Using Machine Learning Algorithms in Nuclear Data Evaluations

uncovering hidden patterns in sensitivity profiles

Michael E. Rising, Pavel Grechanuk, Denise Neudecker, Mike Grosskopf, Wim Haeck, Scott Vander Wiel, Michal Herman, and Patrick Talou

XCP, CCS and T Divisions, Los Alamos National Laboratory

NCSP Technical Program Review
PANTEX, March 26-27, 2019

LA-UR-19-22505

Unraveling the nuclear data problem

- Accurate nuclear data are key to the success of criticality-safety simulations

- Some basic questions:

  - How do we produce accurate nuclear data?
  - How do we know they are accurate or deficient?
  - Can we identify specific deficiencies?
  - Can we eliminate compensating errors?

This Talk:

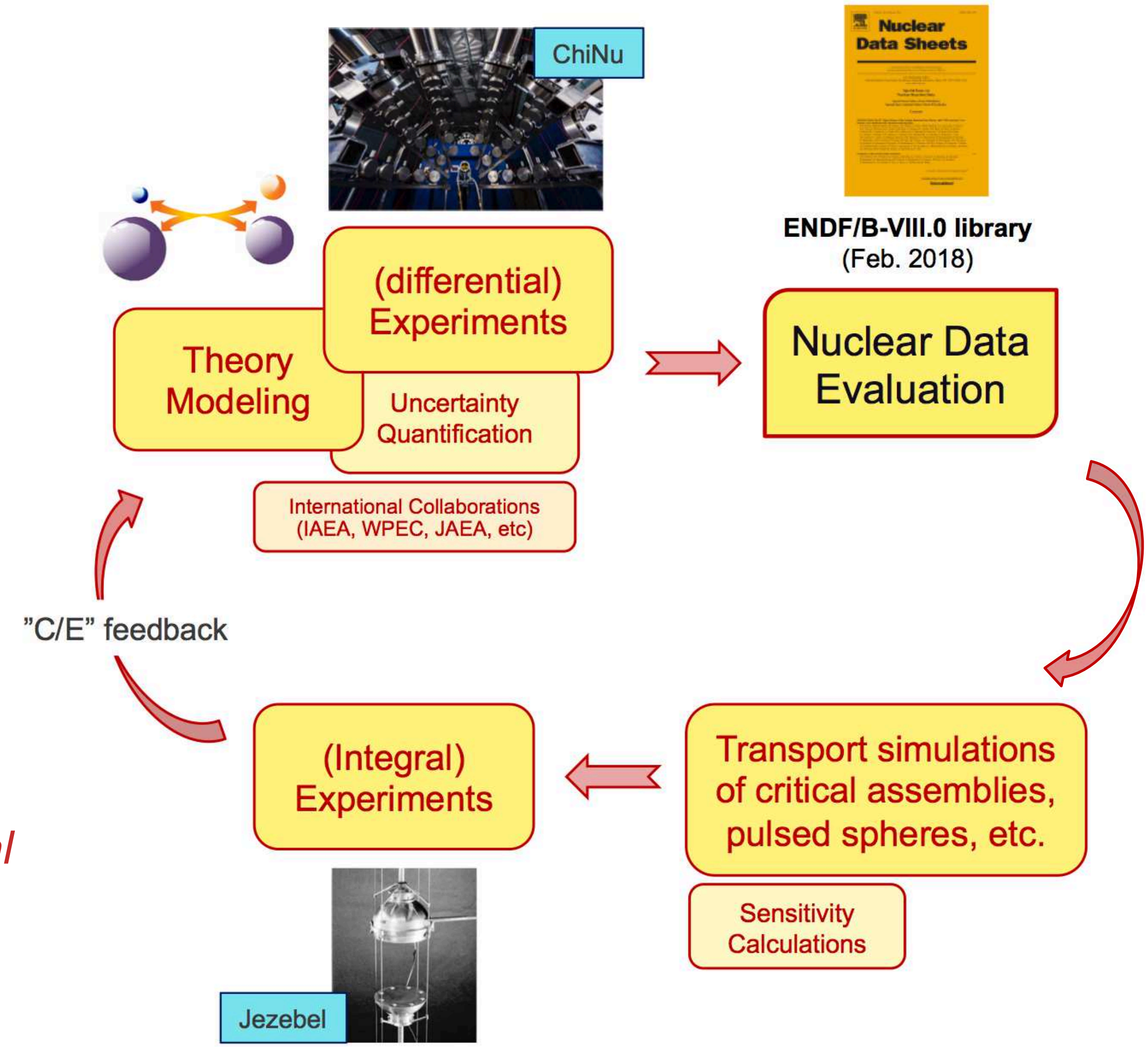
- A very brief tour of the nuclear data evaluation process
- How machine learning algorithms can help in the inverse problem?
- Scoping studies of machine learning in nuclear data
- Perspectives

The nuclear data evaluation process

- Experimental *differential* data, e.g., (n,2n) cross section
- Physics model calc. to describe **all** reactions
- Bayesian statistical techniques, model parameter fitting, etc.

➔ Final product: tabulated data files (mean values + covariance)

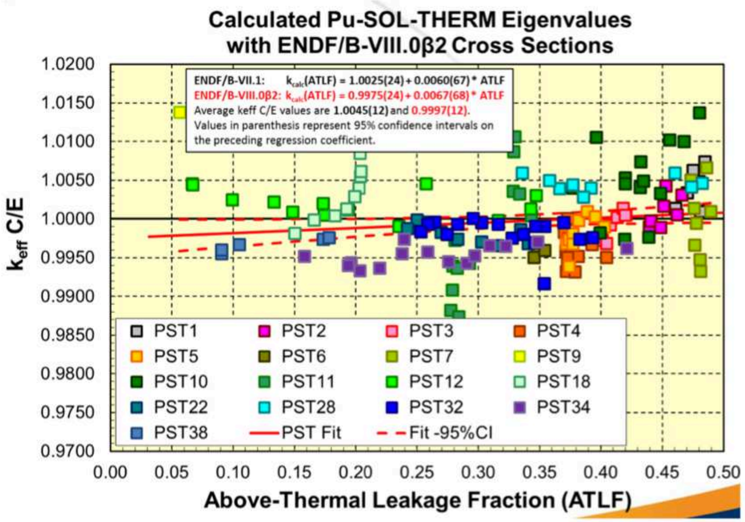
- Feedback loop from *integral* data ("C/E" results)

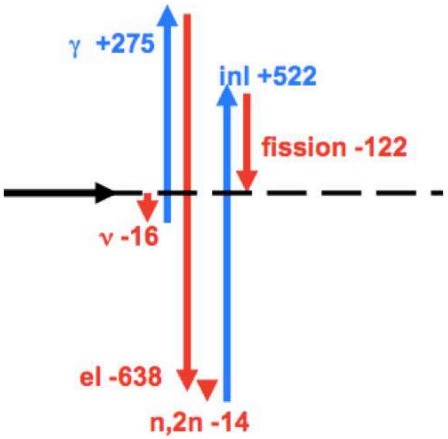


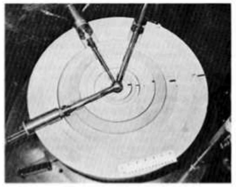
The validation of nuclear data uses simulations of integral data to check C/E values

- Many integral benchmarks are used, but their interpretation is difficult (**inverse problem**)
- We often get the right answers for the wrong reasons
- Lots of room for **compensating errors!**
- The use of integral benchmarks is not reflected in the ENDF covariances

Jezebel critical assembly







Pulsed spheres

Trends/biases along meaningful axes can provide useful feedback on underlying nuclear data

Compensating errors between nuclear reaction channels in the simulation of the Jezebel assembly (ENDF/B-VII.1 vs JEFF-3.2)

Many configurations with reflected materials

Moving beyond the current situation

- Large-scale use of **sensitivity profiles**
  - How does a change in the cross section σ impact the integral *k* value?
  - More generally, how does a change in a feature f impact an integral quantity Q?

$$S_{k,\sigma} = \frac{\Delta k/k}{\Delta\sigma/\sigma}$$

$$S_{Q,f} = \frac{\Delta Q/Q}{\Delta f/f}$$

- Importance of **metadata** in integral benchmarks
  - Geometry, composition, experimental technique, etc.
  - Going beyond (x,y,dy); X_i=experimental features

$$(y, dy) = f(x|X_i)$$

- More **realistic UQ studies** of differential and integral experiments
  - Wrong estimates of uncertainties and correlations can significantly bias statistical results

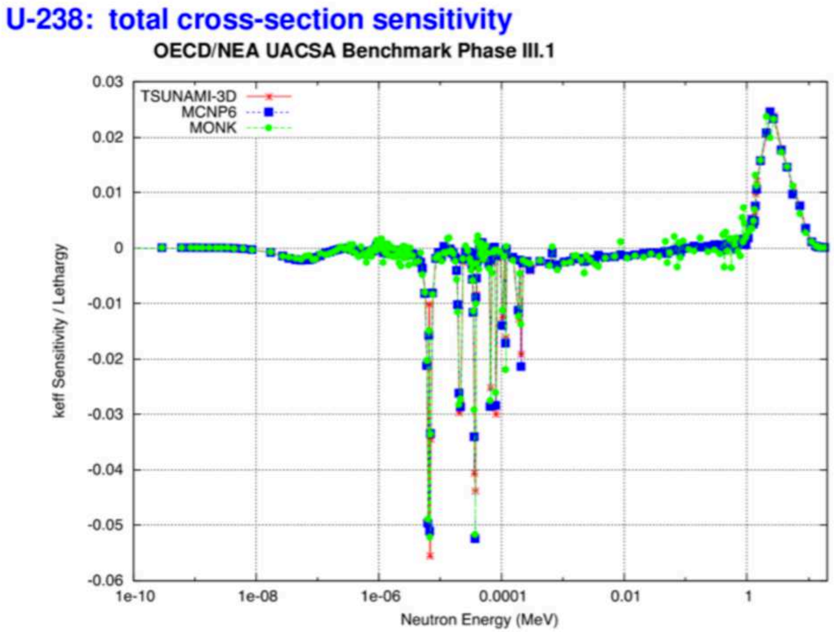
- Use of **machine learning** algorithms to identify hidden patterns and correlations in zoology of integral data

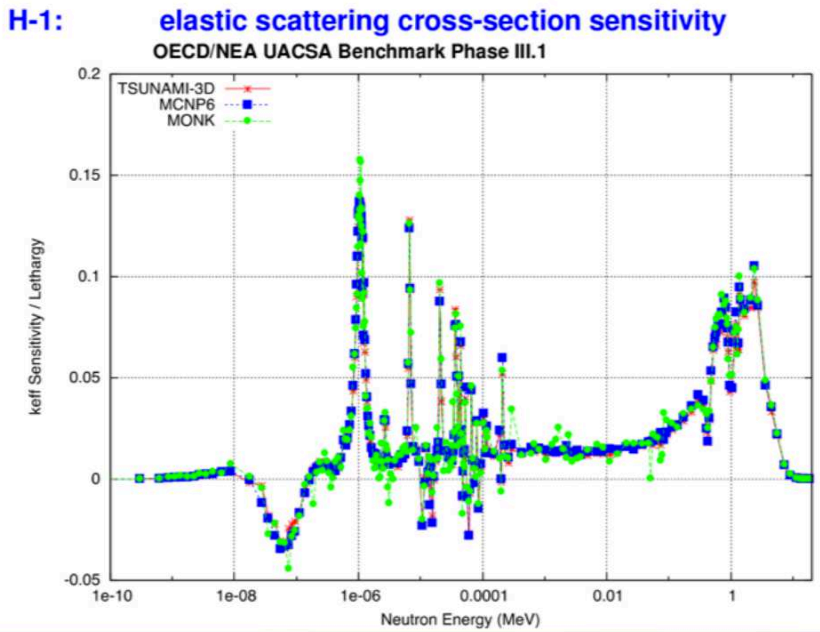
Develop a database/catalog of sensitivity profiles using MCNP6 and Whisper

- Use MCNP6 perturbation/sensitivity features
  - Can compute profiles of k_{eff} – nuclear data sensitivity profiles
  - How does a relative change in the cross section impact k_{eff} of the system?

$$S_{k,\sigma} = \frac{\Delta k/k}{\Delta\sigma/\sigma}$$

  - For a single system, these (energy-dependent) profiles are unique





Recent R&D work on applying machine learning algorithms to Whisper benchmark catalogue

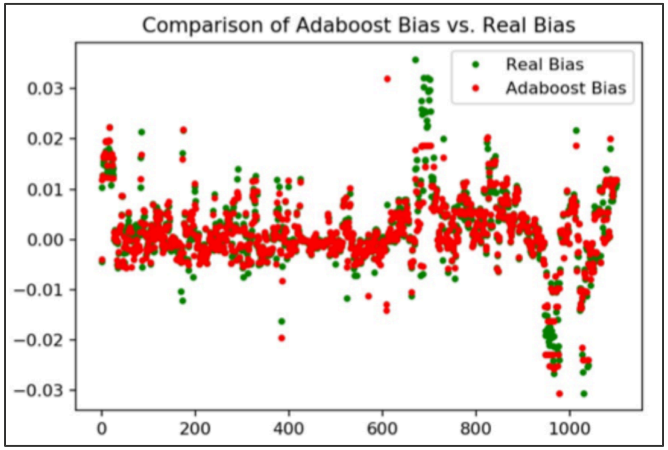
- Oregon State University PhD student, Pavel Grechanuk
  - Began ML research in summer of FY17
    - Explored ML algorithms to predict k_{eff} bias using only sensitivity profiles

$$B_i \approx f(S^i_{k,\sigma})$$

  - Continued ML research in summer of FY18
    - Using the SHAP metric to understand feature importance in predicting bias
    - Explored ML algorithms to cluster benchmarks together
    - Explored the genetic algorithm to perform nuclear data assimilation for the most important nuclear data quantities found in the bias prediction and clustering steps
  - Future work under an established subcontract between XCP-3 and OSU supported through both NCSP and ASC/PEM-NP
- This R&D work has led to a larger LANL team effort, spanning several groups (XCP, T, CCS), to develop robust ML tools for nuclear data evaluation

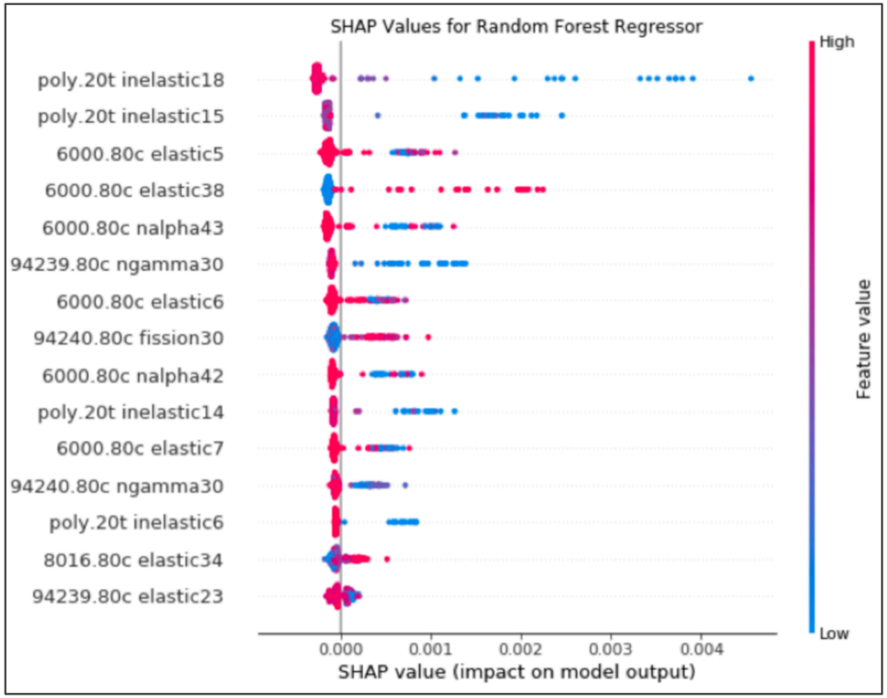
Bias prediction and feature importance

- With the bias known for all Whisper-1.1 catalogue cases, the generalized model predictions are promising



- What are the underlying hidden patterns in the nuclear data that are most related to the bias predictions?

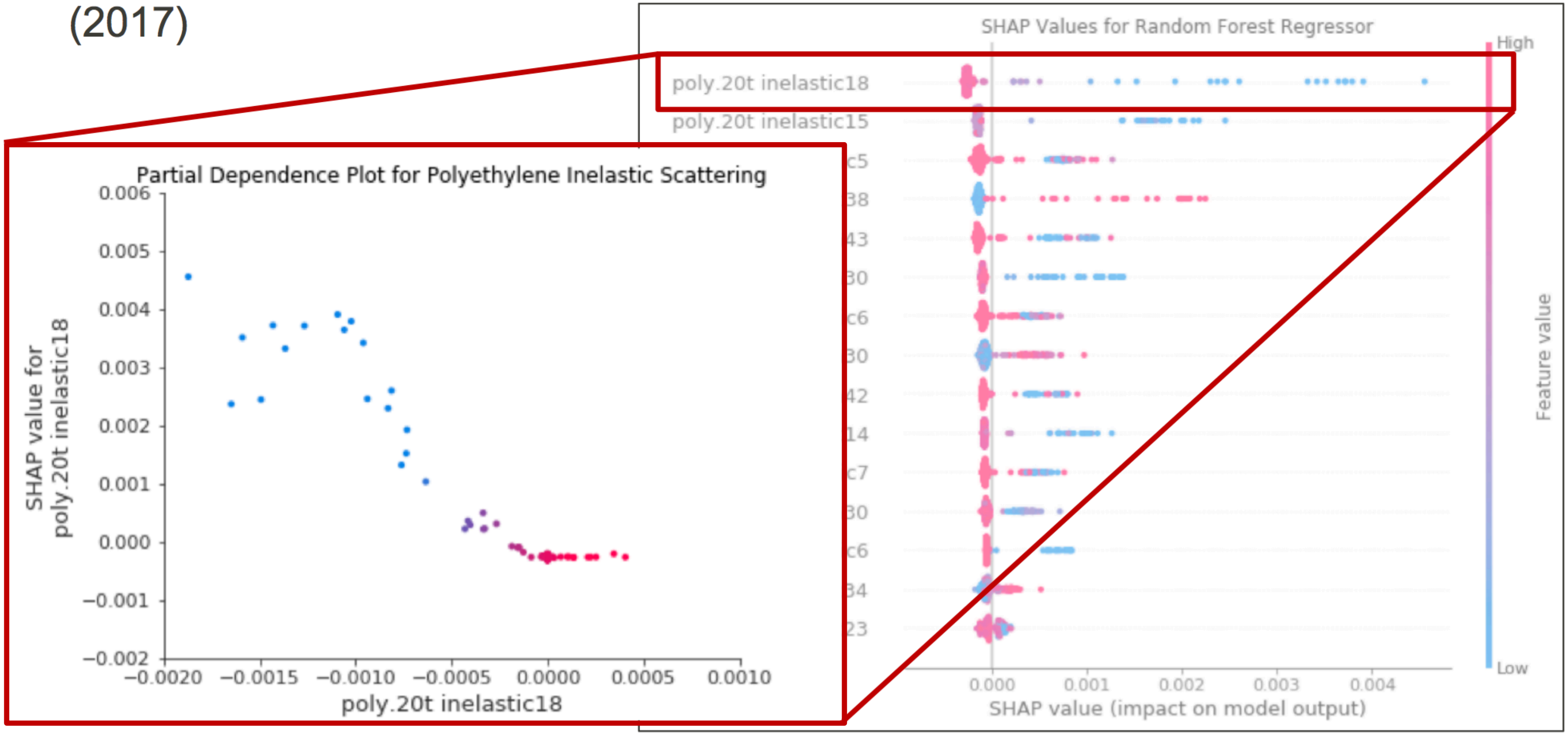
- From the machine learning methods, **feature importance** can be used to identify what nuclear data has a high likelihood to be related to the bias predictions



- Pu benchmarks used in this study

SHAP feature importance

- Shapley Additive exPlanation (SHAP) metric for feature importance
  - For each benchmark, estimate the additive contribution to the predicted bias for each feature
  - "A Unified Approach to Interpreting Model Predictions" Lundberg, Lee (2017)



Clustering benchmarks based on sensitivity profiles

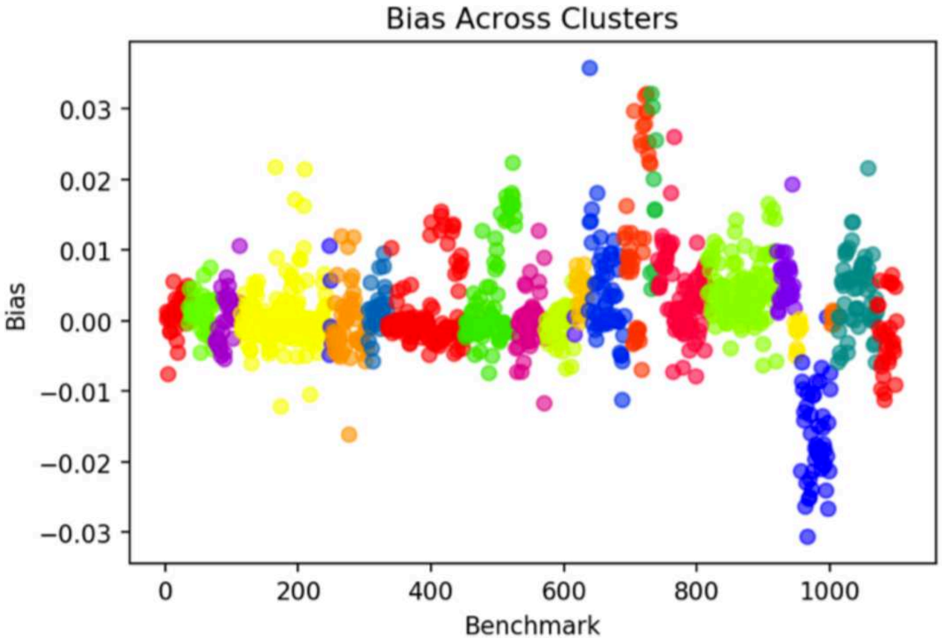
- Clustering is used to find inherent relationships in the data
  - Objects in the same cluster are more similar to each other than those in other clusters
  - Used to find groups of benchmarks that have similar sensitivity profiles, $S^i_{k,\sigma}$

- Can train and test on a few clusters at a time
  - Well populated classes of benchmarks skew the overall model
  - Training and testing on a subset of the data leads to a more specialized and accurate model
  - More accurate model $\leftrightarrow$ More accurate feature importance

- Can use clustering to find similar benchmarks for:
  - Benchmark selection for statistical analysis in Whisper
  - Finding regions in sensitivity space that are sparse (more benchmarks needed, see cluster #11 with mix-comp-fast on next slide)

Clustering Whisper benchmarks

- Finds 24 clusters ranging in population from 2 to 133
  - Segregated mainly based on materials present and spectrum



Cluster	Number of Cases	Benchmark Types
0	33	heu-met-fast
1	41	heu-met-fast, heu-met-mixed
2	38	heu-met-fast
3	133	heu-met-fast
4	5	heu-met-inter
5	54	heu-sol-therm, leu-comp-therm, u233-comp-therm
6	29	heu-met-fast, ieu-met-fast
7	117	leu-comp-therm, heu-comp-therm, heu-met-therm
8	77	heu-comp-therm, leu-comp-therm, heu-sol-therm
9	44	leu-comp-therm, heu-sol-therm
10	43	heu-sol-therm, leu-sol-therm
11	2	mix-comp-fast
12	20	mix-met-fast
13	54	pu-sol-therm, mix-sol-therm, mix-comp-therm
14	39	pu-comp-mixed, pu-sol-therm
15	11	pu-comp-mixed, pu-met-fast
16	75	pu-met-fast, mix-met-fast
17	105	pu-sol-therm, mix-sol-therm, mix-comp-therm
18	26	pu-sol-therm, mix-sol-therm,
19	10	u233-met-fast
20	45	u233-sol-therm, u233-sol-inter
21	10	u233-sol-therm
22	60	u233-sol-therm
23	29	u233-sol-therm, u233-comp-therm

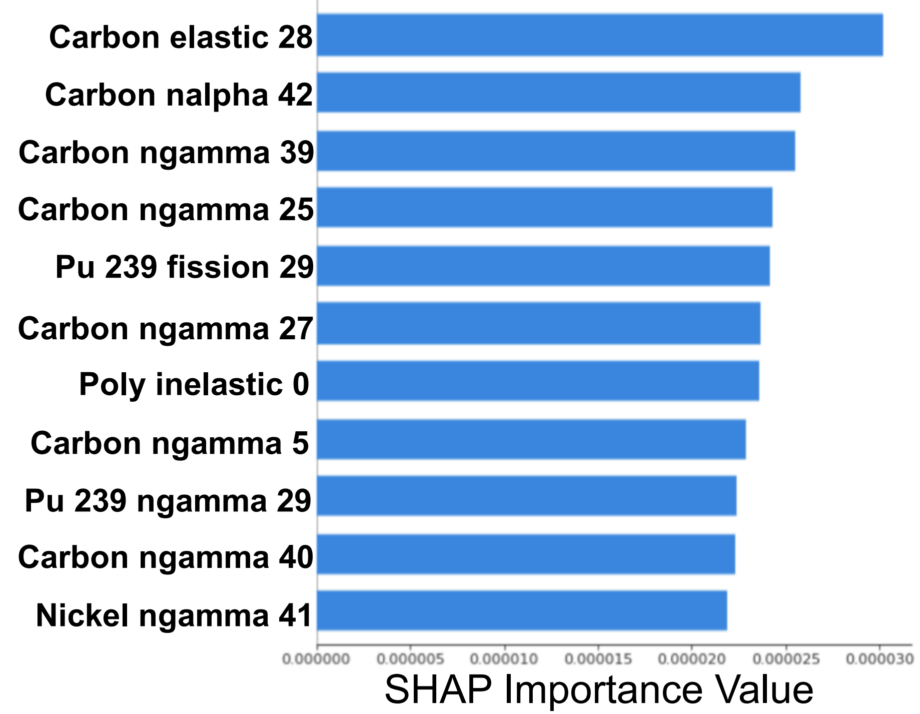
Benchmark clustering, bias predictions, feature importance and nuclear data assimilation

- In FY18, we took a preliminary look at combining several machine learning algorithms together in order to understand where they can be applied within the nuclear data evaluation process
  - Identify a cluster of benchmarks to study, e.g. PMF systems
  - Build a random forest model to predict the k_{eff} bias within this cluster
  - Select the most important features to predicting the bias
  - Apply genetic algorithm to optimize perturbations of the most important features
- We will continue scoping out these methods to find robust methodologies to support nuclear data evaluations

SHAP analysis points to potential deficiencies in ENDF/B-VII.0

- Recent results from Mike G. using Random Forest & SHAP analysis
- This work focuses on finding known and unknown issues in ENDF/B-VII.0 to test ML algorithms capabilities. Once that is successful, we will do production runs validating ENDF/B-VIII.0
- Used all Pu benchmarks
- Carbon: known issue because of questionable benchmarks
- $^{239}Pu(n,\gamma)$ in energy-bin 29 fixed in ENDF8

PRELIMINARY: Questionable nuclear data related to small-scale exp.



Using integral experiment metadata as feature importance

- If only sensitivity profiles are used, only nuclear data observables will be highlighted as related to bias between simulated and experimental benchmark value

- Experiments can be characterized by their MCNP input deck (metadata): geometry, material composition, source definition, analysis method, etc.

- The final result, e.g., difference between simulated and measured k_{eff}, is a function of nuclear data **and** experimental metadata

Conclusions & Future Work

- Using the Whisper-1.1 catalogue of 1100+ criticality safety benchmarks, several machine learning methods were applied to predict k_{eff} bias, cluster similar benchmarks together, find highly important features in the nuclear data and optimize perturbations to important cross sections.

- Need to examine all of the machine learning results more closely, especially the *initial* nuclear data adjustment results
  - Comparison to GLLSM is needed (already underway)
  - Inclusion of the nuclear data covariance is essential
    - The accuracy of the nuclear data covariance data could be problematic

- Using more features of the benchmarks (metadata) is being explored to see if they can help in clustering benchmarks or finding systematic outliers

Acknowledgements

This work was supported by the DOE Nuclear Criticality Safety Program, funded and managed by the National Nuclear Security Administration for the Department of Energy.

Questions ?