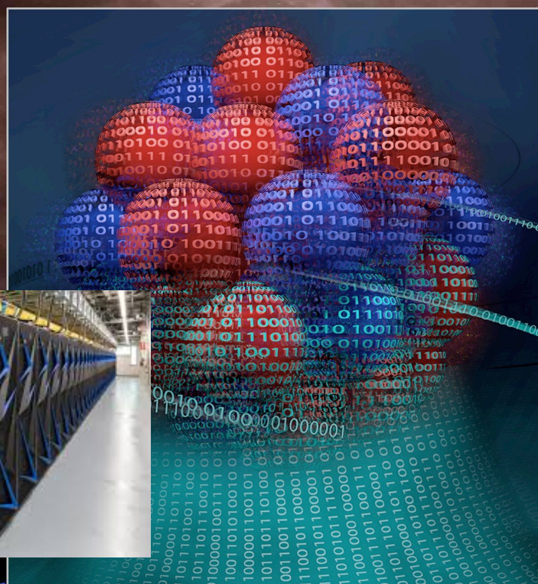
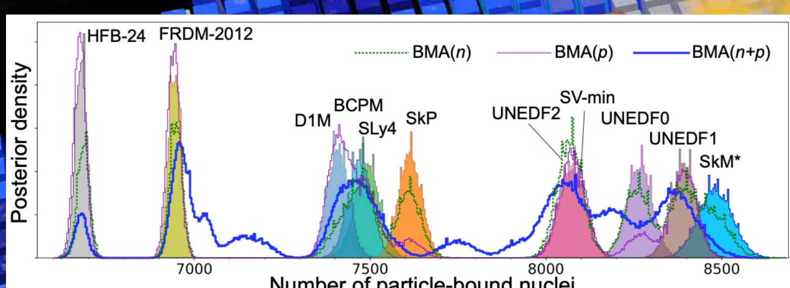
Bayesian Model Mixing: Nuclear Physics Applications

Witold Nazarewicz Michigan State University/FRIB

Advances in Many-Body Theories: from First Principles Methods to Quantum Computing and Machine Learning, ECT*, November 2-6, 2020



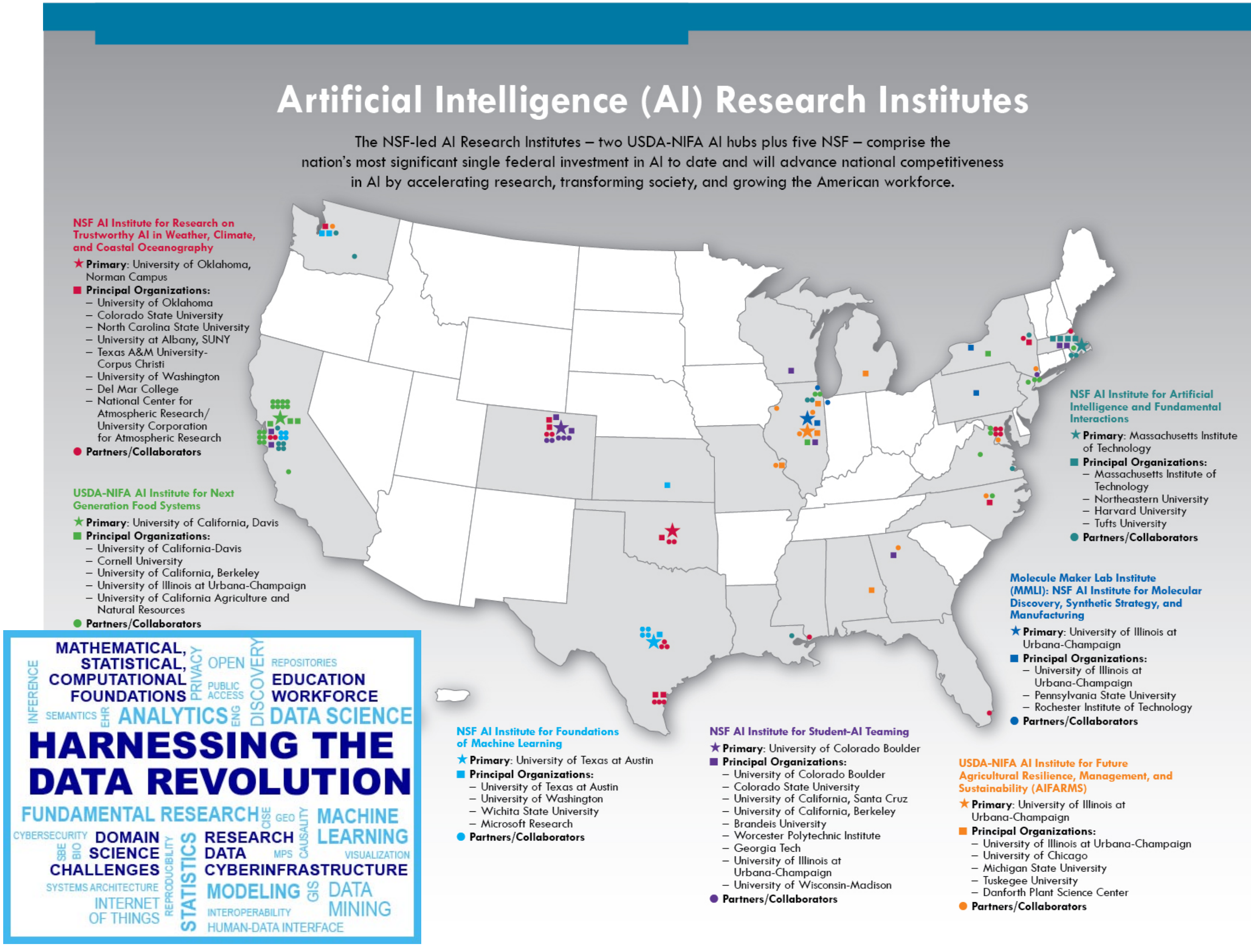


- Preliminaries
- ML initiatives in the US
- Bayesian machine learning
- Bayesian model mixing
- Examples of recent work
- Perspectives

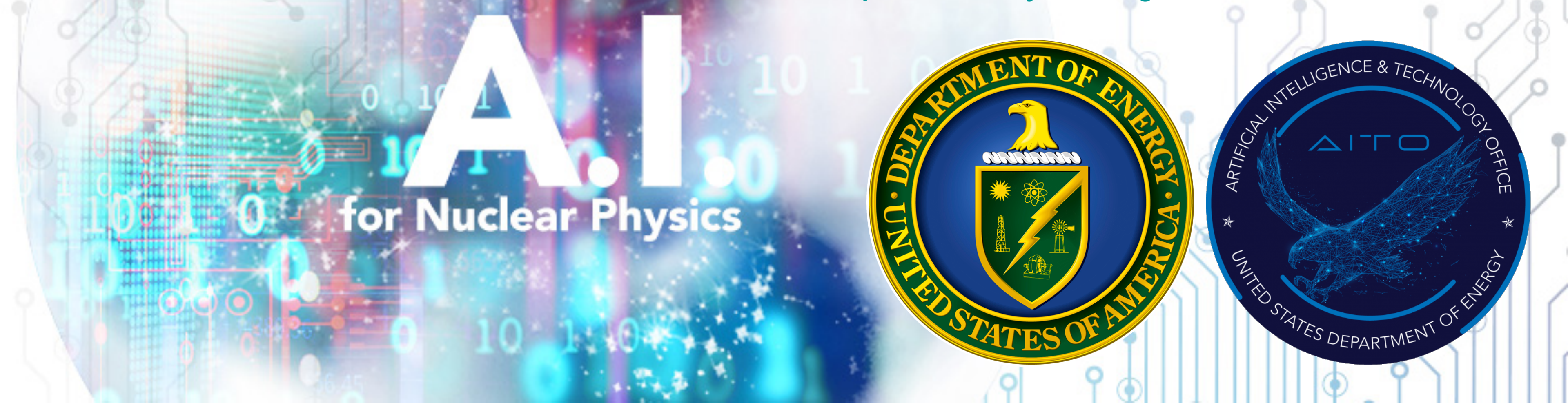
NUCLEI Nuclear Computational Low-Energy Initiative

BAND Bayesian Analysis of Nuclear Dynamics

August 2020: 6 AI NSF institutes ($20 million over five years). In 2021, additional 6 will be selected (in partnership with private sector: Accenture, Amazon, Google and Intel)



https://www.jlab.org/conference/AI2020



March 4-6, 2020, Thomas Jefferson National Accelerator Facility

- Explore the ways in which A.I./ML can be used to advance research in nuclear physics and in the design and operation of large-scale accelerator facilities.
- Explore applications and research needed on several time frames, ranging from immediate benefit.
- The results of the workshop have been summarized in a report* which contains a reasonable assessment of current efforts.

---

*arXiv:2006.0542 and EPJA, in press

Machine learning & low-energy nuclear theory: Why?

ML tools can help us to speed up the scientific process cycle and hence facilitate discoveries

- Enabling fast emulation for big simulations
- Revealing the information content of measured observables w.r.t. theory
- Identifying crucial experimental data for better constraining theory
- Providing meaningful input to applications and planned measurements

ML tools can help us to reveal the structure of our models

- Parameter estimation with heterogeneous/multi-scale datasets
- Model reduction

ML tools can help us to provide predictive capability

- Theoretical results often involve ultraviolet and infrared extrapolations due to Hilbert-space truncations
- Uncertainty quantification essential
- Theoretical models are often applied to entirely new nuclear systems and conditions that are not accessible to experiment

This talk: focus on Bayesian Machine Learning (BML)

Explosion of papers on machine learning in theoretical nuclear structure/reactions

1992 Early neural network applications

2016-2020

- Machine learning for missing data interpolations
- Emulators with neural networks
- Neural networks in ab-initio theory
- Model calibration and sensitivity analysis
- EFT applications
- Network motif studies
- Phase transitions
- Bayesian emulators
- Bayesian neural network extrapolations
- Bayesian uncertainty quantification
- Bayesian model averaging
- Bayesian modeling of neutron stars and EOS
- Experimental design

Many presentations at the 2020 DNP meeting

ABC of Bayesian inference

Bayes' Theorem*:

$$\underbrace{P(A|B)}_{\text{posterior}} = \frac{\overbrace{P(B|A)}^{\text{likelihood}}\overbrace{P(A)}^{\text{prior}}}{\underbrace{P(B)}_{\text{evidence}}}$$

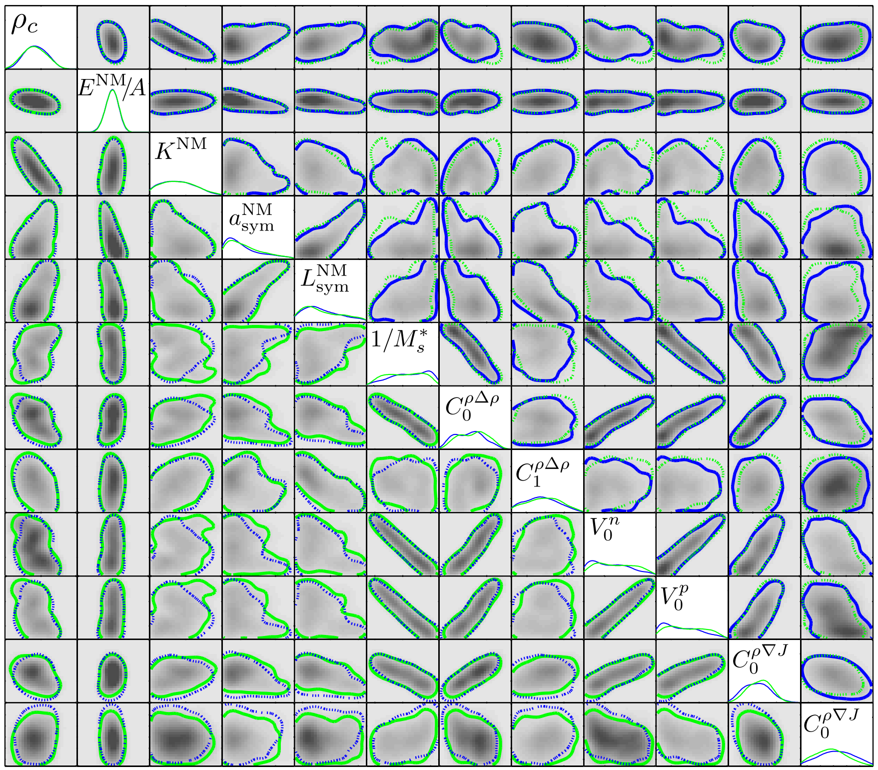
- Posterior: the degree of belief in A after incorporating news that B is true. Posterior probability is obtained from a prior probability, given evidence.
- Likelihood: measures the goodness of fit of a model to a sample of data for given values of the parameters.
- Prior: initial degree of belief in A
- Evidence: probability of B; this factor is the same for all possible hypotheses being considered.

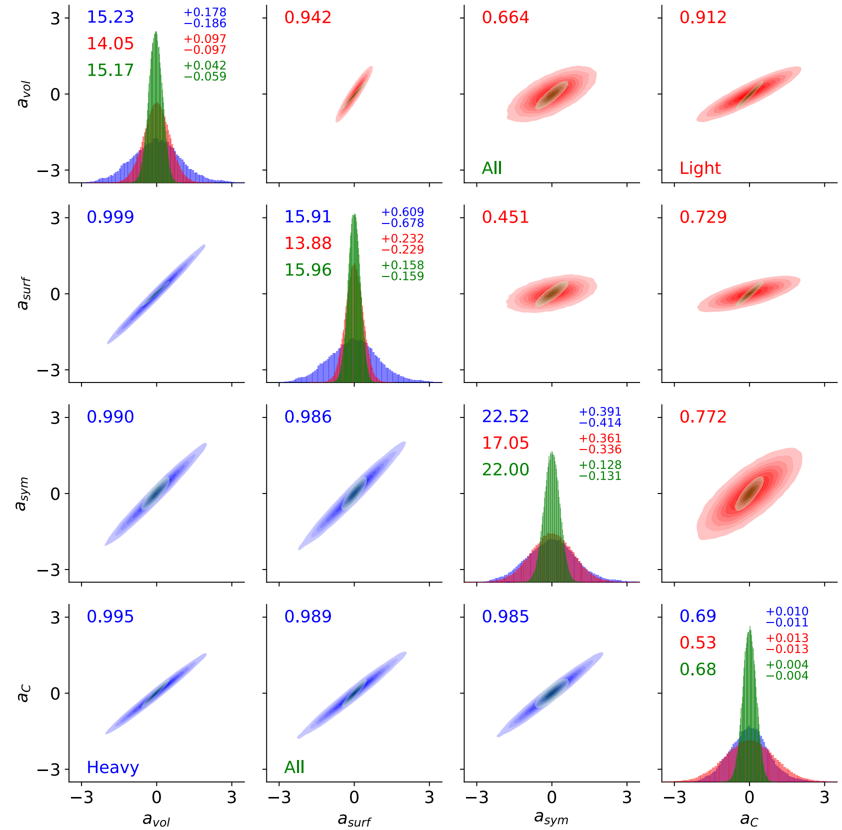
*Thomas Bayes, *An Essay towards solving a Problem in the Doctrine of Chances*, 1763

Emulation and parameter estimation
Probability distribution functions (PDFs)

McDonnell et al.
Phys. Rev. Lett. 114, 122501 (2015)

Kejzlar et al., J. Phys. G (2000)
arXiv:2002.04151





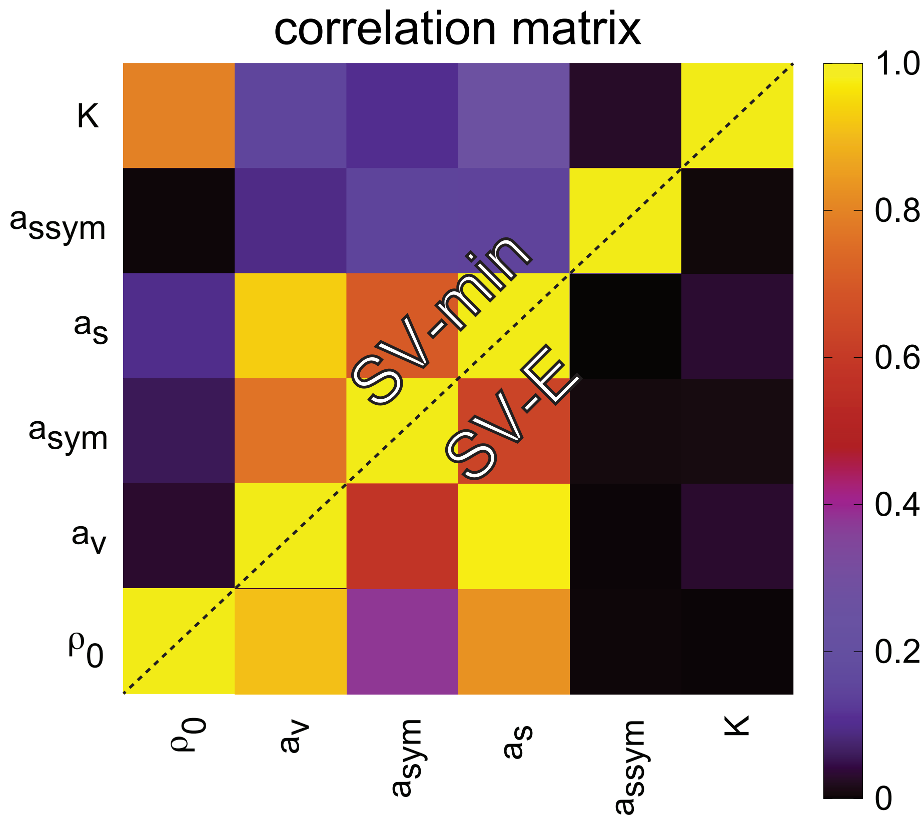
Bivariate marginal estimates of the posterior distribution for the 12-dimensional DFT UNEDF1 parameterization.

Posterior distributions of the model parameters for LDM variants ⇒ LDM is a one-parameter model.

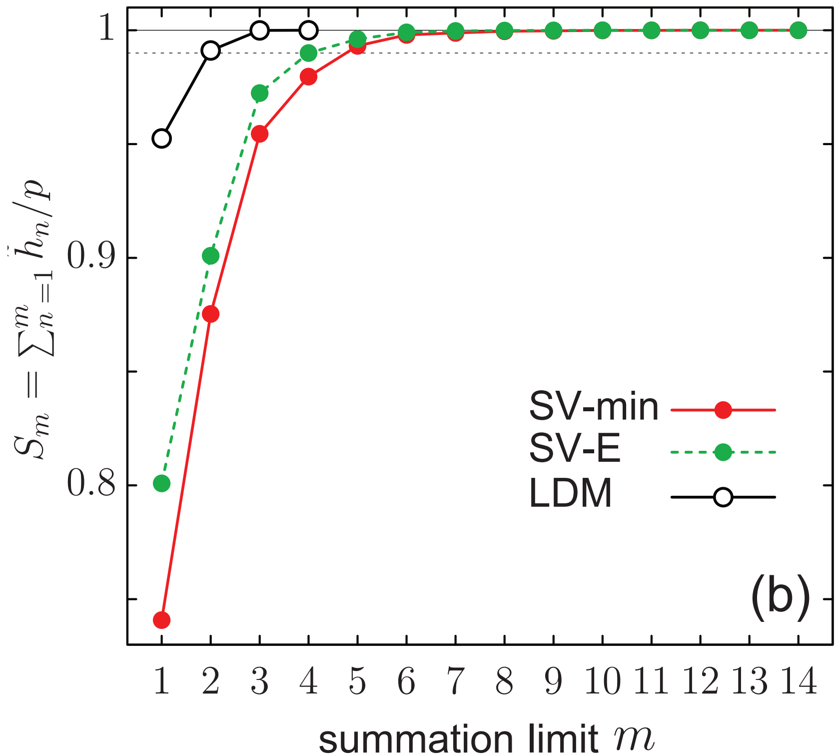
Model calibration and model reduction

Kejzlar et al., J. Phys. G (2000) arXiv:2002.04151

SV-min: informed by masses, sizes, pairing gaps
SV-E: informed by masses only



Model reduction via principal component analysis

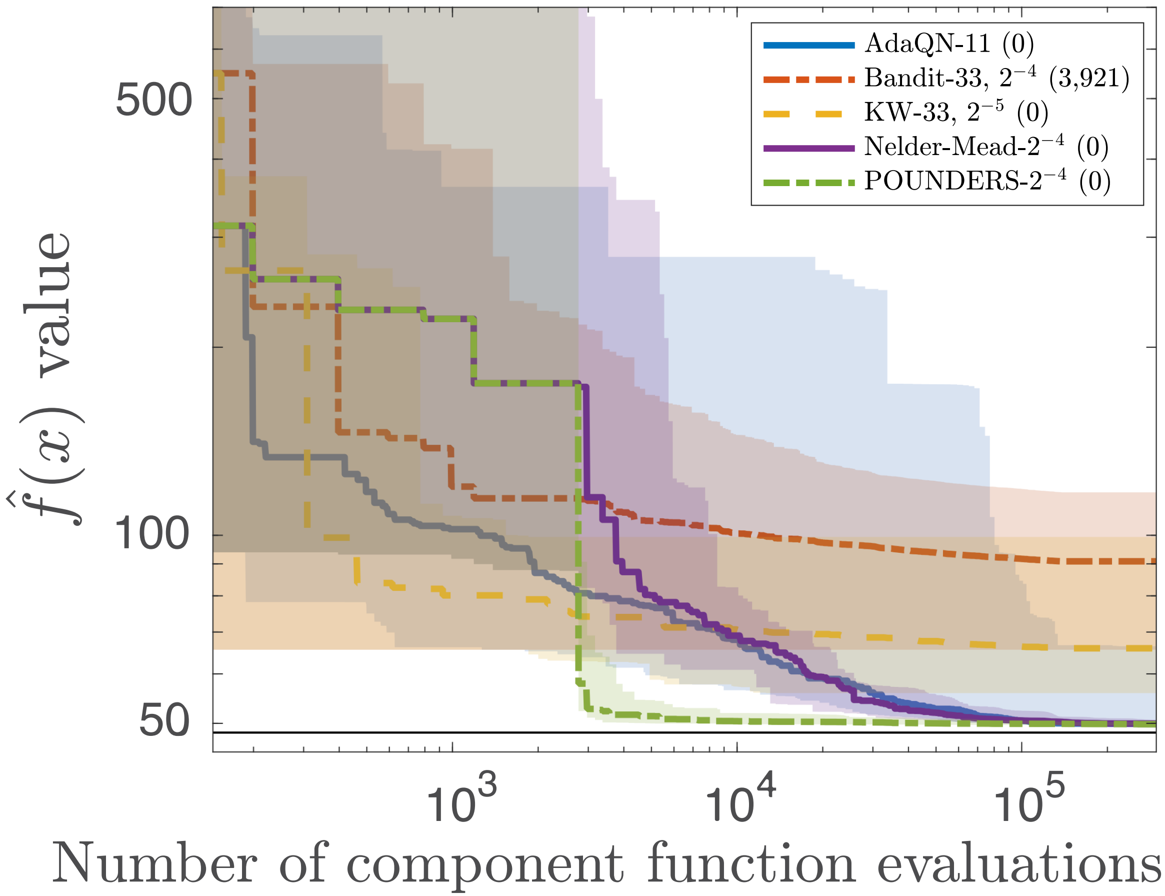


Conclusion: correlations between parameters and observables strongly depend on dataset of fit-observables. Heterogenous datasets are important!

Conclusion: effective number of degrees of freedom is 4-6 for the 14-parameter Skyrme functional. Long way to go!

Optimization and machine learning training algorithms for fitting numerical physics models, R. Bollapragada et al., arXiv:2010.05668 (2020)

The calibration of a computationally expensive nuclear physics model for which derivative information is not available. The performance of optimization-based training algorithms when dozens, rather than millions or more, of training data are available and when the expense of the model places limitations on the number of concurrent model evaluations that can be performed.

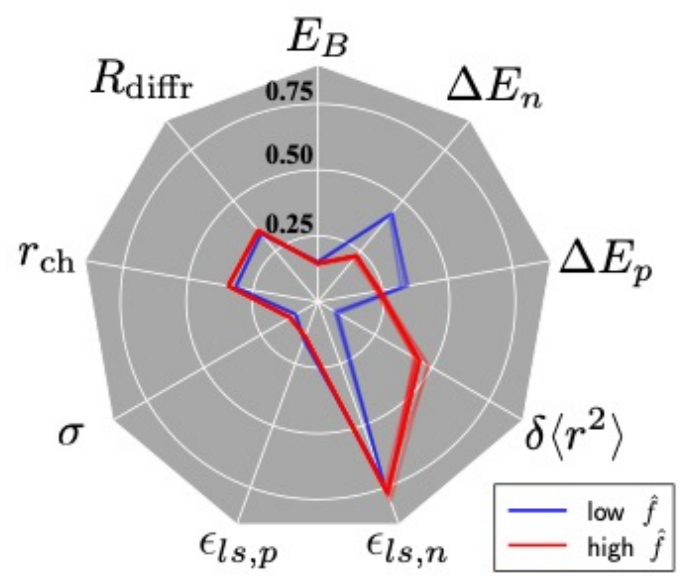


Deterministic algorithms:

- Nelder-Mead, POUNDERS

Stochastic optimization:

- Kiefer-Wolfowitz, Bandit, adaptive quasi-Newton



BML and quantified extrapolations

Residual of an observable O:

$$\delta_{\mathcal{O}}(Z,N) = \mathcal{O}^{\rm exp}(Z,N) - \mathcal{O}^{\rm th}(Z,N)$$ small number!

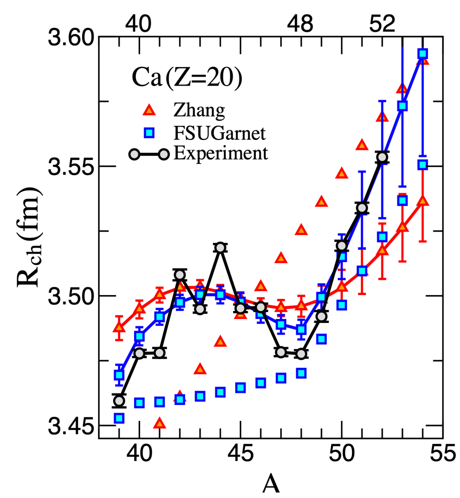
$|\delta_{\mathcal{O}}| \ll |\mathcal{O}|$ Smooth part of the residual represents missing physics (systematic effects)

Estimate of an observable O:

$$\mathcal{O}^{\rm est}(Z,N) = \mathcal{O}^{\rm th}(Z,N) + \delta_{\mathcal{O}}^{\rm em}(Z,N)$$

emulator of the residual

Supervised learning: the nuclear modeling and the choice of priors represent two aspects of the supervision

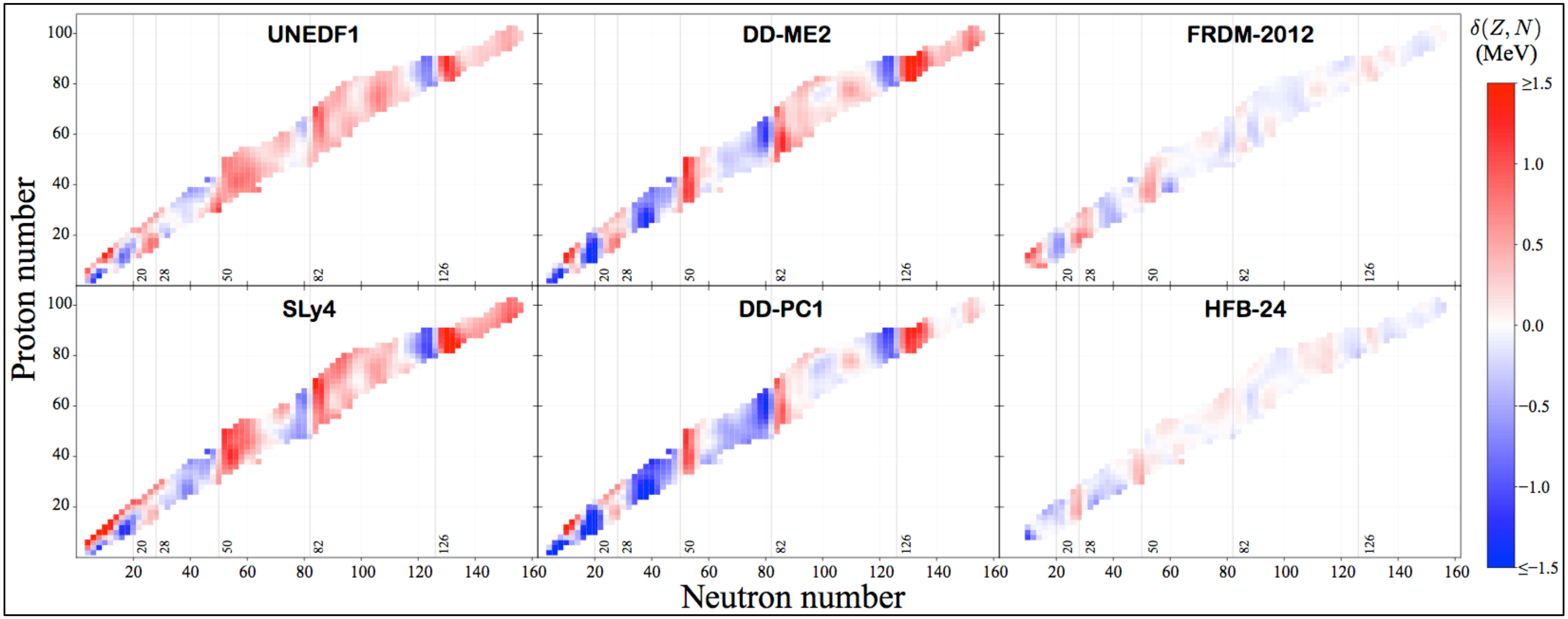


Nuclear radii with BNN
Utama, Chen, and Piekarewicz
J. Phys. G 43 114002(2016)

Residuals (based on data and theory) exhibit patterns

Mass extrapolations with BNN and GP
Neufcourt et al. Phys. Rev. C 98, 034318 (2018)

S_{2n} residuals for models of different fidelity



- This information can be used to our advantage to improve model-based predictions!
- It can also be used to improve models themselves

Bayesian approach

residual $y_i = d(x_i, \theta) + \sigma\epsilon_i$

$(Z,N)_i$

discrepancy model (systematic error) statistical error

- Kennedy and O'Hagan, J. Royal Stat. Soc. B, 63425 (2001)
- Higdon et al., SIAM J. Sci. Comput. 26448–466 (2004)

$$p(y^*|y) = \int p(y^*|y, \underbrace{\theta, \sigma) p(\theta, \sigma}_{\text{marginalization of the model parameters}}|y)\, d\theta d\sigma$$

Prediction of unknown observable *y** given known data *y*

Two statistical models used:

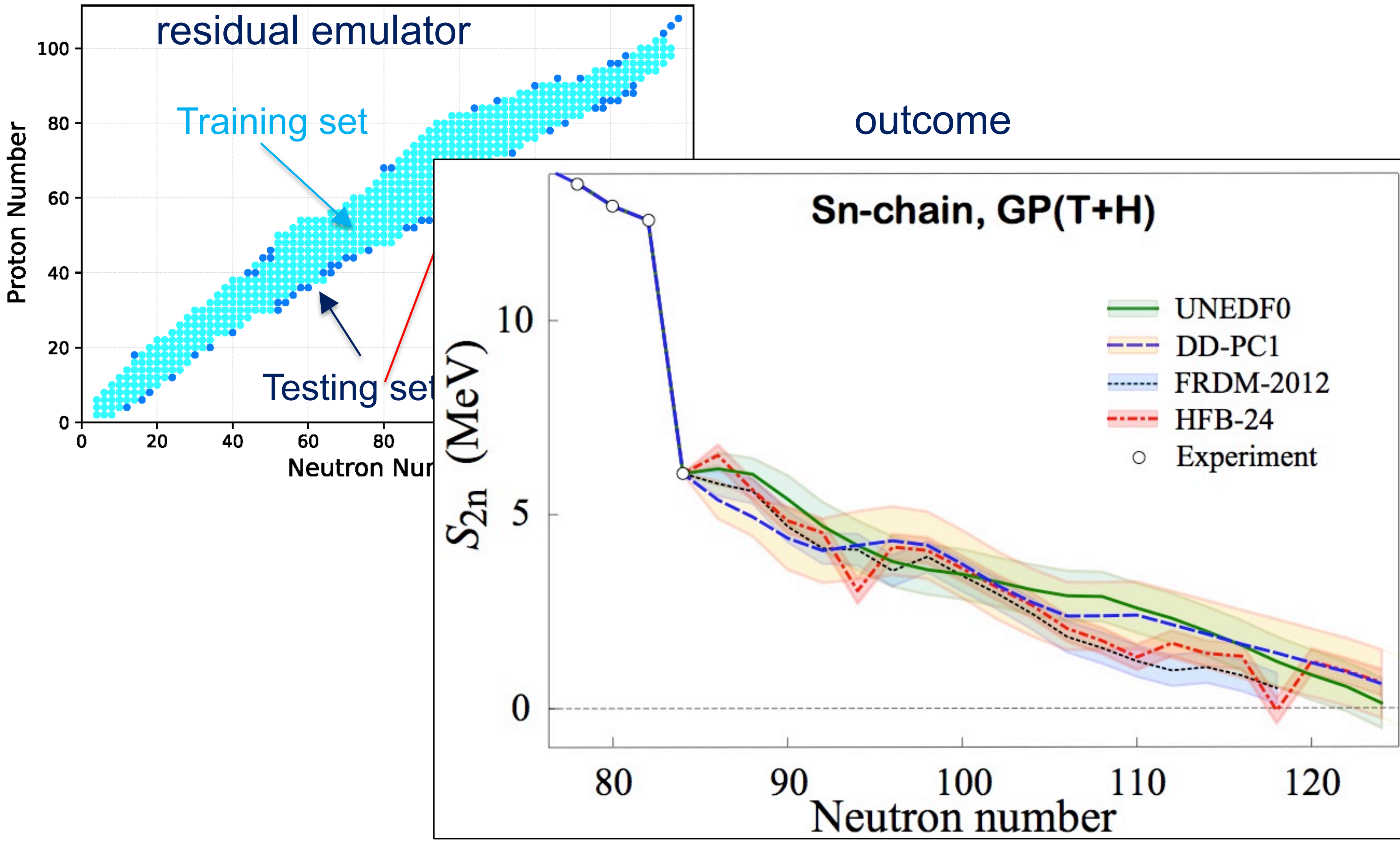
- Gaussian process (**3** parameters)
- Bayesian neural network with sigmoid function (30 neurons, 1 layer; **181** parameters)

100,000+ iterations of an ergodic Markov chain produced by the Metropolis-Hastings algorithm

Some refinements added based on our knowledge of trends (e.g. magic nuclei)

Mass extrapolations with Bayrsian machine learning

Neufcourt et al. Phys. Rev. C 98, 034318 (2018)



Naïve nuclear theorist's approach to a systematic (model) error estimate:

- Take a set of *reasonable* global models M_i, hopefully based on different assumptions/formalism, that satisfy basic theoretical requirements (here comes the expert belief thing).
- Make predictions.
- Compute average and variation within this set
- Compute rms deviation from existing experimental data.

Such a strategy can provide some clues…
⇒ simple BMA

Can we do better?

Bayesian model averaging (assumption: the perfect model is included in the set)

$$p(\mathcal{M}_k|y) = \frac{p(y|\mathcal{M}_k)\pi(\mathcal{M}_k)}{\sum_{\ell=1}^{K} p(y|\mathcal{M}_\ell)\pi(\mathcal{M}_\ell)}$$

Prediction:

$$p(y^*|y) = \sum_{k=1}^{K} p(y^*|y,\mathcal{M}_k)p(\mathcal{M}_k|y)$$

unknown data

Used in many fields

- Weather forecasting
- Political science
- Transportation
- Nuclear physics
- ...

RESEARCH ARTICLE | 1 MAY 2005

Using Bayesian Model Averaging to Calibrate Forecast Ensembles

Adrian E. Raftery; Tilmann Gneiting; Fadoua Balabdaoui; Michael Polakowski

Mon. Wea. Rev. (2005) **133** (5): 1155–1174.

https://doi.org/10.1175/

Bayesian Model Averaging: Theoretical Developments and Practical Applications

Jacob M. Montgomery and Brendan Nyhan

Political Analysis
Vol. 18, No. 2 (Spring 2010), pp. 245-270 (26 pages)
Published By: Cambridge University Press

Bureau of Transportation Statistics

Topics and Geography | Statistical Products and Data | National Transportat

Home » Archive

Application of the Bayesian Model Averaging in Predicting Motor Vehicle Crashes

YAJIE ZOU
DOMINIQUE L
YUNLONG ZHA
YICHUAN PEN

PHYSICAL REVIEW LETTERS

Highlights | Recent | Accepted | Collections | Authors | Referees | Sea
Staff

Neutron Drip Line in the Ca Region from Bayesian Model Averaging

Léo Neufcourt, Yuchen Cao (曹宇晨), Witold Nazarewicz, Erik Olsen, and Frederi Viens

Phys. Rev. Lett. **122**, 062502 – Published 14 February 2019

Statistics > Methodology

[Submitted on 3 Aug 2020 (v1), last revised 24 Aug 2020 (this version, v2)]

Bayesian model averaging for analysis of lattice field theory results

William I. Jay, Ethan T. Neil

Statistical modeling is a key component in the extraction of physical results from lattice field theory calculations. Although
considered for the same lattice data. Model averaging, which amounts to a probability-weighted average over all model var
the perspective of Bayesian statistics, and give useful formulas and approximations for the particular case of least-squares
time separation for fitting a two-point correlation function) as a model selection problem, and study model averaging as a

More advanced: Bayesian model mixing

Assumption: exact model can be represented by an average over models:

$$y^*(x) = \sum_{k=1}^{p} \omega_k^*(x) f_k(x).$$

Bayesian model averaging: exploratory phase

Questions:

- How to choose models?
- How to chose the likelihood?
- How to select model weights?
- How to eliminate “redundant” models?

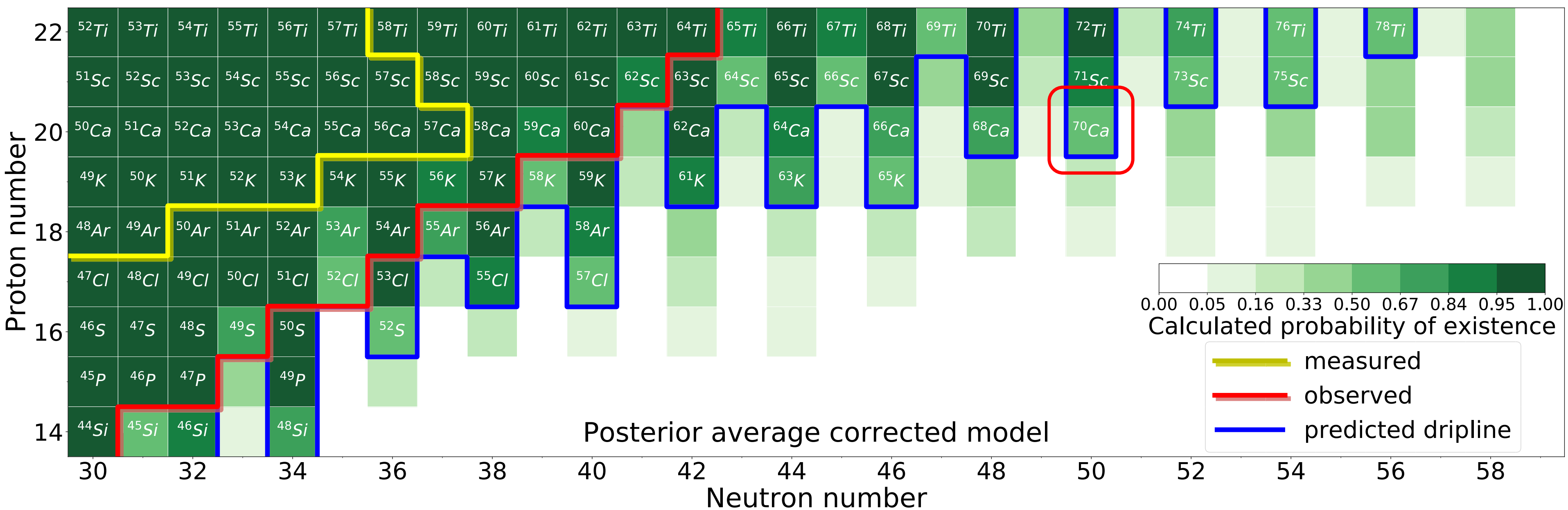
$$p(\mathcal{M}_k|y) = \frac{p(y|\mathcal{M}_k)\pi(\mathcal{M}_k)}{\sum_{\ell=1}^{K} p(y|\mathcal{M}_\ell)\pi(\mathcal{M}_\ell)}$$

Quantified predictions with BMA

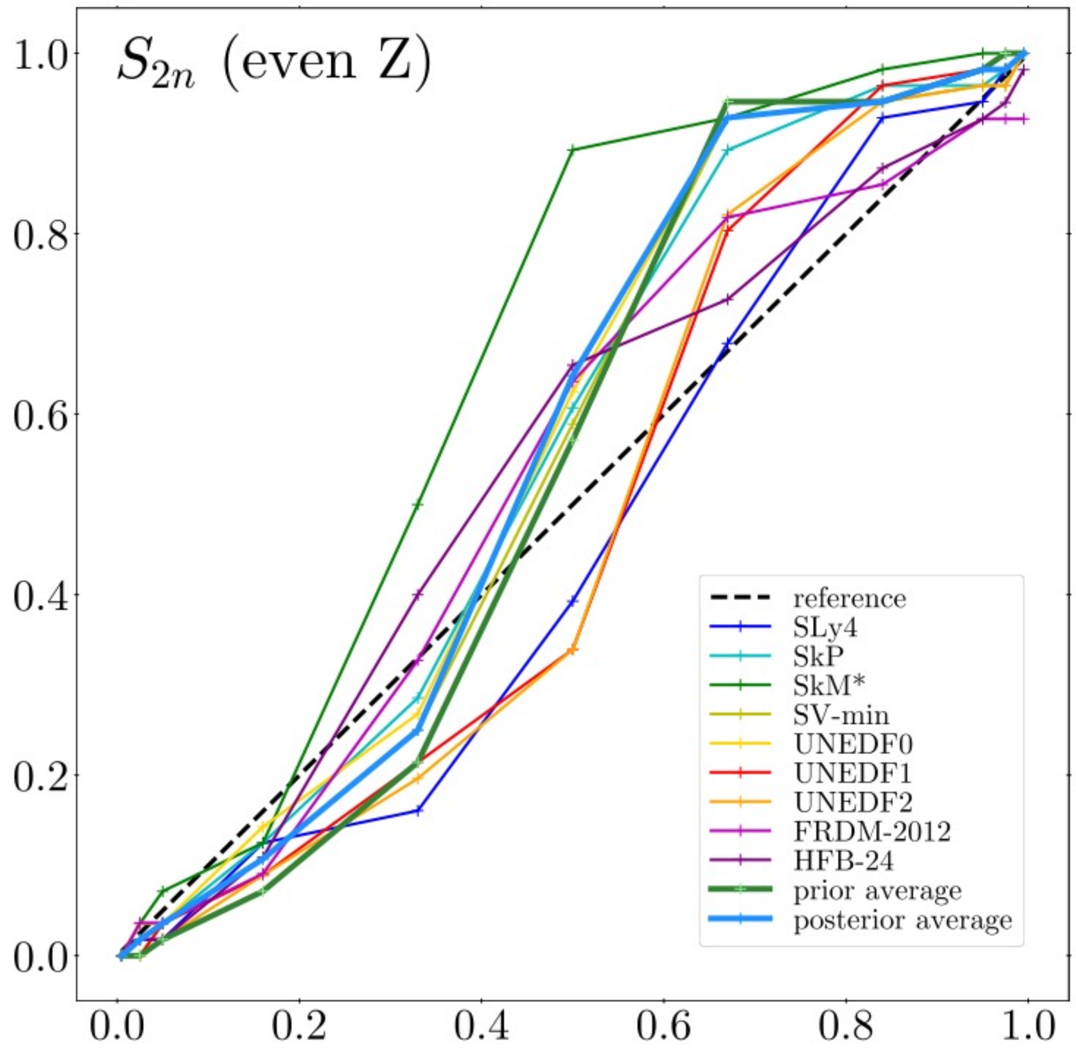
Probability of existence

$$p_{ex}(Z,N) := p(S^{*}_{1n/2n}(Z,N) > 0 | S_{1n/2n})$$

Bayesian model averaging, see L. Neufcourt et al., Phys. Rev. Lett. 122, 062502 (2019)



Diagnostic tools: empirical coverage probability



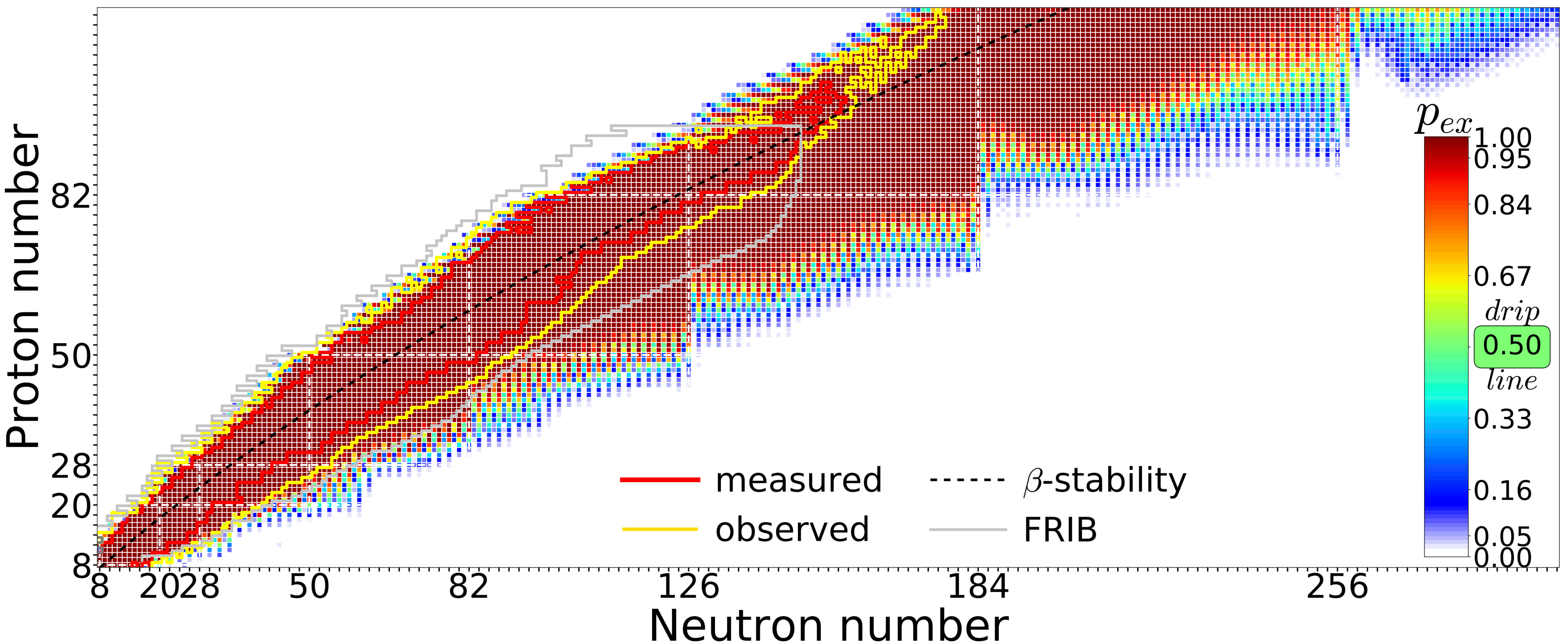
The ECP is a simple metric for assessing the quality of a statistical model's UQ. The ECP curve corresponds to the proportion of the testing data which actually falls inside the predicted credibility intervals (CIs) as a function of the credibility level. For perfect uncertainty quantification, one would obtain a straight line. The matching of the nominal value is overall satisfactory, with an inflection point at the middle of the curve: the CIs are slightly too optimistic at low credibility levels, and slightly too conservative at (most important) high credibility levels.

Quantified limits of the nuclear landscape

Neufcourt et al., Phys. Rev. C 101, 044307 (2020)

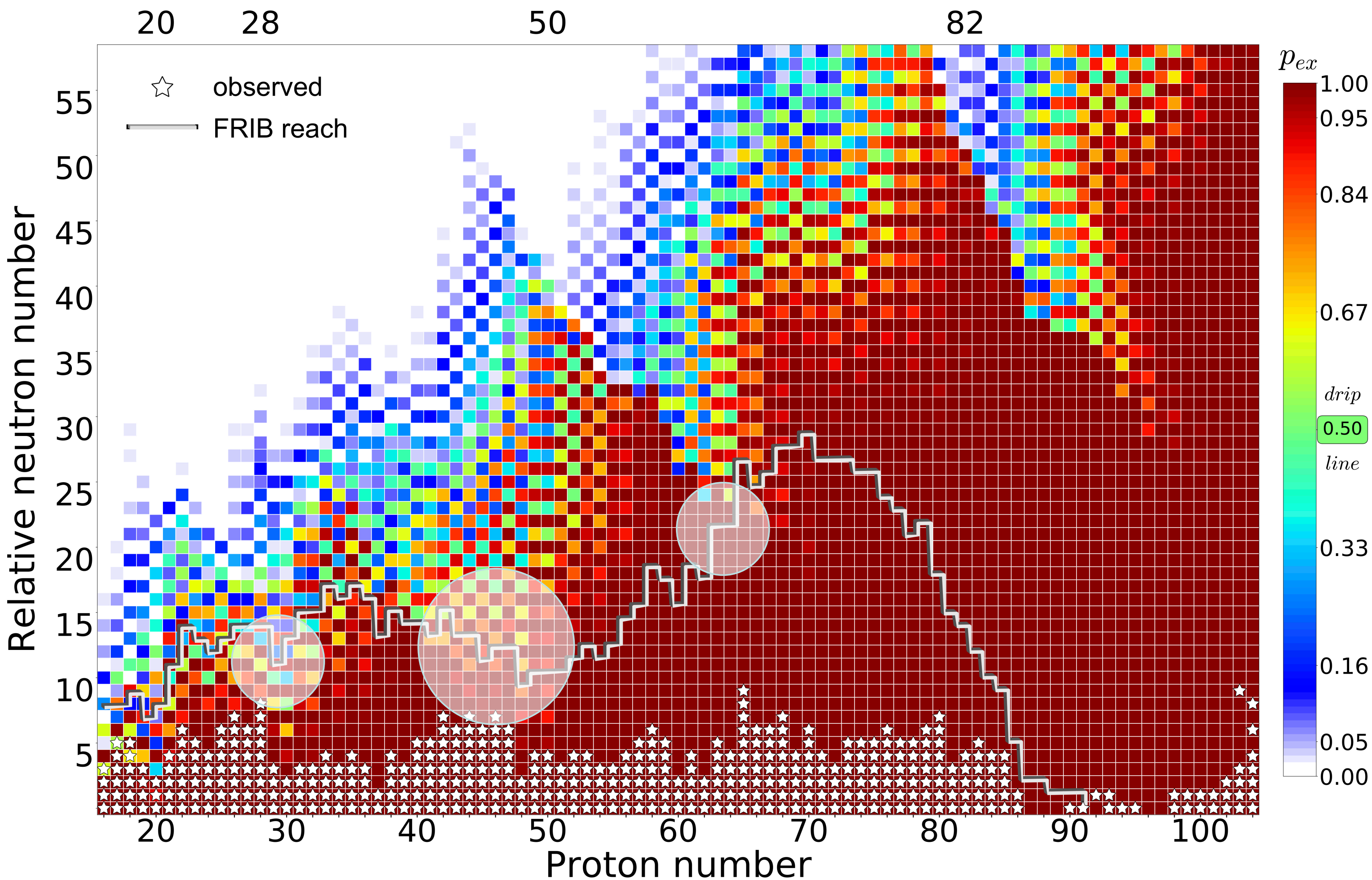
Predictions made with 11 global mass model and Bayesian model averaging

$$p_{\text{ex}} := p(S^*_{1p/2p/1n/2n} > 0 | S_{1p/2p/1n/2n})$$



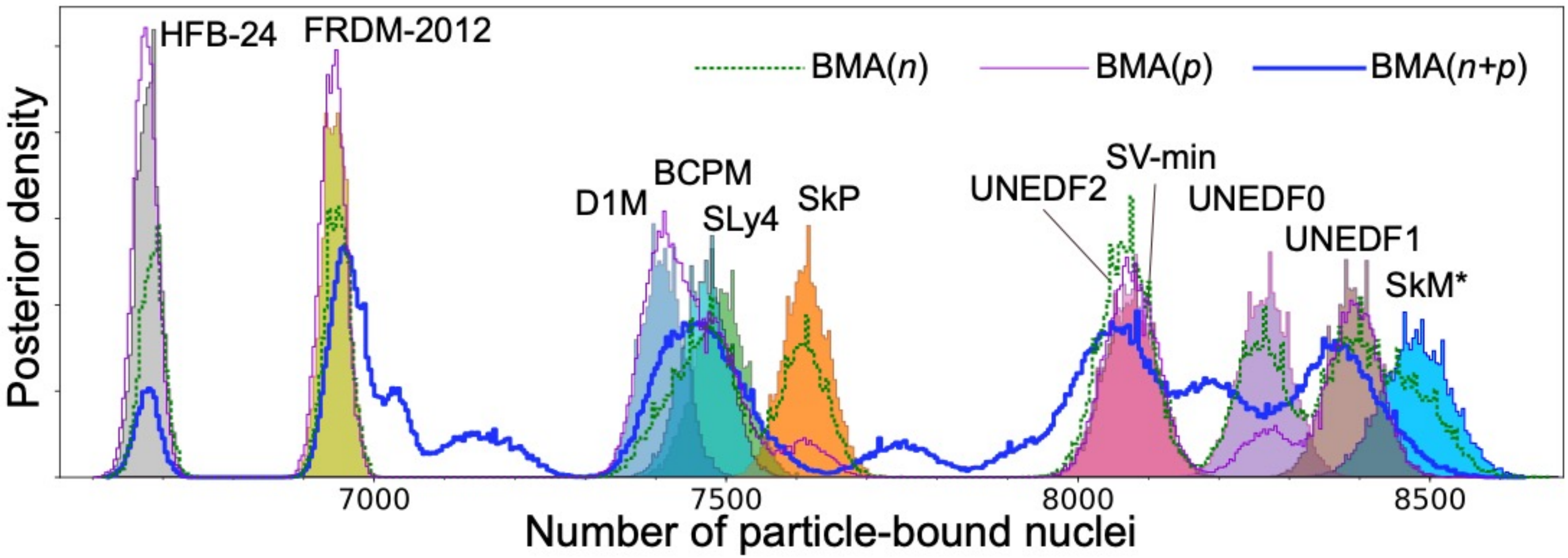
The FRIB production rates estimated with the LISE++ . We assumed the experimental limit for the confirmation of existence of an isotope to be 1 event/2.5 days.

Of particular importance for constraining theory are the existence data for Z=28-30, Z=42-48, and Z=64-66



"0" corresponds is the neutron number of the heaviest isotope for which an experimental separation energy value is available

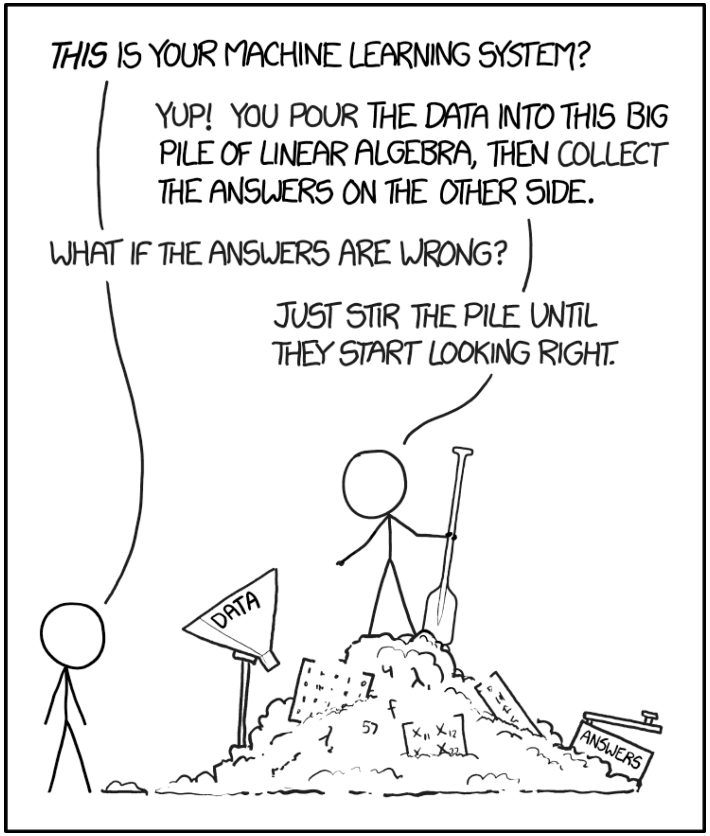
Posterior distribution functions



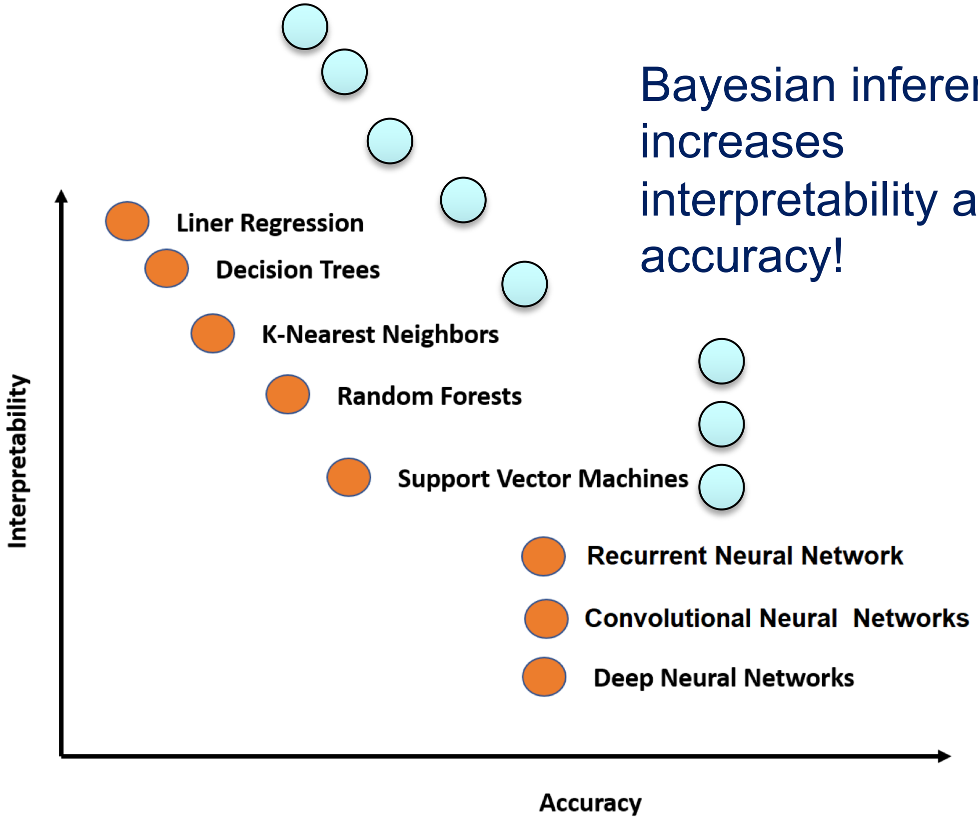
Typical situation: spread of model predictions

Extrapolations and Model Interpretability

Disadvantage of deep learning methods is their “black box” nature: the computationally-advanced methods by which these methods come up with the convolved output is not readily understandable (arxiv.org/1904.08067)



https://xkcd.com/1838/



Bayesian inference increases interpretability and accuracy!

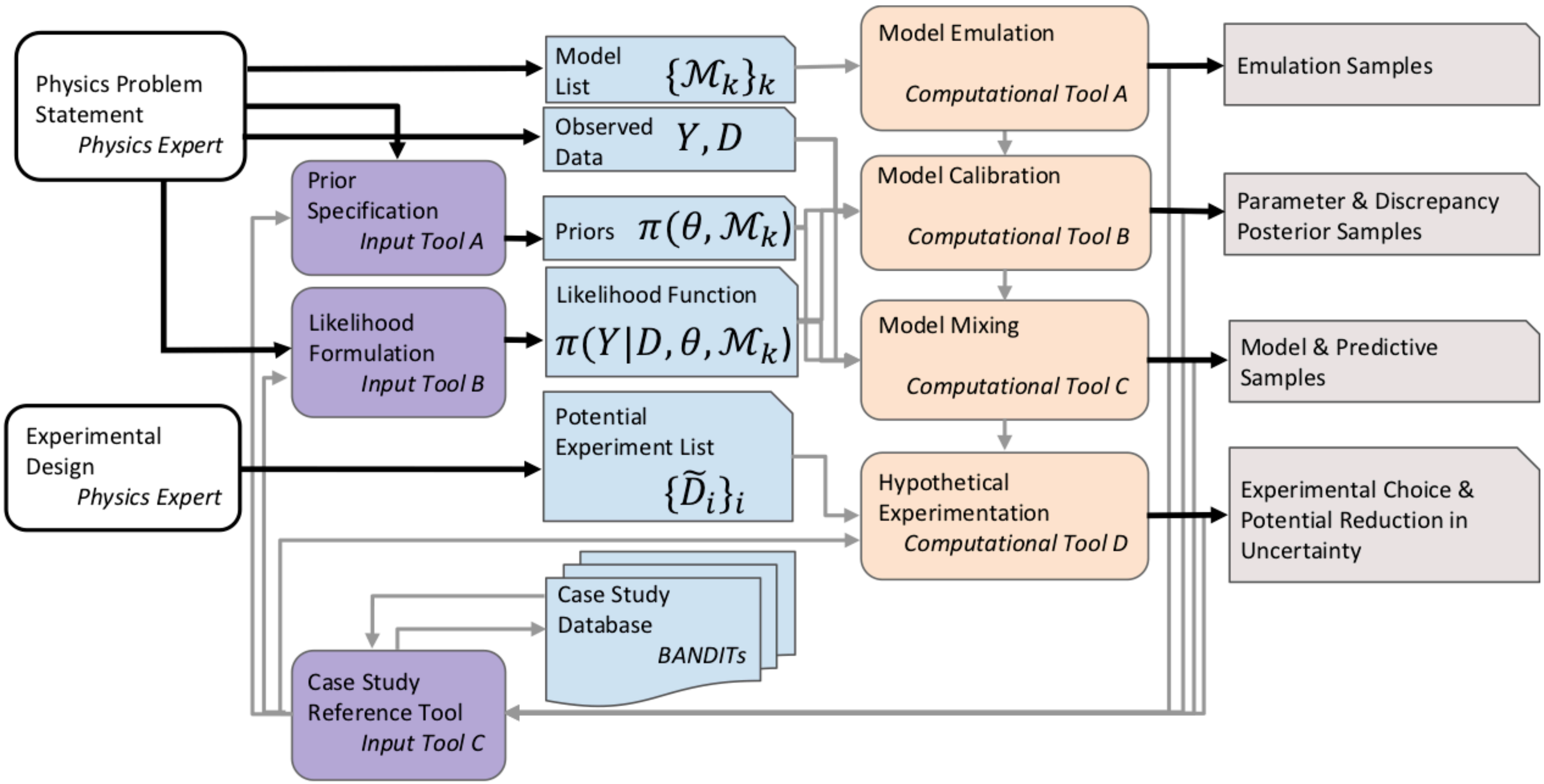
arxiv.org/1904.08067

Cyberinfrastructure for Sustained Scientific Innovation Framework

BAND
Bayesian Analysis of Nuclear Dynamics

https://bandframework.github.io

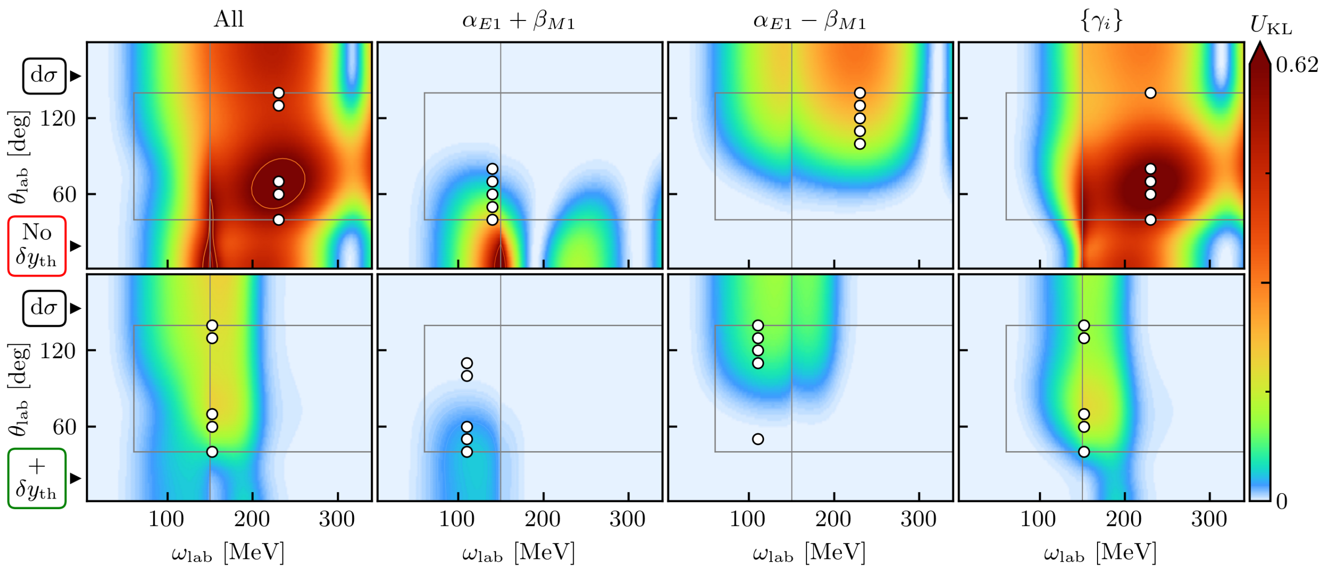
Ohio U.
Michigan State U.
Ohio State U.
Northwestern U.



Experimental design

Beam time and compute cycles are expensive!

- Bayesian experimental design provides a framework in which experiments can be designed using the best experimental and theoretical information available
- *The utility function* is designed to encode the goals of the experiment and the constraints inherent in carrying it out.
- Once the utility function and the possible designs have been specified, the optimal design is simply the scenario that maximizes the expected utility function over the domain of possible designs.



The expected utility of proton differential cross section measurements. The circles show the optimal design kinematics for five measurement points at the same energy but different angles.

Designing optimal experiments: An application to proton Compton scattering
J. A. Melendez et al., arXiv 2004.11307

BAND MANIFESTO (in preparation)

Summary

- Need for Uncertainty Quantification in nuclear physics.
- Much progress in this direction in last few years but still difficult to assess model uncertainty
- To solve many complex problems in the field and facilitate discoveries, multidisciplinary efforts efforts are required involving scientists in nuclear physics, statistics, computational science, and applied math.
- Bayesian Model Mixing provides uncertainty quantification for a nuclear-physics prediction, based on best available nuclear physics knowledge (both experimental and theoretical).
- The community needs to invest in relevant educational efforts.
  - Virtual Nuclear TALENT course on Machine Learning and Data Analysis for NP, ECT*, June 22-July 3, 2020.
  - Information and Statistics in Nuclear Experiment and Theory (ISNET). Virtual, Dec. 14-18, 2020, MSU, https://indico.frib.msu.edu/event/21/
  - The first Winter School on Applications of AI to Topics in NP. Virtual,11-15 January 2021 (CUA+UMD). The School will take place every 1-2 years and that the location will rotate.
  - TALENT 2021,…

Collaborators (current)

Physics

Y. Cao
J. Dobaczewski
D. Furnstahl
S. Giuliani
M. Hjorth-Jensen
Y. Jaganathen
D. Lee
D. Phillips
P.-G. Reinhard
…

Statistics

S. Bhattacharya
V. Kejzlar
T. Maiti
L. Neufcourt
M. Plumlee
M. Pratola
F. Viens
…

Applied math/CS

J. O'Neal
S. Wild
…

Thank You