithms for big data classification* (Springer Press, 2016).
- Song, Y. Y. & Ying, L. U. Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry* **27**(2), 130 (2015).
- Yao, Y., Wang, J., Xie, M., Hu, L. Q. & Wang, J. Y. A new approach for fault diagnosis with full-scope simulator based on state information imaging in nuclear power plant. *Annals of Nuclear Energy* **141**(3), 107274 (2020).
- Qi, B., Zhang, L., Liang, J. & Tong, J. Combinatorial techniques for fault diagnosis in nuclear power plants based on Bayesian neural network and simplified Bayesian network-artificial neural network. *Frontiers in Energy Research* **10**, 920194 (2022).
- Wang, H., Peng, M. J., Abiodun, A., Xia, H. & Wang, X. X. Advanced fault diagnosis method for nuclear power plant based on convolutional gated recurrent network and enhanced particle swarm optimization. *Annals of Nuclear Energy* **151**, 107934 (2020).
- Lee, G., Lee, S. J. & Lee, C. A convolutional neural network model for abnormality diagnosis in a nuclear power plant[J]. *Applied Soft Computing* **99**, 106874 (2020).
- Liu, Y., Hu, R., Kraus, A., Balaprakash, P. & Obabko, A. Data-driven modeling of coarse mesh turbulence for reactor transient analysis using convolutional recurrent neural networks. *Nuclear Engineering and Design* **390**, 111716 (2022).
- Liu, Y., Wang, D., Sun, X., Dinh, N. & Hu, R. Uncertainty quantification for Multiphase-CFD simulations of bubbly flows: a machine learning-based Bayesian approach supported by high-resolution experiments. *Reliability Engineering & System Safety* **212**, 107636 (2021).
- Wu, X., Kozłowski, T. & Meidani, H. Kriging-based inverse uncertainty quantification of nuclear fuel performance code BISON fission gas release model using time series measurement data. *Reliability Engineering & System Safety* **169**, 422–436 (2018).
- Wang, P., Zhang, J., Wan, J. & Wu, S. A fault diagnosis method for small pressurised water reactors based on long short-term memory networks. *Energy* **239**, 122298 (2022).
- Margate, S. *History of the Pressurised Water Reactor* (Springer Press, 2022).
- Oka & Yoshiaki. *Nuclear Reactor Design* (Springer Japan Press, 2014).
- Po, L. C. Conceptual design of an accident prevention system for light water reactors using artificial neural network and high-speed simulator. *Nuclear Technology* **206**(3), 505–513 (2020).
- Gong, H., Zhu, T., Chen, Z., Wan, Y. & Li, Q. Parameter identification and state estimation for nuclear reactor operation digital twin. *Annals of Nuclear Energy* **180**, 109497 (2023).
- Gong, H., Cheng, S., Chen, Z. & Li, Q. Data-enabled physics-informed machine learning for reduced-order modeling digital twin: application to nuclear reactor physics. *Nuclear Science and Engineering* **196**(6), 668–693 (2022).
- Gong, H., Cheng, S., Chen, Z., Li, Q. & Quilodrán-Casas, C. An efficient digital twin based on machine learning SVD autoencoder and generalised latent assimilation for nuclear reactor physics. *Annals of Nuclear Energy* **179**, 109431 (2022).
- Qi, B., Xiao, XY., Liang, JJ., Po, LC. & Tong, JJ. An open time-series simulated dataset covering various accidents for nuclear power plants. *figshare*, <https://doi.org/10.6084/m9.figshare.c.6238473> (2022).
- Akcasuh, Z. *Mathematical Methods in Nuclear Reactor Dynamics* (Elsevier Press, 2012).
- Stan, M. Multi-scale models and simulations of nuclear fuels. *Nuclear engineering and technology* **41**(1), 39–52 (2009).
- Bruens, N W S., Brukx, J F L. M., Latzko, D G. H., Vriesema, B. & Kerlin, T. W. Modeling of Nuclear Steam Generator Dynamics. Report No.CONF-750977-4099056 (United States, 1975).
- Akcasuh, Z. *Mathematical methods in nuclear reactor dynamics* (Elsevier Press, 2012).
- Kleinstreuer, C. *Two-phase Flow: Theory and Applications* (Routledge Press, 2017).
- Todreas, N. E., & Kazimi, M. S. *Nuclear systems volume I: Thermal hydraulic fundamentals* (CRC Press, 2021).
- Yeoh, G. H. & Tu, J. Y. Two-fluid and population balance models for subcooled boiling flow. *Applied mathematical modelling* **30**(11), 1370–1391 (2006).

35. Lausterer, G. K., Franke, J. & Eitelberg, E. Mathematical modelling of a steam generator, *Digital Computer Applications to Process Control*. 411–417 (1980).
36. Po, L. C. PCTRAN: a transient analysis code for personal computers. *Nuclear Engineering International* 33(406), 47–50 (1988).
37. IAEA, *PCTRAN Generic Pressurised Water Reactor Simulator Exercise Handbook* (IAEA Press, 2019).
38. Yamazaki, M. *Fukushima Nuclear Accident Analysis Report* (Tokyo Electric Power Company Press, 2012).
39. Tokyo Electric Power Company, Reactor Core Status of Fukushima Daiichi Nuclear Power Station Unit 1, <http://www.tepco.co.jp/en/press/corp-com/release/11051509-e.html> (2011).
40. Po, L. C. Fukushima Daiichi crisis: Simulation Events at unit 1. <https://www.neimagazine.com/features/featureevents-at-unit-1> (2011).
41. Ahn, J. *et al. Reflections on the Fukushima Daiichi nuclear accident: Toward social-scientific literacy and engineering resilience* (Springer Nature Press, 2015).

## Acknowledgements

The research was supported by the Innovation Funds of CNNC-Tsinghua Joint Center for Nuclear Energy R&D (Lingchuan Fund) and a grant from the National Natural Science Foundation of China.

## Author contributions

Ben Qi: methodology, implementation, data curation, writing- original draft preparation. Xingyu Xiao: data processing, technical validation, writing – review & editing. Jingang Liang: conceptualization, resources, supervision, writing – review & editing. Li-chi Po: methodology, software, supervision, writing – review & editing. Ligu Zhang: supervision, resources, writing – review & editing. Jiejuan Tong: supervision, resources, writing – review & editing.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to J.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022