LA-UR-21-26262
Approved for public release; distribution is unlimited.

Title: Questionable Benchmarks

Author(s): Thompson, Nicholas William; Rising, Michael Evan; Neudecker, Denise; Hutchinson, Jesson D.; Kahler, Albert Comstock III; Clark, Alexander Rich; Herman, Michal W.; Alwin, Jennifer Louise; Grosskopf, Michael John; Haeck, Wim; Vander Wiel, Scott Alan; Cutler, Theresa Elizabeth

Intended for: WPNCS SG-8, 2021-07-05/2021-07-09 (Paris, France)

Issued: 2021-07-02

Los Alamos NATIONAL LABORATORY

Questionable Benchmarks

Nicholas Thompson, Michael Rising, Denise Neudecker, Jesson Hutchinson, A. (Skip) Kahler, Alex Clark, Michal Herman, Jennifer Alwin, Michael Grosskopf, Wim Haeck, Scott Vander Wiel, Theresa Cutler

July 5th, 2021

Questionable Benchmarks

- Effort at LANL to identify benchmarks with questionable keff results or uncertainties
- Selection of benchmark classes were investigated (ie HMF, PMF, etc)
- Multiple methods used
- Brief review of some benchmarks were done, with subjective comments of quality of benchmark and uncertainty analysis
- Results are being written as a journal paper, LANL report is >80 pages so not all results can be presented here

Some caveats

- Not official work of the ICSBEP subgroup or SG-8
- Questionable doesn't mean "bad", it means it should be investigated further
- Meant at input for those groups to help triage which benchmarks to look at first
- This wasn't the most thorough review and should not be seen as a final judgement
- Original intent of these benchmarks may have been very different than how they are used today – some benchmarks may be very useful, but may not be useful in a validation suite
- Any statements should not be viewed as negative towards the benchmark authors – standards have changed over time

List of Methods

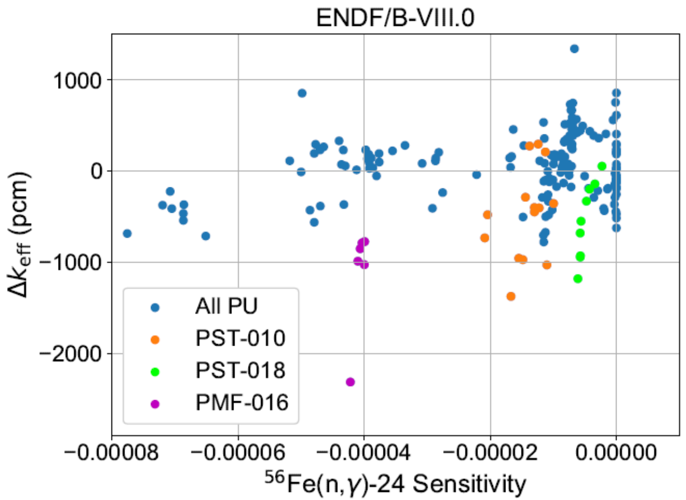
- Random Forest/SHAP
- Whisper
- Simulation/Experiment (C/E) keff >1.02 or <0.98
- 50 pcm uncertainty cutoff
- Histograms
- Similarity coefficients
- Related benchmarks

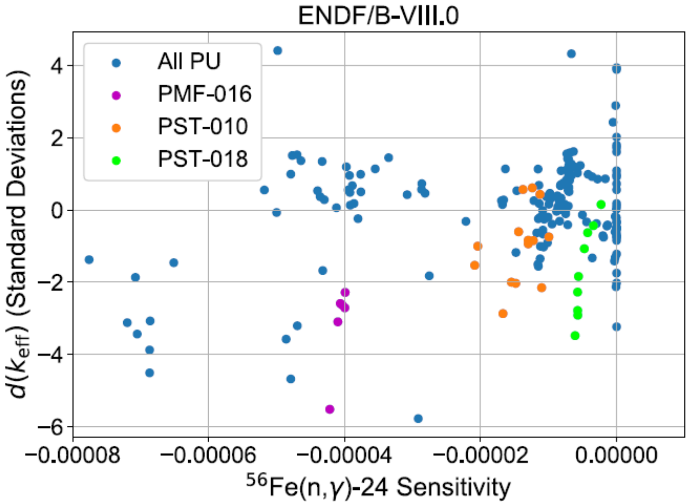
Methods: Random Forest and SHAP

- Random Forest and SHAP are machine learning techniques
- Random Forest models the bias in Δkeff between experimental and simulated keff values (ENDF/B-VIII.0, MCNP-6.2) as a function of keff sensitivity.
- SHAP (SHapley Additive exPlanations) is used to assess the importance of each ‘feature’ (Sensitivity values and measurement attributes for each energy group)

Methods: Random Forest and SHAP

- Some examples of results
- PMF-016 has most of it's results ~1000 pcm for Δkeff (experiment – simulation) but one results is >2000 pcm different.

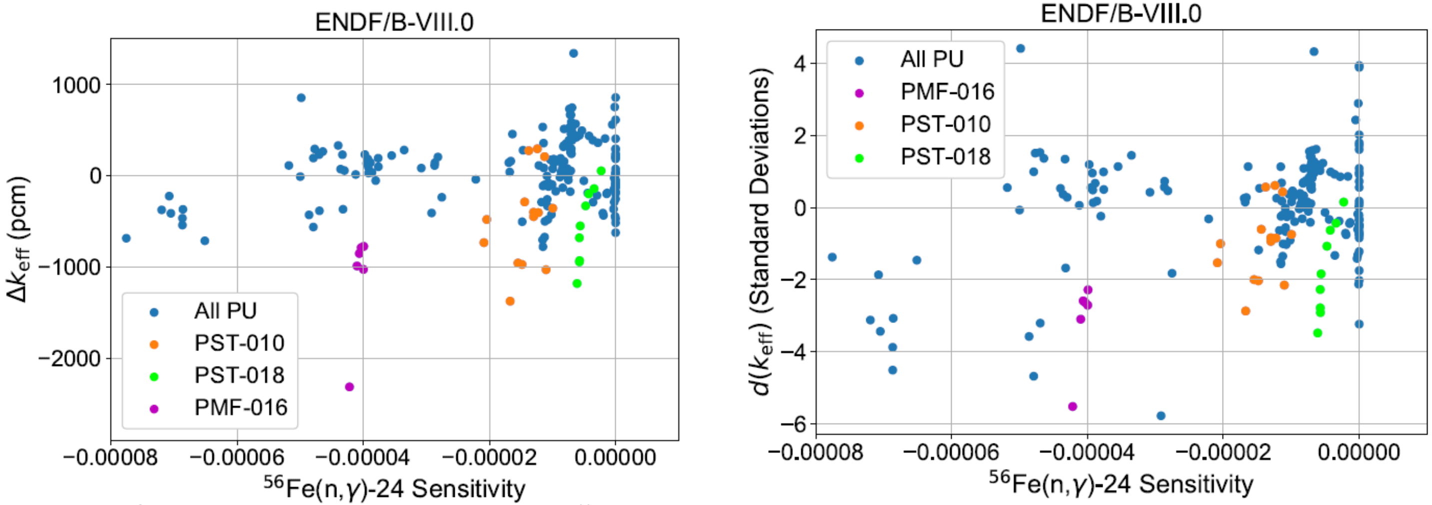




Plots from D. Neudecker et al., "Enchancing Nuclear Data Validation Analysis by Using Machine Learning," Nuclear Data Sheets 167:36-60 (2020).

Methods: Random Forest and SHAP

- Some examples of results
- PMF-010 and PST-018 have large spreads in Δkeff
- Could be a nuclear data issue, could be a model/benchmark issue



Plots from D. Neudecker et al., “Enchancing Nuclear Data Validation Analysis by Using Machine Learning,” Nuclear Data Sheets 167:36-60 (2020).

Methods: Whisper

- Whisper computes a χ^2_{min} related to the degree to which the linear regression can fit the benchmark experimental data within the nuclear data covariances.
- The benchmark with the largest $\chi^2_{diag,i}$ is removed until a certain threshold χ^2_{min} value is reached, removed benchmarks are classified as questionable, remaining are accepted.

Methods: Simulation/Experiment keff <0.98 or >1.02

- These limits are somewhat arbitrary, but any benchmark with large differences between experimental and simulation keff values should be investigated.
- This relates to +/- 2000 pcm difference between experiment and simulation
- Could be nuclear data issues, could be issues with the benchmark or benchmark model

Methods: Cut-off of 50 pcm keff uncertainty

- All benchmarks with keff uncertainties below 50 pcm were listed as questionable
- Very likely uncertainty analysis is lacking for these benchmarks
- These low uncertainties may be justified, but a more thorough review is needed to determine that

Methods: Histograms of Classes

- For each benchmark “class” (ie PST, PMF, HMF…) a histogram of keff uncertainties for those benchmarks were generated.
- The benchmarks with the lowest uncertainties for each “class” were identified as questionable

Methods: Similarity Coefficients

- Similarity coefficients (based on sensitivities) for all pairs of benchmarks were generated.
- For those benchmarks which are extremely similar to each other (similarity coefficient of >0.99), keff uncertainties were compared.
- If keff uncertainties differed by >50%, the pair of benchmarks were listed as questionable
- Again, this doesn’t mean that one of those benchmarks is “wrong” or “bad”, but is a method for triaging benchmarks to investigate further

Methods: Related Benchmarks

- If a questionable benchmark is found by another method, and it was part of a series of experiments undertaken by the same group with the same or very similar equipment in the same time period, those other benchmarks are included as questionable.
- Likely the same issues exist in both sets

Results: Benchmarks with questionable keff values

Method	Benchmarks
ML	PST: 004, 006, 007, 010, 018, 028; PMF: 014, 016; HST: 001, 025, 050; HMF: 003, 005, 025, 051, 057, 072, 084, 092, 093; MMF: 007; UST: 013, 015; LCT: 005, 022, 024, 025, 028;
Related Benchmarks	PMF: 012, 013, 015 (related to PMF014); PST: 001 (related to PST007), 005 (related to PST004);
C/E	PCF: 001, 002; PCI: 001; PCM: 002; PCT: 002; PMF: 008, 015, 018, 026, 028, 035, 045; PMI: 002; PST: 001, 002, 007, 008, 009, 010, 011, 013, 015, 017, 019, 020, 024, 025, 031, 040; HCF: 003; HCI: 003, 004, 005, 006; HCM: 002; HCT: 002, 006; HMF: 003, 007, 021, 041, 052, 053, 057, 060, 064, 067, 070, 073; HMI: 001, 005; HMM: 005; HMT: 001, 003, 006, 011, 012, 026, 027; HSI: 001, 002; HST: 001, 003, 004, 006, 014, 015, 016, 017, 018, 019, 020, 021, 025, 026, 037, 038, 039, 044, 047, 048, 050; ICF: 002; ICI: 003, 004, 005; ICM: 001, 002, 003; ICT: 001, 005, 011; IMF: 005, 009, 013; IST: 001; LCT: 003, 004, 010, 012, 015, 017, 028, 033, 045, 066, 068, 069; MCF: 001; MCI: 006; MCT: 012, 013, 014; MMF: 007, 008; MMI: 001; MMT: 001; MST: 001, 002; SMF: 004, 011, 014; UCT: 001; UMF: 001, 002, 003, 006; USI: 001; USM: 002; UST: 002, 003, 004, 012, 013, 014, 015, 016, 017;

Results: Benchmarks with questionable keff values

Whisper	HCT: 002-015–017, 002-021–022; HMF: 007-006, 007-032–034, 007-036–037, 007-039–040, 025-005, 038-001, 040-001, 051-001, 051-003, 051-009, 051-014–018, 057-001, 057-003, 057-005, 064-001, 064-003, 090-001–002, 092-003, 093-001, 094-002; HMI: 006-001, 006-004; HMT: 012-001, 014-001; HST 001-009—010; LCT: 007-003, 027-002–004; LST: 007-001, 007-003; MMF: 007-002; MST: 001-003–004, 001-011, 003-001–004; 003-006–007; PCM: 001-001–003, 002-001–006, 002-010–016; PMF: 016-001, 039-001, 041-001; PST: 009-003, 010-001, 010-013, 012-006–007, 018-002, 028-001–009; UMF: 004-002; UST: 001-008, 001-010–011, 001-016, 001-018, 001-020–022, 001-030, 001-032, 012-003, 012-006, 015-008–009, 015-018–019, 016-016, 016-021, 016-023, 016-031–033;

Results: Benchmarks with underestimated keff uncertainties

Method	Benchmarks
ML	PST: 022, 028, 032; PMF: 005, 008, 011, 020, 035, 041, 044; HST: 011; HMF: 007, 038, 051, 088, 090, 091, 092, 093, 100; IMF: 001, 002; MMF: 004, 005,010; UMF: 002, 004; UST: 012, 016; LCT: 027; LST: 04;
Related Benchmarks	PMF: 010, 018; UMF: 003, 005; MMF: 001 (related to PMF005, UMF002, UMF004); PMF: 022, 023, 025, 029, 030, 036; MMF: 009 (related to PMF: 035, 41 and MMF: 010); PST: 012, 029, 030, 031 (related to PST: 022, 028, 032); UST: 017 (related to UST: 012, 016); HCT: 009, (related to HCM003); LCT: 043, 044, 046, 054, 058, 067, 077, 082, 083, 084, 088, 089, 090, 091, 092;
Cut-off of 50 pcm	LMT: 005; HCT: 009; HCM: 003, 004; HMF: 047, 051, 059, 069, 071, 074, 076, 099; IMF: 001; LCT: 046, 054, 058, 067, 088, 092; HMT: 011; HMM: 014, 020;

Results: Benchmarks with underestimated keff uncertainties

Histograms	PMF: 001, 008, 011; HMF: 047, 051, 059, 071, 076; IMF: 001; MMF: 009, 010; HMT: 011; LMT: 005, 007; HMM: 014, 020; ICI: 004; HCT: 009, 021; ICT: 011, 013, 014, 015; LCT: 046, 054, 058, 067, 088, 092; MCT: 017; HCM: 003; ICM: 002; USI: 001; PST: 012, 027, 028, 030, 033; HST: 011, 022, 024, 025, 028, 034, 049; IST: 002, 003; LST: 004; UST: 012, 013, 015, 016, 017; MST: 006;
Similarity Coefficients	HMF: 051-009-d, 051-014-d, 051-015-d, 051-016-d, 051-017-d, 051-018-d, 069-001-s; HMT: 031-001-s: IMF: 001-003-ideal, 001-004-ideal; LCT: 007-002, 007-003, 007-004, 007-006, 007-007, 007-009, 007-010, 027-002, 027-003, 027-004, 079-006, 079-007, 079-008, 079-009, 079-010; PST: 028-001-d, 028-002-d, 028-003-d, 028-004-d, 028-005-d, 028-007-d, 028-008-d, 28-009-d, 032-001-d, 032-002-d, 032-003-d, 032-013-d;

Brief review of ‘questionable’ benchmarks

- One hour reviewing each benchmark identified as ‘questionable’
- Looked for obvious problems in description, uncertainty analysis, modeling
- Not a recommendation to the ICSBEP – more in-depth analysis is needed
  - Can be a starting point
- Examples on the next two slides of the review that was performed

Identifier	# of Cases	Year	$\delta k^e_{\rm eff}$ (pcm)	Related Benchmarks	Subjective Opinion on $k_{\rm eff}$ ^{e}	Comments	w
PST004 (4.1.1)	13	1994	350–470	PST: 005	Poor–Average	Performed early 1950s (limited information)	0.75
PST006 (4.1.2)	3	1996	350	-	Poor–Average	Performed early 1950s (limited information)	0.75
PST007 (4.1.3)	11	2004	470	PST: 001	Poor–Average	Performed early 1950s (limited information); study hydrogen content	0.33
PST010 (4.1.4)	14	2004	350–480	-	Poor–Average	Study hydrogen content	0.5
PST018 (4.1.5)	9	2006	300–340	PST: 001, 007, 020, 021	Average–Good (4–9), Suspect (1–3)	Study extrapolation of solution volume to criticality	0.33 (1–3), 1 (4–9)
PST028 (4.1.6)	18 (15 acceptable)	2004	120	PST: 022, 029–032	Average–Good	Study UQ of alloy materials, Mn-unc likely overstated	0.5
PMF012 (5.1.1)	1	1996	210	PMF: 013–015	Below average	Model highly homogenized, $k^e_{\rm eff}$ not based fully on experiment	0.33

Los Alamos NATIONAL LABORATORY

8.1.9 HEU-MET-FAST-072

HMF072 describes three critical configurations containing cylindrical HEU discs with interleaved carbon steel (iron), with and without polyethylene, surrounded by a thick copper reflector. The experiment was conducted at the Los Alamos Critical Experiment Facility (LACEF) in 2002. Irrespective of any interstitual material, these HEU/copper reflected configurations are commonly referred to as "Zeus" criticals.

The evaluation was performed by D.K. Hayes of Los Alamos National Laboratory in 2004. Revision 1 is dated September 30, 2006. The revision history states that this was a "technical" revision without further detail. The MCNP models for these critical configurations are derived from the Revision 1 description.

The uncertainty analysis is reasonably complete, consistent with modern expectations, although we note one omission below and so could benefit from a re-review. In particular Section 2.0 notes that the MCNP sensitivity calculations performed at the time of evaluation are for a total of 6,000,000 active neutron histories. The stochastic uncertainty from those jobs is likely near 200 pcm, resulting in a near 300 pcm combined uncertainty when attempting to evaluate the impact of calculations with and without a specific perturbation. These calculations should be rerun for a much greater neutron history number. It also seems that the uncertainty due to polyethylene attributes needs a second look. The Case 1 uncertainty is ~240 pcm while for Case 3 it is ~690 pcm, with poly-ethylene related uncertainty being the sole difference. The polyethylene mass uncertainty provided in Table 22, which translates

Summary

- Investigation and initial review of ‘questionable benchmarks’ was performed at LANL
- Used multiple methods including machine learning techniques to identify ‘questionable benchmarks’ and/or benchmarks with low uncertainties
- Those benchmarks were then reviewed for obvious errors
- Not a recommendation to ICSBEP, but can be used as a starting point for a more comprehensive review
- Journal paper being written on this work.

Questions?

- Nicholas Thompson, nthompson@lanl.gov