Featured in Physics

Linear Regression and Machine Learning for Nuclear Forensics of Spent Fuel from Six Types of Nuclear Reactors

Shengli Chen,[1] Tianxiang Wang,[1] Zhong Zhang,[1] Runfeng Li,[1] Su Yuan,[1] Ruiyi Zhang,[1] Cenxi Yuan,[1,*] Chunyu Zhang,[1] and Jianyu Zhu[2]

[1] *Sino-French Institute of Nuclear Engineering and Technology, Sun Yat-sen University, Zhuhai, Guangdong 519082, China*

[2] *China Academy of Engineering Physics, Center for Strategic Studies, Beijing 100088, China*

(Received 18 August 2022; revised 7 November 2022; accepted 1 February 2023; published 9 March 2023)

The illicit trafficking of radioactive materials, especially weapon-grade uranium or plutonium, is a significant security threat. Nuclear forensics helps trace the illicit trafficking of radioactive materials. The present study develops the methods for the forensics of the possible origins of fuels irradiated in nuclear reactors, which are the most powerful sources producing radioactive materials, including plutonium. Three key factors are significant for irradiated fuel forensics, namely, initial ^{235}U enrichment, burnup, and the type of irradiation nuclear reactors. The methods for the first two are determined based on experimental data of six nuclear-reactor technologies and are further verified using the neutron-transport-depletion coupling simulation of the two major commercial reactor technologies, a pressurized-water reactor (PWR) and a boiling-water reactor (BWR). In addition, three machine-learning techniques are applied to discriminate between a PWR and a BWR, which are quite similar in neutronic properties, with nice accuracy and generalization ability. In summary, the presently determined methods provide a reliable pathway to predict the origins of spent nuclear fuels.

DOI: 10.1103/PhysRevApplied.19.034028

I. INTRODUCTION

Nuclear nonproliferation demands being able to distinguish between civilian and military nuclear activities [1–4]. In addition to direct military activities, illicit trafficking of radioactive materials, especially weapon-grade highly enriched uranium (HEU) or plutonium, is a significant security threat. As reported in the International Atomic Energy Agency's (IAEA's) Incident and Trafficking Database (ITDB), at least 320 incidents from 1993 to 2021 were (likely) induced by the trafficking or malicious use of nuclear and other radioactive materials [5]. Among these 320 incidents, 12 are related to HEU, and 3 are related to plutonium [5]. Nuclear forensics is intended to respond to how, when, and where radioactive materials are produced [6–8].

The public part of the IAEA's ITDB [5] does not allow knowledge of whether the majority of trafficking or malicious use is related to low-enriched uranium. The nuclear reactor is the most powerful source of radioactive materials due to its high neutron flux and stable operation for a long period. Spent fuel from nuclear reactors is an important source for the illicit trafficking of nuclear and other radioactive materials. Therefore, many studies are performed on the identification of spent fuel. The origins that need to be determined are (i) the type of nuclear reactor where the fuel was irradiated [9–17], (ii) the ^{235}U enrichment of the fresh fuel [15,18] (more generally, isotopic compositions [19]), and (iii) the exposure of the fuel in the reactor (known as burnup) [13,15,17,18]. For the last of these, because the determination of fuel burnup is also important for the operation of nuclear reactors, more studies can be found in the literature (e.g., Refs. [20,21]) and are not repeated here.

Excluding the regular quantification of burnup for known reactor types and fuel enrichment in the nuclear industry, most of the above-cited works were based on neutron-transport-depletion coupling simulation data rather than experimental data of isotopic concentrations. The only exception is Ref. [17], where both experimental measurements and simulation data were used. Due to the limitations of neutron-transport-depletion coupling simulations, more studies based on experimental data are still required for nuclear forensics. Even though for the burnup current experimental studies are limited to a few fuel enrichments, the extension to other fuel enrichments, especially high enrichments, needs to be verified. Therefore, the present work investigates the methods for predicting nuclear-reactor type, fuel enrichment, and burnup of spent fuel using experimental data collected in

*yuancx@mail.sysu.edu.cn

the SFCOMPO-2.0 database [22]. Some neutron-transport-depletion coupling simulation data are calculated using the OpenMC code [23] to verify the present methods.

All methods and theoretical models used in the present study are explained in Sec. II. The trained models based on the SFCOMPO-2.0 database for predicting the origins of spent fuel are presented in Sec. III. Section IV provides the results of the verification of the determined models using simulation data. The main conclusions of the present work are summarized in Sec. V.

II. METHODS AND THEORETICAL MODELS

A. Philosophy of using relative isotopic concentrations

The absolute quantity of an isotope does not help to identify the source of spent fuel. The use of the ratio of two isotopic concentrations can reduce the influence of the density of fresh fuel. In addition, some implicit relationships between several isotopic concentrations can be more easily shown by the ratios of concentrations. Particularly, some nonlinearities are weakened, and a few nonlinear terms may be eliminated numerically in specific ranges (e.g., see the reasoning of a simplified case in Ref. [18]). This is essential for nuclear forensics, including the identification of fuel enrichment and burnup (which are shown herein) and distinguishing between civilian and military nuclear activities [2,3].

B. Summary of SFCOMPO-2.0 data

The SFCOMPO-2.0 database collects experimental data on isotopic composition in spent fuels [22]. Experimental data are measured for 750 fuel samples in 296 fuel rods and 116 assemblies irradiated in 44 nuclear reactors of 8 reactor types. The numbers of samples in each nuclear-reactor type are summarized in Table I. There are more samples with actinide data than those including concentrations of fission products. Furthermore, the inclusion of fission products does not result in a better binary linear model of fuel-enrichment forensics. Thereby, models using only actinides are studied based on experimental data from the SFCOMPO-2.0 database. The present study aims to determine forensics models insensitive to the detailed setup of reactors. Thus, SFCOMPO-2.0 data are not distinguished by the setup and operating history.

Here, favorable isotopes for the forensics of enrichment and burnup are selected among the seven major actinides: ^{235}U, ^{236}U, ^{238}U, ^{238}Pu, ^{239}Pu, ^{240}Pu, and ^{242}Pu. With the selected major actinides, the total number of samples with the corresponding measured data are also included in Table I. ^{241}Pu is not included due to its relatively short half-life: the 87.7-year half-life means that the concentration of ^{241}Pu is sensitive to the cooling time of spent fuel. Thus, the prediction based on ^{241}Pu is uncertain. Because of the few available experimental data for CANDU and MAGNOX, these two reactor types are no longer considered in the present study.

TABLE I. Number of samples irradiated in different reactor types collected in the SFCOMPO-2.0 database.

Reactor type	Total	Selected
Advanced gas-cooled reactor (AGR)	63	60
Boiling-water reactor (BWR)	249	157
Canada deuterium uranium reactor (CANDU)	31	0
Magnox alloy graphite-moderated gas-cooled reactor (MAGNOX)	4	4
Pressurized water reactor (PWR)	295	215
Reaktor Bolshoy Moshchnosti Kanalniy[a] (RBMK)	41	41
Vodo-Vodynaoi Energetichesky Reactor[b] (VVER-440)	47	47
VVER-1000	20	20
Total samples	750	430

[a]High-power channel-type reactor.
[b]Water-water energetic reactor.

Long-lived actinides are considered because the present study determines the models insensitive to the detailed setup of reactors. They are not necessarily important for other indicators. For example, the combinations of long-lived actinides in the fuel pellet are not so helpful in determining the realistic cooling time because their natural decay consumptions are negligible (since cooling time $\ll$ half-lives).

In addition to the uranium fuel, the uranium-plutonium mixed-oxide (MOX) fuel is used in nuclear reactors [24]. Some MOX-fuel-based experimental data are also available in the SFCOMPO-2.0 database [22]. Due to the largely different initial compositions of major actinides in uranium and MOX fuels (especially the concentrations of Pu isotopes [25]), developing general models applicable to both fuels is challenging. Åberg Lindell *et al.* found methods to discriminate irradiated MOX fuel from UO_2 fuel in a PWR [26] (even though their study is also based on simulated data). Therefore, the present work focuses only on samples of uranium fuel, which is still the main nuclear fuel in the world.

C. Linear regression for fuel enrichment and burnup

Fuel enrichment can be simultaneously determined using the signature γ rays of short-lived fission products [27]. Such a method is unsuitable for spent fuel unloaded from nuclear reactors due to the relatively long "shut-down" time for spontaneous decay of the short-lived isomers. It is necessary to develop predictive models using long-lived isotopes. Figure 1 shows the mass ratios of $R_1 = {}^{235}\mathrm{U}/{}^{238}\mathrm{U}$ and $R_2 = {}^{236}\mathrm{U}/{}^{238}\mathrm{U}$ versus real values of

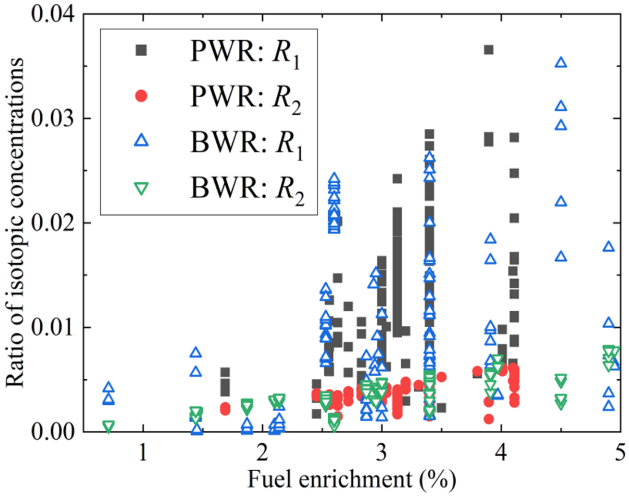

FIG. 1. Selected ratios of $R_1 = {}^{235}\mathrm{U}/{}^{238}\mathrm{U}$ and $R_2 = {}^{236}\mathrm{U}/{}^{238}\mathrm{U}$ versus real values of fuel enrichment for both PWR and BWR samples.

fuel enrichment for both SFCOMPO-2.0 PWR and BWR samples. Because of the different burnup levels and other effects, there is no evident correlation between one of the two ratios and ^{235}U enrichment [28]. Therefore, models need to be developed for forensics.

It is shown that binary linear regression is suitable to describe the relationship between fuel enrichment (or burnup) and the ratios of some selected isotopes in a PWR [18]. Therefore, binary linear regression is still used in the present work for various reactor types. This implies that the fuel enrichment (denoted by e) or burnup (denoted by b_u) is given by

$$e\ (\text{or}\, b_u) = c_0 + c_1 R_1 + c_2 R_2, \tag{1}$$

where R_1 and R_2 are ratios of two selected isotopic concentrations, and the coefficients c_i are determined by fitting experimental data collected from the SFCOMPO-2.0 database.

It is noticeable that the fitting coefficients strongly depend on the data used; this makes the extrapolation of fitting laws uncertain. However, the fitting functions are quite close in the interpolated data range. Consequently, one should cover as large a range as possible to perform the fitting.

One may be interested in the relative contributions of the two concentration ratios to the linearity of the forensics models. For a binary linear model, the weighting factor of each variable (i.e., R_1 or R_2) is

$$w_i = \frac{(c_i \Delta R_i)^2}{\sum_{i=1}^{2} (c_i \Delta R_i)^2}, i = 1, 2, \tag{2}$$

where ΔR_i is the standard deviation of R_i, and c_i are the fitting coefficients described above.

D. Discrimination against nuclear-reactor type

Because the reactor type is not a continuous variable like enrichment and burnup, the above linear regression is not applicable for discriminating nuclear-reactor type if the spent fuel was irradiated. Among various reactor types, distinguishing between a PWR and a BWR is the most challenging because of their similar neutron spectra in the core. Moreover, the SFCOMPO-2.0 database collects only a quite limited number of experimental data for reactors other than PWR and BWR types. Thus, the present work focuses on the discrimination of reactor type between a PWR and a BWR.

1. Logistic regression

Logistic regression (LGR) [29,30] is a machine-learning technique widely used for classification. Because the present study classifies only a PWR or a BWR, a sigmoid function is applied by assigning a PWR for function values smaller than 0.5 and a BWR otherwise. Let denote the experimental data sets by $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)\}$, where $\mathbf{x}_i$ $(i \in [\![1, 2, \ldots, n]\!])$ are the feature vectors (i.e., the ratios of isotopic concentrations in our study) and $y_i \in \{0, 1\}$ represent a PWR or a BWR. A weighting vector $\boldsymbol{\theta} = [\theta_0, \theta_1, \ldots, \theta_m]^T$, where m (=2 in the present study) is the size of $\mathbf{x}_i$, is set to account for the relative contribution of each feature element, i.e., for each data set $(\mathbf{x}_i, y_i)$, a linear combination [31],

$$\theta_0 x_{i,0} + \theta_1 x_{i,1} + \ldots + \theta_m x_{i,m} = \boldsymbol{\theta}^T \boldsymbol{x}_i, \tag{3}$$

is used in the sigmoid function

$$h_{\boldsymbol{\theta}}(\mathbf{x}_i) = (1 + \exp(-\boldsymbol{\theta}^T \mathbf{x}_i))^{-1}, \tag{4}$$

which refers to the probability:

$$P(y = 1 | \mathbf{x}_i; \boldsymbol{\theta}) = h_{\boldsymbol{\theta}}(\mathbf{x}_i). \tag{5}$$

Considering all experimental data, the likelihood is as follows:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} P(y_i | \mathbf{x}_i; \boldsymbol{\theta}) = \prod_{i=1}^{n} (h_\theta(\mathbf{x}_i))^{y_i} (1 - h_\theta(\mathbf{x}_i))^{1-y_i}. \tag{6}$$

To avoid overfitting, the $L2$-regression (also known as ridge regression) term is applied in the present work. The weighting vector $\boldsymbol{\theta}$ is optimized by maximizing the following function by iteration:

$$L_{\mathrm{LGR}}(\boldsymbol{\theta}) = \prod_{i=1}^{n} (h_{\boldsymbol{\theta}}(\mathbf{x}_i))^{y_i} (1 - h_{\boldsymbol{\theta}}(\mathbf{x}_i))^{1-y_i} + \frac{1}{2m} \sum_{j=1}^{m} \theta_j^2. \tag{7}$$

2. Support vector machine

The support vector machine (SVM) [29,30] is derived from statistical learning theory. The philosophy of the SVM is to find a hyperplane:

$$\mathbf{N}^T\mathbf{x}_i + b = 0, \tag{8}$$

where $\mathbf{N}$ is a normal vector, as the decision boundary for a given data set. One can then define

$$y = \begin{matrix} 1, & \mathbf{N}^T\mathbf{x}_i + b > 0, \\ -1, & \mathbf{N}^T\mathbf{x}_i + b < 0; \end{matrix} \tag{9}$$

$+1$ and -1 represent the BWR and PWR, respectively, in the present study. The distance between two classes is called the margin and is given by

$$d = \frac{2}{\| \mathbf{N} \|}. \tag{10}$$

The optimized $\mathbf{N}$ is determined by maximizing d under the condition $y_i(\mathbf{N}^T\mathbf{x}_i + b) \geq 1$.

The Lagrange function

$$\mathcal{L}(\mathbf{N}, b, \boldsymbol{\alpha}) = \frac{1}{2}||\mathbf{N}||^2 - \sum_{i=1}^{n} \alpha_i (y_i(\mathbf{N}^T\mathbf{x}_i + b) - 1), \tag{11}$$

with $\alpha_i \geq 0$, which is introduced to solve the above problem. The optimization is equivalent to

$$\min_{\mathbf{N},b} \left\{ \max_{\alpha_i \geq 0} \{\mathcal{L}(\mathbf{N}, b, \boldsymbol{\alpha})\} \right\}, \tag{12}$$

which can be further expressed by

$$\max_{\alpha_i \geq 0} \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \right\} \tag{13}$$

($\mathbf{N}$ and b are eliminated because minimization over $\mathbf{N}$ and b implies null partial derivatives) under the condition

$$\sum_{i=1}^{n} \alpha_i y_i = 0. \tag{14}$$

3. Multilayer perceptron

The multilayer perceptron (MLP) [30,32] imitates neurons in an organism. The main structure of the MLP can be divided into three layers, input layer, hidden layer(s), and output layer, as shown in Fig. 2. Here, two hidden layers are used, and the output for $\mathbf{x}_i$ is given by

$$f(\mathbf{x}_i) = s\{\mathbf{b}_3 + \mathbf{w}_3 s[\mathbf{b}_2 + \mathbf{w}_2 s(\mathbf{b}_1 + \mathbf{w}_1 \mathbf{x}_i)]\} \tag{15}$$

where $\mathbf{w}_j$ are matrices of weighting factors, $\mathbf{b}_j$ are the biases, and s is the activation function. Because there are only two classes, i.e., a PWR and a BWR, the output layer can be assigned as a scalar, e.g., the probability of a PWR or a BWR.

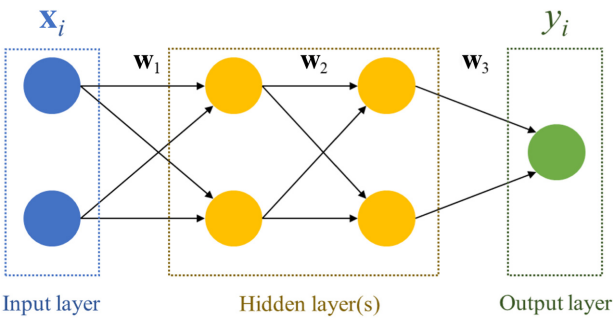

FIG. 2. Schematics of the MLP used in the present work.

The error back-propagation (BP) algorithm is used to train the MLP by determining the weighting matrices and biases. In the BP algorithm, for each datum, the gradient descent method is used to find the parameters minimizing the lost function. The present cost function is

$$g(\hat{y}, y) = -(y \log \hat{y} + (1 - y) \log(1 - \hat{y})), \tag{16}$$

where $\hat{y} = f(\mathbf{x})$ and y represent the predicted and real labels, respectively. This cost function is selected because $g(\hat{y}, y) \geq 0$ for $0 \leq \hat{y}, y \leq 1$ and $g(y, y) = 0$. With the $L2$ regression to avoid overfitting, the final cost function becomes

$$L_{\mathrm{MLP}}(\mathbf{w}, \mathbf{b}) = -\sum_{i=1}^{n} g(\hat{y}_i, y_i) + \frac{1}{2} \sum_{j=1}^{3} ||\mathbf{w}||_j^2, \tag{17}$$

where $||\mathbf{w}||_{j2}$ denotes the $L2$ norm of the weighting matrix, $\mathbf{w}_j$.

4. Evaluation of models

To evaluate the reliability of the models for discrimination, one defines the accuracy rate as

$$\mathcal{A} = \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i} (f(\mathbf{x}_i)), \tag{18}$$

where the δ_{y_i} function is defined by

$$\delta_{y_i}(f(\mathbf{x}_i)) = \begin{matrix} 1, & f(\mathbf{x}_i) = y_i, \\ 0, & f(\mathbf{x}_i) \neq y_i. \end{matrix} \tag{19}$$

The corresponding error rate is $\mathcal{E} = 1 - \mathcal{A}$. These two indicators can give a preliminary evaluation of the models. However, they are not sufficient for completely evaluating the models, especially when training data are dominated by

TABLE II. Confusion matrix.

	Predicted label	
Real label	0 (negative) [PWR]	1 (positive) [BWR]
0 [PWR]	True negative (TN)	False positive (FP)
1 [BWR]	False negative (FN)	True positive (TP)

one of the classes. Therefore, the precision and recall rates are defined based on the confusion matrix

$$\mathcal{P} = \frac{T^+}{T^+ + F^+}, \qquad \mathcal{R} = \frac{T^+}{T^+ + F^-}, \tag{20}$$

where the symbols are defined in Table II. Due to the difficulty of getting both high precision and recall rates in practice, one defines the so-called F_1 score for balancing the precision and recall:

$$F_1 = \frac{2\mathcal{P}\mathcal{R}}{\mathcal{P} + \mathcal{R}}. \tag{21}$$

The highest F_1 score is used as the criterion to choose the best binary combination of isotopic ratios for each discrimination method.

The generalization ability of the trained models is quantified by the so-called area under the curve (AUC). Here, the curve refers to the receiver operating characteristic (ROC) curve, which plots the recall rate, $\mathcal{R}$ (also called the true positive rate, TPR), versus the false positive rate (FPR, equal to $F^+/[F^+ + T^-]$).

The tenfold cross-validation method is applied to test each machine-learning model. As illustrated in Fig. 3, complete data sets are divided into ten different subsets with the same sample numbers in the tenfold cross-validation method. The training-test process is repeated 10 times. At each time, nine subsets are used to train models, and the remaining one is used as the test dataset to validate the trained models. The definitive forensics model is the

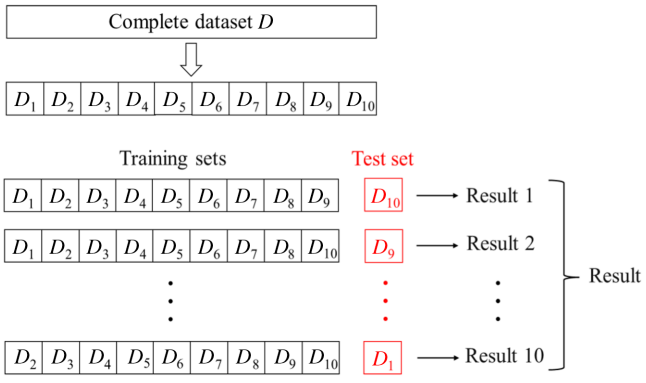

FIG. 3. Illustration of the tenfold cross-validation method.

TABLE III. Key parameters for neutronic simulations.

Parameter	PWR	BWR
Fuel material	UO_2	UO_2
Cladding material	Zircaloy-4	Zircaloy-4
Coolant	Water	Water
Coolant density (g/cm^3)	0.741	0.760 (with 42% void)
Fuel-pellet diameter (mm)	9.29	12.10
Cladding inner diameter (mm)	9.48	12.42
Cladding outer diameter (mm)	10.72	14.30
Fuel-rod pitch (mm)	14.1	18.8
Active height (cm)	364	371

average of the ten models determined during the tenfold cross-validation training-test processes.

E. Theoretical modeling for supplementary verification

The above nuclear forensics models are trained using experimental data from the SFCOMPO-2.0 database. Neutron-transport-depletion coupling simulations are performed to verify the models supplementarily. The trained models are supposed to be insensitive to the detailed setup of the reactors, since the SFCOMPO-2.0 database includes various experimental data with different setups for each considered reactor type. Because the objective is to verify the above-trained models using calculated isotopic concentrations, the neutronic simulations are performed in a unit cell for both a PWR and a BWR. The key parameters for neutronic simulations are summarized in Table III. It is preliminarily shown that the isotopic concentrations are not sensitive to other parameters (e.g., moderator temperature and boron concentration) [33]. The calculations are performed using the OpenMC stochastic code [23] and the ENDF/B-VII.1 nuclear data library [34].

The neutron-transport-depletion coupling simulations depend on both the computation code and nuclear data. For verified codes, nuclear data contribute to the main biases [35]. Both nuclear reaction data and the depletion chain can make the simulation data biased. Therefore, one may find that the prediction based on simulated isotopic concentrations slightly differs from that based on measured concentrations. Nevertheless, due to the relatively reliable nuclear data evaluated recently, the bias of the concentrations for the major isotopes is no longer significant and not very sensitive to nuclear data evaluation. Consequently, simulation data may have a slight bias but remain acceptable for verifying the present models.

III. MODELS TRAINED WITH EXPERIMENTAL DATA

For simplification, in the following descriptions, the isotopic concentrations are expressed by the symbols of

TABLE IV. Coefficients in the fuel enrichment (in %) model.

Reactor	c_0	c_1	c_2	R^2	w_1	w_2
PWR	0.457	73.52	476.1	0.921	0.466	0.534
BWR	0.374	75.30	519.9	0.976	0.324	0.676
VVER-440	1.229	50.19	376.0	0.954	0.211	0.789
VVER-1000	0.724	68.40	441.6	0.927	0.323	0.677
AGR	0.635	55.47	487.4	0.871	0.199	0.801
RBMK	1.622	17.17	65.21	0.735	0.312	0.688

the corresponding isotopes; the ratio of two isotopes refers to the mass ratio.

A. Models for forensics of fuel enrichment

Analysis of SFCOMPO-2 data shows that fuel enrichment can be well described by the binary linear function

$$e = c_0 + c_1 \frac{^{235}\mathrm{U}}{^{238}\mathrm{U}} + c_2 \frac{^{236}\mathrm{U}}{^{238}\mathrm{U}}, \tag{22}$$

where the coefficients and the coefficients of determination (denoted by R^2) are given in Table IV. Equation (22) is a first-order approximation by considering the depletion of uranium in the reactor core because a constant proportion (i.e., the ratio of neutron capture to absorption of ^{235}U) of consumed ^{235}U is transmuted to ^{236}U and e is the percentage of ^{235}U in the fresh fuel. A comparison of the predicted values using Eq. (22) and the real values is intuitively shown in Fig. 4. One can verify from both Fig. 4 and R^2 in Table IV that the trained model, Eq. (22), successfully describes fuel enrichment for both a PWR and a BWR. It is noted that the relatively low value of R^2 for the RBMK is due to the few available enrichments for all 41 samples. Figure 4 demonstrates that the prediction based on isotopic concentrations is reliable for the RBMK.

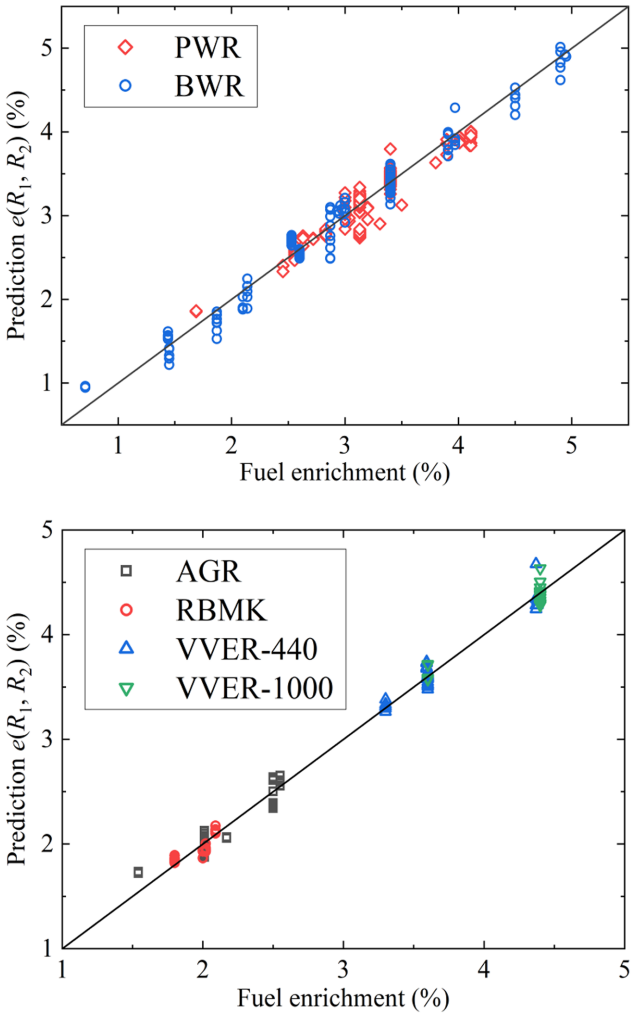

FIG. 4. Trained model-based predictions versus real values of fuel enrichment for both a PWR and a BWR (upper), and AGR, RBMK, VVER-440, and VVER-1000 (lower) samples collected in the SFCOMPO-2.0 database.

B. Models for the forensics of fuel burnup

As was the case for the PWR discussed in Ref. [18], several binary combinations of the ratios of isotopic concentrations can give a good description of burnup. Analysis using samples with the selected seven major actinides shows that ^{236}U/^{238}U is also an important term for predicting fuel burnup. A PWR and BWR share the same best combination of R_1 and R_2 (^{236}U/^{238}U and ^{242}Pu/^{238}U). This combination is expected to be suitable for the two VVER reactor technologies. This can be verified against the fitting results given in Table V [36]. Consequently, the combination (^{236}U/^{238}U, ^{242}Pu/^{238}U) is supposed to be

TABLE V. Coefficients in the fuel burnup (in GWd/tU) models ($R_1 = {}^{236}\mathrm{U}/{}^{238}\mathrm{U}$ for all reactors) by considering all satisfactory samples collected in the SFCOMPO-2.0 database.

Reactor	R_2	c_0	c_1	c_2	R^2	w_1	w_2
PWR	^{242}Pu/^{238}U	0.009	4836	21 790	0.989	0.328	0.672
BWR	^{242}Pu/^{238}U	−0.337	5711	16 540	0.970	0.358	0.642
VVER-440	^{242}Pu/^{238}U	−0.790	4892	21 080	0.983	0.452	0.548
VVER-1000	^{242}Pu/^{238}U	2.549	3752	23 720	0.994	0.314	0.686
AGR	^{240}Pu/^{239}Pu	−1.658	2243	20.60	0.986	0.147	0.853
RBMK	^{240}Pu/^{239}Pu	−0.141	622.3	25.49	0.975	0.024	0.976

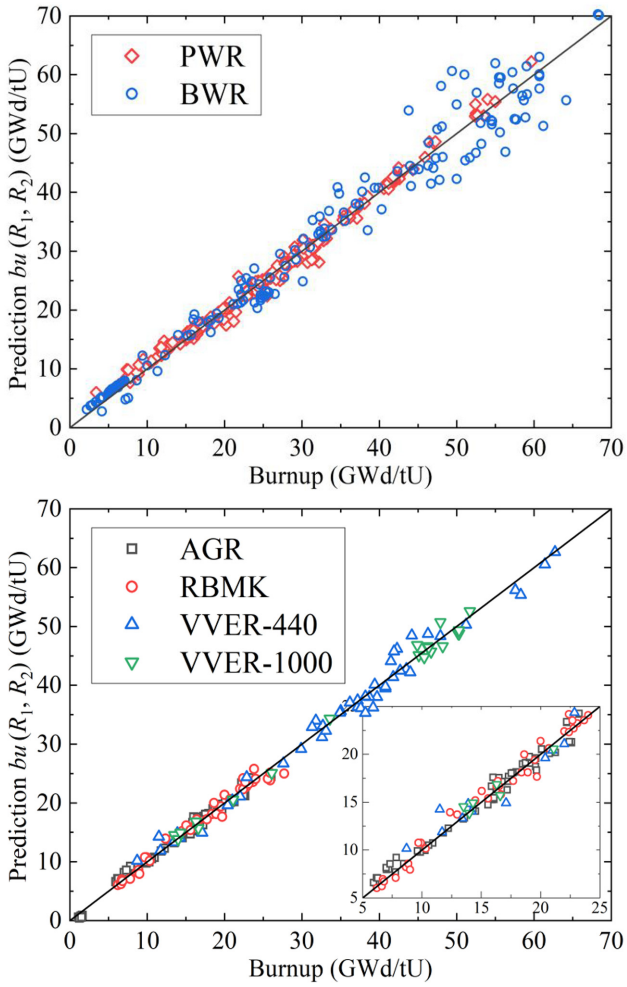

FIG. 5. Trained model-based predictions versus real values of burnup of fuel-irradiated PWR and BWR samples (upper) as well as AGR, RBMK, VVER-440, and VVER-1000 samples (lower) collected in the SFCOMPO-2.0 database.

predictable for the burnup of spent fuel from light-water reactors.

While the neutron moderator in the PWR, BWR, and VVER is light water, that in the AGR and RBMK is graphite. The different neutron moderators result in different neutronic behaviors, and thus, make the characteristics of forensics different. One can get many combinations of (R_1, R_2) with which the binary linear regressions are quite good (i.e., $R^2 \geq 0.97$) for the AGR and RBMK. To get the models as close as possible to those of light-water-moderated reactors, R_1 is forced to be $^{236}U/^{238}U$. ($^{236}U/^{238}U$, $^{240}Pu/^{239}Pu$) is shown to be suitable for describing the burnup of fuel irradiated in both the AGR and RBMK, in which the neutrons are moderated by graphite.

Based on the above analysis, the selected isotopes for describing burnup are ($^{236}U/^{238}U$, $^{242}Pu/^{238}U$) for water-moderated reactors and ($^{236}U/^{238}U$, $^{240}Pu/^{239}Pu$) for graphite-moderated reactors. All samples with the above isotopic concentrations in the SFCOMPO-2.0 database are used to get the predictive models. The corresponding plots of prediction versus real values are illustrated in Fig. 5. The fitting statuses are summarized in Table V.

Jaffke *et al.* developed a tool for thermal neutron-fluence diagnostics in a very low burnup regime for natural uranium samples using $^{235}U/^{238}U$ and $^{236}U/^{235}U$ [37]. Because fuel burnup and thermal neutron fluence are strongly correlated in thermal neutron reactors, the presently determined bilinear models are expected to extend the burnup region for diagnosing the thermal neutron fluence.

C. Models for forensics of nuclear-reactor type

Figure 6 shows the best MLP model prediction of a PWR and a BWR using isotopic concentrations. 95% PWR and 91% BWR are correctly identified. As presented in Sec. II D 4, many other indicators should also be used to evaluate the reliability of the trained model. However, none of the three models (i.e., LGR, SVM, and MLP) optimizes all indicators. The highest F_1 score is used as the criterion to choose the best-trained models. For each method, the top-three binary combinations with the highest F_1 scores are tabulated in Table VI. For the discrimination between a PWR and a BWR, it is found that the isotopes getting the highest F_1 score are the same (i.e., ^{239}Pu, ^{240}Pu, and ^{242}Pu) for the three methods. While LGR and SVM use

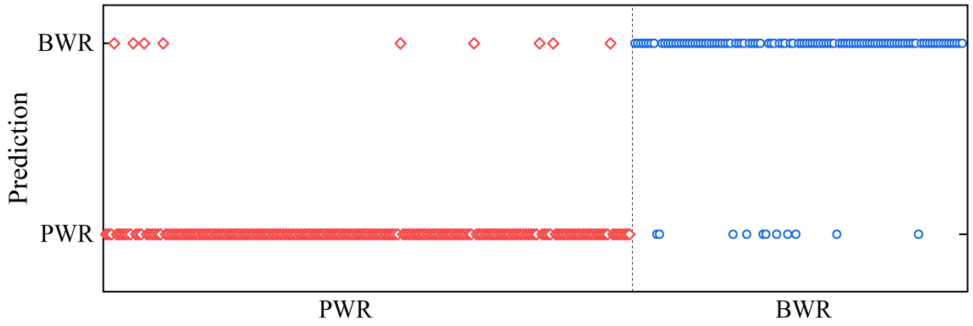

FIG. 6. MLP model-based prediction of a PWR and a BWR.

TABLE VI. Top-three binary combinations with the highest F_1 score.

Method	R_1	R_2	F_1
LGR	$^{239}Pu/^{240}Pu$	$^{240}Pu/^{242}Pu$	0.860
	$^{235}U/^{236}U$	$^{240}Pu/^{242}Pu$	0.857
	$^{239}Pu/^{240}Pu$	$^{236}U/^{242}Pu$	0.839
SVM	$^{239}Pu/^{240}Pu$	$^{240}Pu/^{242}Pu$	0.893
	$^{236}U/^{242}Pu$	$^{240}Pu/^{239}Pu$	0.834
	$^{235}U/^{240}Pu$	$^{236}U/^{242}Pu$	0.827
MLP	$^{239}Pu/^{240}Pu$	$^{239}Pu/^{242}Pu$	0.917
	$^{235}U/^{236}U$	$^{240}Pu/^{242}Pu$	0.881
	$^{242}Pu/^{236}U$	$^{240}Pu/^{239}Pu$	0.871

$^{239}Pu/^{240}Pu$ and $^{240}Pu/^{242}Pu$ as bases, MLP is optimized by using $^{239}Pu/^{240}Pu$ and $^{239}Pu/^{242}Pu$.

The confusion matrices for the three methods are displayed in Fig. 7. Due to the quite similar ROC curves for the three methods, Fig. 8 shows an example of the MLP model. The indicators for the three methods using the corresponding best bases (i.e., the binary combinations with the highest F_1 scores) are tabulated in Table VII. In the present study, LGR is shown to be less suitable than the other two methods, of which the predictions remain unchanged for parameters perturbed by up to ±5%. MLP has a nice performance for discriminating between a PWR and a BWR using $^{239}Pu/^{242}Pu$ and $^{239}Pu/^{240}Pu$ as two basic components. For all three models, the generalization ability is ensured by the high AUC values. In practice, one can add even more concentration ratios to identify the reactor technology if the depleted fuel is irradiated.

D. Discussion on forensics models

Notably, the forensic models of fuel enrichment and burnup depend on reactor technology, which means the forensics of fuel enrichment and burnup require knowledge of the reactor type. Therefore, the forensics of the reactor type should be performed before the identification of enrichment and burnup. Again, we emphasize that accuracy cannot be ensured for the extrapolation of the fitted binary linear forensics models.

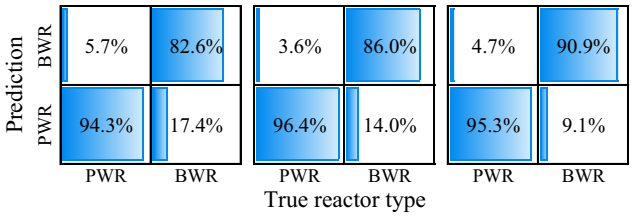

FIG. 7. Confusion matrices of the LGR (left), SVM (middle), and MLP (right) models for the classification of a PWR and a BWR.

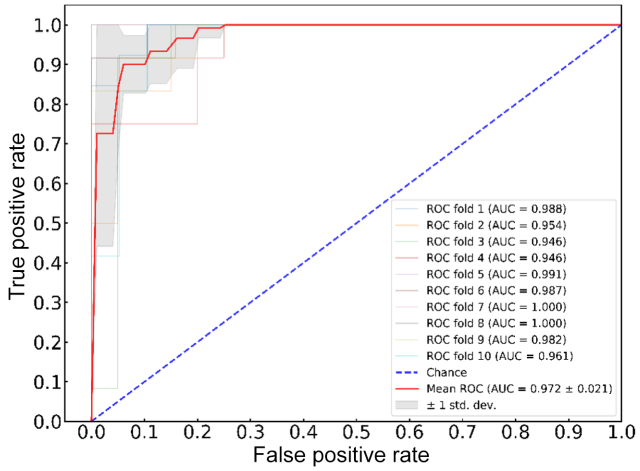

FIG. 8. ROC curves of the MLP model. Dashed blue line (i.e., chance) indicates the ROC line of the random classification.

The parameters quantifying the reliability of the forensic models for discriminating between a PWR and a BWR are given in Sec. III C. For the forensics of enrichment, the average deviation between model prediction and the actual values is about 0.1% ^{235}U, although the positive and negative deviations are compensated for numerically. Therefore, the predicted ^{235}U enrichment is assigned to be model calculation ±0.1% (1σ uncertainty). Similarly, the model uncertainty of burnup is 0.8 GWd/tU for the five reactor types, excluding the BWR. That of the BWR is 2.4 GWd/tU. The dispersion for a burnup level above 43 GWd/tU, shown in Fig. 5, is an important reason for the large deviation of the burnup model prediction. The dispersion can be induced by the deficiency of the binary linear model and/or the quality of measured data.

Table VIII summarizes the nine BWR samples with the largest deviations between model predictions and measured burnup. The measured ^{238}U concentration in the FDN-1|2F1ZN3|A9|UM sample is questionable because the 2.1% enrichment implies 97.9% initial mass concentration of ^{238}U, whereas the measured concentration at 64 GWd/tU is 98.6%. The intercomparison of MON-1|MTB099|A1 rod shows the contradiction between ^{242}Pu concentration and burnup: ^{242}Pu concentration is an increasing function of burnup for UO_2 fuel. However, S1 and S2 have higher ^{242}Pu concentrations than S3 and S4 (2.73 and 2.63), of which

TABLE VII. Indicators for the trained model of discriminating between a PWR and a BWR.

Method	Accuracy	Precision	Recall	F_1 score	AUC
LGR	0.898	0.908	0.827	0.860	0.959
SVM	0.924	0.945	0.860	0.893	0.976
MLP	0.937	0.936	0.910	0.917	0.972

TABLE VIII. Top-nine BWR samples with the largest deviations between model predictions and measured burnup (in GWd/tU).

SFCOMPO-2.0 sample[a]	e (%)	^{236}U[b]	^{238}U	^{242}Pu	Expt.	Model	Diff.
FDN-1|2F1ZN3|A9|UB	2.1	2.672	926.3	1.969	61.2	51.30	9.90
FDN-1|2F1ZN3|A9|UM	2.1	2.861	**986.1**	2.35	64.18	55.65	8.53
FDN-1|2F1ZN3|A9|UT	2.1	2.867	929.0	1.663	56.32	46.90	9.42
MON-1|MTB099|A1|S1	1.45	1.520	940.9	**2.798**	47.98	58.07	10.09
MON-1|MTB099|A1|S2	1.45	1.708	932.7	**2.923**	55.	61.96	6.96
MON-1|MTB099|B1|S1	1.87	**2.645**	944.3	2.184	43.8	53.91	10.11
MON-1|MTB099|G3|S3	2.87	**3.503**	932.3	**1.193**	50.00	42.29	7.71
MON-1|MTB099|H1|S1	1.87	2.420	937.5	**2.585**	50.98	60.01	9.03
MON-1|MTB099|H8|S1	2.14	2.687	937.1	**2.525**	49.40	60.61	11.21

[a]The name of SFCOMPO-2.0 sample is constituted from reactor|assembly|rod|sample.
[b]The isotopic concentrations (in mg/g) are normalized by the initial uranium mass. Analysis in the text focuses mainly on data given in bold.

the burnups (55.7 and 57.3 GWd/tU) are higher. Similar contradictions are found for the ^{236}U concentration of MON-1|MTB099|B1|S1 versus the other four MON-1|MTB099|B1 samples, ^{236}U and ^{242}Pu concentrations of MON-1|MTB099|G3|S3 versus S1, and the ^{242}Pu concentrations of MON-1|MTB099|H1|S1 versus S3 and S4 and MON-1|MTB099|H8|S1 versus S2–S4. Such contradictions need to be further investigated by both simulation interpretation of experiments and the remeasuring of isotopic concentrations and burnup.

To analyze the binary linear models in detail, the weighting factors of the two isotopic ratios are calculated in Tables IV and V for enrichment and burnup, respectively. For the enrichment models, the weighting factors of ^{235}U/^{238}U are smaller than those of ^{236}U/^{238}U. This implies that ^{236}U/^{238}U contributes more to the binary linearity of the enrichment forensics model. Similarly, ^{236}U/^{238}U contributes less to the linearity of the burnup model than ^{242}Pu/^{238}U (^{240}Pu/^{239}Pu for AGR and RBMK). For the AGR and RBMK, the weighting factors of ^{240}Pu/^{239}Pu are much higher than those of ^{242}Pu/^{238}U, especially for the RBMK. Least-squares linear fittings of burnup versus ^{242}Pu/^{238}U result in $R^2 = 0.955$ for the RBMK and 0.943 for the AGR. However, since the ^{236}U/^{238}U ratio is important for the forensics of enrichment, it is not necessary to exclude it from the burnup models (even for the RBMK). Moreover, because the correlations between two bases (i.e., isotopic ratios here) are neglected in the present analysis, the base with a smaller weighting factor has an additional implicit contribution to the linearity.

IV. VERIFICATION OF MODELS USING SIMULATION DATA

Because of the simplifications (especially constant power during the whole computational period) used in neutronic simulations, many more linear relationships can be obtained between the enrichment (or burnup) and isotopic concentration ratios. In addition, the fitting coefficients are certainly changed if simulation data are used to perform the binary linear fittings. Therefore, in this section, simulation data are used to verify the experimental-data-trained models determined in Sec. III rather than repeating the training of models for the PWR and the BWR.

A. Simulation verification for the fuel-enrichment model

Based on the isotopic concentrations from numerical simulations, fuel enrichments are calculated or predicted using the models and the corresponding parameters determined in Sec. III A. Figure 9 shows such predicted enrichments versus realistic values used in the OpenMC neutron-transport-depletion coupling simulations for both the PWR and BWR. The ^{235}U enrichment in UO_2 fuel can be well predicted by the presently determined models, which are trained using independent experimental data. Therefore, the models for predicting fuel enrichment using

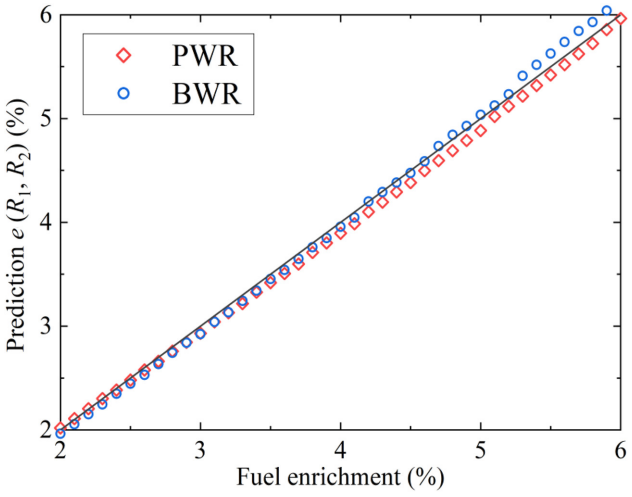

FIG. 9. Model-based predictions versus real values of fuel enrichment for both a PWR and a BWR using simulation data.

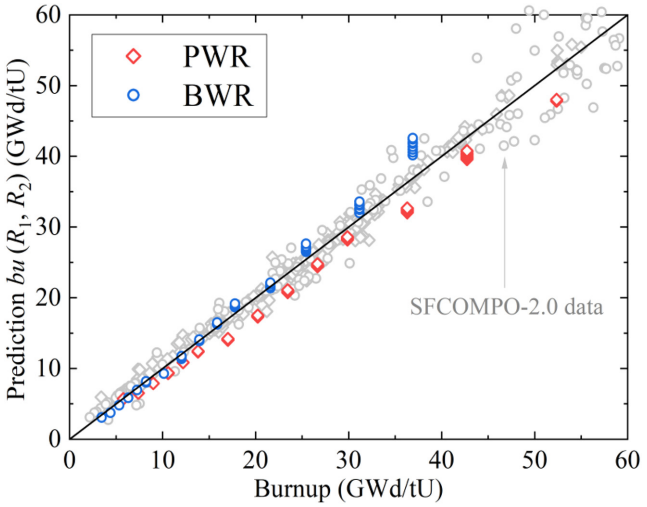

FIG. 10. Model-based predictions versus real values of burnup for both a PWR and a BWR using simulation data. SFCOMPO-2.0 data are included for comparison.

$^{235}U/^{238}U$ and $^{236}U/^{238}U$ as two fingerprints are verified for the PWR and the BWR.

B. Simulation verification for the burnup model

As explained in Sec. IV A, Fig. 10 illustrates the burnup predicted using the models determined in Sec. III B versus the burnup levels given by the OpenMC simulations. The predictions based on isotopic concentrations are close to those calculated by independent simulations. The deviations are generally within the same order of magnitude as those of SFCOMPO-2.0 data, as shown in gray in Fig. 10 for comparison. Consequently, the independent transport-depletion coupling simulations verify the burnup model and the corresponding parameters obtained in the present work within the uncertainties of the models.

V. CONCLUSIONS

The present work investigates models for the nuclear forensics of fuel enrichment, burnup, and reactor type, according to experimental data on isotopic concentrations acquired from the SFCOMPO-2.0 database. Binary linear models are determined for predicting ^{235}U enrichment in UO_2 fuel using $^{235}U/^{238}U$ and $^{236}U/^{238}U$ ratios in the irradiated fuel as two fingerprints. The binary linear expression is shown to be quite suitable for identifying enrichment for water-moderated reactors (PWR, BWR, and two VVER technologies) and graphite-moderated reactors (AGR and RBMK). For the identification of burnup, the combination ($^{236}U/^{238}U$, $^{242}Pu/^{238}U$) can be used as the fingerprints for fuel irradiated in the PWR, BWR, and two VVERs, while ($^{236}U/^{238}U$, $^{240}Pu/^{239}Pu$) is better for that in both the AGR and RBMK (graphite moderator). Both the enrichment and burnup models are verified using neutron-transport-depletion coupling modeling of a PWR and a BWR. ^{239}Pu, ^{240}Pu, and ^{242}Pu are the featured isotopes to discriminate between a PWR and a BWR through the three methods (i.e., LGR, SVM, and MLP). While LGR and SVM use $^{239}Pu/^{240}Pu$ and $^{240}Pu/^{242}Pu$ as fingerprints, MLP is optimized by using $^{239}Pu/^{240}Pu$ and $^{239}Pu/^{242}Pu$. In both trained SVM and MLP models, the 0.97 AUC values ensure their generalization ability. In summary, the models determined in the present work can reasonably predict the origin of spent fuel.

ACKNOWLEDGMENT

This work is supported by the National Key Research and Development Program of China under Grant No. 2018YFB1900405, the National Natural Science Foundation of China under Grant No. 12205390, the Guangdong Major Project of Basic and Applied Basic Research under Grant No. 2021B0301030006, the International Sci & Tech Cooperation Program of Guangdong Province No. 2019A050510022, the Fundamental Major Sci & Tech Special Project of Guangdong Province No. 2019B030302011, the Fundamental Research Funds for the Central Universities under Grant No. 22qntd1801, and computational resources from Sun Yat-sen University.

[1] R. Carr, F. Dalnoki-Veress, and A. Bernstein, Sensitivity of Seismically Cued Antineutrino Detectors to Nuclear Explosions, Phys. Rev. Appl. **10**, 024014 (2018).
[2] K. Hain, P. Steier, M. B. Froehlich, R. Golser, X. Hou, J. Lachner, T. Nomura, J. Qiao, F. Quinto, and A. Sakaguchi, $^{233}U/^{236}U$ Signature allows to distinguish environmental emissions of civil nuclear industry from weapons fallout, Nat. Commun. **11**, 1 (2020).
[3] T. Hopp, D. Zok, T. Kleine, and G. Steinhauser, Non-natural ruthenium isotope ratios of the undeclared 2017 atmospheric release consistent with civilian nuclear activities, Nat. Commun. **11**, 1 (2020).
[4] J. Petrović, A. Göök, and B. Cederwall, Rapid imaging of special nuclear materials for nuclear nonproliferation and terrorism prevention, Sci. Adv. **7**, eabg3032 (2021).
[5] International Atomic Energy Agency, IAEA Incident and Trafficking Database (ITDB) – 2022 Factsheet, 2022.
[6] K. Mayer, M. Wallenius, and I. Ray, Nuclear forensics—a methodology providing clues on the origin of illicitly trafficked nuclear materials, Analyst **130**, 433 (2005).
[7] International Atomic Energy Agency, *Nuclear Forensics in Support of Investigations* (IAEA, Vienna, 2015).
[8] M. J. Kristo, in *Handbook of Radioactivity Analysis*, 4th ed., edited by M. F. L'Annunziata (Elsevier, San Diego, CA, 2020), Vol. 2, pp. 921–951.
[9] G. Nicolaou, Determination of the origin of unknown irradiated nuclear fuel, J. Environ. Radioact. **86**, 313 (2006).
[10] M. Robel and M. J. Kristo, Discrimination of source reactor type by multivariate statistical analysis of uranium

and plutonium isotopic concentrations in unknown irradiated nuclear fuel material, J. Environ. Radioact. **99**, 1789 (2008).

[11] G. Nicolaou, Identification of unknown irradiated nuclear fuel through its fission product content, J. Radioanal. Nucl. Chem. **279**, 503 (2009).

[12] A. E. Jones, P. Turner, C. Zimmerman, and J. Y. Goulermas, Classification of spent reactor fuel for nuclear forensics, Anal. Chem. **86**, 5399 (2014).

[13] J. M. Osborn, E. D. Kitcher, J. D. Burns, C. M. Folden, and S. S. Chirayath, Nuclear forensics methodology for reactor-type attribution of chemically separated plutonium, Nucl. Technol. **201**, 1 (2018).

[14] J. Su, J. Wu, and S. Hu, Identification of unknown spent nuclear fuel with factor analysis for nuclear forensic purpose, Ann. Nucl. Energy **126**, 43 (2019).

[15] J. Su, J. Wu, and S. Hu, Signature selection for the identification of unknown spent nuclear fuel samples, Ann. Nucl. Energy **131**, 325 (2019).

[16] J.-H. Su, J. Wu, and S.-D. Hu, Identification of spent nuclear fuel with multivariate analysis based on database (in Chinese), Acta Phys. Sin. **68**, 090204 (2019).

[17] E. D. Kitcher, J. M. Osborn, and S. S. Chirayath, Characterization of plutonium for nuclear forensics using machine learning techniques, Ann. Nucl. Energy **170**, 108987 (2022).

[18] Z. Zhang, J. Yang, W. Qiu, J. Shi, C. Yuan, C. Zhang, J. Zhu, J. Su, S. Chen, and Y. Ge, Linear relationship among nonlinear transport-depletion problem: Forensics of enrichment and burnup of PWR, Ann. Nucl. Energy **173**, 109121 (2022).

[19] I. Lantzos, C. Kouvalaki, and G. Nicolaou, Plutonium fingerprinting in nuclear forensics of spent nuclear fuel, Prog. Nucl. Energy **85**, 333 (2015).

[20] J.-S. Kim, Analysis of the irradiated nuclear fuel using the heavy atom and neodymium isotope correlations with burnup, Nucl. Eng. Technol. **29**, 327 (1997).

[21] J. S. Kim, Y. S. Jeon, S. D. Park, Y.-K. Ha, and K. Song, Analysis of high burnup pressurized water reactor fuel using uranium, plutonium, neodymium, and cesium isotope correlations with burnup, Nucl. Eng. Technol. **47**, 924 (2015).

[22] F. Michel-Sendis, I. Gauld, J. S. Martinez, C. Alejano, M. Bossant, D. Boulanger, O. Cabellos, V. Chrapciak, J. Conde, I. Fast, *et al.*, SFCOMPO-2.0: An OECD NEA database of spent nuclear fuel isotopic assays, reactor design specifications, and operating data, Ann. Nucl. Energy **110**, 779 (2017).

[23] P. K. Romano, N. E. Horelik, B. R. Herman, A. G. Nelson, B. Forget, and K. Smith, OpenMC: A state-of-the-art Monte Carlo code for research and development, Ann. Nucl. Energy **82**, 90 (2015).

[24] F. N. von Hippel, Plutonium and reprocessing of spent nuclear fuel, Science **293**, 2397 (2001).

[25] S. Chen and C. Yuan, Transmutation study of minor actinides in mixed oxide fueled typical pressurized water reactor assembly, J. Nucl. Eng. Radiat. Sci. **4**, 041017 (2018).

[26] M. Åberg Lindell, P. Andersson, S. Grape, C. Hellesen, A. Håkansson, and M. Thulin, Discrimination of irradiated MOX fuel from UOX fuel by multivariate statistical analysis of simulated activities of gamma-emitting isotopes, Nucl. Instrum. Methods Phys. Res., Sect. A **885**, 67 (2018).

[27] S. W. Finch, M. Bhike, C. R. Howell, Krishichayan, W. Tornow, A. P. Tonchev, and J. B. Wilhelmy, Measurements of Short-Lived Isomers from Photofission as a Method of Active Interrogation for Special Nuclear Materials, Phys. Rev. Appl. **15**, 034037 (2021).

[28] It is shown in Sec. III A that the binary linear combination of these two ratios can describe fuel enrichment well.

[29] A. Smola and S. V. N. Vishwanathan, *Introduction to Machine Learning* (Cambridge University Press, United Kingdom, Cambridge, 2008).

[30] Z. Zhou, *Machine Learning (in Chinese)* (Tsinghua University Press, 2016).

[31] It should be noted that the scalar $\theta_0 x_{i,0}$ is a bias. It is added to get a better prediction model and is denoted in this form to simplify the notation.

[32] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. (Prentice Hall, Upper Saddle River, N.J, 1998).

[33] S. Chen, C. Yuan, and D. Guo, Radial distributions of power and isotopic concentrations in candidate accident tolerant fuel U_3Si_2 and UO_2/U_3Si_2 fuel pins with FeCrAl cladding, Ann. Nucl. Energy **124**, 460 (2019).

[34] M. B. Chadwick, M. Herman, P. Obložinský, M. E. Dunn, Y. Danon, A. C. Kahler, D. L. Smith, B. Pritychenko, G. Arbanas, R. Arcilla, *et al.*, ENDF/B-VII.1 nuclear data for science and technology: Cross sections, covariances, fission product yields and decay data, Nucl. Data Sheets **112**, 2887 (2011).

[35] S. Chen, E. Vandermeersch, P. Tamagno, D. Bernard, G. Noguere, and P. Blaise, Uncertainty propagation from n $+^{56}$ Fe nuclear reaction model parameters to neutron multiplication factor, Ann. Nucl. Energy **163**, 108553 (2021).

[36] It is noted that several tens of combinations lead to $R^2 \geq 0.97$ for the two VVER types. One of the reasons is the relatively small quantity of available experimental data in the SFCOMPO-2.0 database.

[37] P. Jaffke, B. Byerly, J. Doyle, A. Hayes, G. Jungman, S. Myers, A. Olson, D. Porterfield, and L. Tandon, Developing Diagnostic Tools for Low-Burnup Reactor Samples, Phys. Rev. Appl. **8**, 044025 (2017).