NUCLEAR SCIENCE AND ENGINEERING

DOI: https://doi.org/10.1080/00295639.2020.1822661

A Preliminary Study on the Use of the Linear Regression Method for Multigroup Cross-Section Interpretation

Cihang Lu and Zeyun Wu*

Virginia Commonwealth University, Department of Mechanical and Nuclear Engineering, Richmond, Virginia 23219

Received May 26, 2020
Accepted for Publication September 9, 2020

Abstract — *Computational modeling and simulations are widely used for evaluation of the performance and safety features of innovative nuclear reactor designs. Multigroup-based deterministic neutronics codes are often employed in these reactor design calculations because they can provide fast predictions of the neutron flux distribution and other neutronics characteristic parameters. Nevertheless, providing accurate multigroup cross sections for deterministic codes is an onerous job, which makes establishing an exhaustive cross-section library computationally prohibitive. Partly because of these reasons, multigroup neutron cross sections are normally stored only at certainty state points in the data library of these deterministic codes, and linear interpolation methodology is commonly utilized to estimate the cross sections at unknown states. However, the applicability of linear interpolation is limited, and the precision of its results is moderate.*

In this paper, we discuss a preliminary feasibility study that we performed on providing more precise multigroup cross sections for deterministic neutronics codes by using the linear regression methodology. Compared to the traditional linear interpolation method, the linear regression approach principally showed improved computational efficiency considering the use of more data in the cross-section library, and constructed hypothesis functions for the responses of interest with a higher order of accuracy. In this study, a case study on Lightbridge Corporation's metallic fuel element was carried out to demonstrate the feasibility and advantages of linear regression in multigroup cross-section interpretation. A reference cross-section library was established through calculations conducted with the Monte Carlo neutronic code Serpent. Because of the preliminary nature of this feasibility study, only the macroscopic total cross section is considered. Linear interpolation and linear regression were both used to estimate cross sections at unknown states based on the data available in the library. By comparing the performance of both methodologies, we demonstrated that the linear regression methodology achieved wider applicability and better precision in cross-section interpretation. Moreover, the linear regression process was finished within 15 s using a single processor core, which indicated that the additional computational burden brought by the implementation of linear regression methodology in the task was acceptable.

Keywords — *Reactor physics, cross-section interpretation, Monte Carlo, linear regression method.*

Note — *Some figures may be in color only in the electronic version.*

I. INTRODUCTION

Through the development of nuclear technology, a large number of designs for advanced and innovative nuclear reactors are being studied. Nowadays, computational modeling and simulation become an indispensable technique for evaluation of the performance and analysis of the safety features of these designs. Continuous-energy-mode Monte Carlo neutron transport codes are widely used for steady-state reactor analysis[1–3] but are rarely employed for transient analysis because of the large amount of computational time required. On the other hand, multigroup-based deterministic

*E-mail: zwu@vcu.edu

codes are more suitable for transient calculations[4,5] in nuclear reactor analysis. In the course of transient analysis, deterministic codes calculate the time-dependent neutron flux distribution in the geometries of interest by using either multigroup neutron diffusion or transport equations. Therefore, the accuracy of the multigroup cross-section data required to solve these equations becomes vital to the success of these calculations.

These group cross sections can be established either with deterministic lattice physics codes using a discrete energy group structure, such as CASMO (Ref. 6) and WIMS (Ref. 7), or with Monte Carlo codes using continuous-neutron-energy structures, such as MCNP (Ref. 8) and Serpent.[9] Between these two approaches, the former one (i.e., the lattice code approach) may be more commonly seen in current academia and industrial practices, particularly for light water reactor analysis.

Generating efficient and accurate multigroup cross sections for deterministic codes can be an intimidating job in reactor analysis. In the lattice code approach, the difficulties of cross-section generation stem from manifold aspects, yet the most prominent challenge is due to energy and spatial self-shielding effects on the neutron flux.[10] Energy self-shielding is primarily caused by neutron interactions with nuclei in the resonance energy range while spatial self-shielding is the result of material heterogeneity and geometric complexities in various reactor designs. The self-shielding effect poses significant challenges for the prediction of the flux shape because it essentially depresses the flux distribution in phase space and increases the complexity of the flux characteristics. The complexity of the problem is amplified when the two self-shielding effects are coupled and intermixed among neighboring energies and locations. Such complexities are often found exaggerated in advanced reactor or small modular reactor concepts because these new reactor designs are usually constructed with highly exotic geometric and material configurations. Examples of such advanced designs include Lightbridge fuel with "clover" designs,[11] transformational challenge reactor (TCR) fuel with square fuel segments with round flow channels,[12] and so on.

In addition to geometric issues, many advanced designs have energy spectra in epithermal and fast ranges, which further amplifies the complexity of the energy self-shielding effect due to a higher degree of anisotropic and inelastic resonance scattering in these reactors. In short, the difficulty of predicting accurate energy spectra of resonances in arbitrary geometry makes the cross-section generation an extremely unwieldy and onerous task in reactor physics calculations with deterministic methodologies.

On the other hand, the Monte Carlo approach can provide the most accurate solution for cross-section generation because it works with continuous-energy cross sections and does not require the use of any local approximations to the flux. In principle, the Monte Carlo method can handle all the difficulties mentioned above. Furthermore, Monte Carlo methods can naturally address many additional important challenges in cross-section generation, including resonance interference in unresolved resonance ranges in fast reactors, upscattering effects in resolved resonance ranges in thermal reactors, anisotropic scattering effects, etc. These challenges may not be directly related to self-shielding effects but are also very difficult to be handled with a deterministic lattice code approach.

The Monte Carlo approach has been recently increasingly employed for cross-section generation. The Serpent code[9] has capabilities to generate coarse-mesh group constants for multigroup diffusion solvers, and the OpenMC code[13] has been recently extended to generate a fine-spatial-mesh multigroup cross section for high-fidelity whole-core transport solvers. However, the accuracy of the Monte Carlo approach comes at the computational expense of converging group constant tallies to acceptably low uncertainties. Therefore, the Monte Carlo approach is generally considered computationally inefficient for design work involving either large reactors and/or thousands of repeated calculations, and thus, it has not been widely adopted as a common multigroup cross-section generation tool either.

Because of all the difficulties addressed above, the multigroup cross-section data library established by these approaches (either lattice code or Monte Carlo code approach) is limited. The nuclear data imported from the library are only for certain conditions of different combinations of stepwise state variables. Some of these variables are used to describe the depletion history, including the burnup (BU), the fuel and moderator temperatures at which the depletion calculation was performed, etc. Other variables are used to describe the actual reactor operation conditions, including the fuel temperature, the moderator temperature, the moderator boron concentration, the moderator density, etc. The deterministic codes use different methods to estimate the cross sections when the condition desired for the calculation is not available in the library. Linear interpolation techniques, with subtle implementation differences, are commonly used in most of the well-known deterministic core physics codes for cross-section interpretation, including many nodal diffusion codes such as PARCS by the University of Michigan,[14,15] DIF3D-K by Argonne National Laboratory,[16] SIMULATE by Studsvik,[17] and DYN3D by the German Institute of Safety Research,[18] as well as some neutron transport codes such as DRAGON by École Polytechnique de Montréal[19] and

core physics modules in the SCALE code system by Oak Ridge National Laboratory.[20]

Linear interpolation consists of fitting linear polynomials within the range of discrete sets of known data for the estimation of unknown data. Thanks to its simplicity, the implementation of this method into deterministic codes is straightforward, and the computational burden brought is small. However, linear interpolation can be used only when data are available on both sides of the unknown data of interest, and its precision is low under certain circumstances. Linear regression, on the other hand, seeks to construct higher-order polynomials based on all the known data. It therefore has wider applicability and can provide higher accuracy than the classical linear interpolation. In this paper, we will discuss these two advantages of using linear regression for the interpretation of multigroup cross sections for deterministic codes, together with the associated computational expanse.

A case study on Lightbridge's metallic fuel element is carried out through the paper to demonstrate the feasibility and advantages of the linear regression method for multigroup cross-section interpretation. It is pointed out that a good discussion on the use of the regression method for cross-section interpretation has been given in the work of Zimin and Semenov.[21] The two main novelties of the work presented in this paper include the following. (1) The work of Zimin and Semenov looked into replacing the neutron cross-section libraries by cross-section functions of several variables, which led to modifications of the library of a deterministic code before its employment. In this study, we looked into a method to estimate the neutron cross section by using existing libraries, which can be applied to any existing deterministic code, with different case-specified libraries, without modifying the library of the code. (2) The work of Zimin and Semenov first treated the dependency of the cross section on BU with a cubic spline interpolation and then the dependencies on other state variables with a multidimensional polynomial. In this study, we considered all the state variables together, which made the process more straightforward while keeping the accuracy of the cross-section prediction.

The rest of the paper is organized as follows. In Sec. II, the fuel element model used as an example for the case study is detailed. In Sec. III, the establishment of the neutron cross-section library used in the current study is discussed. In Sec. IV, both methodologies used for the cross-section interpretation, including the linear interpolation method and the linear regression method, are introduced, and the limitations of the linear interpolation method are emphasized. In Sec. V, comparisons of the cross section estimated for the case problem through both methodologies are presented and discussed. The advantages of the use of linear regression is highlighted. The additional computational burden introduced by the use of linear regression is also discussed. In Sec. VI, conclusions of the current study are summarized, and some foreseeable future endeavors on this subject are offered.

II. CASE STUDY MODEL DESCRIPTION

The current study is focused on a fuel rod of an innovative fuel design (IFD) based on Lightbridge's metallic fuel design.[11] The IFD fuel rod consists of four lobes and has a unique cruciform shape. This special fuel design was selected in the study to demonstrate the wide applicability of the linear regression method (discussed in Sec. IV) for cross-section interpretation. The multipurpose three-dimensional continuous-energy Monte Carlo particle transport code Serpent[9] was used in the current study for the generation of the neutron cross-section library. A top view of the Serpent model of an IFD fuel rod unit cell is shown in Fig. 1. This IFD has a circumscribed diameter similar to the pitch of conventional UO_2 fuel rods and was designed with the goal to perform a one-to-one replacement with conventional UO_2 fuel rods in operating light water reactors. The special cruciform shape of the IFD has the potential to enhance the coolant-fuel heat transfer as well as mitigate the fuel vibration during reactor operation.

As shown in Fig. 1, in the current study, the central displacer was modeled as pure zirconium. The metallic fuel was modeled as a 50 wt% alloy of zirconium and uranium with a ^{235}U enrichment of 19.7 wt%. The cladding was modeled as Zircaloy-4 (Ref. 22), consisting of 1.2 wt% of tin, 0.18 wt% of iron, 0.07 wt% of chromium, and 98.55 wt% of zirconium. The moderator (coolant) surrounding the fuel rod was modeled as pure water. The actual dimensions of the IFD individual fuel rods are proprietary information. Therefore, we adopted the same values as those described in a recent publication on the Lightbridge fuel study.[23]

Periodic boundary conditions were applied to all x, y, and z directions during the Serpent calculations. For every calculation performed, a 1 500 000 neutron population, 160 active cycles, and 60 inactive cycles were used such that the resultant uncertainty was smaller than 10 pcm (per cent mille) for the calculated effective neutron multiplication factor k_{eff} and smaller than 0.5% for the calculated cross sections. Because of the consideration of an infinite medium and the ignorance of soluble ^{10}B as well as other burnable neutron poisons, the initial k_{eff} of the fuel rod was calculated to be around 1.540. The fact that the initial k_{eff} is comparable with that calculated

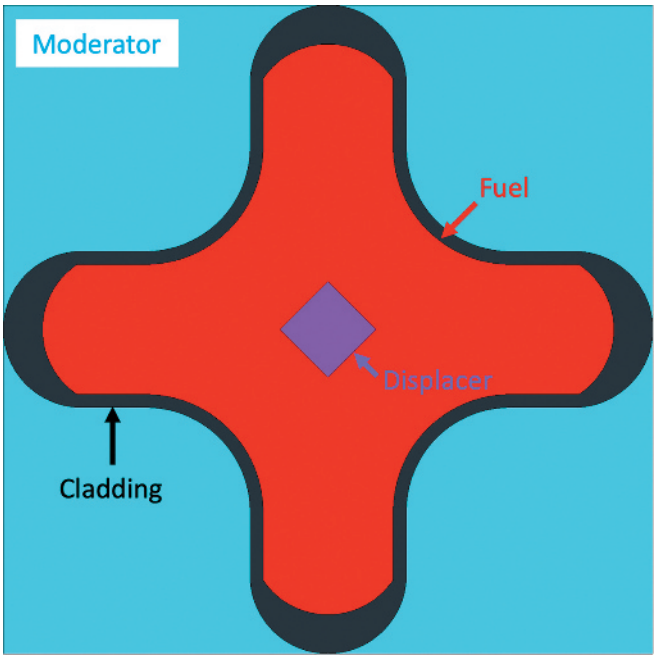

Fig. 1. Serpent model of the Lightbridge IFD fuel rod unit cell.

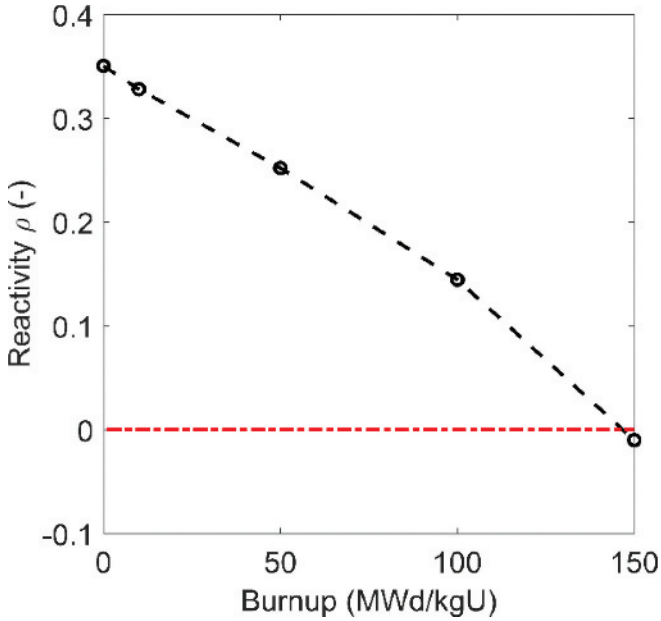

Fig. 2. Reactivity as a function of the BU of the IFD fuel rod.

in an existing reference[24] demonstrated our model to be reasonable. Through a depletion calculation performed with a constant fuel temperature T_{fuel} of 900 K and a moderator temperature T_{mod} of 600 K, we found that the BU of the fuel rod modeled could reach about 150 MWd/kgU, as shown in Fig. 2. The densities of nuclides in the fuel material at five specified BU steps, namely, 0, 10, 50, 100, and 150 MWd/kgU, were calculated by Serpent and stored such that branch calculations could be performed without repeating the depletion process.

III. DATA LIBRARY ESTABLISHMENT

Different neutron cross sections, including fission cross section, capture cross section, scattering cross section, etc., are used in the deterministic codes for the neutron flux calculation. In the current study, only the total cross section was generated and stored in the data library for the demonstration of the capability of linear regression. The total cross section is the sum of many other cross sections and thus considered to be adequate in this preliminary study as a rule of principle to justify the viability of the presented methodology.

The total cross section is dependent on a large number of state variables in real life, including T_{fuel}, T_{mod}, BU, control rod insertion, boron concentration, etc. In the current study, for simplicity, we investigated the impact of only T_{fuel}, T_{mod}, and BU on the neutron cross section. Similar to the BU steps, five steps were selected for T_{fuel} and T_{mod}, respectively, while the water density was calculated by considering T_{mod} as the boiling temperature. The steps of the three variables considered are summarized in Table I. The macroscopic neutron total cross sections calculated in the fuel region in 125 conditions, consisting of different combinations of BU, T_{fuel}, and T_{mod}, were used as the reference library in the current study.

TABLE I

Steps of the Three Variables Considered

State Variables Considered	BU(MWd/kgU)	T_{fuel}(K)	T_{mod}(K)
Different steps considered	0 10 50 100 150	600 750 900 1050 1200	300 400 500 600 625

Based on the fuel model defined, Serpent calculates the critical neutron flux spectra by solving the B1 equations iteratively.[25] The resultant flux spectra are then used to calculate the group cross section by conserving the physical reaction rates because only the neutron cross-section data in the form of discrete neutron energy group structures can be used by multigroup method–based deterministic codes. We employed the widely used SCALE 238-group structure in the current study. The groupwise neutron total cross sections calculated in the fuel region, with T_{fuel} of 750 K and T_{mod} of 400 K, at different BU steps are compared in Fig. 3. Generally, the cross section decreases with BU in the thermal region and increases with BU in the fast region. However, its behavior becomes harder to be predicted in the resonance region.

The percentage difference of the cross section obtained at different BUs with respect to that obtained at 0 MWd/kgU using T_{fuel} of 750 K and T_{mod} of 400 K is shown in Fig. 4. The percentage difference of the cross section obtained at different values of T_{fuel} with respect to that obtain at 600 K using BU of 100 MWd/kgU and T_{mod} of 400 K is shown in Fig. 5. The percentage difference of the cross section obtained at different values of T_{mod} with respect to that obtained at 300 K using BU of 100 MWd/kgU and T_{fuel} of 750 K is shown in Fig. 6.

By comparing Figs. 4, 5, and 6, it could be concluded that BU had a significantly larger impact on the cross section than the other two variables within the range of consideration. Figures 4, 5, and 6 can also provide many other insights about the total cross-section changes with the variables of interest, which are BU, T_{fuel}, and T_{mod} for this specific example. These insights can be effectively captured by a more advanced data interpretation method for the realization of cross sections for unknown states. This simple observation laid the theoretical foundation of this study, in which we use a standard linear regression approach to fulfill the goal of providing more precise cross sections for the downstream deterministic core physics code.

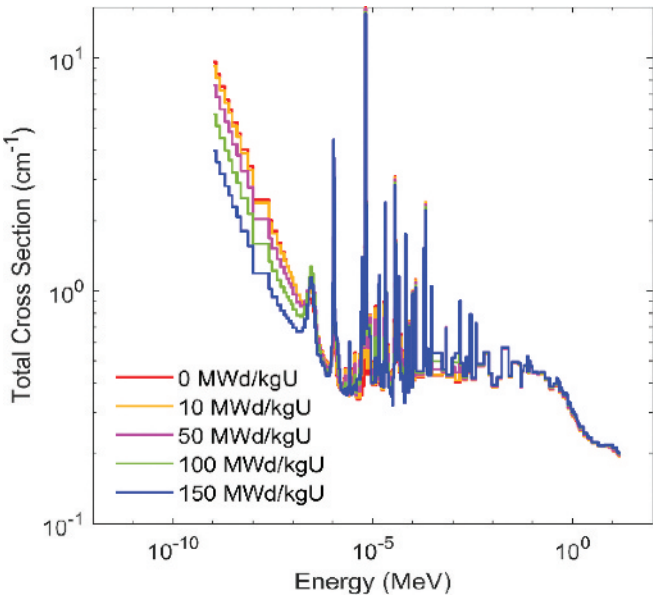

Fig. 3. Groupwise neutron total cross sections at different BU steps.

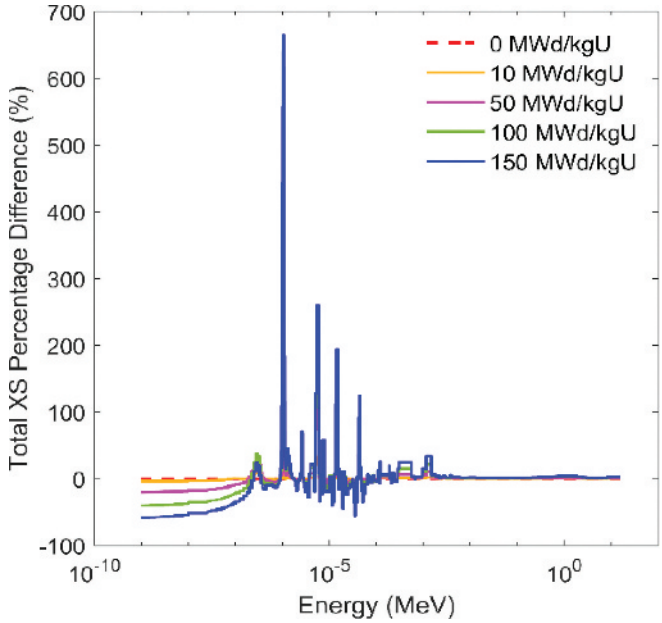

Fig. 4. Percentage difference of cross sections at different BU steps.

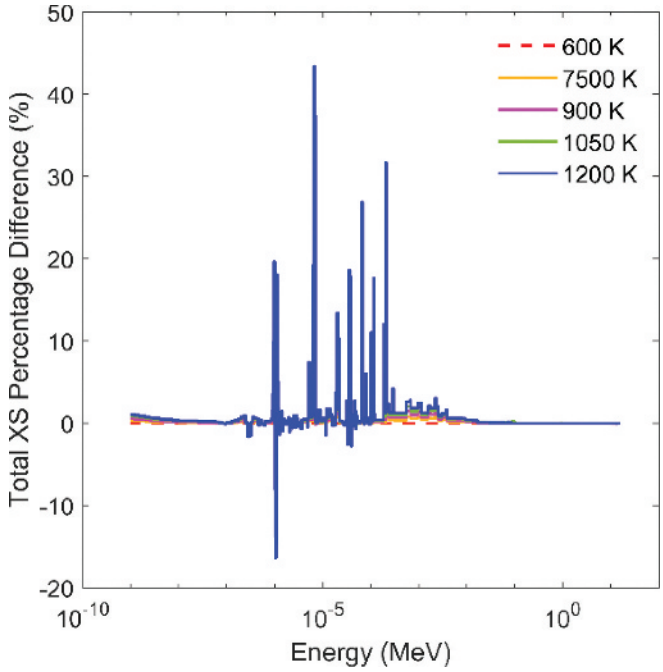

Fig. 5. Percentage difference of cross sections at different T_{fuel} steps.

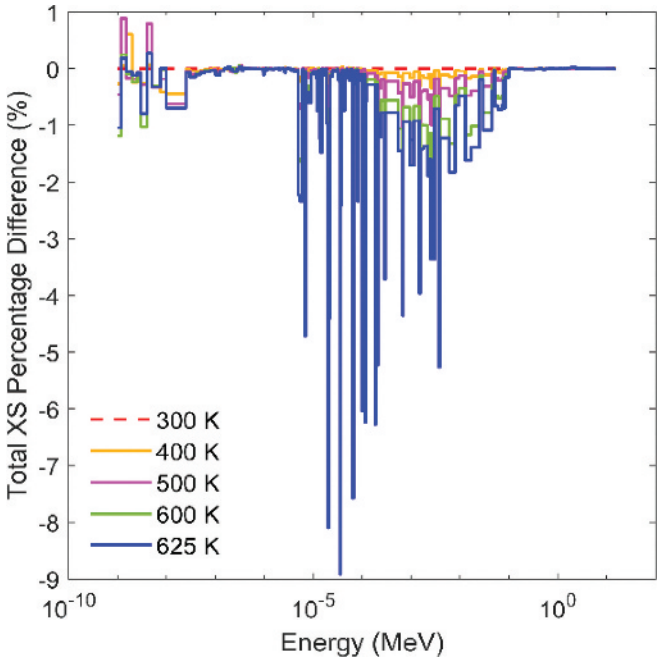

Fig. 6. Percentage difference of cross sections at different T_{mod} steps.

IV. METHODOLOGY

In this paper, the linear regression models[26,27] are developed as the engines for multigroup cross-section interpretation. An overview description as well as many numerical implementation caveats of this linear regression method is discussed in this section. For contrast, the linear interpolation method, which is commonly used for the unknown state cross sections in deterministic core physics codes, is also briefly discussed later in this section. Since the linear interpolation method is really widely understood, the discussion presented here does not focus on the procedure of this method but rather on its differences to the linear regression method.

IV.A. Linear Regression Method

Linear regression is the process to determine the parameter θ of the hypothesis function $h(x)$ such that the cost function $J(\theta)$ is minimized. For simplicity, a univariate example, where the value to be predicted y is assumed dependent on only one variable x, is given for the explanation of this methodology. When the univariate problem is considered, the hypothesis function is the goal polynomial, which is defined as

$$h(\boldsymbol{x}) = \boldsymbol{\theta}^{\boldsymbol{T}} \cdot \boldsymbol{x} \ , \tag{1}$$

where

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix}$$

and

$$\boldsymbol{x} = \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^n \end{pmatrix} \tag{2}$$

are vectors with $n+1$ terms and the hypothesis function is of order n. The least-squares cost function is defined as

$$J(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{m}\left[h\left(\boldsymbol{x}^{(i)}\right) - y^{(i)}\right]^2 , \tag{3}$$

where $\boldsymbol{x}^{(i)}$ and $y^{(i)}$ are the m pairs of variables and observations used for the training process. When the value to be predicted is considered dependent on multiple variables, the sizes of both $\boldsymbol{\theta}$ and $\boldsymbol{x}$ will increase exponentially.

In this study, the neutron cross section is the predicted variable, and we assumed that the cross section is dependent on three system variables: T_{fuel}, T_{mod}, and BU. The variables to be considered in the hypothesis functions of different orders are constructed and summarized in Table II. The list of the variables and the complexity of the hypothesis functions can be adjusted based on the effectiveness of the applications.

A phenomenon that should be avoided when applying linear regression is called overfitting, in which the cost function is minimized whereas the resultant hypothesis function has poor capability for the prediction of the data sets that are not used during the training process. A simple univariate example is illustrated below for a more concrete description of overfitting and for a better explanation of the ways to avoid this phenomenon.

Nine data points were selected randomly from a linear polynomial, $y = -5 + 4x$, within the range of $x \in [1, 6]$. Minor disturbances were added to these data points to mimic the uncertainties associated with the experimental measurement or numerical simulation. These nine data points, as shown in Fig. 7, were then used for the linear regression training such that the resultant hypothesis function could predict other data points

TABLE II

Variables Considered in Hypothesis Functions of Different Orders

Order of Hypothesis Function	0	1	2	3	4
Additional variables	1	BU T_{fuel} T_{mod}	BU^2 T_{fuel}^2 T_{mod}^2 $BU \cdot T_{fuel}$ $BU \cdot T_{mod}$ $T_{fuel} \cdot T_{mod}$	BU^3 T_{fuel}^3 T_{mod}^3 $BU^2 \cdot T_{fuel}$ $BU^2 \cdot T_{mod}$ $T_{fuel}^2 \cdot BU$ $T_{fuel}^2 \cdot T_{mod}$ $T_{mod}^2 \cdot BU$ $T_{mod}^2 \cdot T_{fuel}$ $BU \cdot T_{fuel} \cdot T_{mod}$	BU^4 T_{fuel}^4 T_{mod}^4 $BU^3 \cdot T_{fuel}$ $BU^3 \cdot T_{mod}$ $T_{fuel}^3 \cdot BU$ $T_{fuel}^3 \cdot T_{mod}$ $T_{mod}^3 \cdot BU$ $T_{mod}^3 \cdot T_{fuel}$ $BU^2 \cdot T_{fuel}^2$ $BU^2 \cdot T_{mod}^2$ $T_{fuel}^2 \cdot T_{mod}^2$ $BU^2 \cdot T_{fuel} \cdot T_{mod}$ $T_{fuel}^2 \cdot BU \cdot T_{mod}$ $T_{mod}^2 \cdot BU \cdot T_{fuel}$

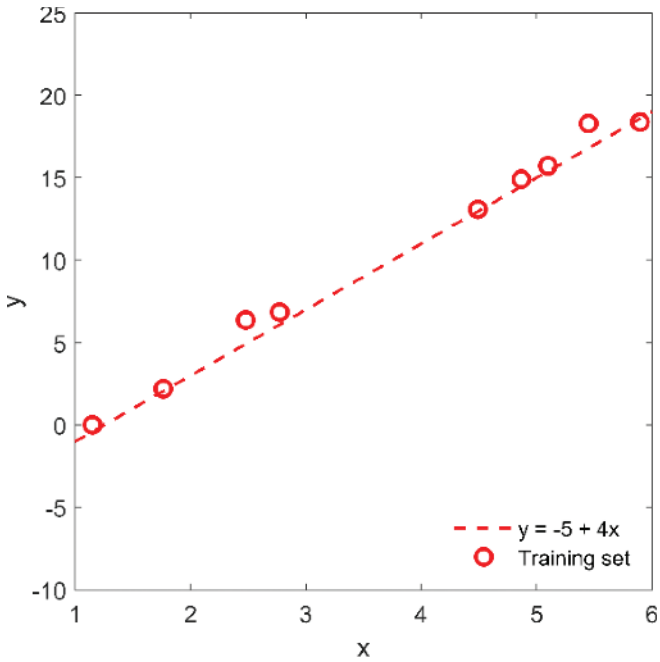

Fig. 7. The nine data points used as the training set.

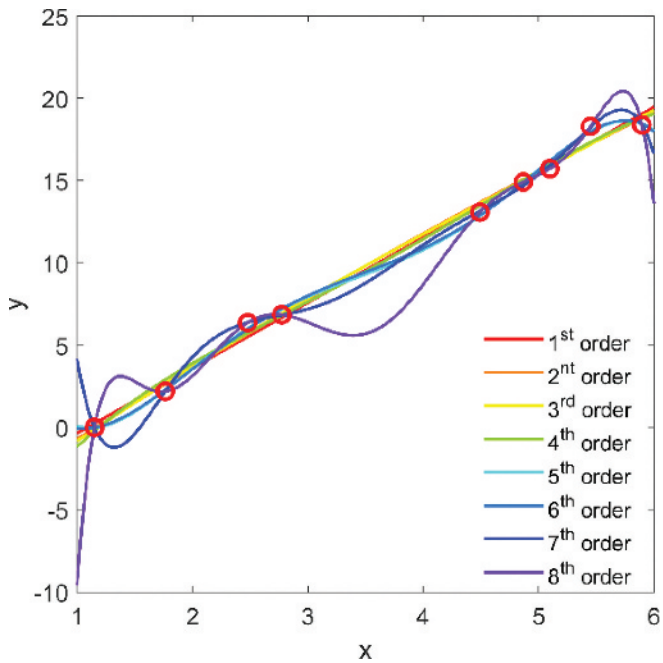

Fig. 8. Comparison of the hypothesis functions of up to eighth order for linear regression.

on the linear polynomial. Linear regression was performed using hypothesis functions with an order up to eight, and the predictions of the eight hypothesis functions are compared in Fig. 8. The value of the cost function decreased when the order of the hypothesis function increased, as shown Fig. 9, and became zero when the hypothesis function of order eight was used. This is because an eighth-order polynomial could be determined when nine points that it passes through were given. Both seventh- and eighth-order hypothesis functions had small values of their cost function but largely deviated from $y = -5 + 4x$ and will therefore give poor predictions for other data points on this original linear

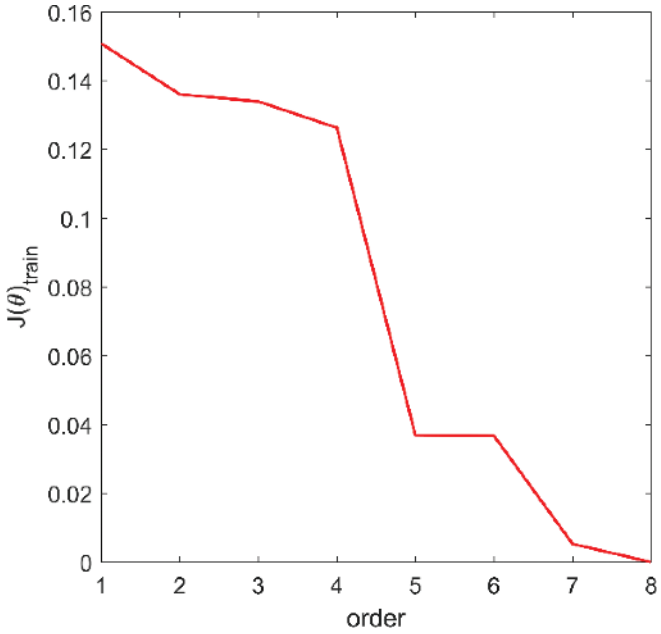

Fig. 9. Values of the cost function for hypothesis function of different orders.

polynomial that were not used for the training process. This phenomenon is what people refer to as overfitting.

The parameters determined by linear regression for the hypothesis functions are summarized in Table III. A common sign of overfitting is that some of the parameters could have an absolute value that is orders of magnitude larger than that of the mean-squared error (MSE) of the predictions.

Overfitting can be generally avoided by using one of the two methods described below. The first one used a modified linear regression methodology such as Ridge regression,[28] Least Absolute Shrinkage and Selection Operator (LASSO) regression,[29] or Elastic Net regression[30] to avoid overfitting. Regularization terms were added into the cost functions of these three alternative regression methodologies to eliminate parameters that are too large. An L2 regularization term was used in Ridge regression, and its cost function is expressed as

$$J_{Ridge}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{m}\left[h\left(\boldsymbol{x}^{(i)}\right) - y^{(i)}\right]^2 + \lambda\sum_{j=1}^{n}\theta_j^2 \ . \tag{4}$$

An L1 regularization term was used in LASSO regression, and its cost function is expressed as

$$J_{\mathrm{LASSO}}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{m}\left[h\left(\boldsymbol{x}^{(i)}\right) - y^{(i)}\right]^2 + \lambda\sum_{j=1}^{n}\left|\theta_j\right| \ . \tag{5}$$

Elastic Net regression can be considered as a linear combination of these two regressions, the cost function of which is expressed as

$$J_{ElasticNet}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{m}\left[h\left(\boldsymbol{x}^{(i)}\right) - y^{(i)}\right]^2 + \lambda\left[\rho\cdot\sum_{j=1}^{n}\left|\theta_j\right| + (1-\rho)\cdot\sum_{j=1}^{n}\theta_j^2\right] \ , \tag{6}$$

where $\lambda > 0$ and $\rho \in [0,\ 1]$ are the hyperparameters that users need to determine for the best performance of these regression methodologies. Ridge regression, LASSO regression, and Elastic Net regression were applied on the data set shown in Fig. 7 using hypothesis functions of order up to eight. The parameters of these three regression methodologies are summarized in Tables IV, V, and VI, and their performance is shown in Figs. 10, 11, and 12. Large parameters were avoided, and hypothesis functions, even of a seventh or eighth order, did not largely deviate from the original linear polynomial.

The second method to avoid overfitting is the use of a subset of data for the validation process. Assuming that

TABLE III

Parameters of the Hypothesis Functions of Different Orders for Linear Regression

Order of $h(x)$	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
1	−4	4							
2	−5	5	−0.1						
3	−6	5	−0.3	−0.02					
4	−10	12	−4	0.7	−0.05				
5	20	−46	37	−13	2	−0.1			
6	13	−30	23	−6	0.4	0.07	−0.01		
7	312	−820	856	−465	144	−26	2	−0.1	
8	−1192	3652	−4649	3217	−1326	334	−51	4	−0.1

TABLE IV

Parameters of the Hypothesis Functions of Different Orders for Ridge Regression

Order of $h(x)$	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
1	−4.30	3.96							
2	−5.21	4.61	−0.09						
3	−2.63	2.03	0.65	−0.07					
4	−0.11	0.29	0.70	0.04	−0.01				
5	0.04	0.13	0.26	0.32	−0.07	0.004			
6	0.82	−2.75	−1.02	4.19	−1.92	0.34	−0.02		
7	−0.54	−0.34	0.04	0.46	0.50	−0.34	0.07	−0.004	
8	−0.03	0.00	0.04	0.11	0.16	0.11	−0.09	0.02	−0.001

TABLE V

Parameters of the Hypothesis Functions of Different Orders for LASSO Regression

Order of $h(x)$	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
1	−4.30	3.96							
2	−5.15	4.58	−0.09						
3	0.00	0.00	1.08	−0.09					
4	−4.71	4.15	0.03	−0.01	−1E-4				
5	−4.57	4.03	0.06	−0.01	−3E-4	2E-5			
6	−1.14	1.16	0.65	−0.02	−0.004	−1E-4	−2E-6		
7	0.00	0.00	0.82	0.01	−0.007	−0.001	1E-5	1.E-05	
8	0.00	0.00	0.24	0.20	−0.01	−0.004	−1E-4	1.E-05	4.E-06

TABLE VI

Parameters of the Hypothesis Functions of Different Orders for Elastic Net Regression

Order of $h(x)$	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
1	−4.30	3.96							
2	−5.21	4.62	−0.09						
3	0.00	0.00	1.07	−0.09					
4	0.00	0.00	0.00	0.32	−0.04				
5	−4.86	4.28	0.01	−0.01	−5E-6	4E-5			
6	−2.76	2.47	0.41	−0.02	−0.002	9E-6	−6E-6		
7	0.00	0.00	0.93	−0.01	−0.01	−4E-4	3E-5	7.E-06	
8	0.00	0.00	0.05	0.28	−0.01	−0.005	−1E-4	2.E-05	5.E-06

besides the nine data points from the training data set, shown by the circles in Fig. 13, we have three more randomly selected data points, shown by the pentagrams, which serve for the validation process. The MSE for the validation data set as a function of the order of the hypothesis functions is plotted in Fig. 14. By combining Fig. 9 and Fig. 14, it can be concluded that the hypothesis of order five may have the best capability for the prediction of the unknown data on the original linear polynomial because it has both a low value of cost function and the lowest MSE for the validation data set.

In this study, we adopted the second method to avoid the overfitting phenomena in the cross-section interpretation problem. This is not only because the second method

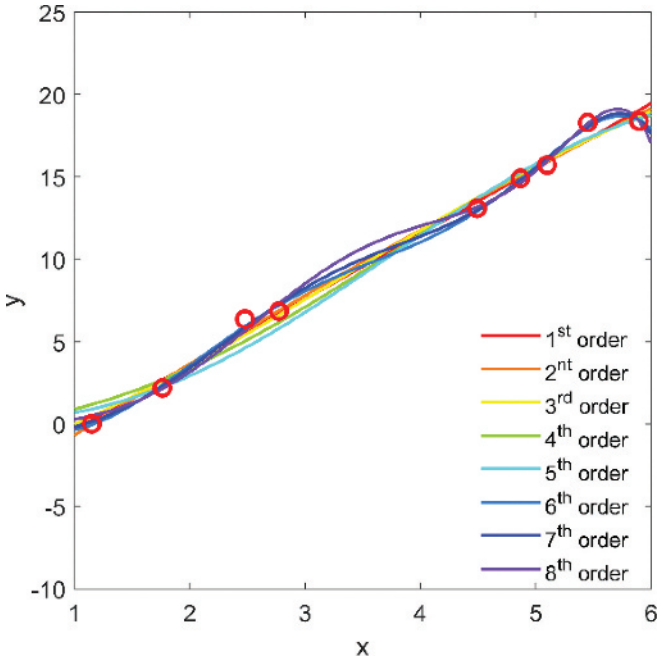

Fig. 10. Comparison of the hypothesis functions of up to eighth order for Ridge regression.

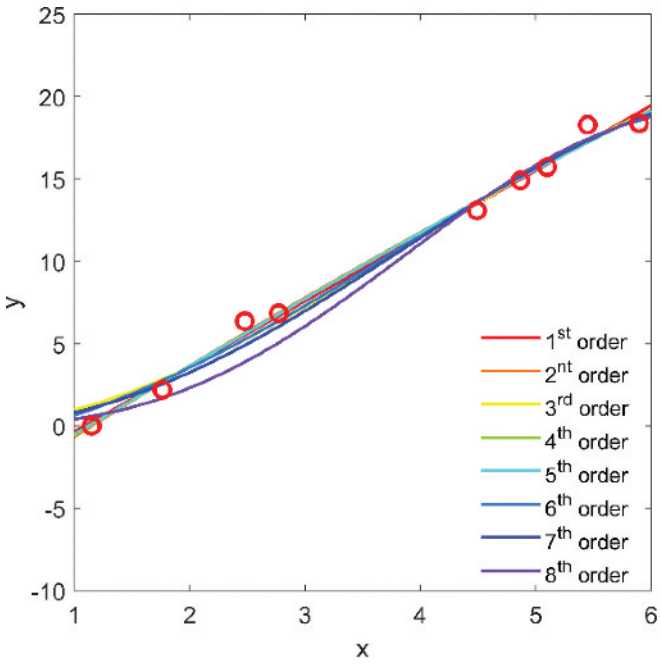

Fig. 11. Comparison of the hypothesis functions of up to eighth order for LASSO regression.

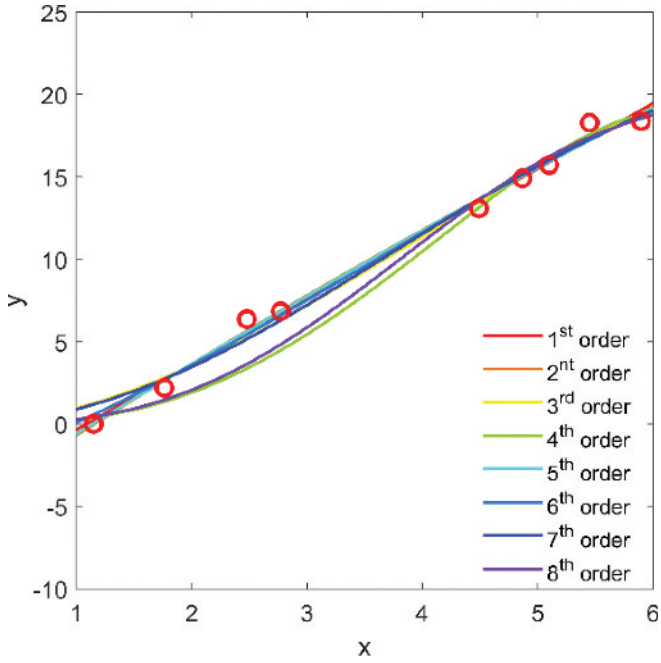

Fig. 12. Comparison of the hypothesis functions of up to eighth order for Elastic Net regression.

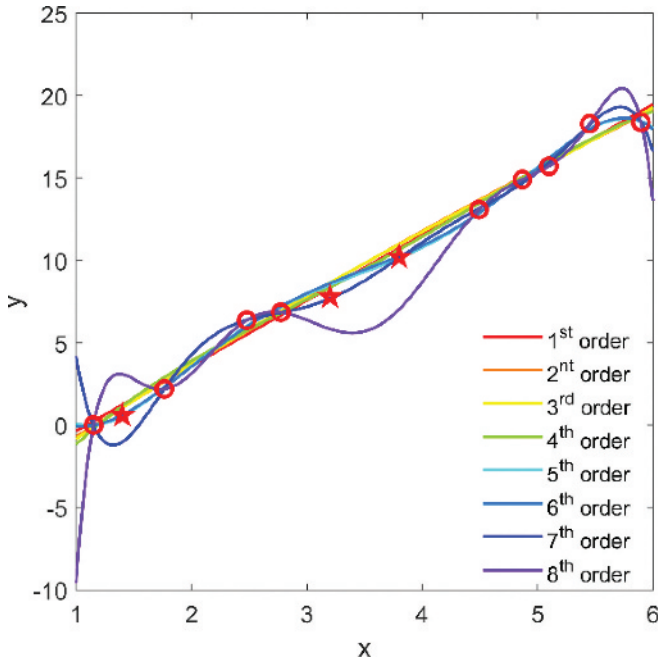

Fig. 13. Validation of the hypothesis functions through a new set of data.

helped to determine the best order of the hypothesis functions to be used, but also the MSE calculated by this method for the validation data set provided a more straightforward evaluation on the performance of the resultant hypothesis functions. Among the 125 data points in the cross-section library generated by Serpent, 100 of them were randomly selected to serve as the training data set, 19 were used as the validation data set, and the 6 remaining data points were used as the first evaluation set to evaluate the performance of the linear regression methodology. Additionally, 60 more data points were generated through Serpent calculation specifically for the

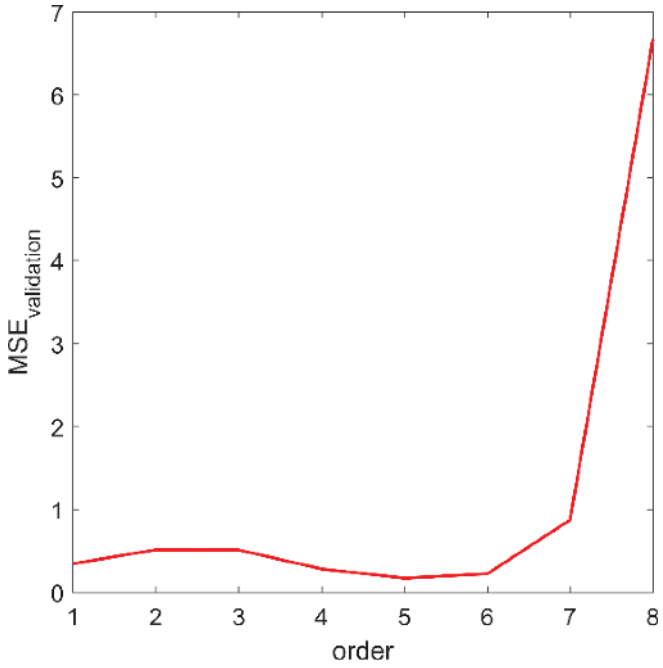

Fig. 14. Mean-squared error for the validation data set as a function of the order of the hypothesis functions.

evaluation of the performance of the linear regression methodology, as summarized in Table VII. The performance evaluation with this second evaluation data set would be more convincing as it was generated separately from the first 125 data points. The distribution of the 185 data points in a three-dimensional view is shown in Fig. 15.

IV.B. Linear Interpolation Method

In the current study, because the cross section was assumed to be dependent on three variables, namely, BU, T_{fuel}, and T_{mod}, the linear interpolation could also be performed according to each of these parameters by fixing the other two. A trilinear interpolation could also be used by averaging the calculation results obtained with each of the three variables. The two main differences between linear interpolation and linear regression are the following:

1. Linear regression uses the data gathered from the whole library to train the polynomial model while linear interpolation uses only the data points adjacent to the ones of interest. The applicability of linear interpolation is therefore limited. Among the six conditions of the first evaluation data set with BU of 100 MWd/kgU, which are shown by the blue dots in Fig. 15, conditions 1, 2, and 3 cannot be predicted through linear interpolation with T_{mod} whereas conditions 1, 3, 4, and 6 cannot be predicted through linear interpolation with T_{fuel}. Only condition 5 can be predicted by trilinear interpolation. Alternative options exist when the adjacent points are not available, including using data from points that are farther away from the point of interest with linear interpolation or using extrapolation. However, these alternative options would further decrease the precision of the interpolation predictions and were therefore not considered in the current study.

2. Linear interpolation trains only first-order polynomials while linear regression employs higher-order polynomials, which makes the data gathered from the whole library to be used more efficiently. First-order polynomials obtained through linear interpolation have low performance in certain circumstances. The curves in Fig. 3 are plotted again in Fig. 16 by zooming to the neutron energy range of [0.13eV, 1.3eV]. Linear interpolation will not work well for neutron energy group 184 because the cross section cannot be linearly correlated to BU. The performance of linear interpolation will be worse for energy group 209 because the variation of cross section as a function of BU is not monotone.

Nonmonotone data can be problematic for the linear interpolation method. Assume that we would like to estimate

TABLE VII

Steps of the Three State Variables Considered for the Second Evaluation Data Set

State Variables Considered	BU(MWd/kgU)	T_{fuel}(K)	T_{mod}(K)
Different steps considered	0 10 50 100 150	675 825 975 1125	350 450 550

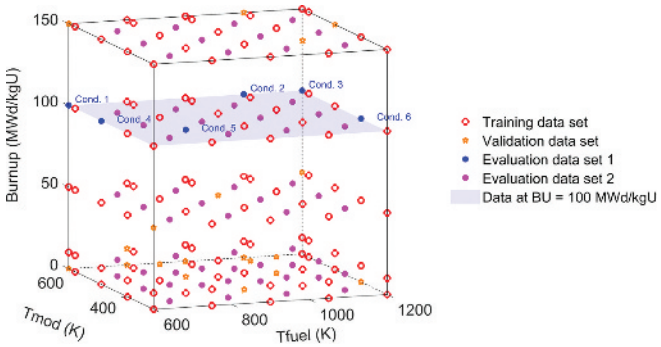

Fig. 15. Distributions of the training and the validation data sets.

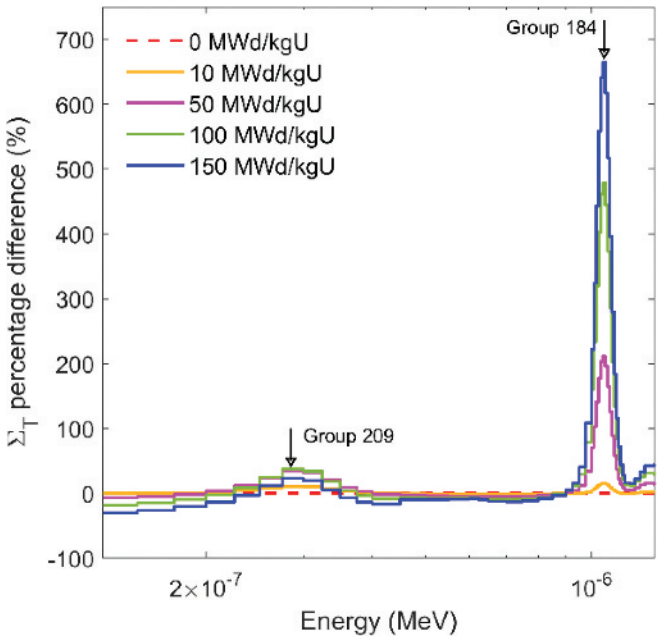

Fig. 16. Percentage difference of cross sections at different BU steps for neutron energy groups 184 and 209.

data y at location x using known data y_1 at location x_1 and y_2 at location $x_2 (x_1 \leq x \leq x_2)$. Using linear interpolation, we can estimate y as

$$\hat{y}_{linear} = y_1 + \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) \ . \tag{7}$$

Or, by using logarithmic interpolation, we can estimate y as

$$\hat{y}_{\log} = y_1 \times e^{\frac{\log(y_2/y_1)}{\log(x_2/x_1)}\log(x/x_1)} \ . \tag{8}$$

It is pointed out that $y_1 \leq \hat{y}_{linear} \leq y_2$ and $y_1 \leq \hat{y}_{log} \leq y_2$. In the case of monotone data, where $y_1 \leq y \leq y_2$, both linear interpolation and logarithmic interpolation can probably provide a good estimation of y. However, in the case of nonmonotone data, where $y \geq$ both y_1 and y_2 or $y \leq$ both $y_1$and y_2, apparently neither linear interpolation nor logarithmic interpolation can give a good estimation of y.

V. CASE STUDY RESULTS

Python, as one of the most popular languages for scientific computing, was used in the current study as the programming language to perform linear regression analysis. Known as an open-source code, Python has largely benefited from packages developed by third parties, including the Scikit-learning[31] package that we used in this study. Consisting of a wide range of state-of-art machine-learning algorithms, this package can be employed to solve both supervised and unsupervised machine-learning problems. Linear regression is considered as one of the supervised problems and was addressed by using Scikit-learning in the current study.

In this study, we considered the cross sections of the 238 energy groups to be independent of each other and conducted the linear regression analysis for the cross section of each energy group separately. Correlations of the cross sections of different energy groups may exist but were not considered in the current study. This is because in terms of the purpose of this study, the simple hypothesis functions built for each energy group were sufficient for the demonstration of the advantages brought by the use of linear regression. The reading of the library file and the generation of the hypothesis functions for the 238 energy groups were completed within 15 s using a single processor core. The hypothesis function of most of the energy groups was determined to have an order of four (to give the lowest MSE). The parameters of the hypothesis function for groups 184 and 209 are plotted in Figs. 17 and 18 as a demonstration of the representative parameters used in the method. These two groups were selected for plots because they were considered the most difficult groups to estimate accurate cross sections by the traditional linear interpolation approach. Although we employed only very standard linear regression methodology with no regularization terms added to the cost functions, none of the parameters were unrealistically large thanks to the use of the validation data set, as described in Sec. IV.A

The performance of the linear regression methodology was then evaluated. Groupwise neutron cross sections

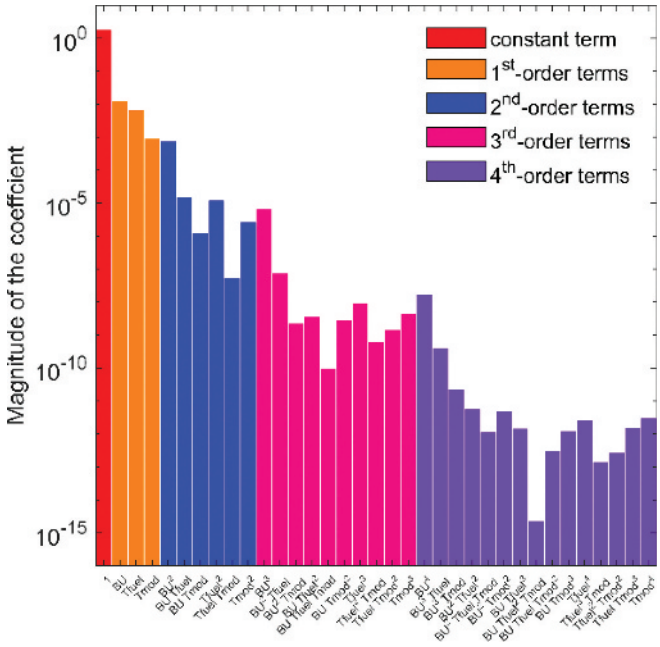

Fig. 17. Parameters of the hypothesis function for group 184.

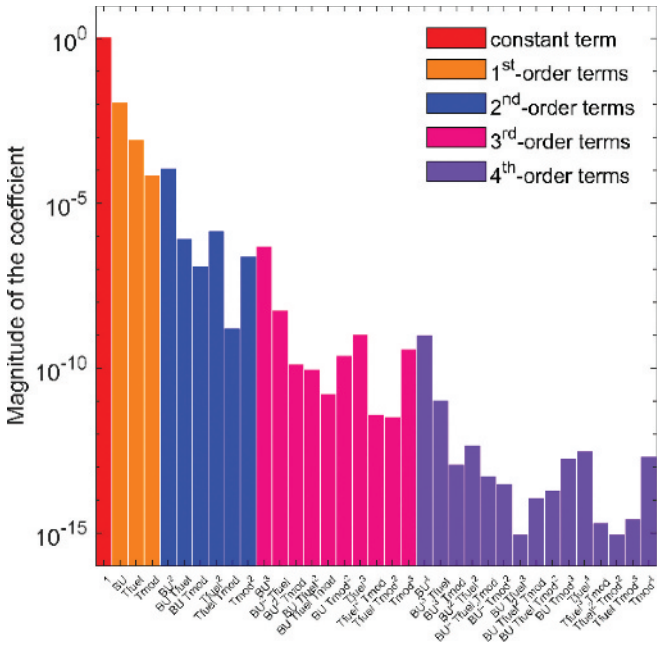

Fig. 18. Parameters of the hypothesis function for group 209.

were calculated for the six conditions of the first evaluation data set with both the linear regression and the linear interpolation methodologies and compared to that generated by Serpent, as shown in Figs. 19 and 20. For both energy groups, the cross sections estimated through linear interpolation with BU had the largest errors, and the underprediction was on the order of 10%. Linear interpolation

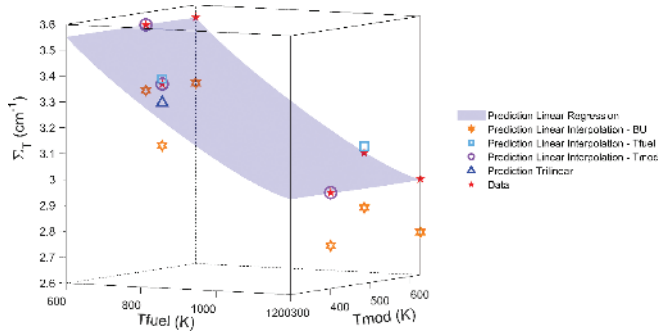

Fig. 19. Comparison of the cross sections for group 184 generated for the first evaluation data set.

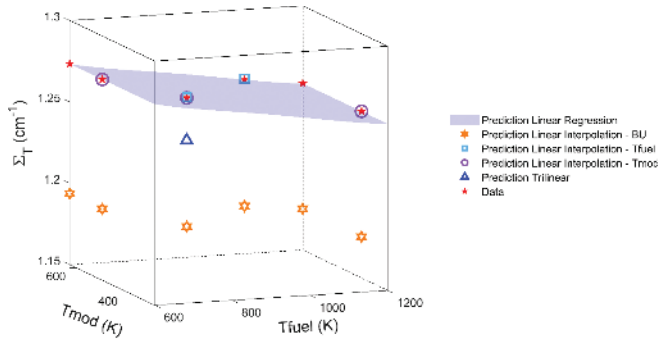

Fig. 20. Comparison of the cross sections for group 209 generated for the first evaluation data set.

with T_{fuel} and T_{mod} had narrower applicability but provided better precision because the impact of these two parameters on the cross section was significantly smaller than that of BU within the range of consideration, as discussed in Sec. III. The trilinear interpolation provided medium precision but had the narrowest application because it is the average of the first three calculations. In contrast, linear regression was applicable at every point of interest, and its prediction agreed well with the Serpent calculation. In Figs. 21 and 22, group 184 and group 209 neutron cross sections calculated with linear interpolation for the second evaluation data set were further compared to that generated by Serpent. The linear regression predictions showed good agreement with the Serpent data for both energy groups at all five BUs.

Groupwise neutron cross sections for all the 238 energy groups were calculated with both methodologies for condition 5 of the first evaluation data set, which had BU of 100 MWd/kgU, T_{fuel} of 750 K, and T_{mod} of 400 K. The percentage errors of the linear regression prediction are shown in Fig. 23 in comparison with that obtained through linear interpolation. No spikes were observed for the linear regression curve, which proved the methodology to work well for

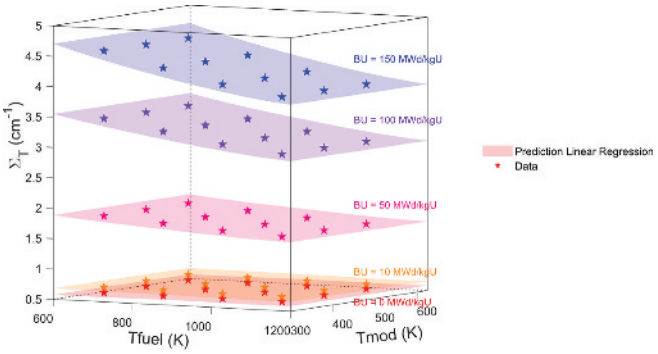

Fig. 21. Comparison of the cross sections for group 184 generated for the second evaluation data set.

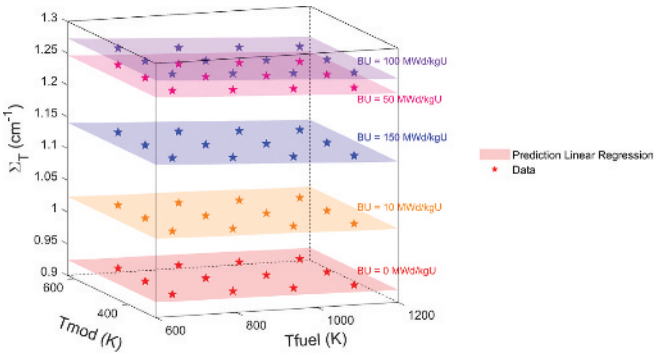

Fig. 22. Comparison of the cross sections for group 209 generated for the second evaluation data set.

all 238 energy groups. The maximum and the averaged prediction errors for the 238 energy groups are summarized in Table VIII. The smaller values of both errors demonstrated superior precision provided by linear regression than linear interpolation.

VI. CONCLUSIONS AND FUTURE WORK

In this study we performed a preliminary investigation on the feasibility of providing multigroup cross sections for deterministic neutronic codes by using linear regression models. With the help of Monte Carlo calculations with Serpent,

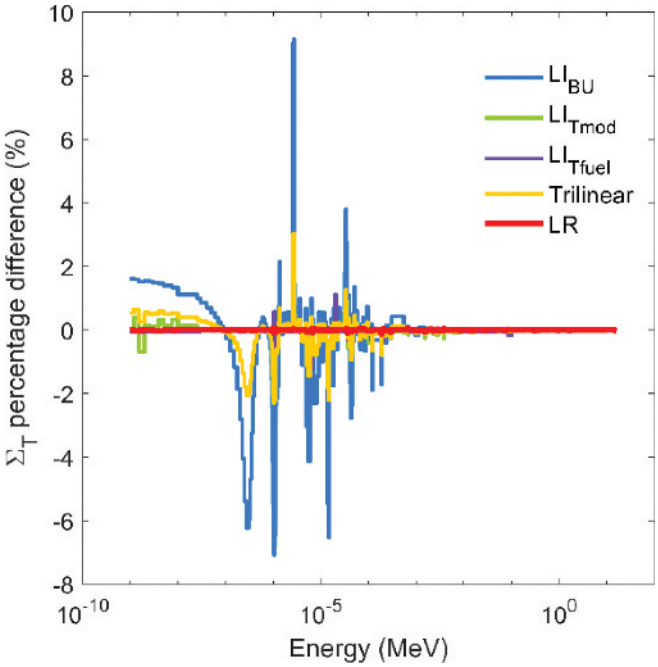

Fig. 23. Comparison of the cross sections generated for condition 5 of the evaluation data set.

we built a data library of total cross section calculated for 185 different conditions. We used 100 data points of them to train the linear regression models and 19 data points for the model validation. The remaining 66 data points were used to test and evaluate the performance of the linear regression methodology in a manner of comparing to the results with the linear regression methodology. We considered the cross sections of the 238 energy groups independent and built linear regression hypothesis functions for them separately. Both the maximum and the averaged prediction errors of linear regression for the 238 groups were smaller than those of linear interpolation, which demonstrated superior precision provided by the linear regression approach.

The superior performance of linear regression over linear interpolation can be explained by the following two theoretical reasons. (1) Linear regression used the data gathered from a wider range in the library to train the polynomial model while linear interpolation used up to

TABLE VIII

Maximum and Averaged Linear Regression and Linear Interpolation Predictions

	Linear Regression	Linear Interpolation, BU	Linear Interpolation, T_{fuel}	Linear Interpolation, T_{mod}	Trilinear
Maximum error (%)	0.13	9.15	1.20	0.69	3.04
Averaged error (%)	0.002	0.520	0.031	0.023	0.191

six points in the current study. (2) Linear interpolation trained only first-order polynomials while linear regression employed higher-order polynomials. The higher-order polynomials used by the linear regression methodology made the data gathered from the whole library to be used more efficiently.

Moreover, we proved in this study that linear regression had wider applicability than linear interpolation. In terms of the computational expense of the linear regression process, the reading of the library file and the generation of the hypothesis functions for the 238 energy groups were completed within 15 s using a single processor core. The short computation time meets expectations as the main computing powers needed for the linear regression are essentially for function evaluations. The effective computation procedure proved the additional computational burden caused by the implementation of the linear regression methodology in cross-section interpretation to be acceptable in practice.

We believe that the linear regression methodology can be implemented into any existing deterministic code to improve its capability of cross-section interpretation, while additional verifications are needed when specific cases are considered. For future work, we will implement the linear regression methodology into one of the popular deterministic codes for cross-section interpretation and use this modified deterministic code for the calculation of benchmark transient problems. The calculation results should further demonstrate the improvements brought by the implementation of linear regression.

Acknowledgments

The authors are grateful to the two anonymous reviewers' constructive suggestions for the work during the preparation of the manuscript.

ORCID

Cihang Lu http://orcid.org/0000-0002-6385-6338
Zeyun Wu http://orcid.org/0000-0002-6114-0352

References

1. Z. WU et al., "Neutronics and Safety Studies on a Research Reactor Concept for an Advanced Neutron Source," *Nucl. Technol.*, **199**, *1*, 67 (2017); https://doi.org/10.1080/00295450.2017.1335146.

2. C. LU et al., "Fully Ceramic Microencapsulated Fuel in Prismatic High Temperature Gas-Cooled Reactors: Analysis of Reactor Performance and Safety Characteristics," *Ann. Nucl. Energy*, **114**, 277 (2018); https://doi.org/10.1016/j.anucene.2017.12.021.

3. C. LU and N. R. BROWN, "Fully Ceramic Microencapsulated Fuel in Prismatic High-Temperature Gas-Cooled Reactors: Design Basis Accidents and Fuel Cycle Cost," *Nucl. Eng. Des.*, **347**, 108 (2019); https://doi.org/10.1016/j.nucengdes.2019.03.022.

4. Z. WU et al., "A Core Design for a Small Modular Boiling Water Reactor with Long-Life Core," *Nucl. Technol.*, **193**, *3*, 364 (2016); https://doi.org/10.13182/NT15-58.

5. N. R. BROWN et al., "Neutronic Evaluation of a PWR with Fully Ceramic Microencapsulated Fuel. Part II: Nodal Core Calculations and Preliminary Study of Thermal Hydraulic Feedback," *Ann. Nucl. Energy*, **62**, 548 (2013); https://doi.org/10.1016/j.anucene.2013.05.027.

6. J. RHODES, K. SMITH, and D. LEE, "CASMO-5 Development and Applications," *Proc. PHYSOR-2006*, Vancouver, British Columbia, Canada, September 10-14, 2006.

7. J. R. DEEN and W. L. WOODRUFF, "WIMS-D4M User Manual, Rev. 0," ANL/RERTR/TM-23, Argonne National Laboratory (1995).

8. C. J. WERNER et al., "MCNP User's Manual Code Version 6.2," LA-UR-17-29981, Los Alamos National Laboratory (2017).

9. J. LEPPÄNEN et al., "The Serpent Monte Carlo Code: Status, Development and Applications in 2013," *Ann. Nucl. Energy*, **82**, 142 (2015); https://doi.org/10.1016/j.anucene.2014.08.024.

10. D. KNOTT and A. YAMAMOTO, "Lattice Physics Computations, Section 3 – Resonance Treatment," *Handbook of Nuclear Engineering*, Vol. 2, Chap. 9, D. G. CACUCI, Ed., Springer (2010).

11. J. MALONE et al., "Lightbridge Corporation's Advanced Metallic Fuel for Light Water Reactors," *Nucl. Technol.*, **180**, *3*, 437 (2017).

12. B. J. ADE et al., "Candidate Core Designs for the Transformational Challenge Reactor," *Proc. PHYSOR 2020*, Cambridge, United Kingdom, March 29–April 2, 2020.

13. P. K. ROMANO and B. FORGET, "The OpenMC Monte Carlo Particle Transport Code," *Ann. Nucl. Energy*, **51**, 274 (2013); https://doi.org/10.1016/j.anucene.2012.06.040.

14. T. J. DOWNAR et al., "PARCS V3.0 – U.S. NRC Core Neutronics Simulator," *Theory Manual*, University of Michigan, Ann Arbor, Michigan (2010).

15. Y. XU and T. J. DOWNAR, "GenPMAXS Code for Generating the PARCS Cross Section Interface File PMAXS," PU/NE-00-20, Rev. 8, Purdue University (2006).

16. T. A. TAIWO, "DIF3D-K: A Nodal Kinetics Code for Solving the Time-Dependent Diffusion Equation in Hexagonal-z Geometry," ANLlNPR-92/17, Argonne National Laboratory (1992).

17. T. BAHADIR and S. LINDAHL, "Studvik's Next Generation Nodal Code SIMULATE-5," *Proc. Advances in Nuclear Fuel Management IV*, Hilton Head Island, South Carolina, April 12–15, 2009, American Nuclear Society (2009).

18. U. GRUNDMANN et al., "DYN3D Version 3.2, Code for Calculation of Transients in Light Water Reactors (LWR) with Hexagonal or Quadratic Fuel Elements – Description of Models and Methods," FZR-434, Forschungszentrum Rossendorf, Institute of Safety Research (2005).

19. G. MARLEAU, A. HÉBERT, and R. ROY, "A User Guide for DRAGON Version 5," IGE-335, École Polytechnique de Montréal (2020).

20. B. T. REARDEN et al., "SCALE Code System," ORNL/TM-2005/39, Oak Ridge National Laboratory (2018).

21. V. G. ZIMIN and A. A. SEMENOV, "Building Neutron Cross-Section Dependencies for Few-Group Reactor Calculations Using Stepwise Regression," *Ann. Nucl. Energy*, **32**, *1*, 119 (2005); https://doi.org/10.1016/j.anucene.2004.06.009.

22. F. ROUGH, "An Evaluation of Data on Zirconium–Uranium Alloys," BMI-1030, Battelle Memorial Institute (1955).

23. T. C. BRITT, "Innovative Fuel Design to Improve Proliferation Management," Master's Thesis, Virginia Commonwealth University.

24. M. D. SHAH et al., "Investigating Material Attractiveness for an Innovative Metallic Fuel Design," *Nucl. Eng. Des.*, **357**, 110385 (2020); https://doi.org/10.1016/j.nucengdes.2019.110385.

25. E. FRIDMAN and J. LEPPÄNEN, "On the Use of the Serpent Monte Carlo Code for Few-Group Cross Section Generation," *Ann. Nucl. Energy*, **38**, *6*, 1399 (2011); https://doi.org/10.1016/j.anucene.2011.01.032.

26. D. C. MONTGOMERY, E. A. PECK, and G. GEOFFREY VINING, *Introduction to Linear Regression Analysis*, John Wiley & Sons (2012).

27. X. YAN and X. G. SU, *Linear Regression Analysis: Theory and Computing*, World Scientific (2009).

28. A. E. HOERL and R. W. KENNARD, "Ridge Regression: Biased Estimation for Non-Orthogonal Problems," *Technometrics*, **12**, 55 (1970); https://doi.org/10.1080/00401706.1970.10488634.

29. R. TIBSHIRANI, "Regression Shrinkage and Selection via the Lasso," *J. R. Stat. Soc.*, **B58**, 267 (1996).

30. H. ZOU and T. HASTIE, "Regularization and Variable Selection via the Elastic Net," *J. R. Stat. Soc.*, **B67**, 301 (2005); https://doi.org/10.1111/j.1467-9868.2005.00503.x.

31. F. PEDREGOSA et al., "Scikit-Learning: Machine Learning in Python," *J. Mach. Learn. Res.*, **12**, 2825 (2011).