TALENT COURSE 11
LEARNING FROM DATA: BAYESIAN METHODS AND MACHINE LEARNING

Lecture 27: Nuclear Physics Context 3

Daniel Phillips
Ohio University
TU Darmstadt
ExtreMe Matter Institute

TALENT Course 11 is possible thanks to funding from the STFC

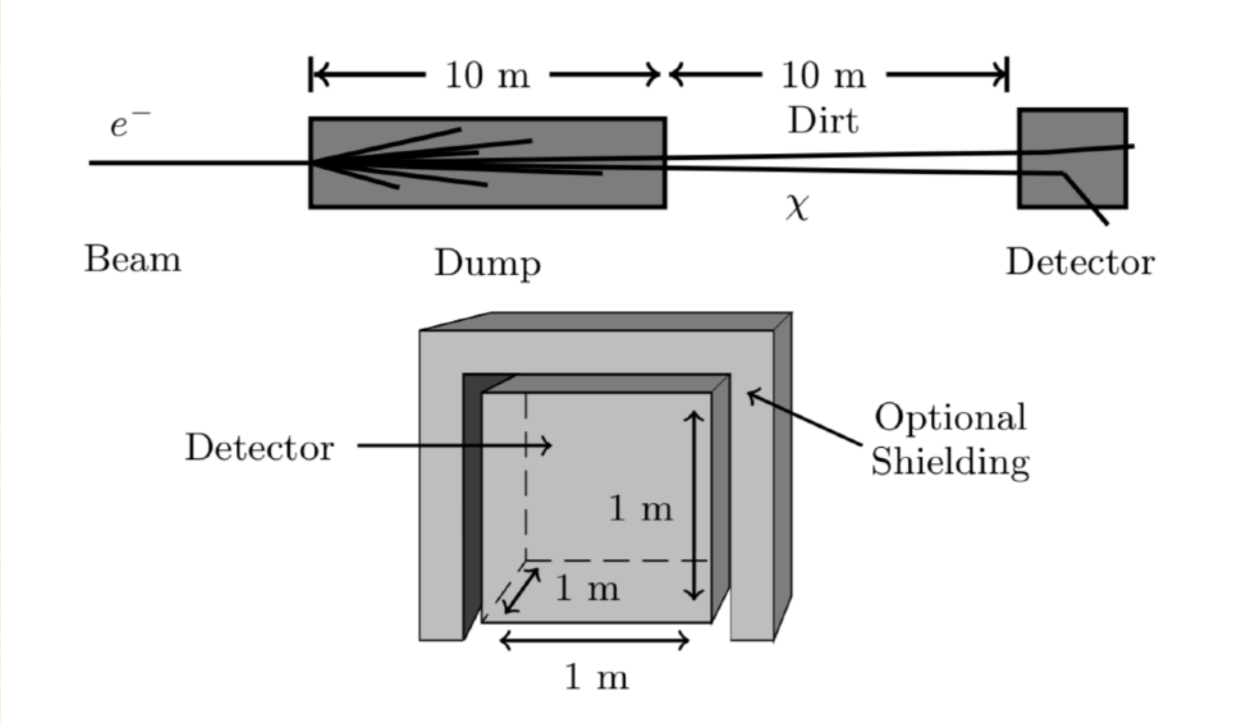
Today

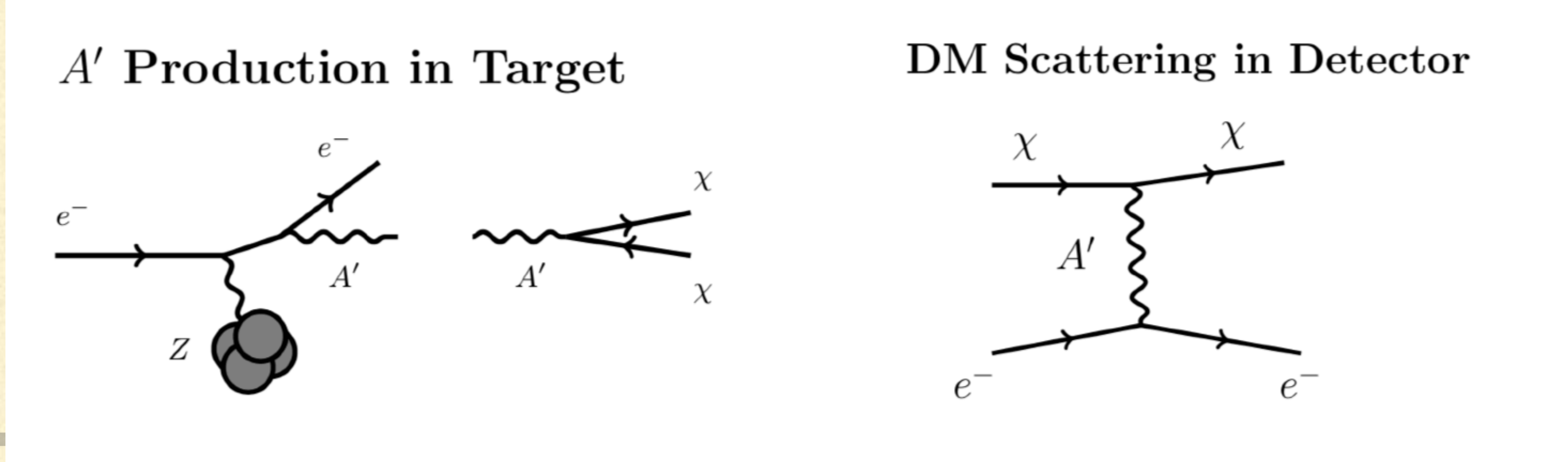
- A window on experimental planning: the BDX experiment at JLab
- MHOU in PDFs
- Is an ANN smarter than EFT? Machine learning for N_{max} extrapolation of ab initio calculations
- Machine learning for nuclear mass model residuals: BNNs, GPs, and UQ for DFT
- Emulate this! Using parameter estimation, sampling, and GP emulators to determine the nuclear EoS at supra-nuclear densities
- Uniform in what? How prior choice will influence the inferences from neutron-star observations

BDX experiment at JLab

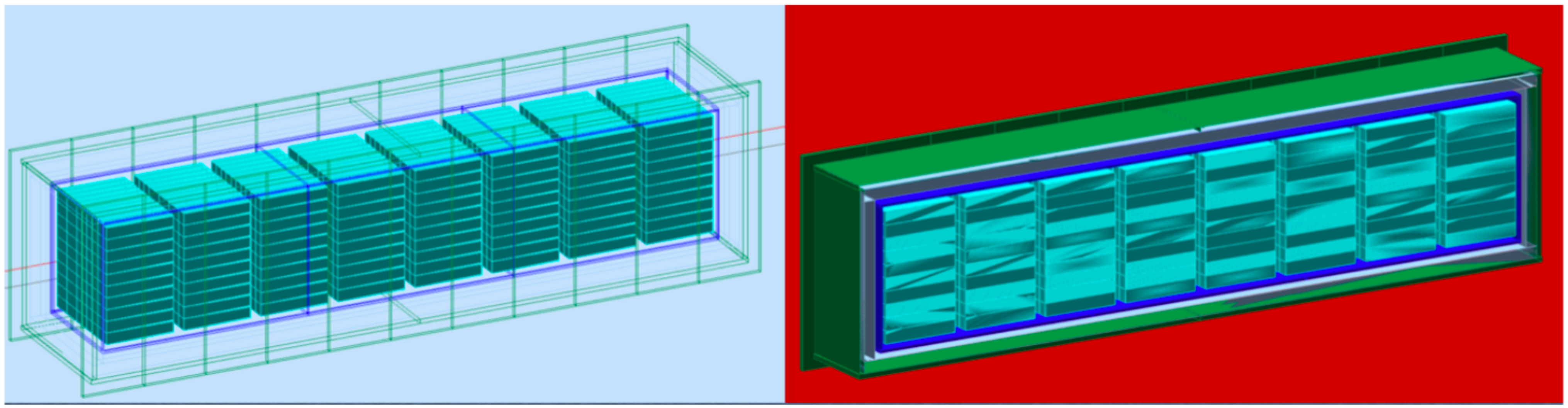
M. Battaglieri et al., PR 12-16-001
arXiv:1607.01390

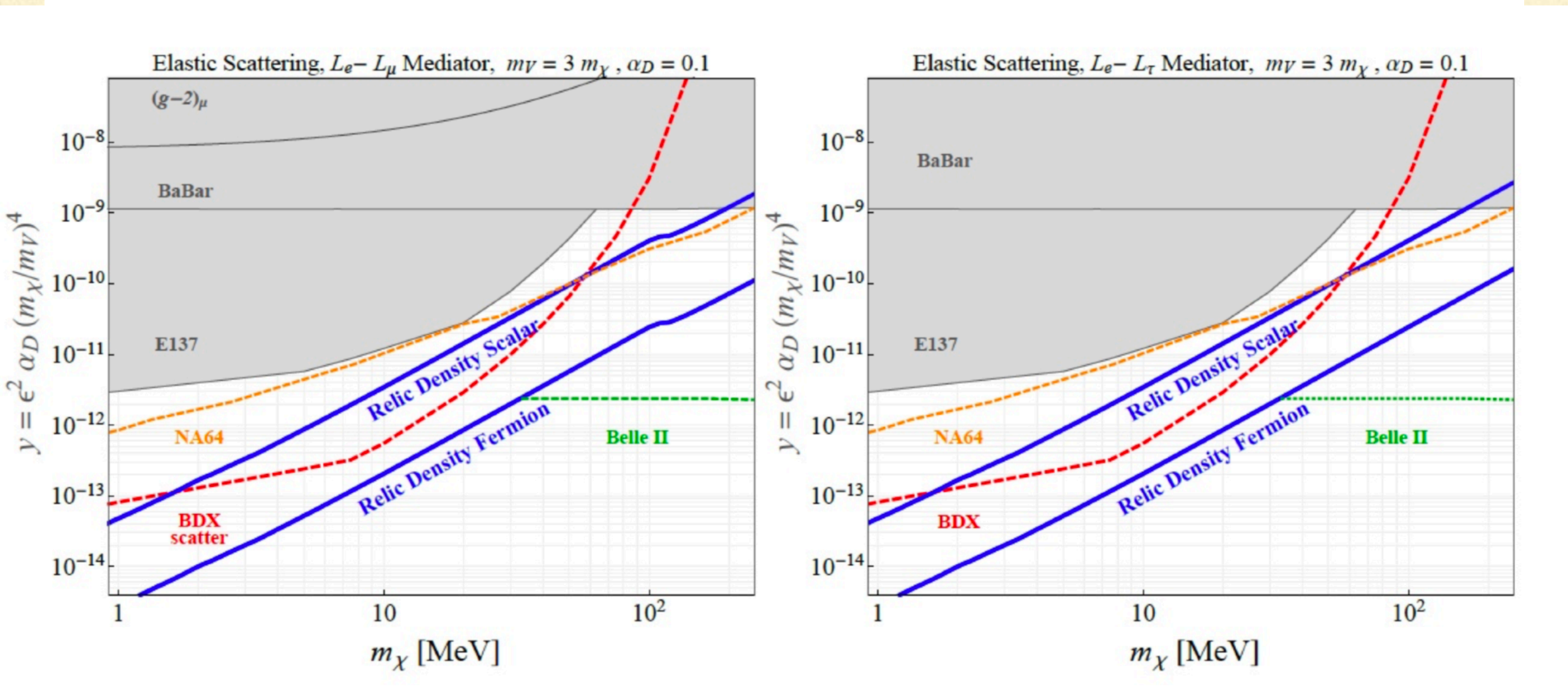
- Look for production of light dark matter particles when electron beam hits beam stop





The tool and the (hoped for) output





Computing the sensitivity

For signal, S, and background, B, how many counts, n, will I see?

$$\mathrm{pr}(n\,|\,S,B) = \frac{(S+B)^n e^{-(S+B)}}{(S+B)\,!}$$

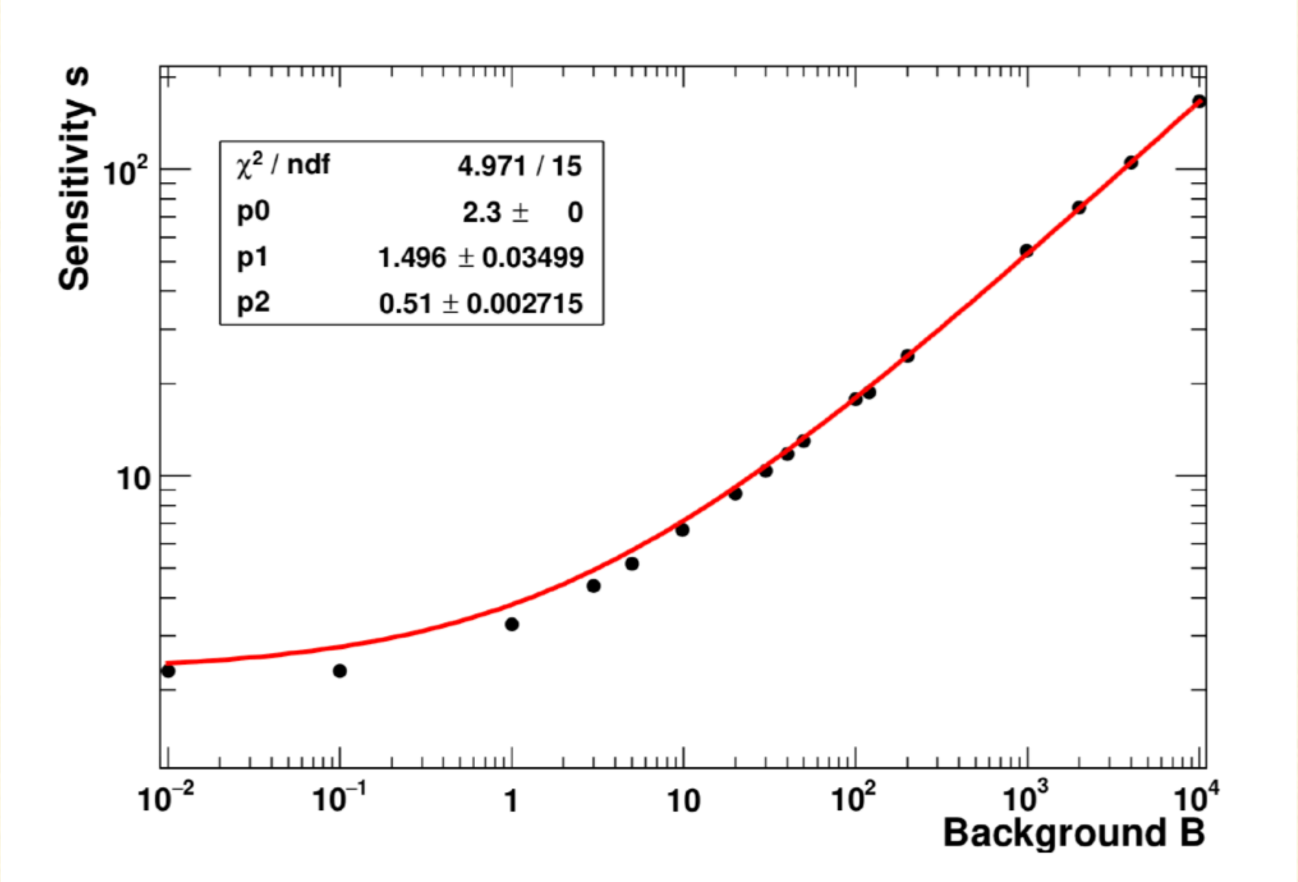
Use Bayes' theorem to compute pr(S|n,B)

Upper limit $S_{up}(n,B,\alpha)$:

$$\int_0^{S_{up}(n,B,\alpha)} dS\,\mathrm{pr}(S\,|\,n,B) = 1-\alpha$$

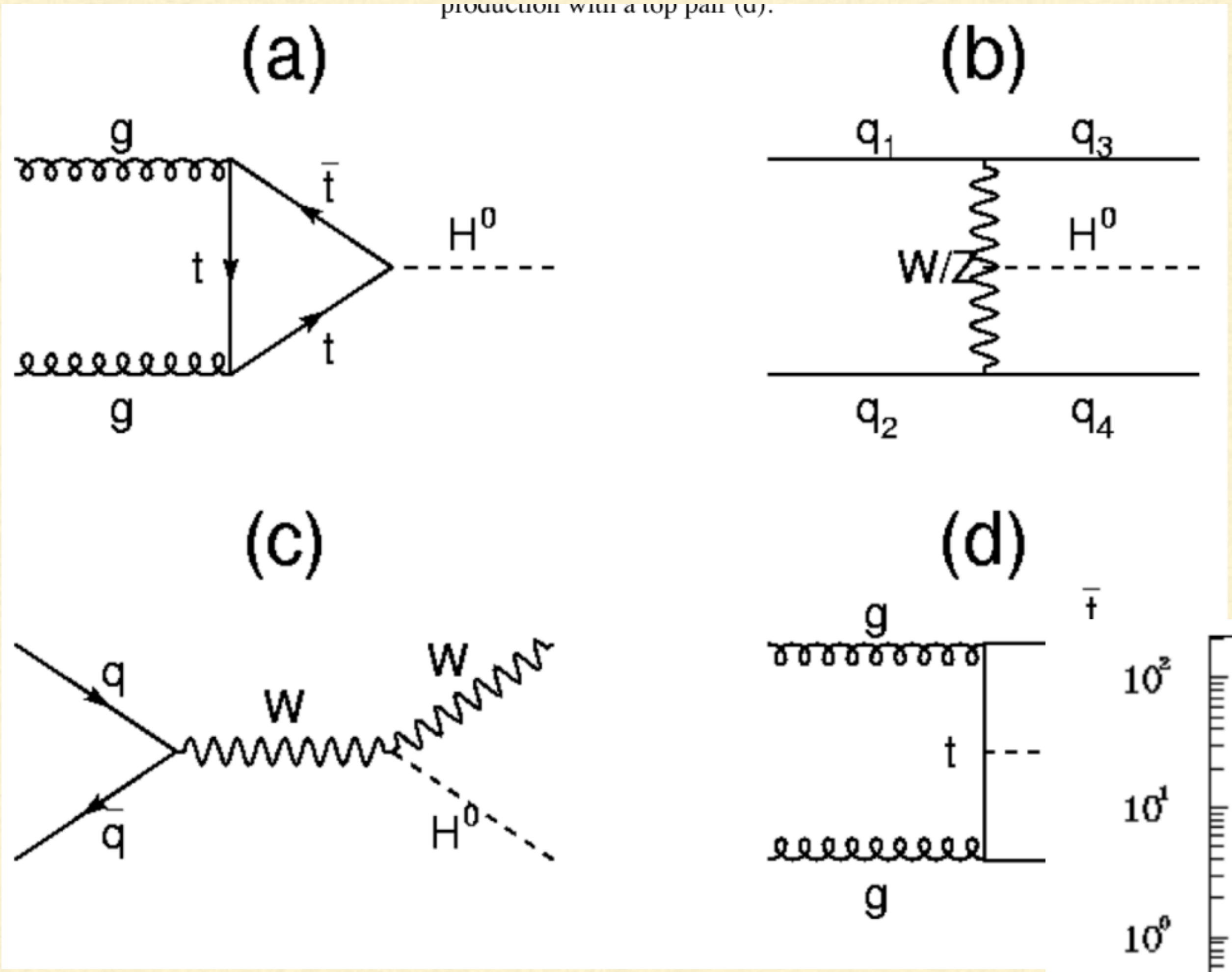
Sensitivity s: average of upper limits reported by identical experiments if S=0

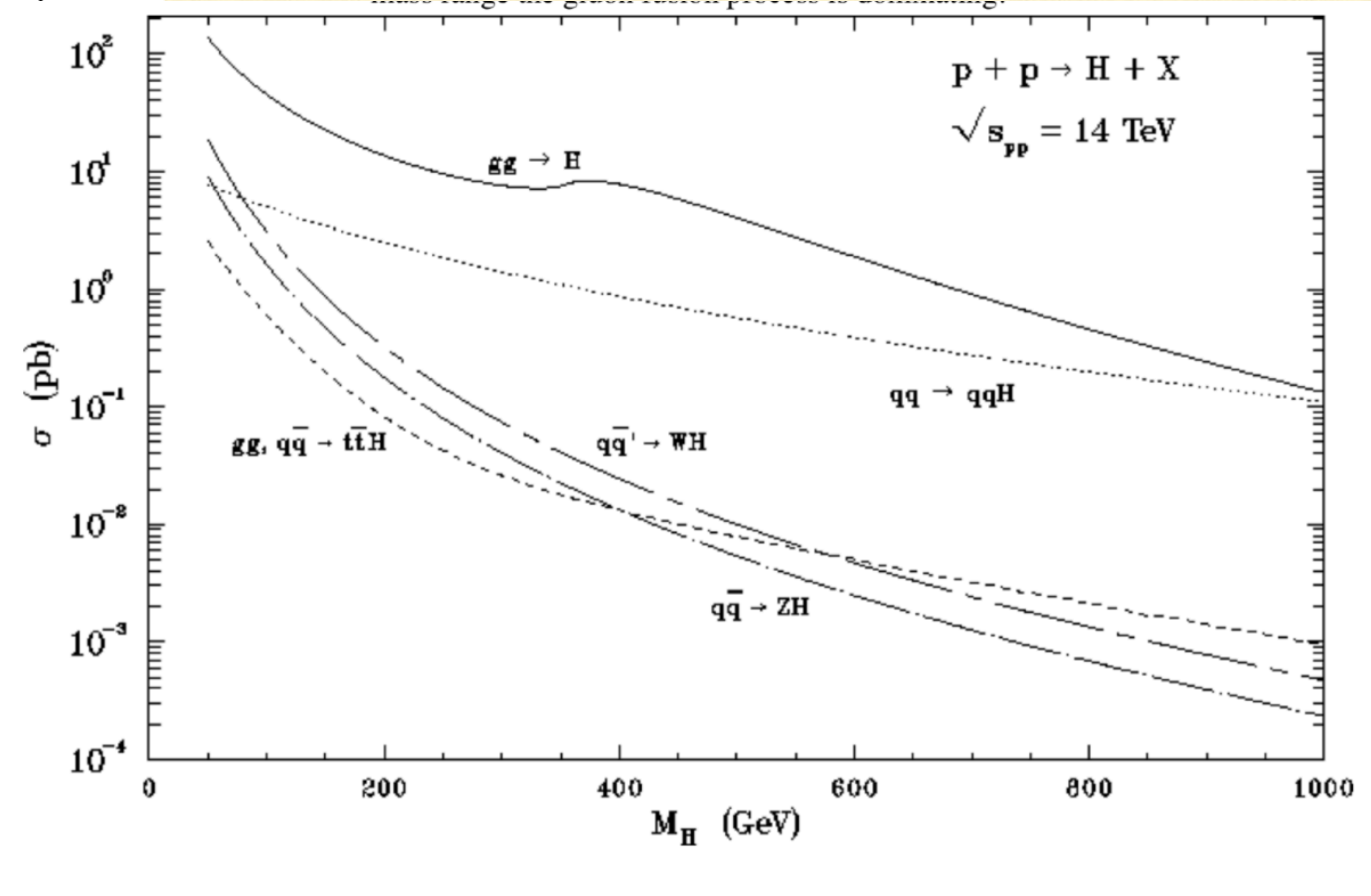
$$s(\alpha,B) = \sum_{n-0}^{\infty} P(n\,|\,B,S=0)S_{up}(n,B,\alpha)$$



Making Higgses from gluons

Egede, thesis, Lund University





The problem: extracting PDFs from data

$$F(Q^2) = C(\alpha_S(Q^2)) \otimes f(Q^2)$$

- Measure F, compute C up to given pQCD order, extract f
- Data errors in F
- But C only computed up to fixed order in $\alpha_S(Q^2)$
- So f inherits errors of $O(\alpha_S^2)$ (for example)
- This can be estimated by varying the scale at which C is calculated

$$\bar{F}(Q^2, \mu^2) = C(\alpha_S(\mu^2), \mu^2/Q^2) \otimes f(Q^2)$$

- 2819 data go into extraction of PDFs in the study I'll describe

MHOU in PDFs

NNPDF collaboration, arXiv:1906.10698

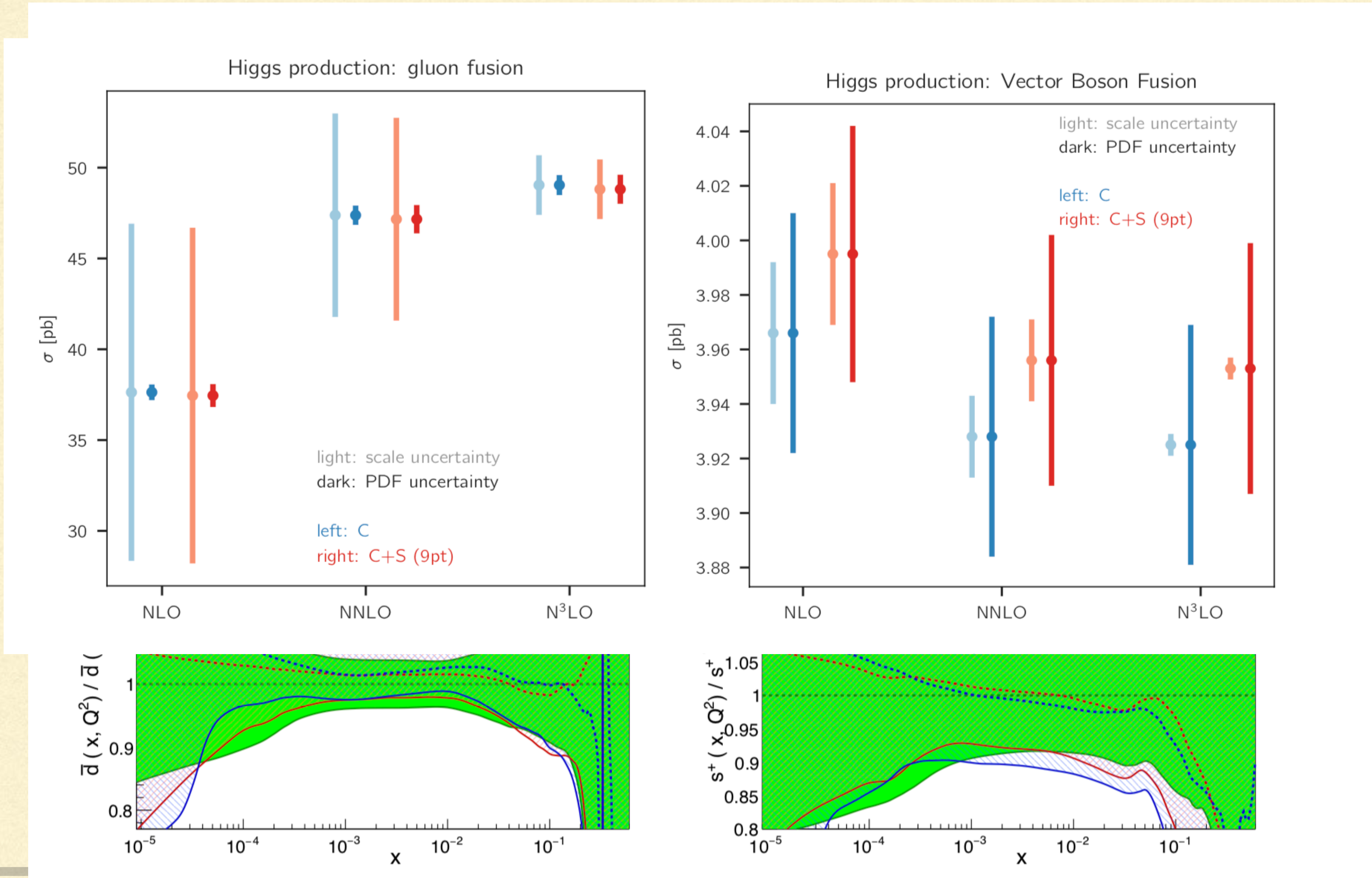
$$P(T|D) \propto \exp\left(-\frac{1}{2}(D_i - T_i)(C+S)^{-1}_{ij}(D_j - T_j)\right)$$

$$S_{ij} = \langle(\mathscr{T}_i - T_i)(\mathscr{T}_j - T_j)\rangle = \langle\Delta_i\Delta_j\rangle$$

- Assess S_{ij} by varying factorization and pQCD scale in calculation of T_i's.

$$\begin{aligned}S^{(9\mathrm{pt})}_{i_1 j_2} = \tfrac{1}{24}\big\{&2\big(\Delta^{+0}_{i_1} + \Delta^{++}_{i_1} + \Delta^{+-}_{i_1}\big)\big(\Delta^{+0}_{j_2} + \Delta^{++}_{j_2} + \Delta^{+-}_{j_2}\big)\\ &+2\big(\Delta^{-0}_{i_1} + \Delta^{-+}_{i_1} + \Delta^{--}_{i_1}\big)\big(\Delta^{-0}_{j_2} + \Delta^{-+}_{j_2} + \Delta^{--}_{j_2}\big)\big\}\\ &+3\big(\Delta^{0+}_{i_1} + \Delta^{0-}_{i_1}\big)\big(\Delta^{0+}_{j_2} + \Delta^{0-}_{j_2}\big)\big\}.\end{aligned}$$

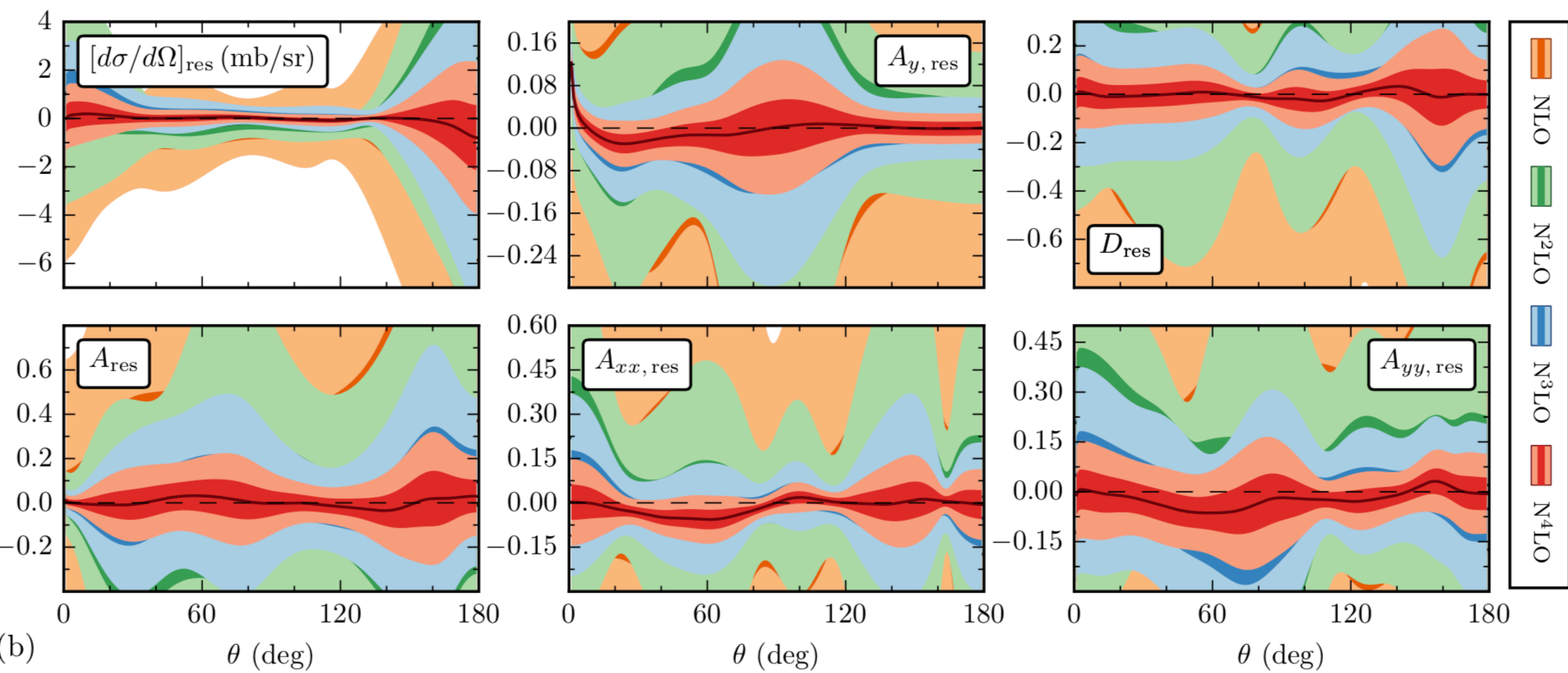
Results


Higgs production: gluon fusion
σ [pb]
50
45
40
35
30
light: scale uncertainty
dark: PDF uncertainty
left: C
right: C+S (9pt)
NLO
NNLO
N³LO
Higgs production: Vector Boson Fusion
σ [pb]
4.04
4.02
4.00
3.98
3.96
3.94
3.92
3.90
3.88
light: scale uncertainty
dark: PDF uncertainty
left: C
right: C+S (9pt)
NLO
NNLO
N³LO
d̄ (x, Q²) / d̄ (
1
0.9
0.8
10⁻⁵
10⁻⁴
10⁻³
10⁻²
10⁻¹
x
s⁺ (x, Q²) / s⁺
1.05
1
0.95
0.9
0.85
0.8
10⁻⁵
10⁻⁴
10⁻³
10⁻²
10⁻¹
x

MHOU for NN observables

Melendez, Furnstahl, Wesolowski, PRC, 2017

EKM R=0.9 fm potential

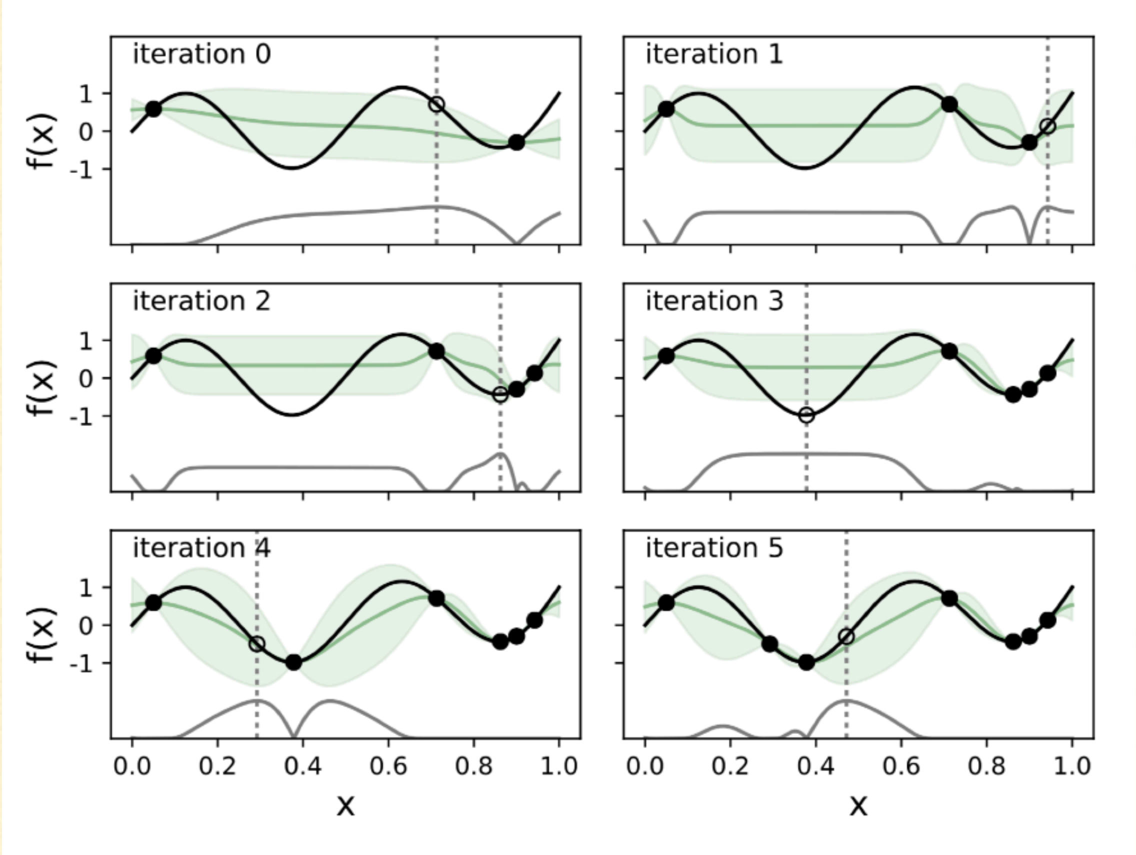


E_{lab}=250 MeV

Bayesian optimization

Ekström et al., JPG to appear (2019)

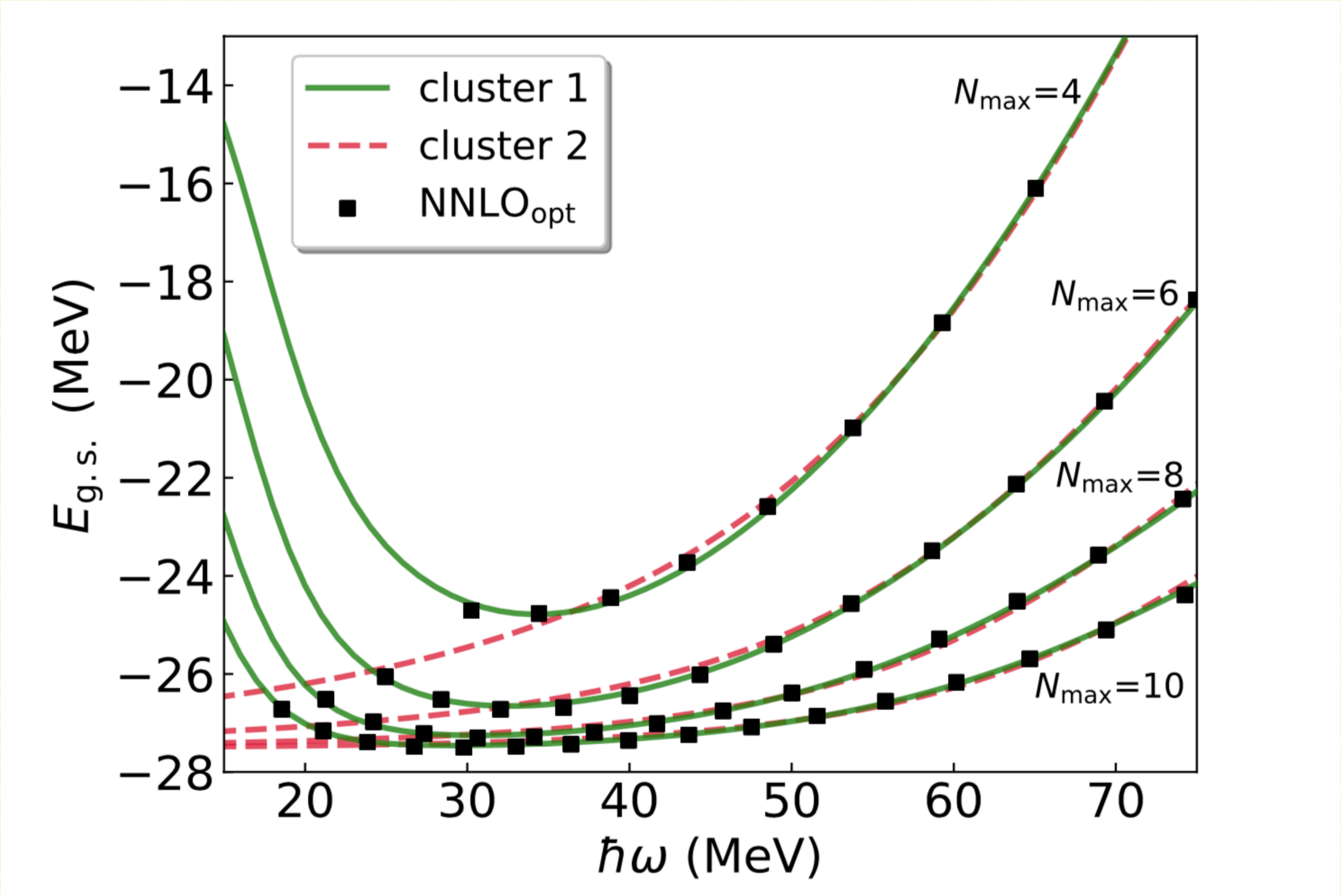
- BO: Prior knowledge/belief is everything
- BO will never find a narrow minimum nor determine a precise value of the optimum. Switch to, e.g., POUNDers at some point
- Acquisition function more important than GP kernel
- High-dimensional domains challenging (dimensional reduction!)



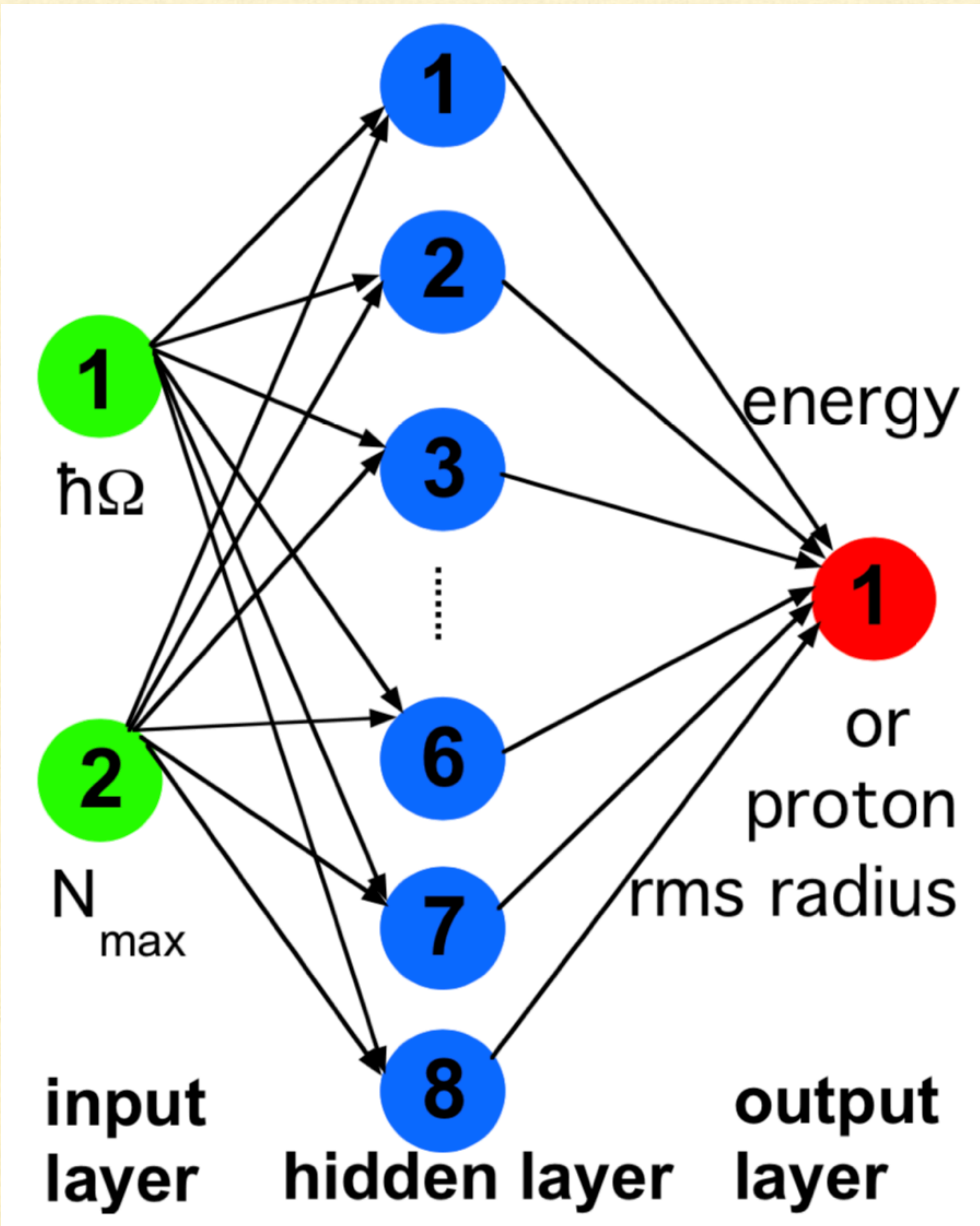
Using ANNs for N_{max} extrapolation

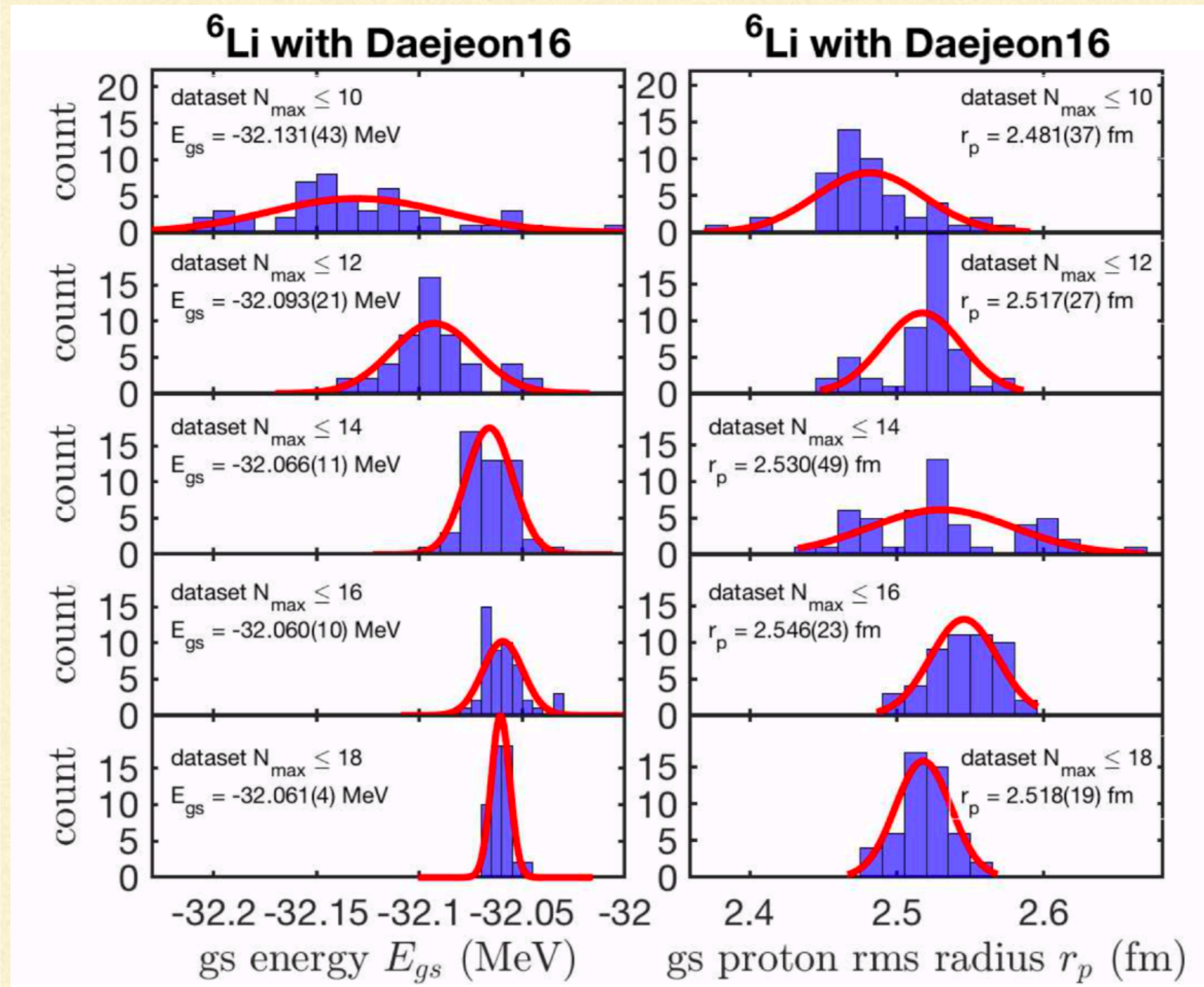
Negoita, Vary, et al., Phys. Rev. C 99, 054308 (2019)

Jiang, Hagen, Papenbrock, arXiv:1905.06317

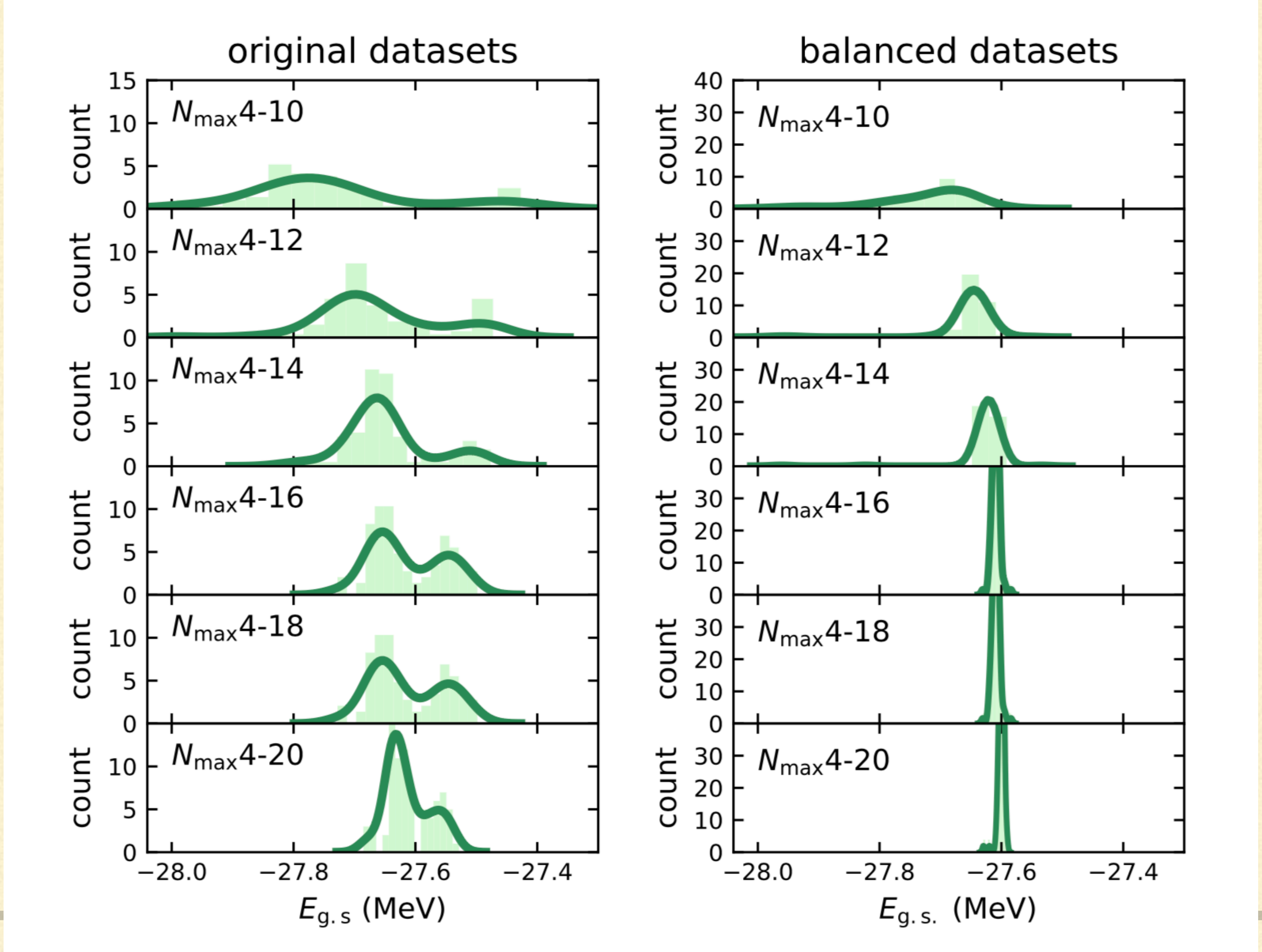


One ANN implementation

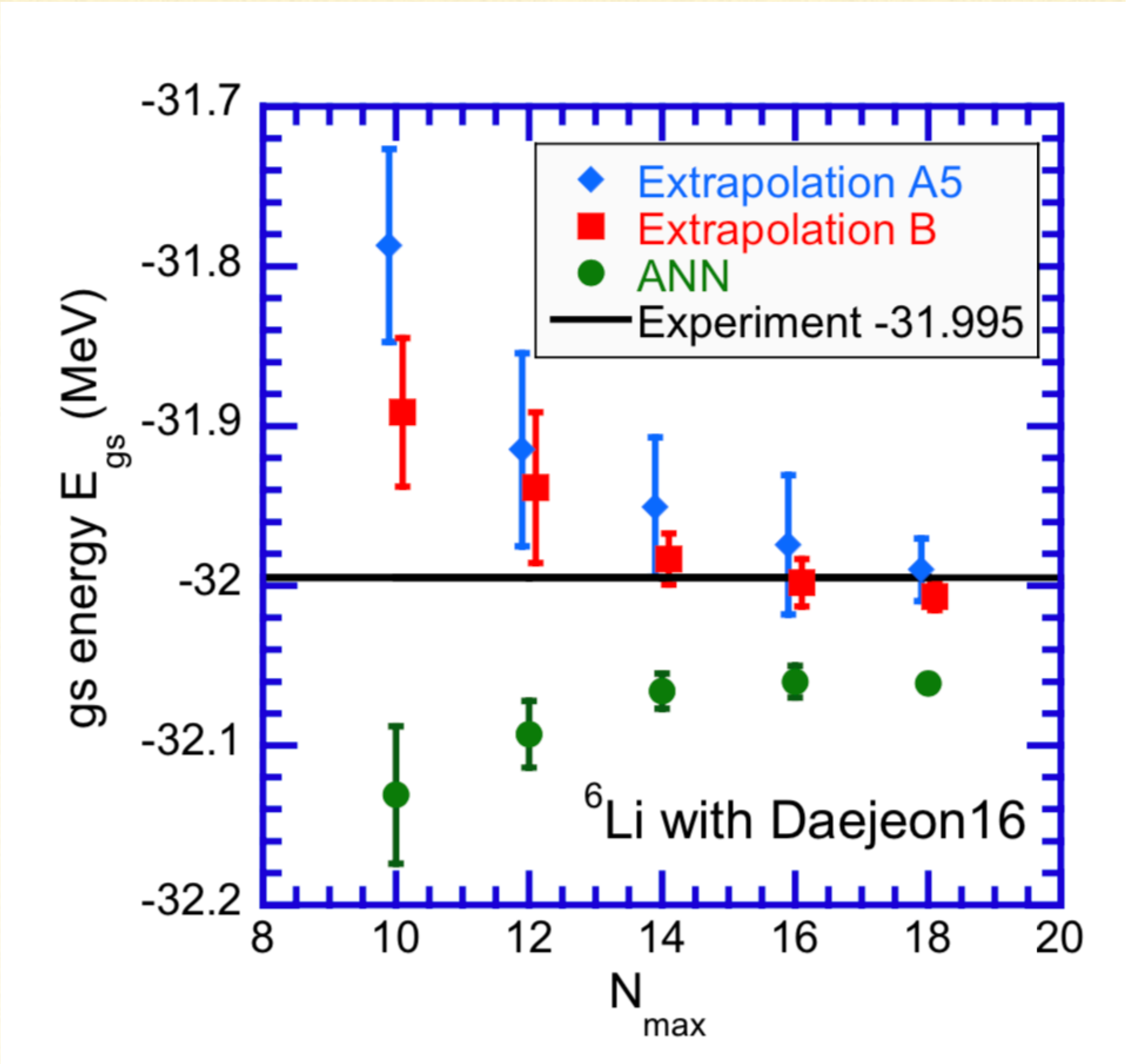


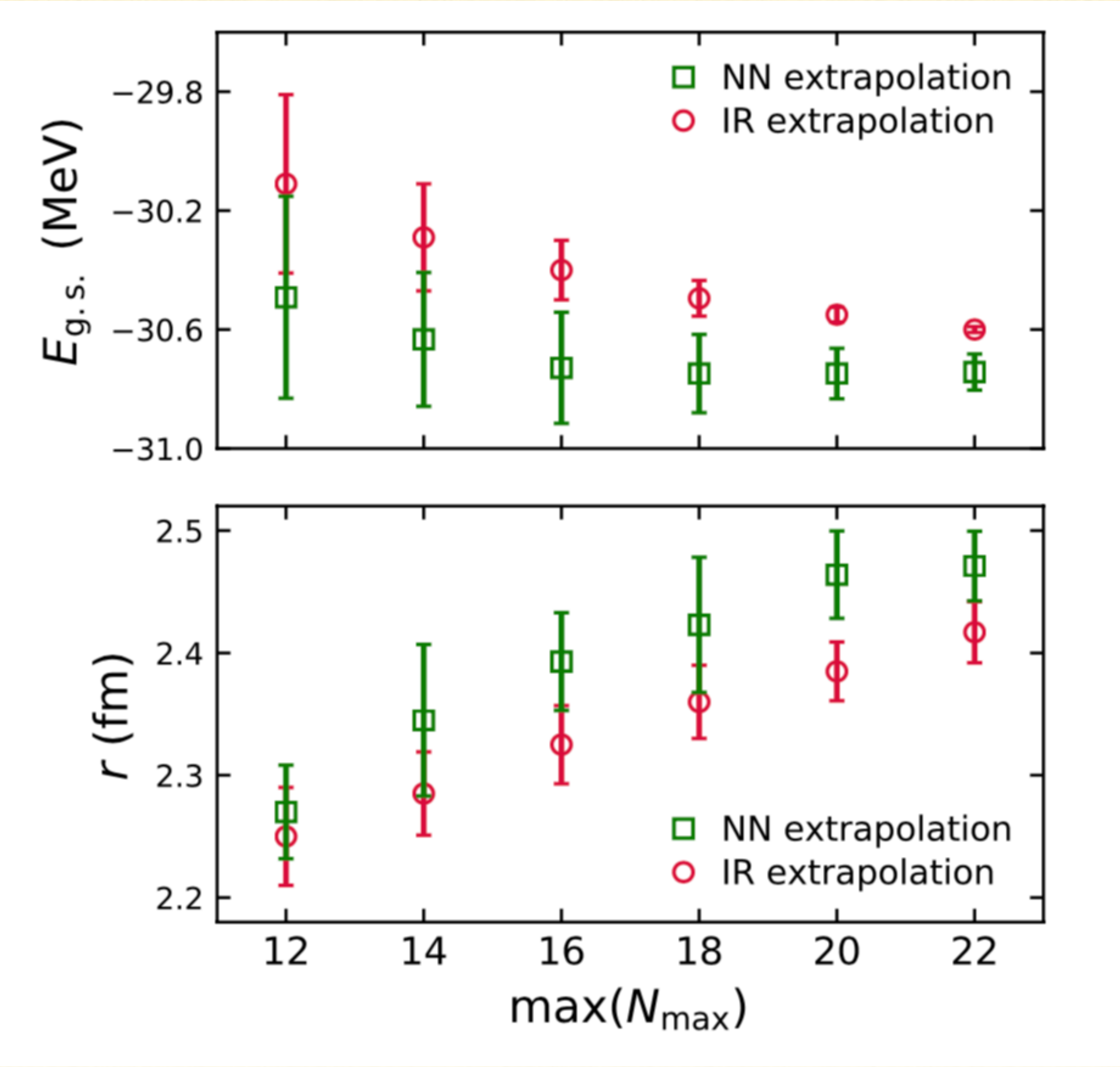


Another implementation

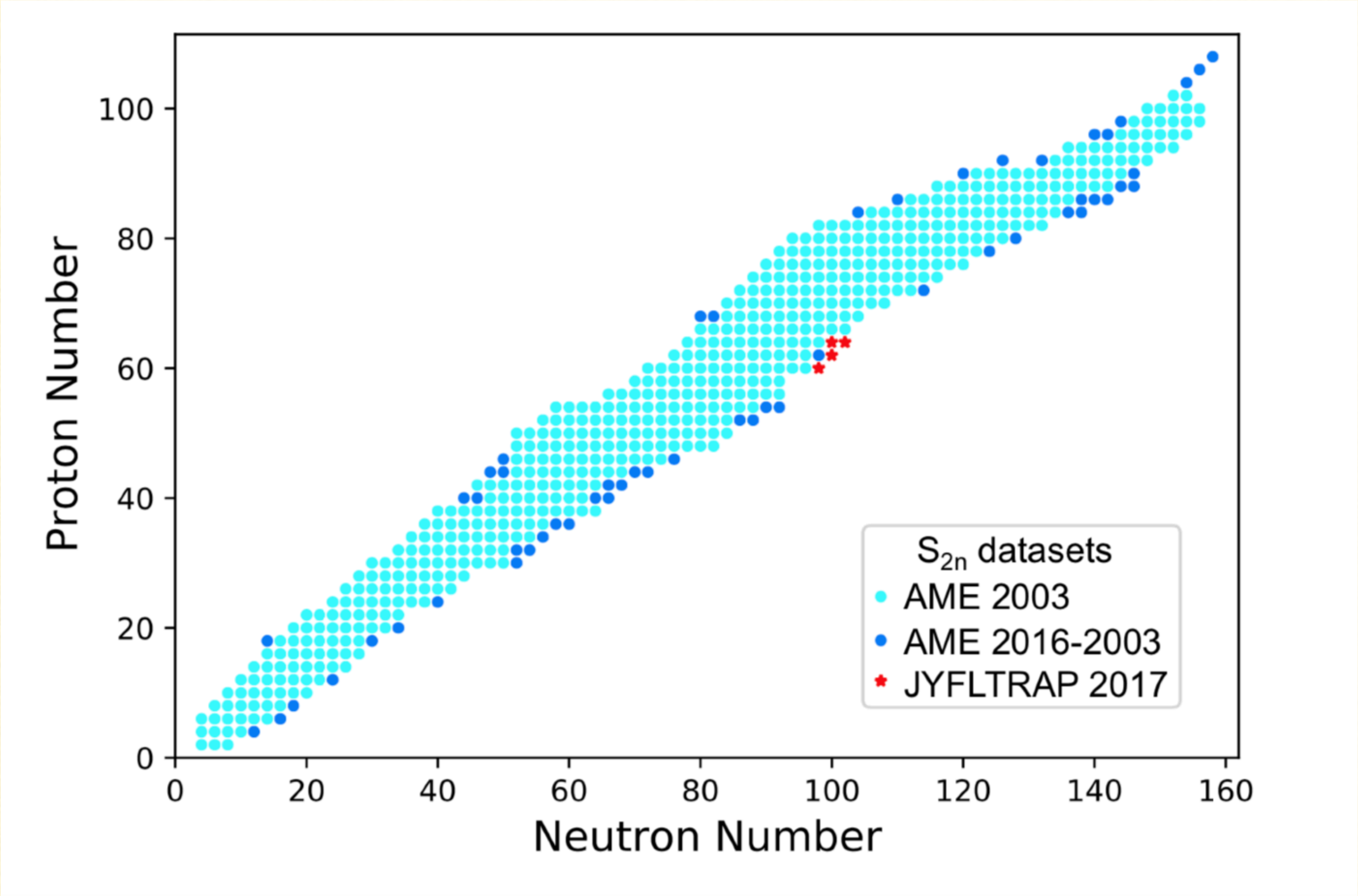


ANN Extrapolation





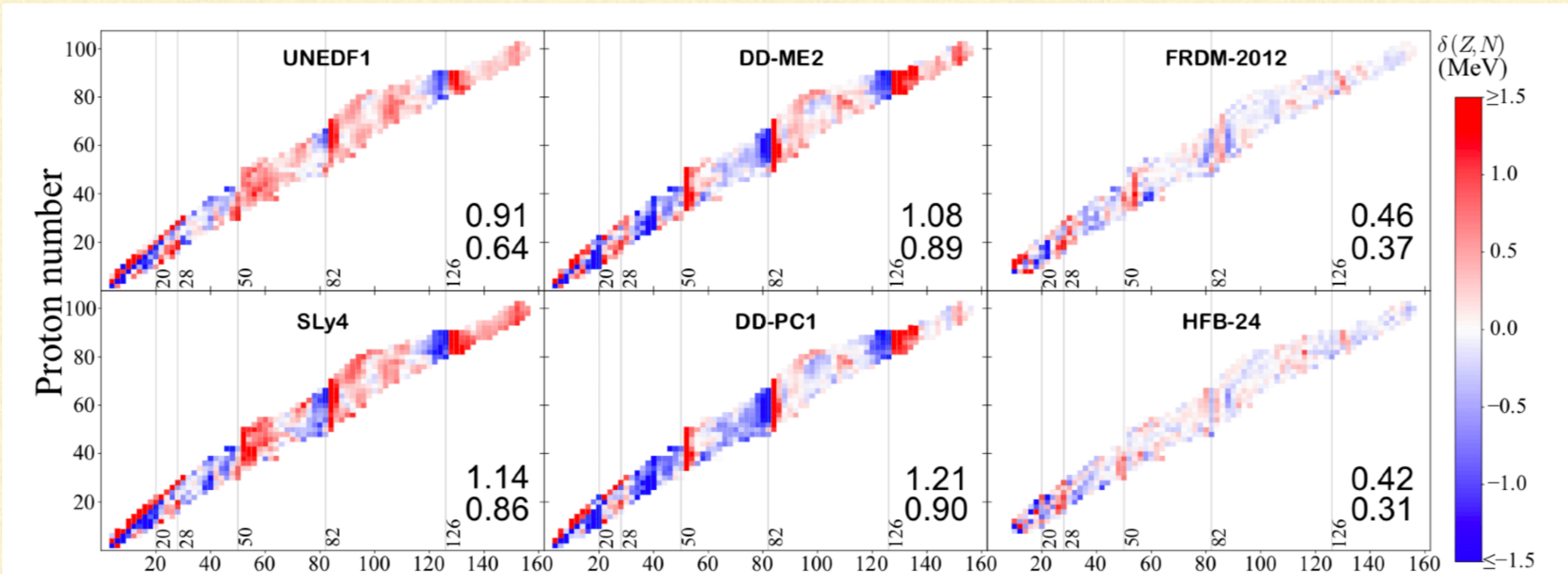
The nuclear landscape

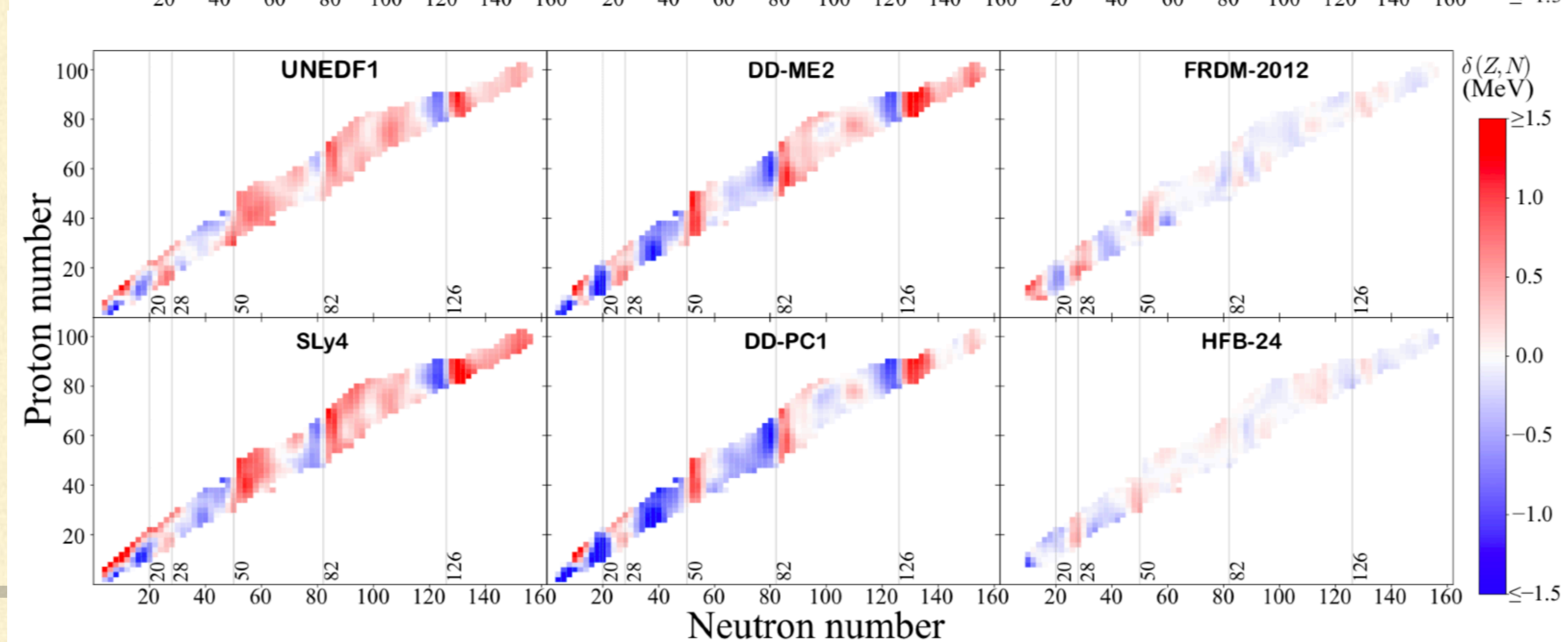


Neufcourt, Cao, Nazarewicz, and Viens, PRC 98, 034318 (2018)

- Masses predicted by a range of DFTs. But generally come without any UQ.

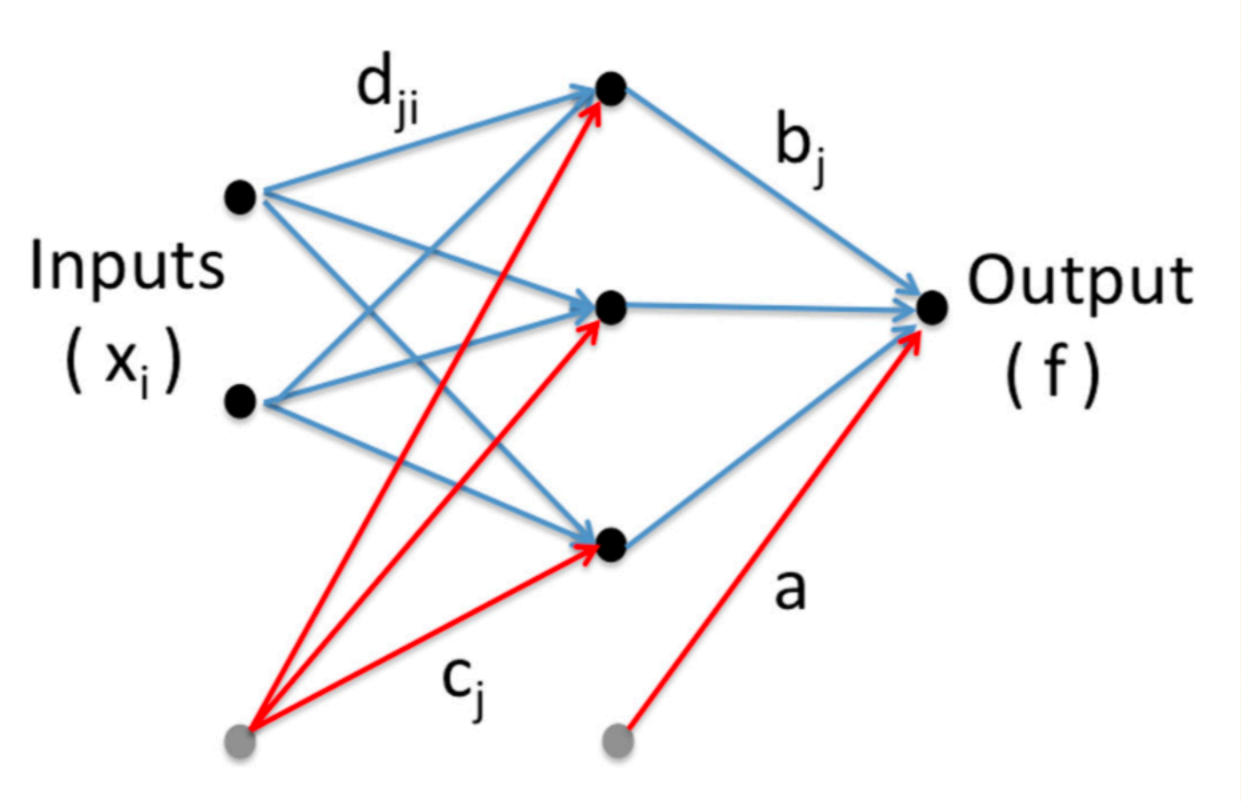
Residuals


UNEDF1
DD-ME2
FRDM-2012
SLy4
DD-PC1
HFB-24
0.91
0.64
1.08
0.89
0.46
0.37
1.14
0.86
1.21
0.90
0.42
0.31
Proton number
δ(Z,N) (MeV)
≥1.5
1.0
0.5
0.0
−0.5
−1.0
≤−1.5


UNEDF1
DD-ME2
FRDM-2012
SLy4
DD-PC1
HFB-24
Proton number
Neutron number
δ(Z,N) (MeV)
≥1.5
1.0
0.5
0.0
−0.5
−1.0
≤−1.5

Machine learning for mass-model residuals

Utama and Piekarewicz, PRC 96, 044308 (2017)



- Use DFT to predict nuclear masses
- (In this case mainly DZ10 model.)
- Compute residuals
- BNN: input (Z,A); output residual
- Train on a randomly selected subset of the nuclear masses between ^{40}Ca and ^{240}U.
- Validation set: remainder
- Goal: predict masses of as-yet-unmeasured nuclei

BNN details

$$\mathrm{pr}(w \mid \{t(x_i)\}) \propto \mathrm{pr}(\{t(x_i)\} \mid w)\mathrm{pr}(w)$$

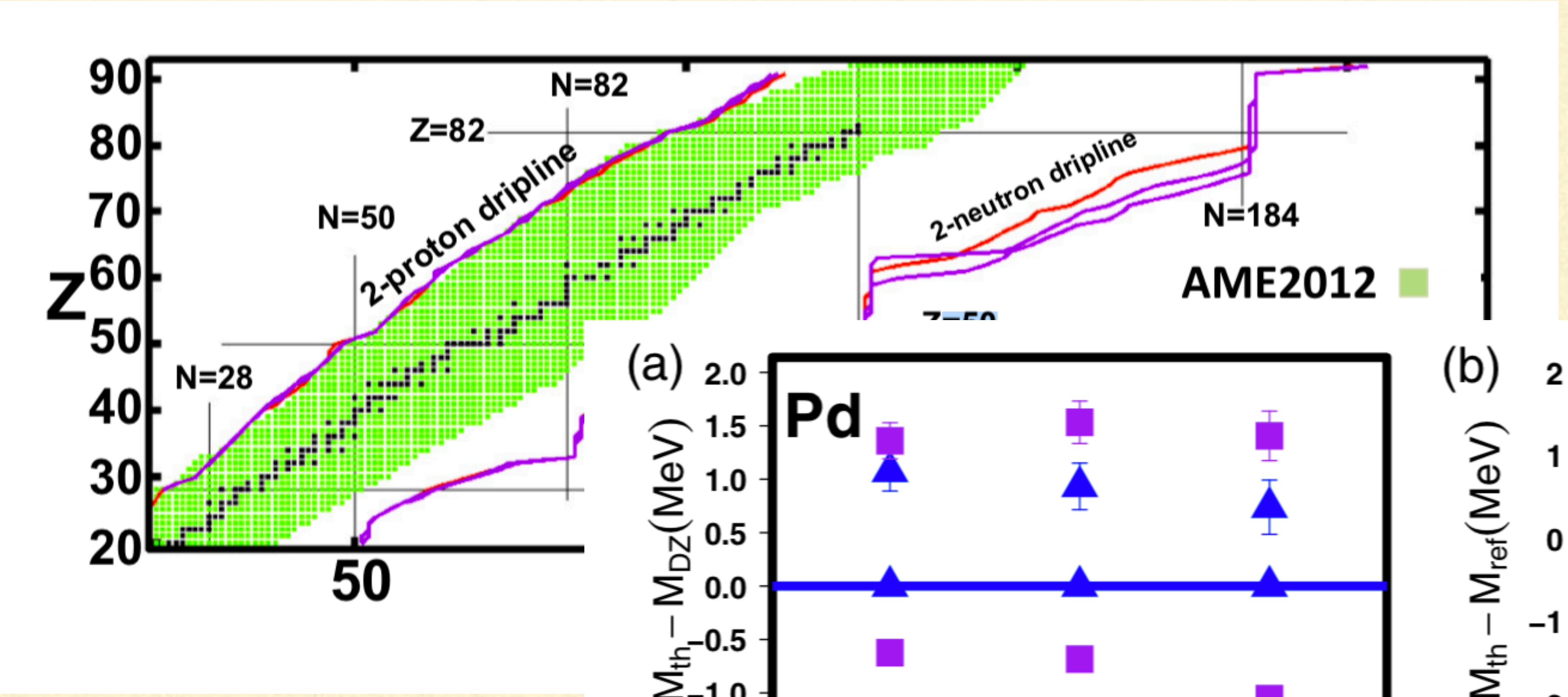
$$\mathrm{pr}(w \mid \{t_i\}) \propto \exp(-\chi^2(w)/2)$$

$$f(x, w) = a + \sum_{j=1}^{H} b_j \tanh\left(c_j + \sum_{i=1}^{I} d_{ji} x_i \right)$$

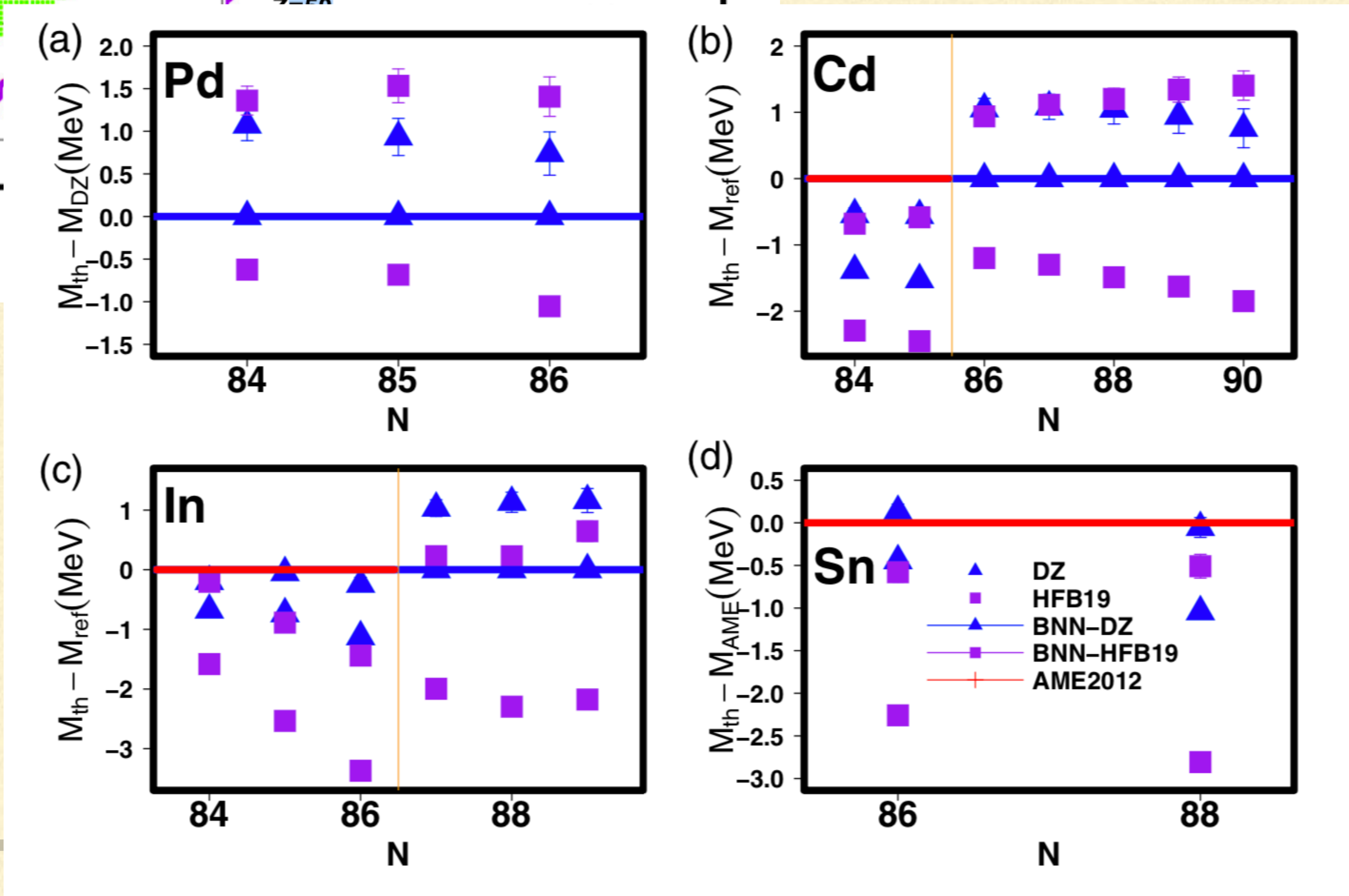
I: number of inputs (two here)
H: number of hidden nodes
Parameters: $\{a, b_j, c_j, d_{ij}\}$

Compute posterior by MCMC sampling "exact"

Results



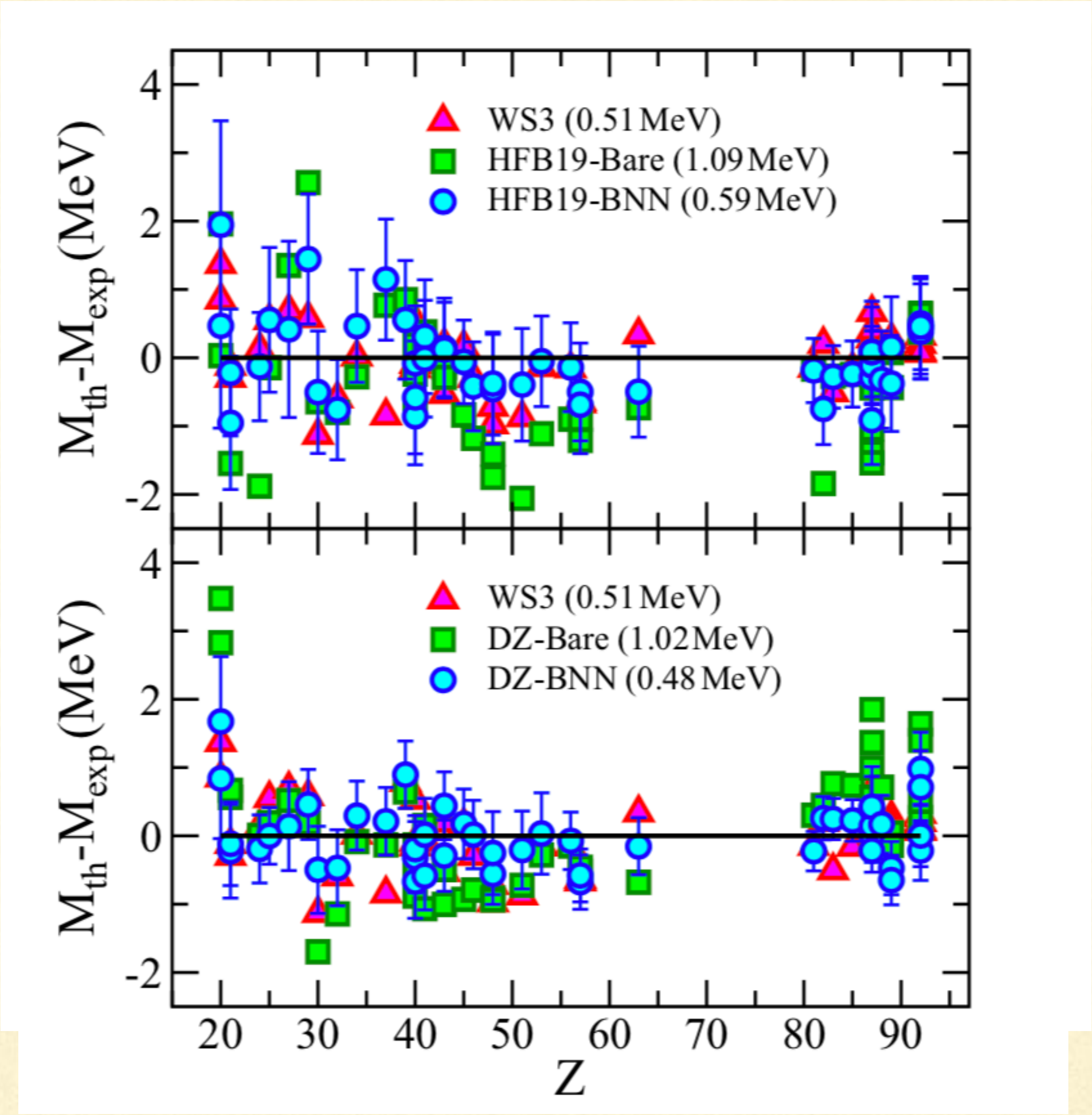
N=82
Z=82
2-proton dripline
N=50
2-neutron dripline
N=184
AME2012
N=28
Z
50



(a)
Pd
$M_{th} - M_{DZ}$(MeV)
N
(b)
Cd
$M_{th} - M_{ref}$(MeV)
N
(c)
In
$M_{th} - M_{ref}$(MeV)
N
(d)
Sn
$M_{th} - M_{AME}$(MeV)
DZ
HFB19
BNN-DZ
BNN-HFB19
AME2012
N

Validation

Utama and Piekarewicz, PRC 97, 014306 (2018)



- New nuclei in AME2016 not in previous evaluations
- Marked improvement in rms error

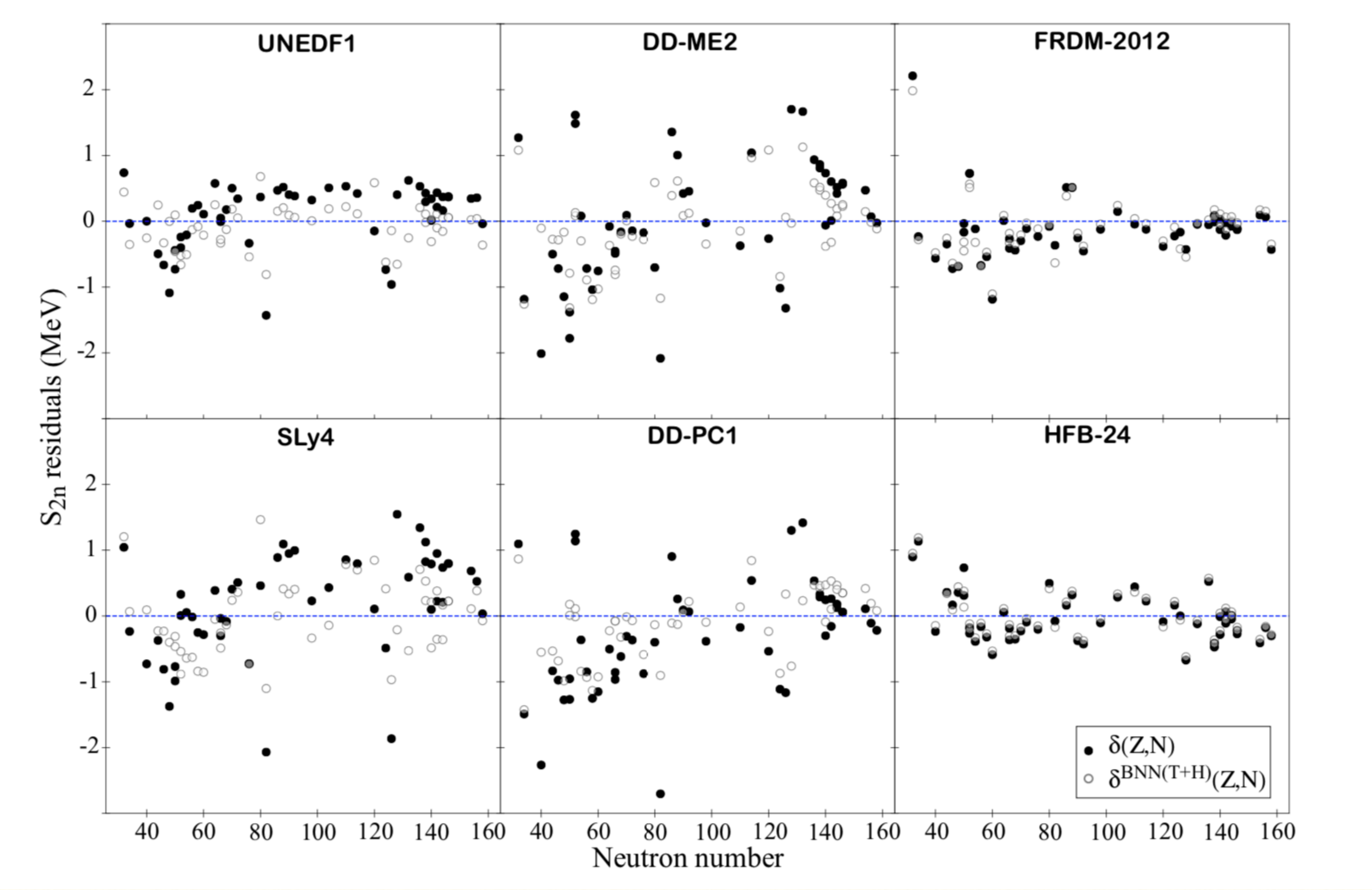
A different BNN

Neufcourt, Cao, Nazarewicz, and Viens, PRC 98, 034318 (2018)

- Criticize activation function choice: "hyperbolic tangent has linear tails which cannot vanish simultaneously, raising potential issues in the case of a bounded linear extrapolation".

The number of parameters in a BNN is key to the model's performance. With about 500 data points, taking $H = 30$ neurons leads to much better performance than higher (or lower) H. As mentioned earlier, increasing the number of layers beyond $L = 1$ decreases performance. This is almost certainly due to the small amount of data which, as we explained, is a non-negotiable aspect of this type of nuclear theory UQ study. The number of parameters for an ANN containing L layers with H hidden neurons in each layer is given by $(1 + |x|)H + [H(H + 1)]^{L-1} + (H + 1)|y|$, where $|x|$ and $|y|$ are the respective dimensions of the network data input and outputs. With $|x| = 2$ (or 4 in the refinement described in Sec. IV C) and $|y| = 1$, this results in 121 (or 181) parameters; adding one layer would add 120 parameters at once. There exists an unwritten rule of thumb in statistics, by which the ratio of data to parameters needed to have a hope of estimating parameters in a statistically significant way in linear regressions (e.g., with 95% confidence/credibility on most parameters), should be bounded below by 10 in a classical frequentist setting, and should be bounded below by 3 in a Bayesian setting when there is no expectation of showing that the output is insensitive to the priors. With about 500 datapoints, this explains why one cannot use more than one BNN layer in our study, and why a frequentist ANN

BNN results


UNEDF1
DD-ME2
FRDM-2012
SLy4
DD-PC1
HFB-24
S_{2n} residuals (MeV)
Neutron number
δ(Z,N)
$\delta^{BNN(T+H)}(Z,N)$

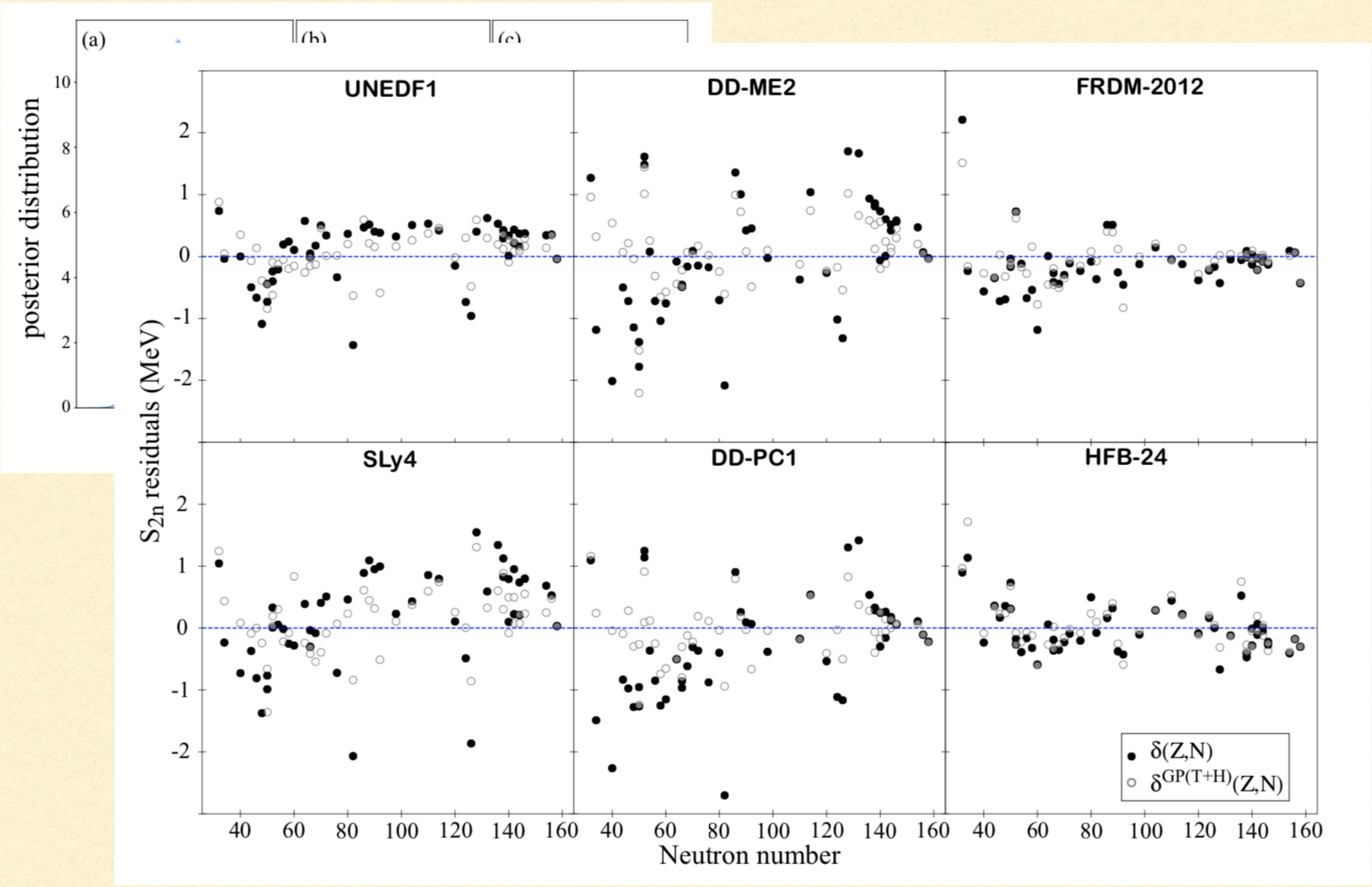
Comparison to GP for residuals

- Differences:
  - BNN is ultimately deterministic
  - Noise is assumed uncorrelated
  - More parameters in BNN, therefore GP has UQ advantage

$$f(x;\theta) \sim GP(0, k_{\eta,\rho_Z,\rho_Z})$$

$$k_{\eta,\rho_N,\rho_Z}(Z,N;Z',N') = \eta^2 \exp\left[-\frac{(Z-Z')^2}{2\rho_Z^2} - \frac{(N-N')^2}{2\rho_N^2}\right]$$

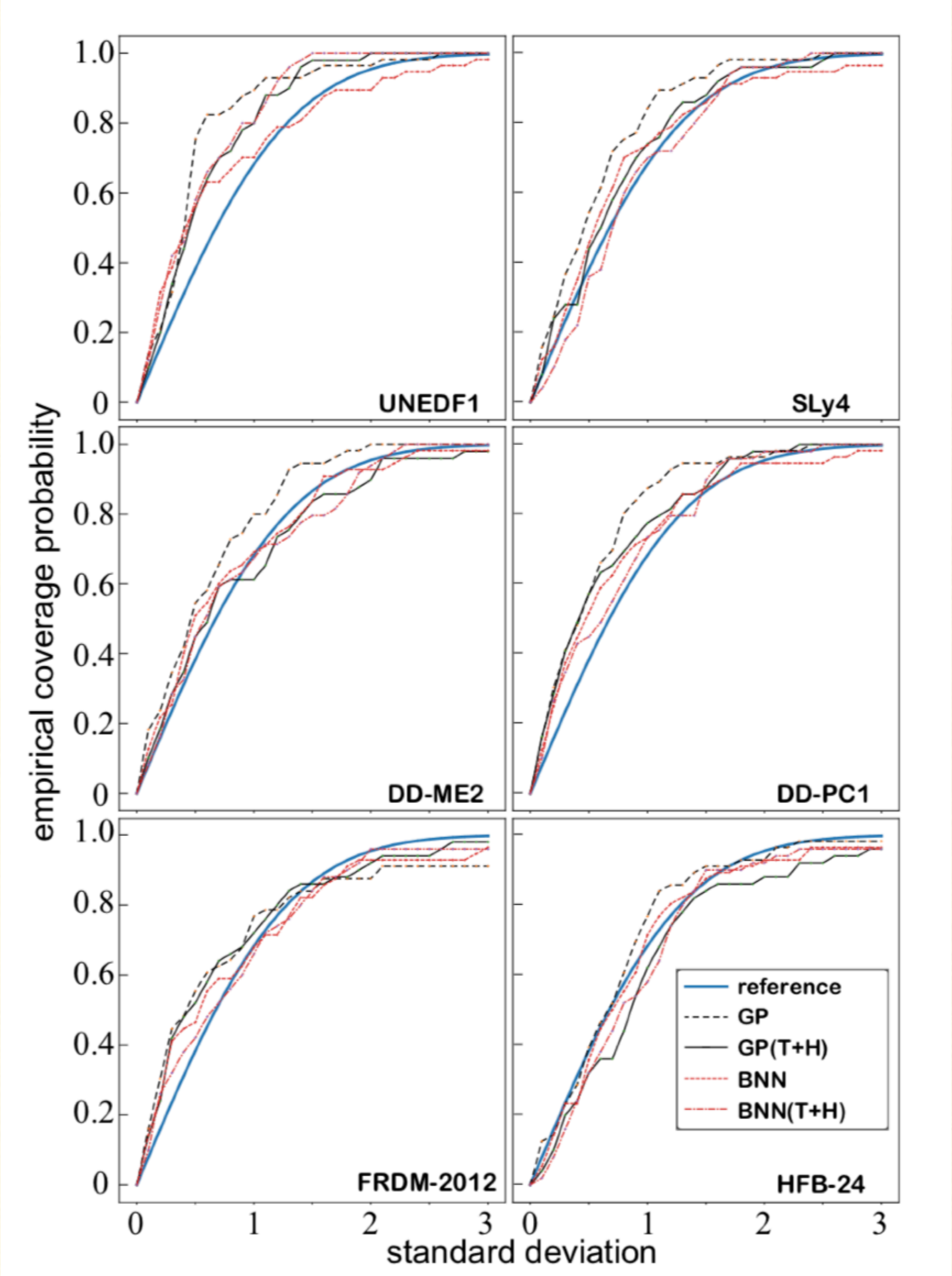
Results



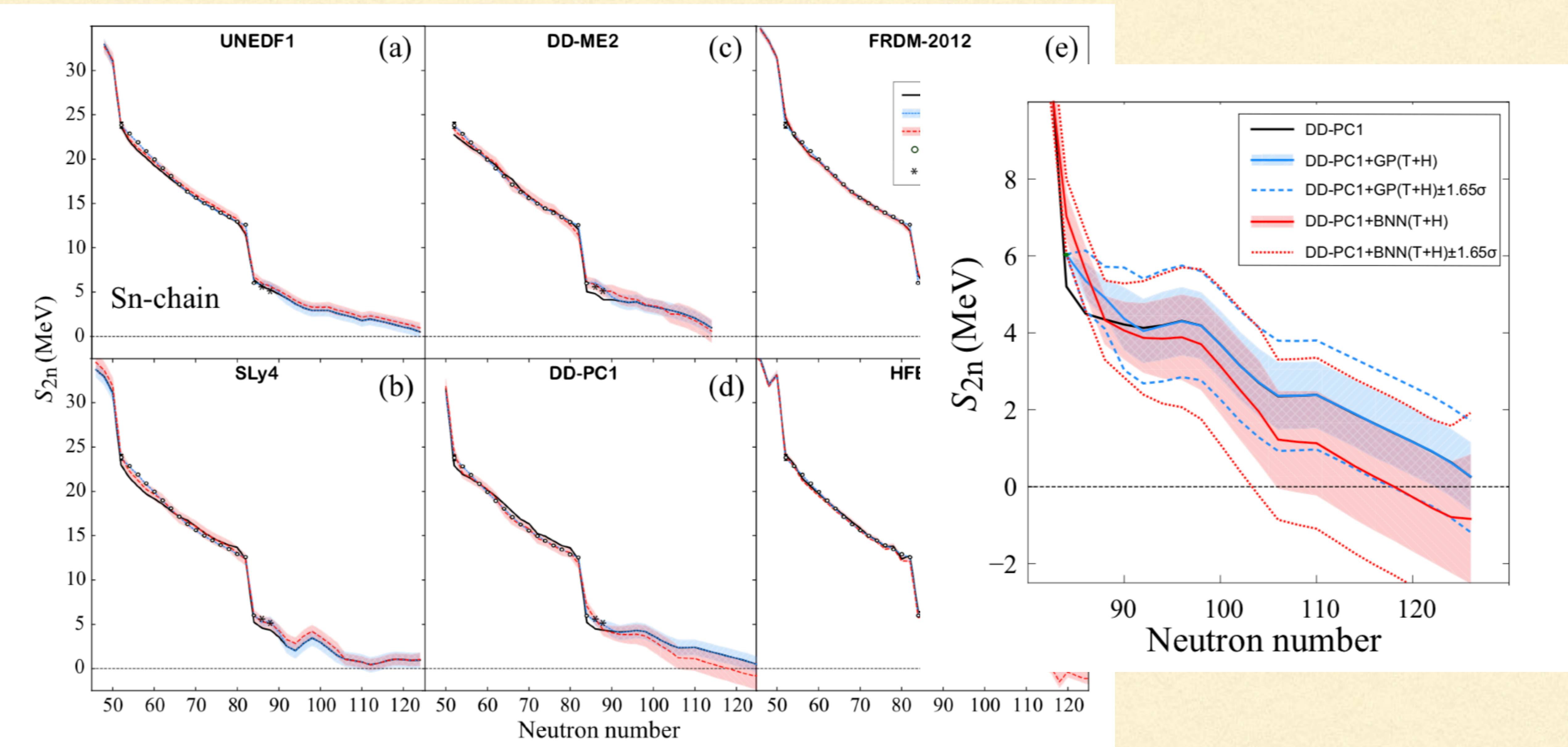
posterior distribution
(a)
(b)
(c)
UNEDF1
DD-ME2
FRDM-2012
SLy4
DD-PC1
HFB-24
S_{2n} residuals (MeV)
Neutron number
δ(Z,N)
$\delta^{GP(T+H)}$(Z,N)

Improvement & Model checking

Model	Std	T	H	$T+H$
SkM*	0.96(23)	0.96(23)	0.49(52)	0.49(52)
1.25/1.01	0.99(20)	0.81(35)	0.73(28)	0.53(47)
SLy4	0.82(13)	0.82(13)	0.52(35)	0.52(35)
0.95/0.80	0.91(3)	0.82(14)	0.71(11)	0.56(30)
SkP	0.75(11)	0.75(11)	0.38(39)	0.38(39)
0.84/0.62	0.76(9)	0.74(12)	0.59(5)	0.45(27)
SV-min	0.70(10)	0.70(10)	0.32(34)	0.33(34)
0.78/0.49	0.72(8)	1.35(−73)	0.50(−1)	0.43(12)
UNEDF0	0.73(6)	0.73(6)	0.34(37)	0.34(37)
0.78/0.54	0.87(−12)	0.73(7)	0.55(0)	0.46(16)
UNEDF1	0.61(8)	0.61(8)	0.34(30)	0.34(30)
0.66/0.49	0.79(−20)	0.74(−12)	0.53(−10)	0.32(33)
NL3*	0.84(29)	0.84(29)	0.46(47)	0.45(47)
1.19/0.86	1.10(7)	0.90(24)	0.83(4)	0.69(20)
DD-MEδ	0.73(35)	0.74(35)	0.55(42)	0.55(42)
1.13/0.96	1.08(4)	0.91(19)	0.89(7)	0.75(22)
DD-ME2	0.71(32)	0.71(31)	0.63(34)	0.62(34)
1.04/0.95	1.00(4)	1.32(−27)	0.90(5)	0.61(36)
DD-PC1	0.79(28)	0.79(28)	0.46(50)	0.46(50)
1.10/0.91	1.00(9)	1.33(−22)	0.85(7)	0.54(41)
FRDM-2012	0.57(9)	0.57(9)	0.36(25)	0.36(26)
0.63/0.49	0.61(4)	0.72(−15)	0.48(2)	0.45(7)
HFB-24	0.40(−1)	0.40(−1)	0.40(−8)	0.40(−8)
0.40/0.37	0.59(−48)	0.44(−10)	0.37(1)	0.35(6)



Extrapolation



EoS at supra-nuclear densities

Pratt, Sangaline, Sorensen, and Wang, PRL 114, 202301 (2015)

- Data from high-energy Au-Au RHIC collisions (100A + 100A GeV)
- And from Pb-Pb LHC collisions at 1.38A + 1.38A TeV
- 14-parameter model, we focus here on EoS:

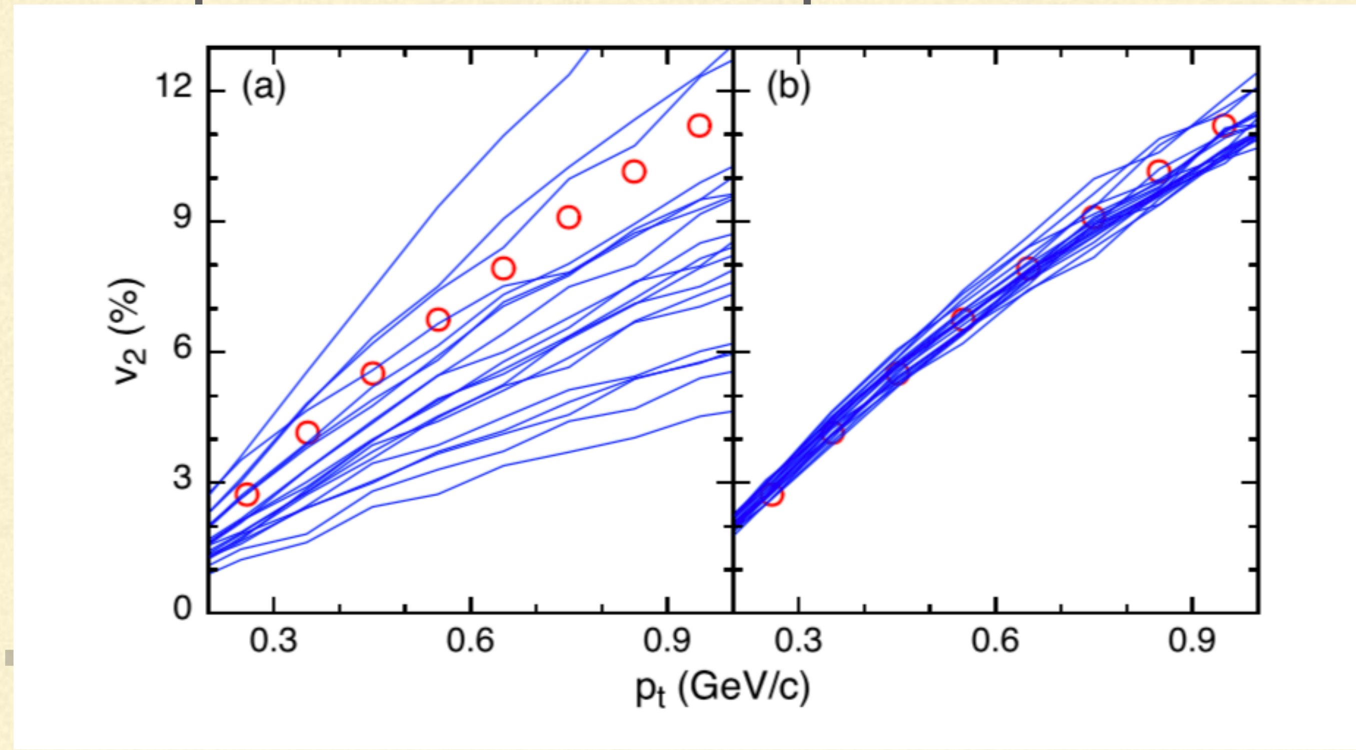
$$c_s^2(\epsilon) = c_s^2(\epsilon_h) + \left(\frac{1}{3} - c_s^2(\epsilon_h)\right)\frac{X_0 x + x^2}{X_0 x + x^2 + X'^2}$$

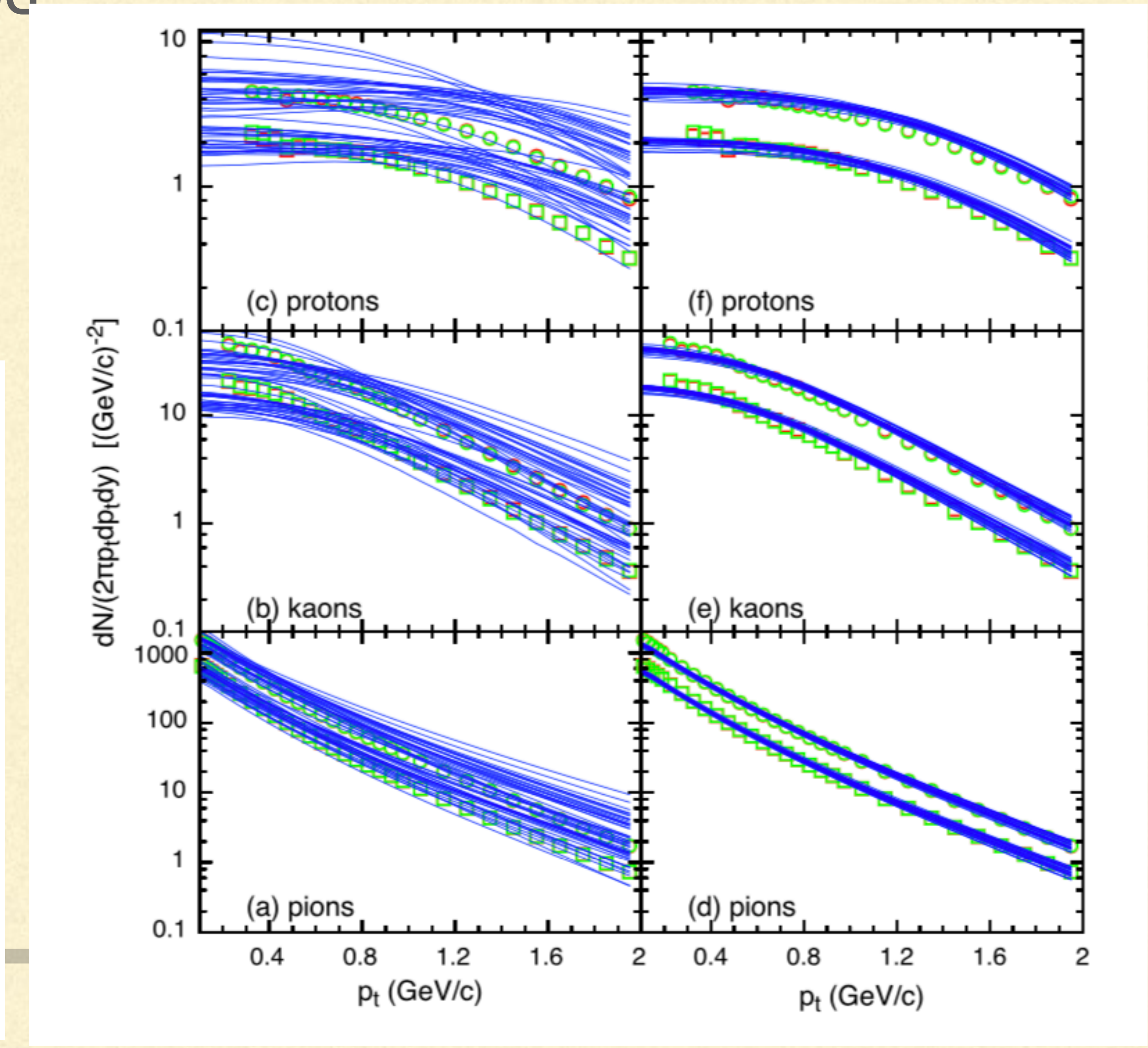
$$X_0 = X' R c_s(\epsilon)\sqrt{12}; x = \ln(\epsilon/\epsilon_h)$$

- Thirty observables: 15 from each of RHIC and LHC. Related to elliptic flow, spectra, femtoscopic source sizes for different centralities
- Expensive to evaluate that many observables in a 14-D space.

Emulation

- Thirty observables reduced to 14 principal components (99.9% of variance)
- Interpolate PCs using a GP
- 50 more points chosen a/c to likelihood
- Retrain emulator
- Repeat those two steps three times

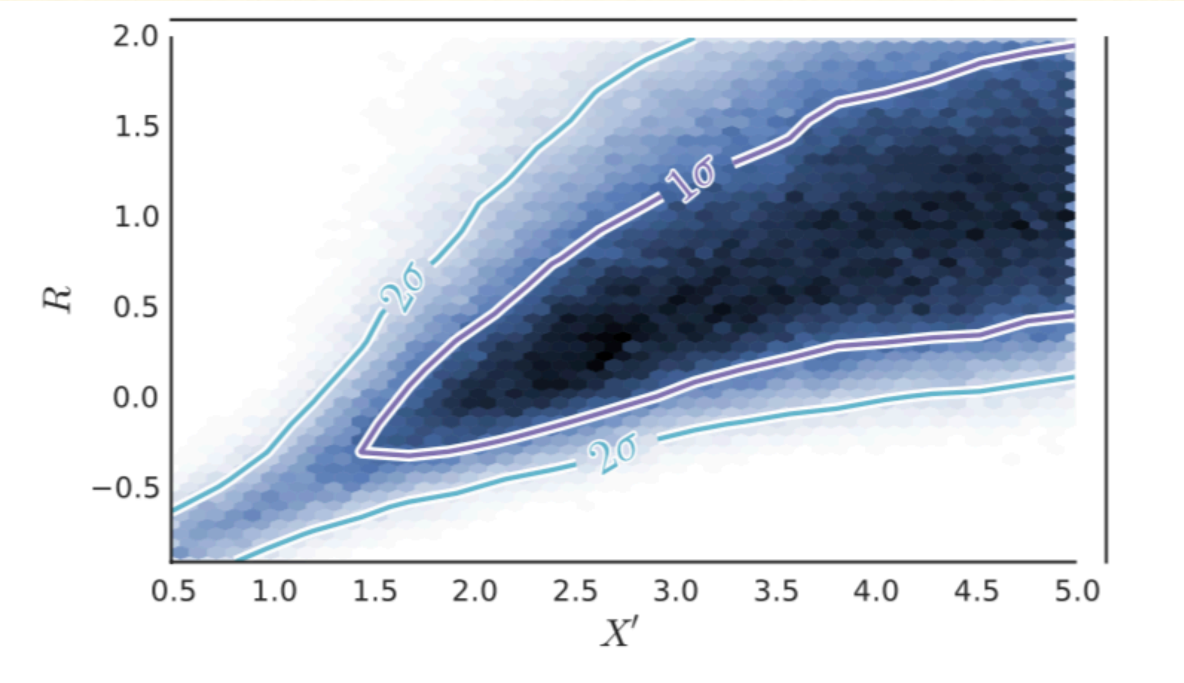


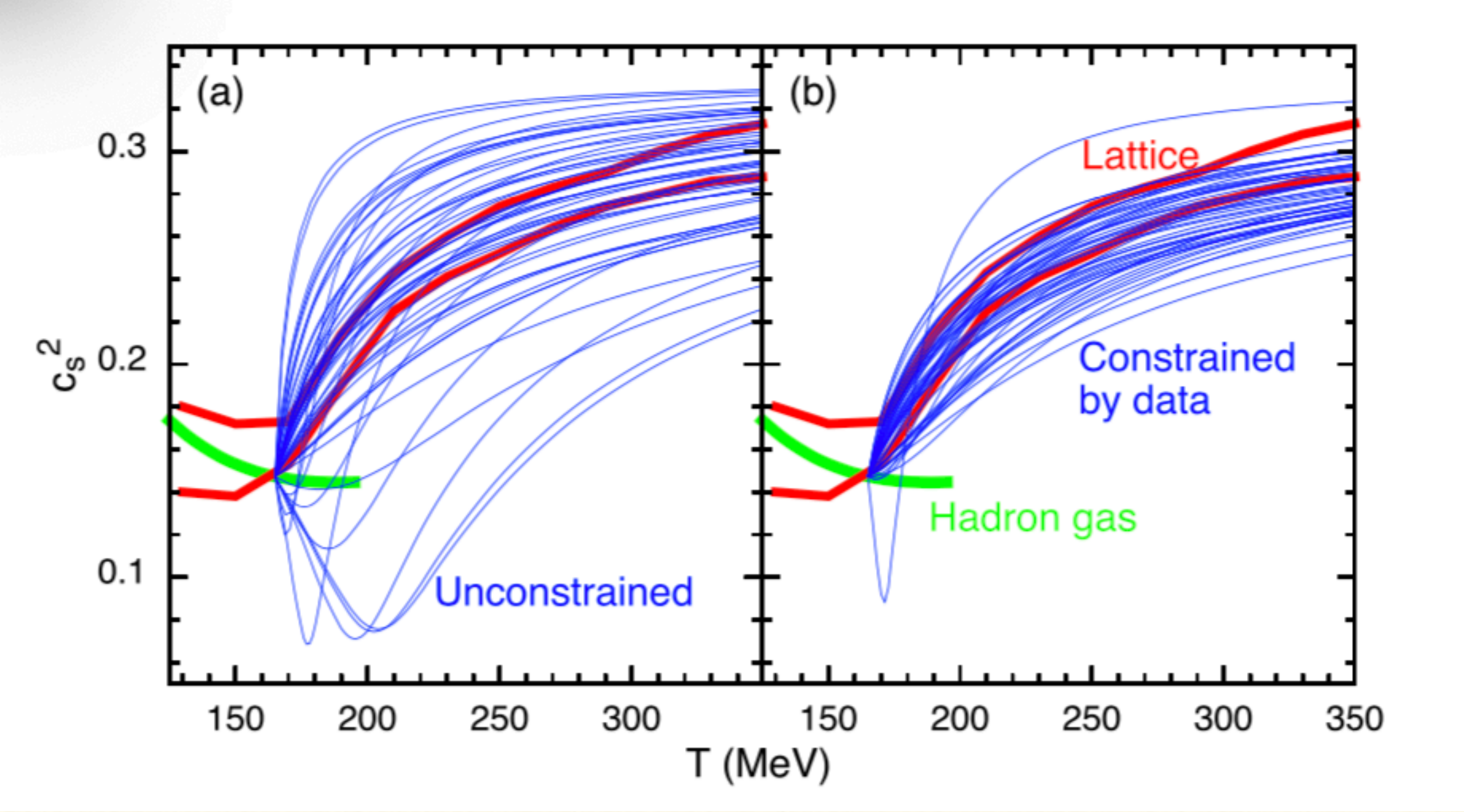


Results for EoS

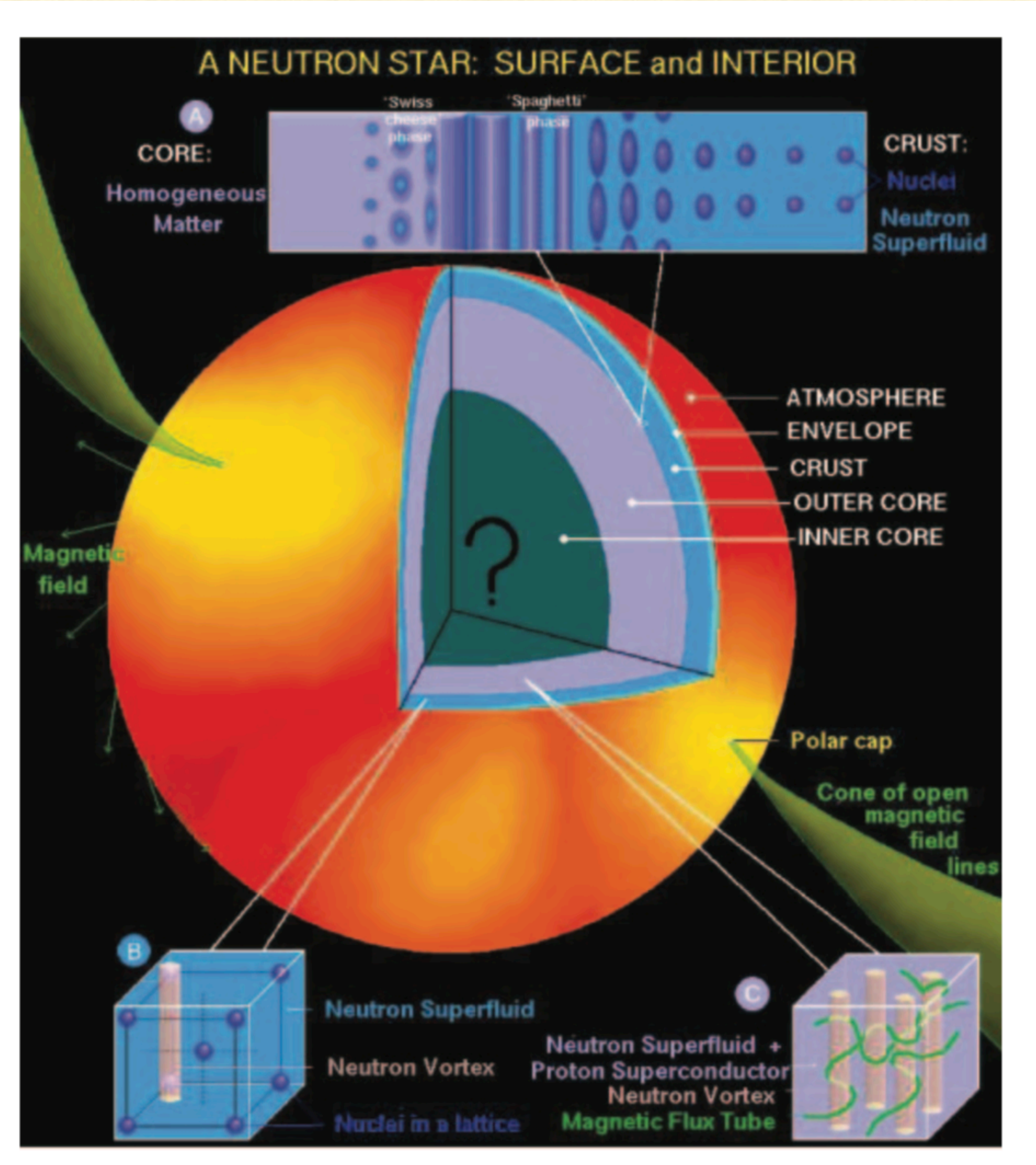
$$c_s^2(\epsilon) = c_s^2(\epsilon_h) + \left(\frac{1}{3} - c_s^2(\epsilon_h)\right) \frac{X_0 x + x^2}{X_0 x + x^2 + X'^2}$$

$$X_0 = X' R c_s(\epsilon)\sqrt{12};\, x = \ln(\epsilon/\epsilon_h)$$





Neutron stars and the EoS of dense matter



- EoS: p(ε)
- Derive sound speed from it

$$c_s^2 = \frac{dp}{d\epsilon}$$

- Rigidity of matter affects radius of the star that gets "built" for a given mass
- p(ε)⇒M(R)

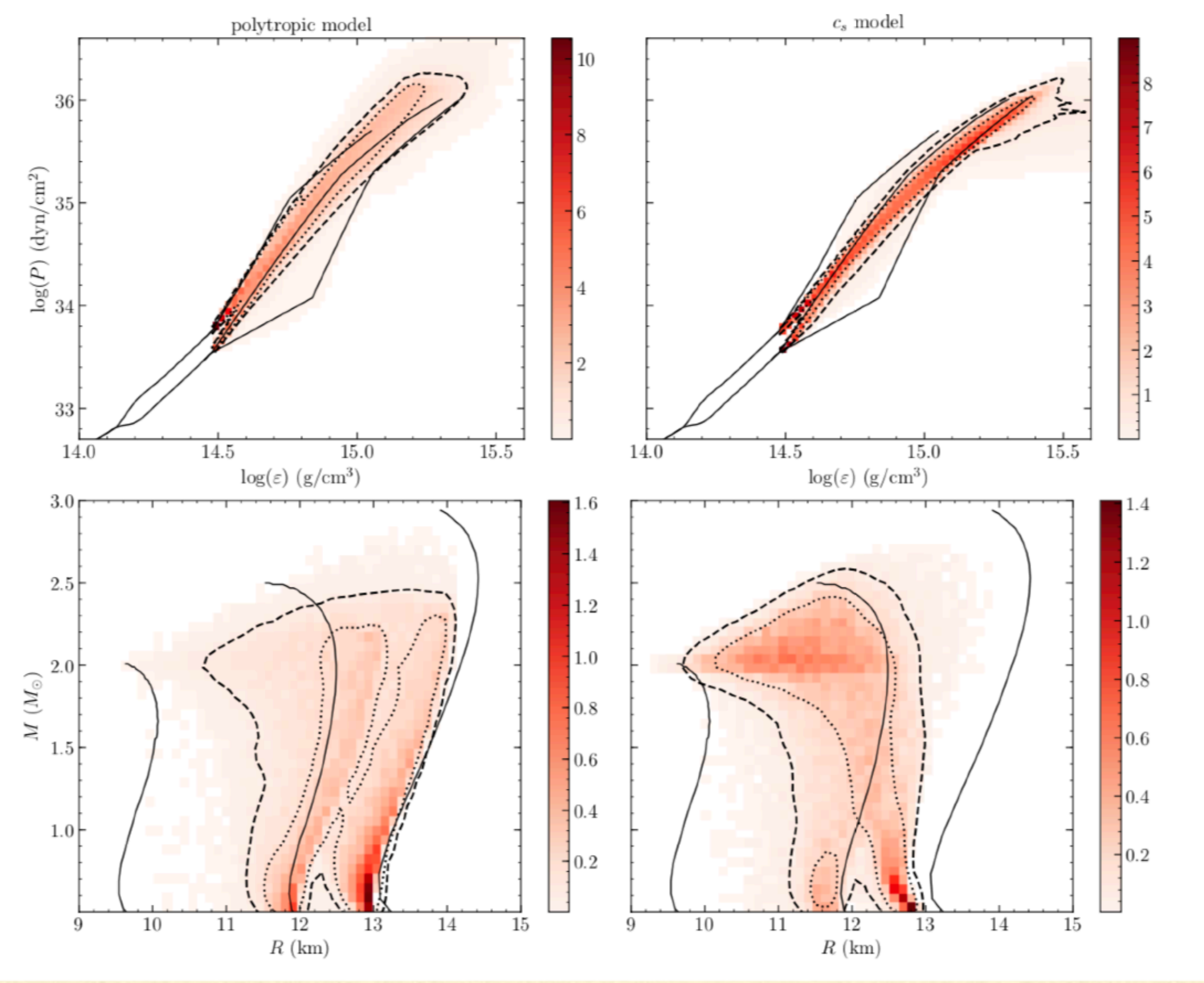
Lattimer, Prakash, Science (2004)

Flat in what? Priors and the mass-radius relation

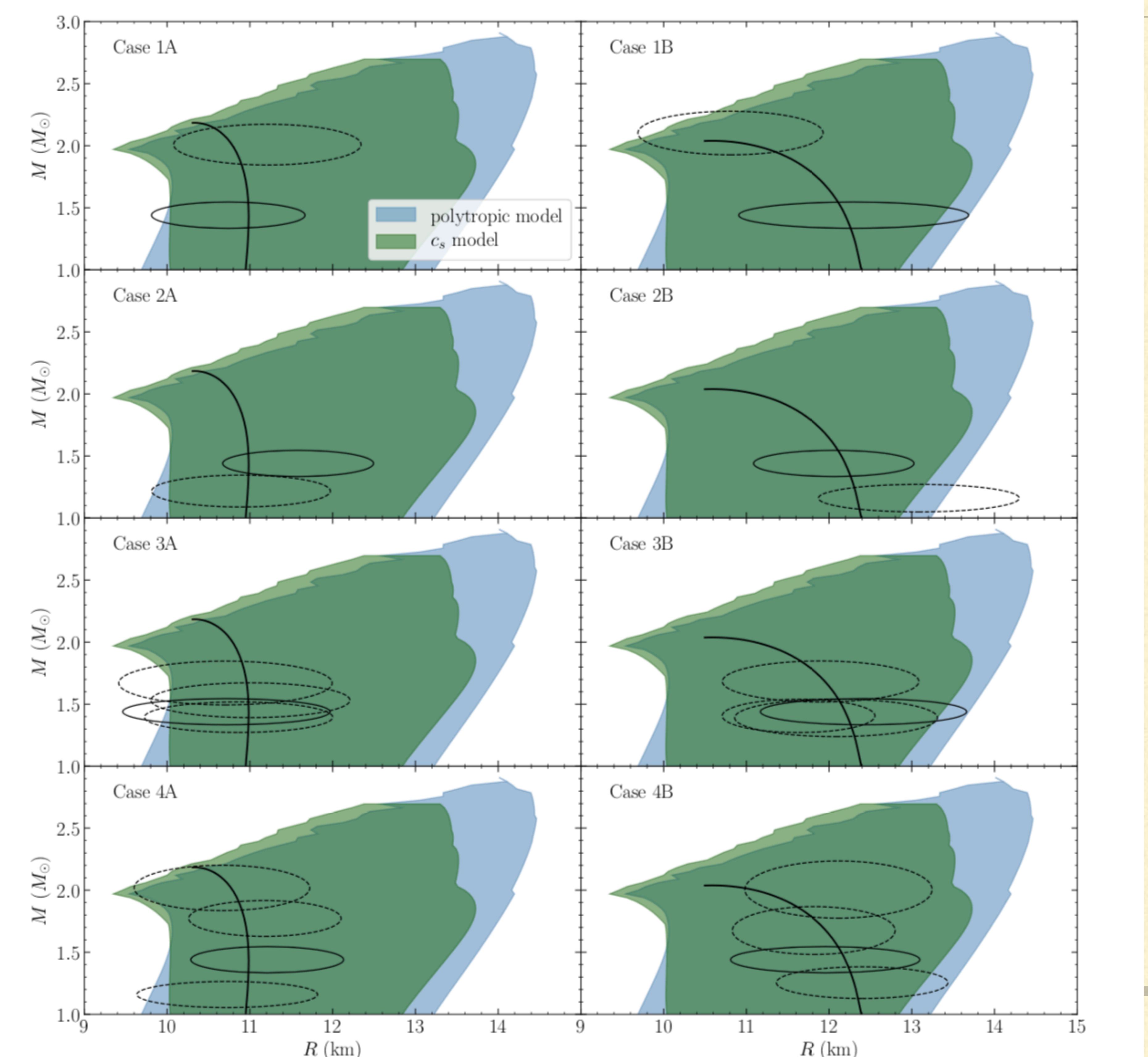
Greif et al., MNRAS (2019)

- Two parameterizations: PP and CS
- CS: $\frac{c_s^2}{c^2} = a_1 \exp\left(-\frac{(x-a_2)^2}{2a_3^2}\right) + a_6 + \frac{\frac{1}{3} - a_6}{1 + \exp(-a_5(x-a_4))}$
- $x = \epsilon/\epsilon_0$ and integrate to get p(ε)
- Sample uniformly in parameters

Goal: understand what mass & radius observations from NICER will teach us about the nuclear EoS

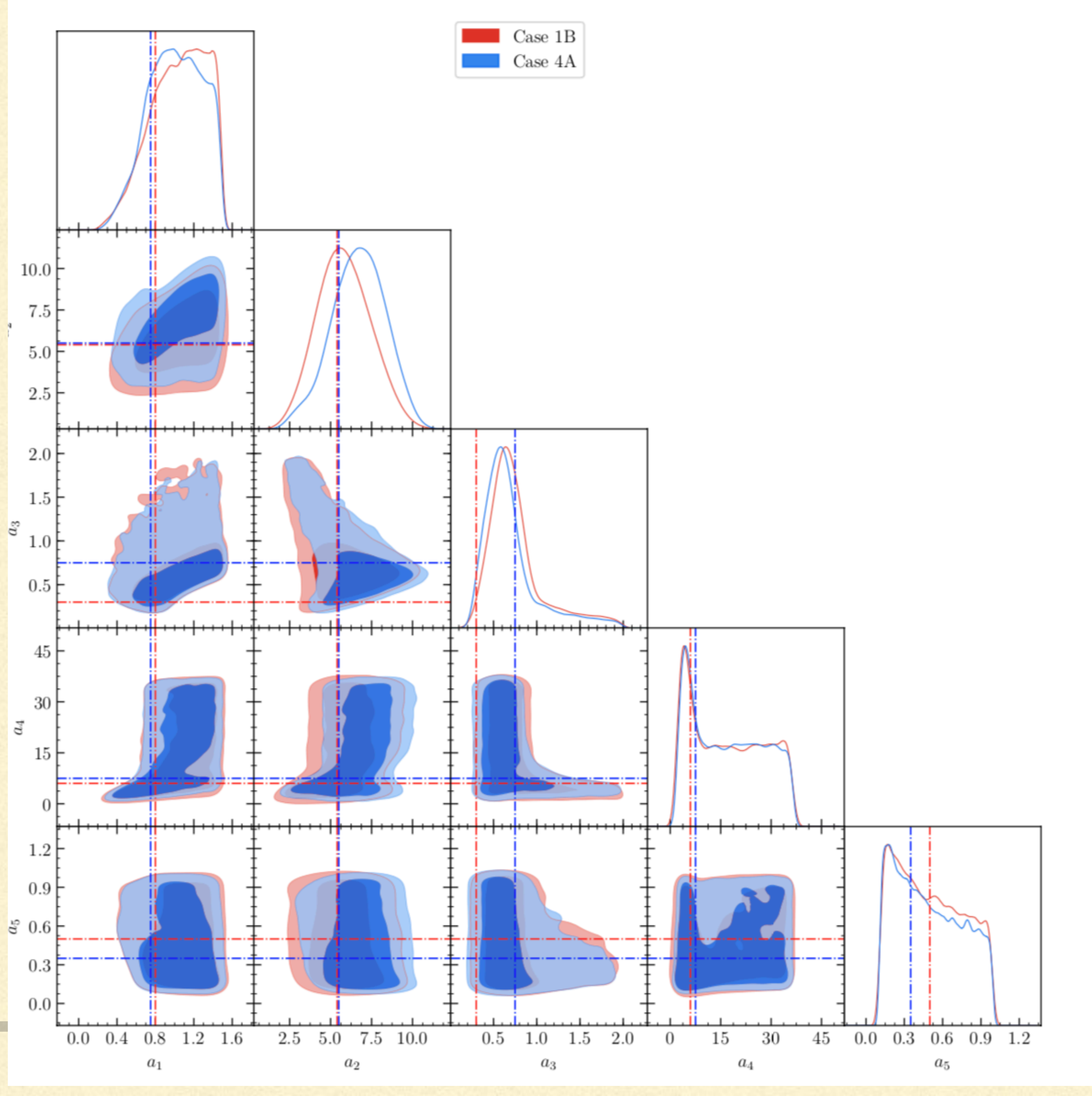


NICER futures

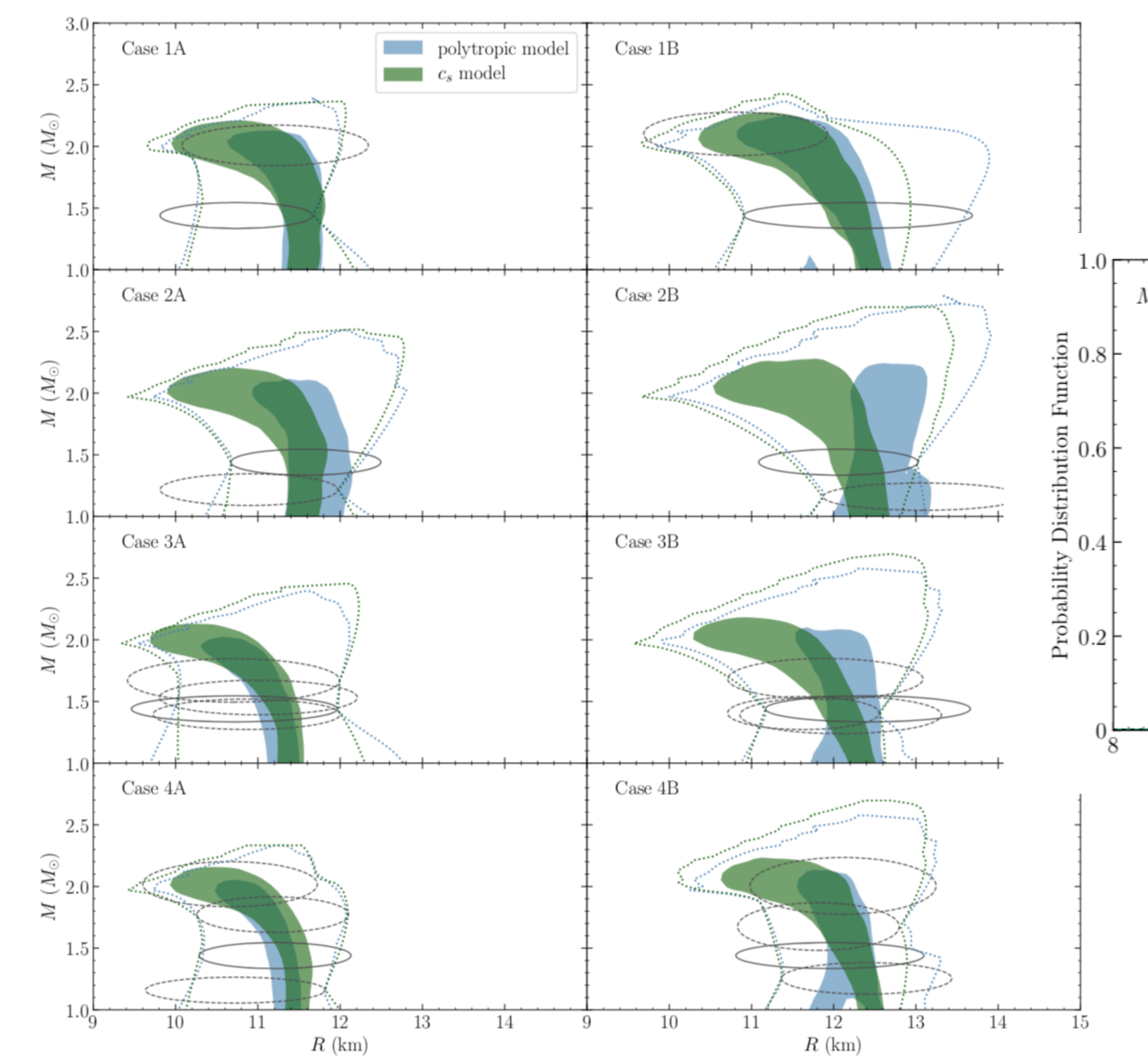


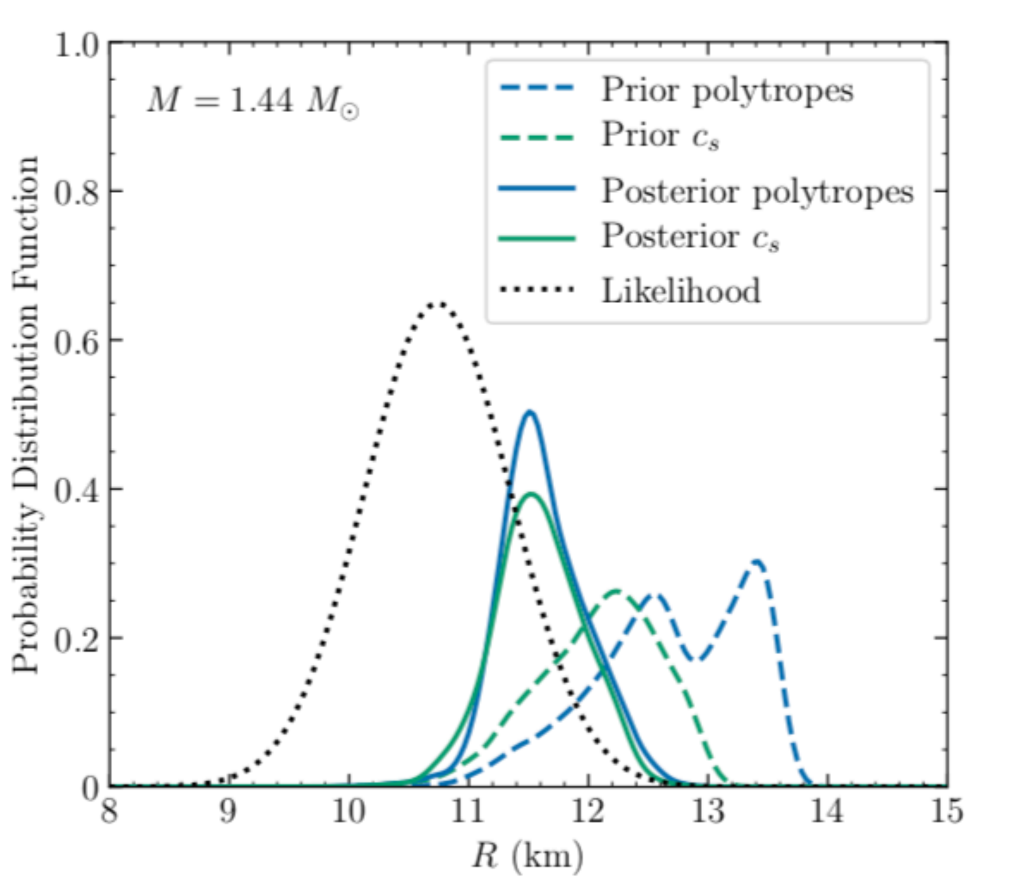
NICER parameter constraints

- Sampling shows how different NICER outcomes would reduce space of parameters
- This case for CS model



Mass-radius posteriors





Lessons

- Bayesian methods are useful for setting bounds in presence of background
- Theory errors matter, sometimes
- When using ML methods choice of training set is crucial
- And validation (model checking) is essential
- GP > BNN?
- Calculation too costly? Try an emulator
- Make sure you understand how your prior affects your posterior