Reviewers' comments:

Reviewer #1: In U-ZrHx fueled TRIGA reactor calculations, the reactivity as well as other important reactor performance characteristics such as mean neutron generation time and fuel temperature coefficient is largely affected by the thermal neutron scattering cross-section, which is solely determined by the thermal scattering law. However, the vibration frequency distributions (also known as phonon spectra) of H and Zr in the fuel mixture contain all the information to compute the thermal scattering law. Thus it became crucial to develop a mathematical model that is capable of capturing the frequency distribution for H and Zr in the ZrHx fuel. This paper suggested such a feasible model called parameterized phonon spectrum (PPS) model based on some formerly developed models with the intention to provide a general framework to account for various ZrHx fuels with different H concentrations in different type TRIGA reactors. One primary task for the PPS model is to calibrate the main parameters in the model and quantify the uncertainties of the outputs associated with the model in reactor calculations. This paper developed three approaches (direct, GPR and BMARS emulator-based calibration) to perform the parameter calibrations for the PPS model. All the approaches were successfully implemented and applied to a case problem of TAMU TRIGA reactor calculations. The viability and accuracy of the calibration framework are tested with surrogated experiments (MCNP simulations), the uncertainty ranges for the QoIs in the application have been significantly reduced with the implementation of calibration approaches.

Based on my review, the overall technical demonstration of the manuscript meets the publishing standards of the NSE journal. The work presented in the manuscript is significant, the material is relevant, and the approaches can be helpful to high-order TRIGA reactor calculations. However, some revisions must be made to enhance the overall quality of the manuscript in both technical and written aspects (please see my detailed review comments about suggested revisions), therefore my review decision for the manuscript is "publish with revisions".

My general comments for the manuscript are as follows:

1. Though I don't have any big challenges to understand most of the technical parts when I read through the manuscript, my impression is the English writing shown in the manuscript (including the abstract part) can be improved by precluding ambiguous words and using more precise terms in the text. I believe this goal can be easily achieved through carefully proof-reading and multiple reviewing the manuscript. Some examples of the unsatisfied English presentation are given below:

Page 2, the second paragraph in Section 1.1 (line 8 - 13) can be better organized to make it more understandable to readers

Page 4, line 50, "… a scattering law which associates with a specific set of …" can be changed to "… a scattering law that determines a specific set of …"

Page 4, line 53, "Here we give a brief overview of this work." can be changed to "Here we briefly summarize these works."

Page 16, line 233 - 237, this sentence is too verbose as MCMC is well known standard method in statistics.

Hope these examples give the authors some insights how to enhance the written quality of the article in the revised version.

2. The main body of the manuscript was to introduce two emulators (GPR and BMARS) to help calibrating the parameters in the PPS model with the intention of reducing computational overheads and saving time. The calibration framework and the methodology implemented in the work are relatively easy to follow (which is good), however, the physics discussed for the problem is not very natural to follow if the reader has insufficient nuclear physics knowledge to thermal scattering (at which I believe most of nuclear engineers probably are not strong). Since this paper seeks publication in ANS NSE journal, I suggest the authors spending efforts in the revised version on introducing more fundamentals on this aspect to make the paper more amiable to readers with nuclear engineering background, because once the physics behind the problem is well understood, one will realize the importance of the work and have interest to investigate the methodologies. To better illustrate my point here, I raise up some questions to give the authors some ideas that what relevant materials may be needed in the introduction section.

Why are phonon spectra of H and Zr so influential to thermal scattering cross-section (I knew the math equations indicate the reasons, but what are the reasons in physics?) Does it only apply to TRIGA reactors or it is a general issue in all reactors such as LWRs? Why do the phonon spectra for H have branches but the spectra for Zr has not? How Debye temperature defined, what is it really? etc.

3. Reference citations should be coped with greater care. First, the format of references shown in the reference section does not follow the requirements of the NSE paper, please correct it. Second, some references are not cited in the order as their appearances in the text, and some references are missed. Some examples of the inappropriate reference citations are summaries in the following:

Page 2, line 7, this is the first place to appear Ref. 9 (Slaggie et al. CF model).

Page 2, line 25, this is the first place to appear Ref. 10 (Mattes et al. DG model).

Page 3, line 47, it is not appropriate to cite the whole book of Ref. 8 (Bell and Glasstone book), please specify the chapter number of the page number of the book in the citation

Page 4, line 61, it should be Ref. 3 (Evan et al. paper).

Page 5, line 78, it should be cited as [5, 6, 7].

Page 5, line 78, this is the first place to appear LHS design, please add reference and I don't think Ref. 5/6/7 is the original paper for LHS design. (LHS citation can also be used in Page 8, line 141.)

Page 6, line 85, this is the first place to appear NJOY (Ref. 17?) and MCNP, please add references.

Page 6, line 87, this is the first place to appear ANOVA, please add reference. And since ANOVA is not very familiar code to reactor analyst, it is better to add more description on it here.

Page 15, line 229 (the third line after section title 4.2), add Ref. 15 right after "Firedman".

Page 15, three lines after the one mentioned above, remove Ref. 15 in the citation after "different orders"

My specific comments on particular revisions suggested for the manuscript are summarized in the following. (Please note these suggestions are based on my review points on the manuscript, the authors may or may not need to respond all of them.)

1. Page 4, line 52. The subtitle "previous work" can be replaced by "former work" or "historical work" to eliminate the confusion to the authors pervious work described in the incoming subsection (Page 5, line 80)
2. Page 4, line 65. Since the PPS model presented by the authors has significant relationship to DG model (Ref. 10), in my opinion, it is natural here to give a more substantial introduction on DG model rather than just give a sentence description, such that the reader will clearly see the essential improvements imposed in the PPS model.
3. Page 5, line 67 - 79. It will be more straightforward if the authors can insert a table here to summarize the seven parameters (and the range values as well) chosen for the PPS model. In addition, it will be brilliant here if the authors can dig into the discussion to explain why these parameters are selected and what physical motivation behind the selection (if there is any).
4. Also in page 5, line 67 -79. T_DH is denoted as Debye temperature, but T_DZr is denoted as the peak position of the spectrum. If I understand correctly, they are essentially the same meaning here. However, using inconsistent terminologies additionally impose ambiguity in the context. Thus I suggest making some modifications on them. (Note, this is just my personal impression, but I do notice a couple of other places in the manuscript have the same problem.)
5. Page 6, figure 1. What is IKE? (I think you mean DG model, please be consistent with the text description), what is ENDF-VII? (I think you mean Slaggie's entered force model, please mention it in the text.) Why are the case numbers for H spectra different from the ones for Zr spectra? If no particular reasons, I suggest using same case numbers as examples for both H and Zr spectra.
6. Page 7, line 124. Why run 250 MCNP simulations while the sample size is 256 (shown in page 12, line 198)? I guess because the authors used the 250 simulations for calibration and the remaining 6 simulations for testing. If this is the case, I would say the description of calibration framework presented here is rather vague, though the diagrams shown in Fig. 2 help understanding the whole picture of the calibration, it is difficult to follow the text. Thus I suggest re-writing this part to clearly introduce the process of the calibration framework. I also think the part in line 102 - 114 in this paper should be moved and integrated into Section 2.1.
7. Page 8, figure 2. Please make the font size (too small to read) and type of the text in the boxes to be consistent with the text shown in the manuscript.
8. Page 9, line 152 (the first line in the page). Change " 2 LHS" to "two LHS" to make it be consistent with the previous appeared term.
9. Page 13, line 216, the period symbol after McClarren should be deleted.
10. Page 14, the second line in the page. What is X'? Please explain it in the text.
11. Page 14, Eq. 12, should the first (X_i - X_j) in the bracket should be transposed?
12. Page 14, the three lines right after Eq. 13 should be move the place right after Eq. 14.
13. Page 15, figure 7, to be precise, I suggest changing the caption of the figure to "GPR 1D data fitting for the logistic function". Similar revision can be applied to figure 8.
14. Page 16, Eq. 16, explain the beta_i in the equation.
15. Page 20, line 299. After the discussion of unanimous picking b and p as the most influential parameters in the PPS model from three different calibration approaches, it will be enlightening if some physics interpretations of the significance of these two parameters can be provided here.
16. Page 20, line 299. It will also be very neat if the authors can insert a comparative table here to summarize the results of b and p yielded from three different calibration approaches. Brief discussion on the results can be offered following the table.
17. Page 20, line 314. After the section 4.5 and 4.6, some discussions on the advantageous features of GPR and BMARS calibration approaches over direct calibration approach is necessary, because this is the original motivation to invoke this work. The discussion can be focused on the aspects of computational efforts saving and higher accuracy obtaining, etc. It will be more convincing if quantitative results can be provided in the discussion.
18. Page 24, line 335. BMARS results shown in Fig. 16 should also be discussed here as I don't see any other places discussing them.
19. Page 26, table 1. The tables should be numbered in Roman numeral as required by the format of NSE papers.

Reviewer #2: Authors primarily present the latest research on the uncertainty quantification (UQ) of ZrHx thermal scattering data - S(alpha, beta) nuclear data. In this paper, based on the previously developed parameterized phonon spectra (PPS) model, authors proposed a general calibration framework by applying two emulation based calibration methods to determine the optimal values of the parameters used in the PPS model. The presented research is novel and important to the UQ work for neutron interaction cross-section data, especially for the thermal scattering nuclear data.

However, the manuscript is not acceptable in the current form. Authors' writing style makes the manuscript not easy to follow and in some places, only authors know what the presented material means. Therefore, more clarifications and substantial editorial revisions must be provided before it can be accepted. The comments are listed below in detail and they are in the order of appearance.

(1) In the abstract, "Moreover, we extend the test to the reactivity at at a different temperature, 293.6 K, and obtained close

results to the surrogate experiment."

"at" appears more than once.

What is the other temperature that authors first study? There is no mention of temperature before, why "a different temperature"?

(2) Line 25, "based on Mattes's Debye-plus-Gaussian (DG) model". Add a reference after "model".

(3) Line 29-30, "for ZrHx that do not depend on the model," This is hard to follow. What do you mean by "not depend on the model"? What "model"?

(4) Line 35, "a optimal set" should be "an optimal set".

(5) Line 61, "More recently, Mattes et al. developed a new simplified model in which Zr is treated as free gas". The reference is Mattes et al. 2005 report. However, the model was developed in 1970 and 1971. "More recently" seems not accurate.

(6) Line 66, "Though the DG model with such simple mathematical form lacks the details of the one from solid dynamics computations," It is hard to follow "the details of the one". What "details"? What do you mean by "the one"?

(7) After Eqs. (5)-(7), in the explanation of each parameter, the range of each parameter is given. How are these ranges determined? What is the distribution of each parameter in its own range? These should be given.

(8) Line 76-79, should describe the LHS sampling design in detail here. At least enough information should be given for readers to understand how you sample each parameter within its own range.

What is the sequence number? How to understand the different results at different sequence number that shown in Figure 1? The larger the sequence number is, the better the result is? Necessary interpretation should be given.

(9) Figure 1 b seems not for H. There is no IKE model presented in the Figure 1b but the caption includes the IKE.

(10) From Figure 1, it is obvious that the spectra obtained from ENDF-VII, IKE and PPS are different. Which one is more reliable? What is the reason causing the difference? Is there a benchmark one to compare with? Authors didn't make the point clear why putting spectra based on three different models together. What do authors want to tell by plotting them together?

(11) Line 81, "We uniformly sampled parameters over the seven dimensional input space by using LHS designs". Do you assume that each parameter is "uniformly" distributed in each dimension? Again, this is related to comment (8). If the LHS is explained clearly, the confusion would go away.

(12) Lines 87-89, "A method combining ANOVA analysis with regression based cross validation was developed to investigate the sensitivity of QOIs to parameters and the interactions between the parameters."

What is "ANOVA"? This is first time mentioned. A relatively detailed description of the "method combining ANOVA analysis with regression based cross validation" should be provided. This method (technique) is mentioned again in Line 94 so a minimum interpretation is necessary. Otherwise, authors keep telling something that only they know.

(13) Lines 92-93, "a new and effective sampling strategy which separately sampled the two parameters together from the other five parameters"

What is the old strategy? LHS? What does it mean by "separately sample"? Again, a short and concise description is needed for this "new and effective sampling strategy". A reference is not enough. Also, the reference 7 is a conference proceeding, not available to the public.

(14) Line 96, "to be sensitive to the parameters". You mean all the seven parameters or just two main parameters?

(15) Line 97, "we ran simulations with ENDF-VII thermal scattering data of ZrHx as a surrogate of experiment". Since ENDF-VII only has the thermal scattering data of ZrH2, this means the calibration work is done only for ZrH2 phonon spectrum. Authors may point out that although only ZrH2 phonon spectrum is used to demonstrate the feasibility of the calibration framework, the framework is applied to the phonon spectrum calibration of any type of ZrHx or other materials.

(16) Line 105, "or inherently ignores regions of the parameter space". This doesn't make sense. Please rephrase it to make it clearer.
(17) Line 106, you need a comma "," between "calibration" and "Bayesian".

(18) Line 110, "with both the original calibration method and". What is "the original calibration method"? Where is described

before? Please clarify.

(19) Line 124, "We ran 250 MCNP simulations under this condition. In addition, another 250 runs were taken". Why you need 250 "surrogate experiments"? Or this "250 run" is not for the "surrogate experiment" but for the simulations with PPS models?

(20) Line 130, "propagated through" reads weird. Usually we say the uncertainty is propagated through. Is there a better way to express your meaning here?

(21) Line 132-133, "In Step 5, the modeler needs to calibrate plausible PPS parameter values (or value sets) with QOIs generated in Step 4."

These QoIs generated in Step 4 are based on PPS parameters or ENDF data? It seems based on PPS parameters from Figure 2. How can you compare the QoIs based on the same parameters (PPS vs PPS)? Some clarification is needed here.

(22) More clarification and discussion about Figs. 3 and 4 are needed.

What does each notation represent "V1, V2, V3, …"?
What does "the histogram of each parameter" mean? What can we learn from these histograms? Why you plot them?
What does "the pair-wise correlation for each pair of parameters" mean? What does "each number" mean in the figure? Should tell readers something you learn from these numbers.

If a figure or table show something, necessary discussion should be provided on what you learn or what conclusion you draw from these results.

(23) Line 164, it is better to say "The simulation results are made with scattering data 'based on the phonon spectra constructed with the PPS parameters'". Again, "propagated from" reads weird.

On the same line, how is "the score of a realization" calculated? At least, some qualitative description about the score definition and calculation should be given.

(24) Lines 199-200, "that the high score region can only be located near the location of the samples: other regions of parameter space are effectively ignored."

This is hard to follow. What do you mean by "region"? Where do you show this "high score region"? Please rephrase the sentence to make it clearer.

(25) Line 207, what are these "outside discrete points"? Please make it clearer.

(26) In Equation 12, capital lambda is already used for mean generation time. Maybe use a different notation here?

(27) Line 264, "One of benefit" should be "the benefits". Also, this sentence reads a little weird. Please rephrase it.

(28) Last comment but may need authors to think about more: authors use MCNP simulations to generate "experimental" and "simulation" results for the calibration purpose. QoIs predicted by Monte Carlo methods always have the statistical uncertainties. During the calibration, how can authors distinguish the uncertainty between the statistical uncertainty and the PPS parameter uncertainty? This is important to correctly quantify the final optimal values of the parameters. If a deterministic code is used for this work, there is no such statistical error issue. But a Monte Carlo code does. Authors should provide more comments/discussions on this in the conclusion section.