UPPSALA UNIVERSITET

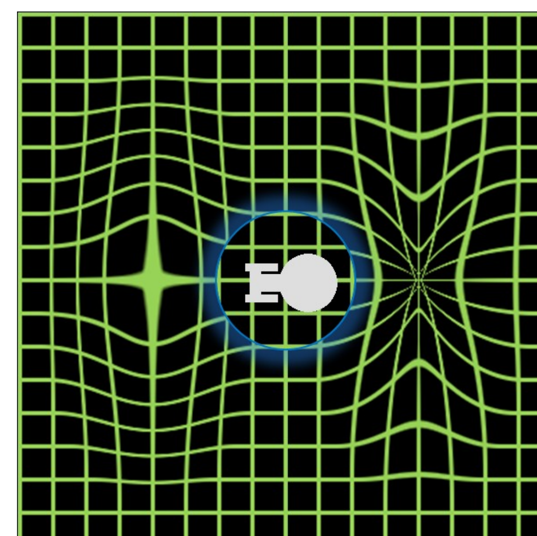
Dynamic time warping Gaussian processes for nuclear physics needs

(calibration, nuclear data evaluation, parameter estimation, uncertainty quantification)

Georg Schnabel, Henrik Sjöstrand

Department of Physics and Astronomy
Uppsala University

georg.schnabel@physics.uu.se

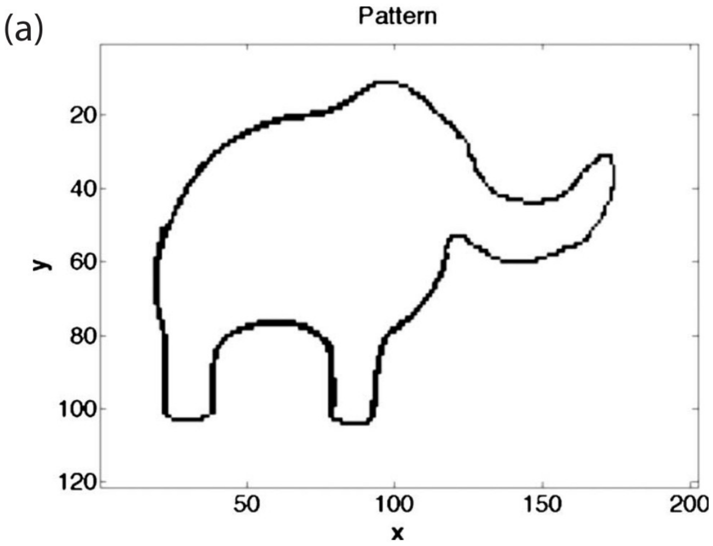


WONDER, 8-12 Oct 2018, Aix-en-Provence, France

What did Johnny think...

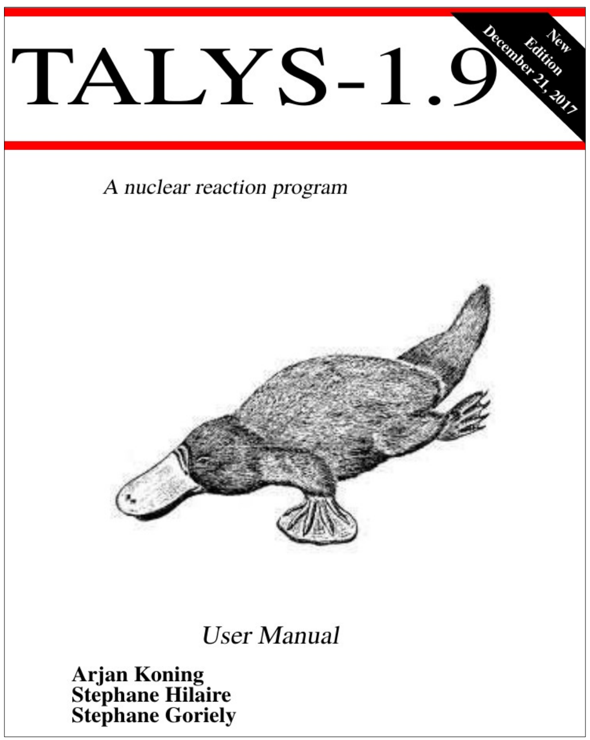
John von Neumann
(1903 - 1957)

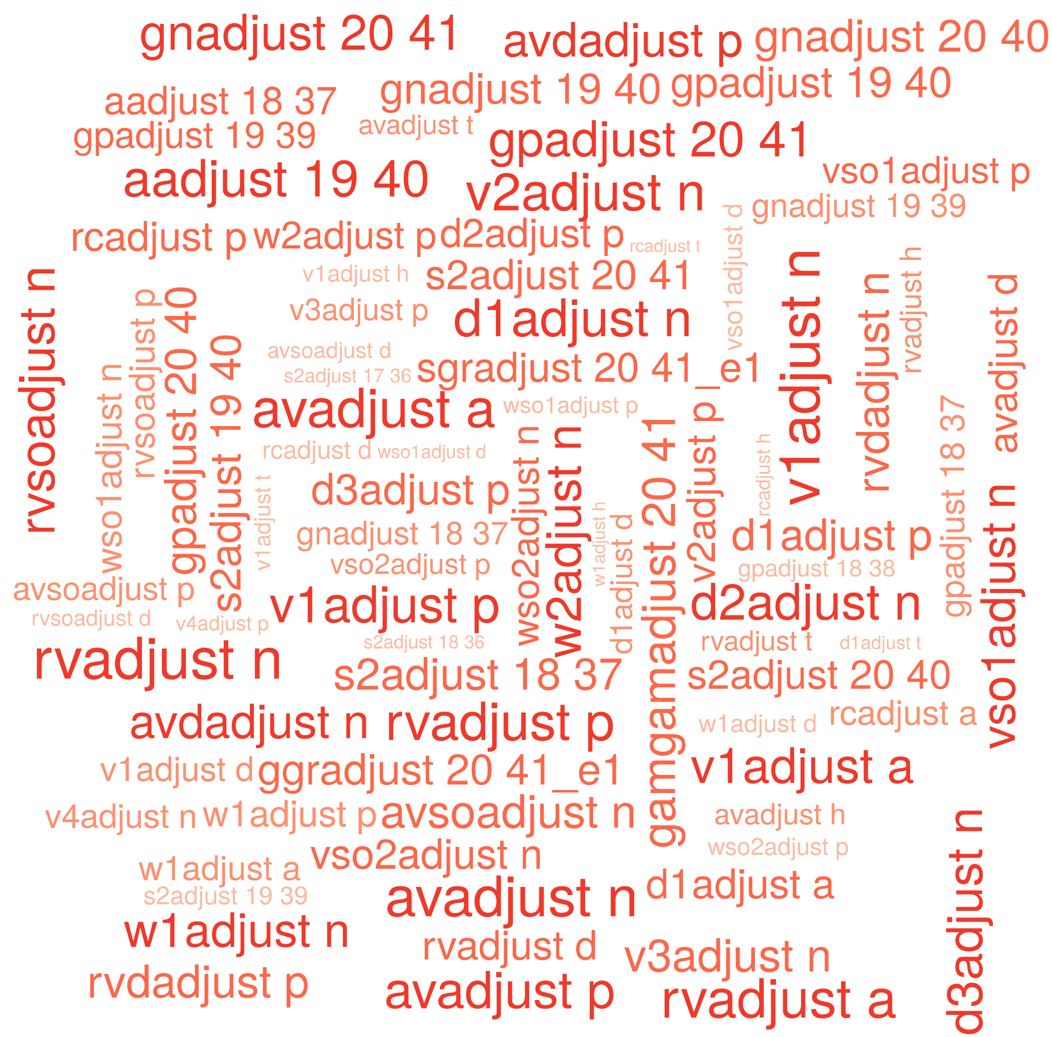
"With four parameters I can fit an elephant, and with five I can make him wiggle his trunk."



"**Drawing an elephant with four complex parameters**"
Jürgen Mayer, Khaled Khairy, and Jonathon Howard,
Am. J. Phys. 78, 648 (2010), DOI:10.1119/1.3254017.

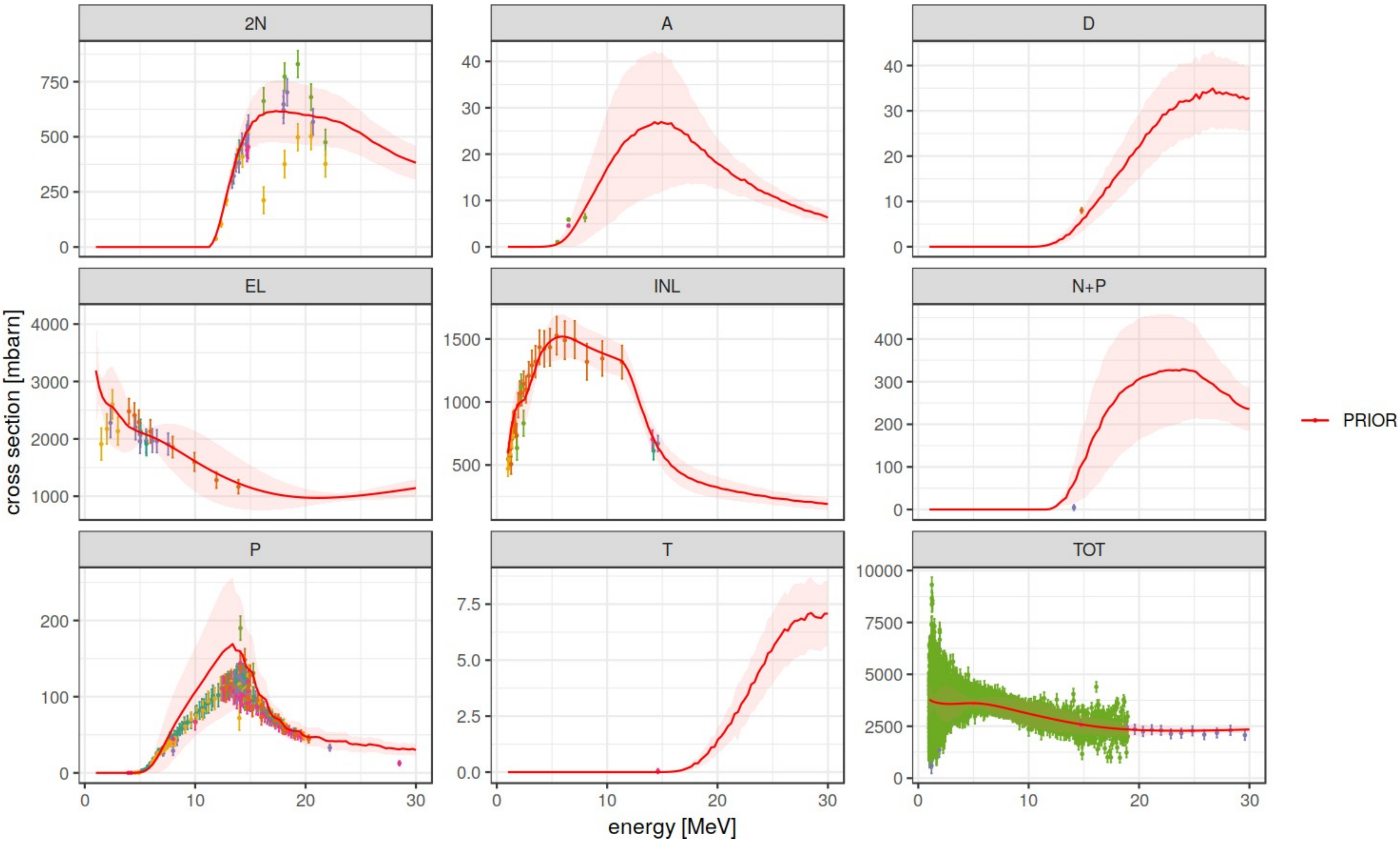
Swiss army knife of nuclear data



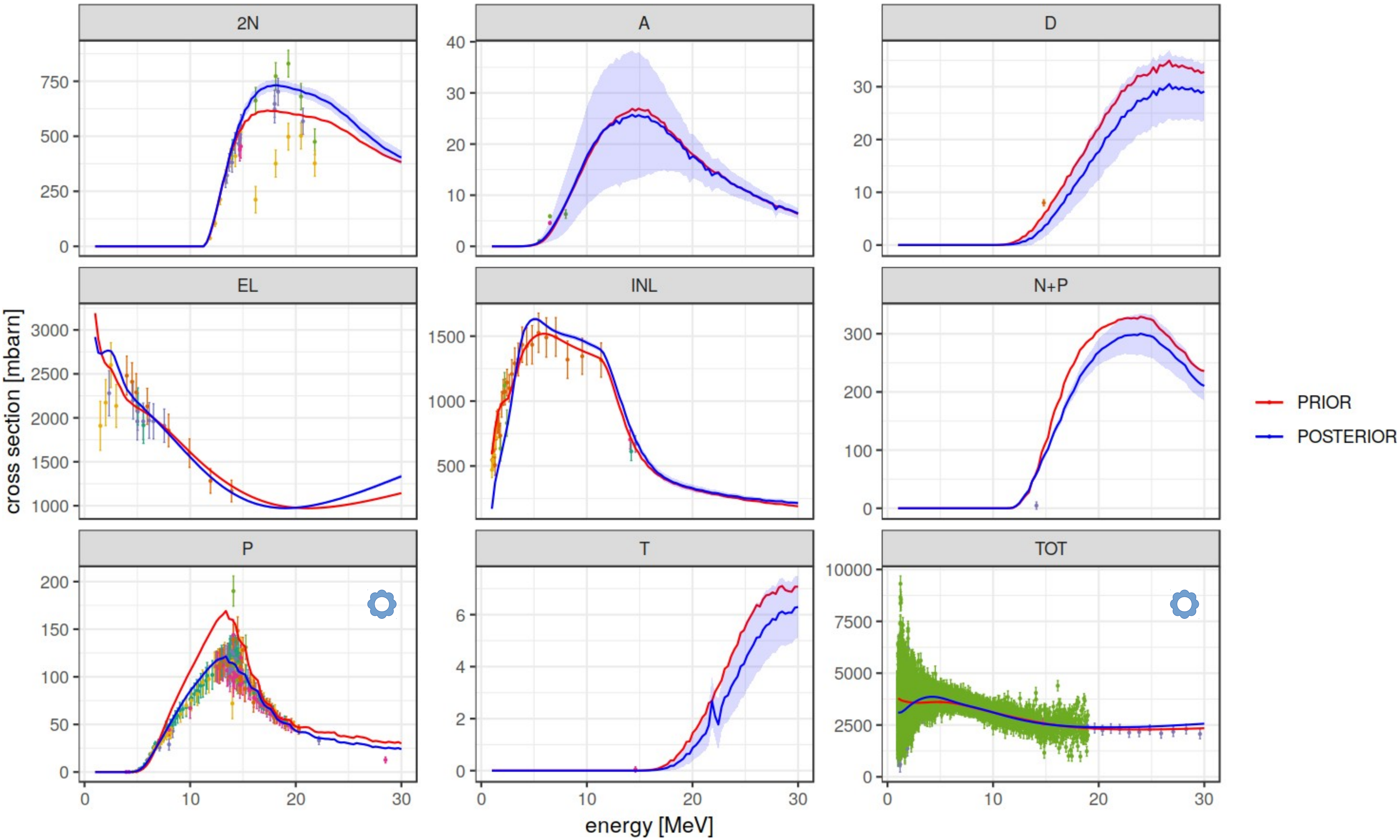


What can we fit with ~350 TALYS parameters?

^{56}Fe differential cross sections (n, ...)

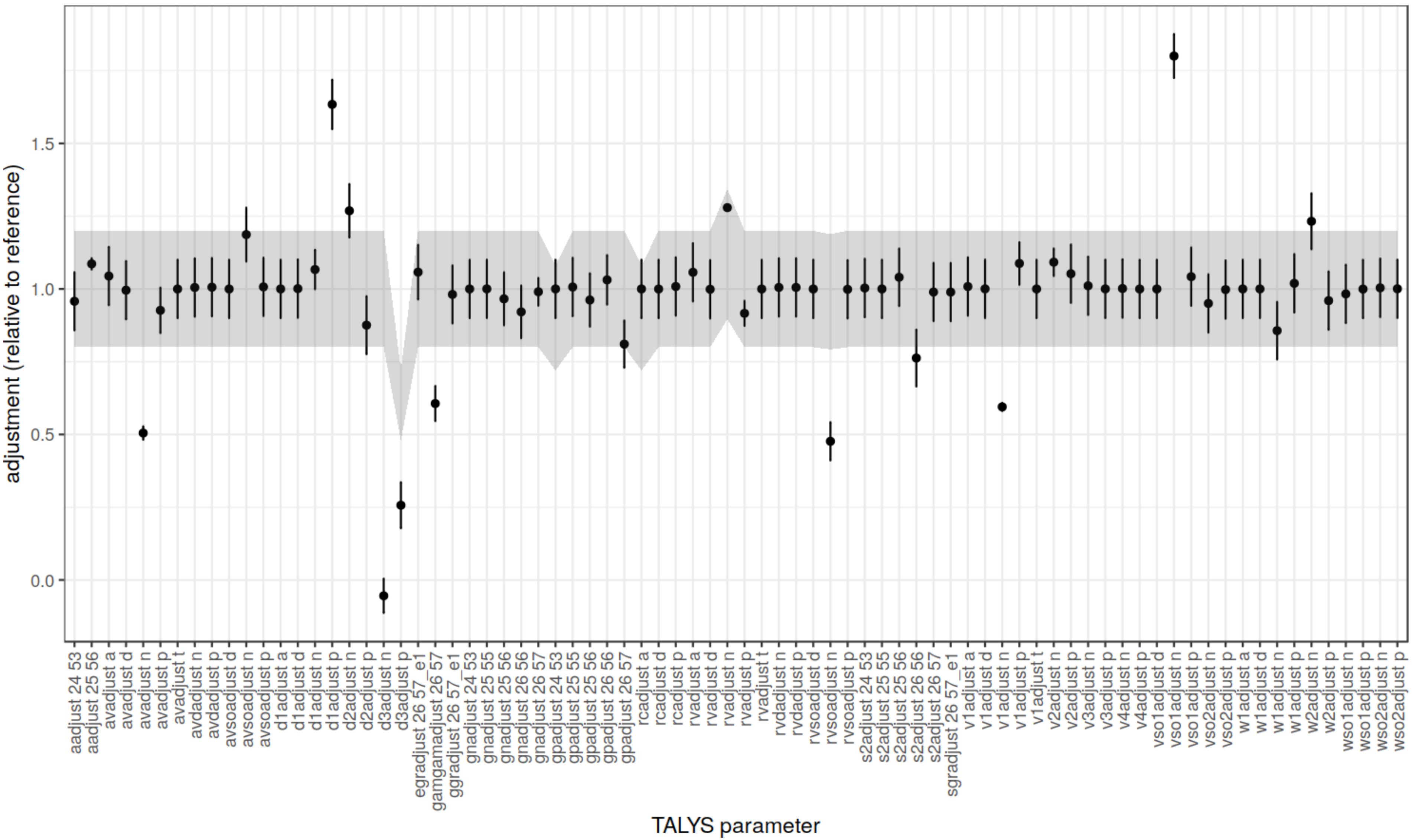


^{56}Fe --- update with (n,tot) and (n,p)

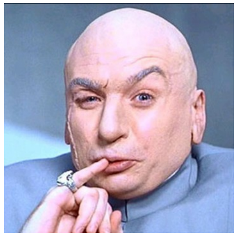


(e.g. EMPIRE-Kalman, SAMMY Bayesian update, ...)

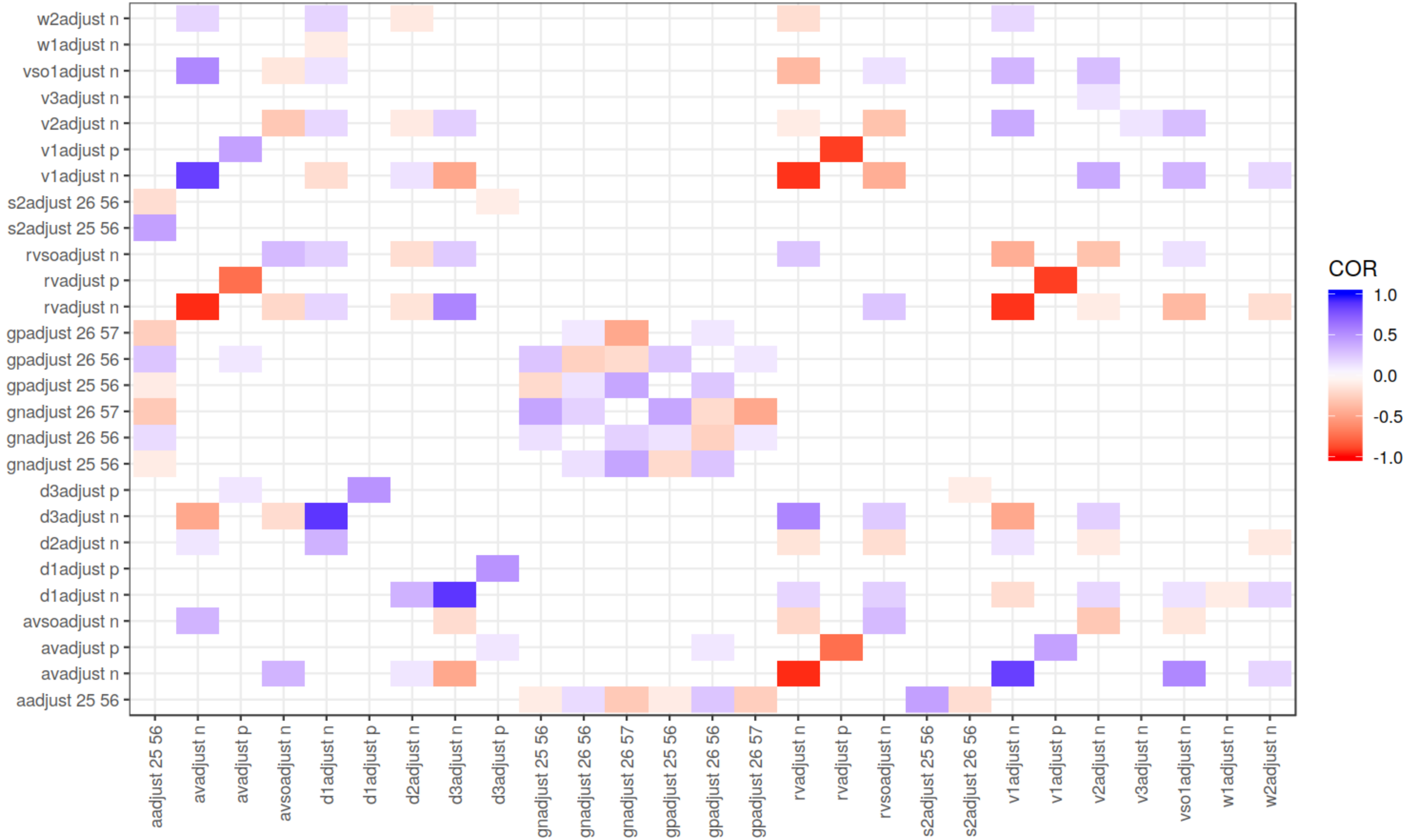
Parameter outlaws



Evil correlations

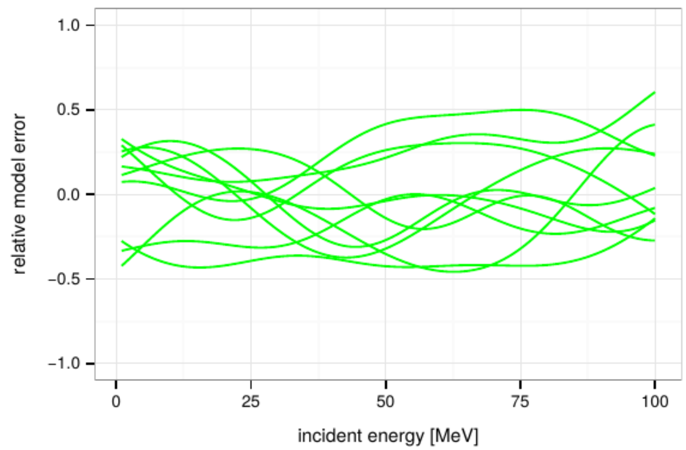


Mike Myers as Dr. Evil in Austin Powers



Take away

- Reasonable (not too low) uncertainties and parameter estimates of physics models can in general not be obtained by a χ^2 fit (or GLS, EMPIRE-Kalman, etc. without model defects)
- Otherwise we risk running the model outside sound specifications and losing its predictive power
- We need **model defects**! (Especially if we can fit an elephant or two)



Pigni, M.T., Leeb, H., 2003. Uncertainty Estimates of Evaluated 56Fe Cross Sections Based on Extensive Modelling at Energies Beyond 20 MeV, in: Proc. Int. Workshop on Nuclear Data for the Transmutation of Nuclear Waste. GSI-Darmstadt, Germany.

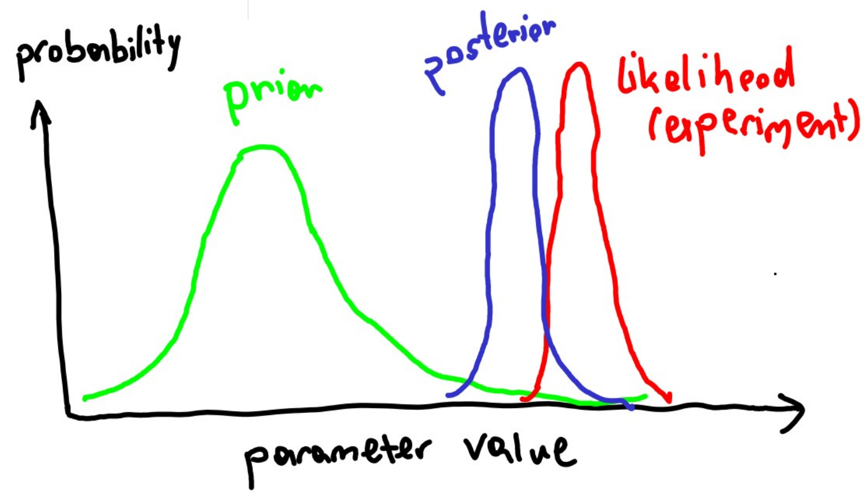
Don’t break the law!

$$\chi^2 = (\vec{\sigma}_{\text{exp}} - \mathcal{M}(\vec{p}_{\text{ref}}))^T M^{-1} (\vec{\sigma}_{\text{exp}} - \mathcal{M}(\vec{p}_{\text{ref}}))$$

Side note
Serves as criterion to remove experimental outlieres in TSURFER module of SCALE. Option called $\Delta\chi^2$-filtering there

$$M = S K_{\text{par}} S^T + K_{\text{exp}}$$

$$M = S K_{\text{par}} S^T + K_{\text{def}} + K_{\text{exp}}$$



How to choose K_{def}?

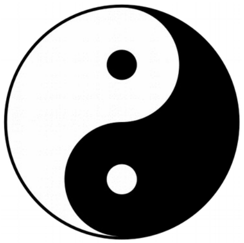
$$\text{if } K_{\text{def}} \to \infty \text{ then } \chi^2 \to 0$$

$$\pi(\mathcal{D}) = \int \ell(\mathcal{D}\,|\,\vec{p})\,\pi(\vec{p})d\vec{p}$$

$$M = SK_{\text{par}}S^T + K_{\text{def}} + K_{\text{exp}}$$

Marginal likelihood maximization

$$\log \pi(\mathcal{D}) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|M| - \frac{1}{2}\chi^2(M)$$



Brief motivation of criterion

$$M = \begin{pmatrix} \delta^2 & 0 & \cdots \\ 0 & \delta^2 & \cdots \\ 0 & 0 & \ddots \end{pmatrix}$$

$$-\frac{1}{2}\chi^2(M) = -\frac{1}{2\delta^2}\sum_{i=1}^{N}(\text{exp}_i - \text{mod}_i)^2$$

$$-\frac{1}{2}\log|M| = -\frac{N}{2}\log\delta^2$$

$$\mathbb{E}\left[-\frac{1}{2}\chi^2 M\right] = -\frac{N\delta^2_{\text{true}}}{2\delta^2}$$

$$\delta^2 = \delta^2_{\text{true}} + \Delta$$

$$\log(\delta^2_{\text{true}} + \Delta) = \log(\delta^2_{\text{true}}) + \frac{\Delta}{\delta^2_{\text{true}}} - \frac{1}{2}\frac{\Delta^2}{\delta^4_{\text{true}}} + \dots$$

$$\frac{\delta^2_{\text{true}}}{\delta^2_{\text{true}} + \Delta} = 1 - \frac{\Delta}{\delta^2_{\text{true}}} + \frac{\Delta^2}{\delta^4_{\text{true}}} + \dots$$

$$\log\pi(\mathcal{D}) = \text{const} - \frac{N}{4}\frac{\Delta^2}{\delta^4_{\text{true}}} \rightarrow \text{maximal for } \Delta = 0, \text{ i.e. } \delta^2 = \delta^2_{\text{true}}$$

Cyrille De Saint Jean, Noguere, G., Habert, B., Iooss, B., 2008. A Monte Carlo approach to nuclear model parameter uncertainties propagation. Nuclear Science & Engineering 161, 363.

New stuff

- Use the data from neighboring isotopes to learn about the apriori performance of the nuclear model on a per energy basis (*)
- Replace the amplitude and length-scale hyperparameter in the GP by an amplitude function and a metric function

(*) This point by itself not new, see PhD thesis of Denise Neudecker and [2]

Related work

[1] Schnabel, G., 2017. Estimating model bias over the complete nuclide chart with sparse Gaussian processes at the example of INCL/ABLA and double-differential neutron spectra arXiv:1803.00928 (accepted for EPJ-N)

[2] Leeb, H., Neudecker, D., Srdinko, T., 2008. Consistent Procedure for Nuclear Data Evaluation Based on Modeling. Nuclear Data Sheets 109, 2762–2767. https://doi.org/10.1016/j.nds.2008.11.006

Interesting to discuss commonalities and differences but ... … mind the chairman

One essential difference:
Here we account for model systematics (in 1st order)

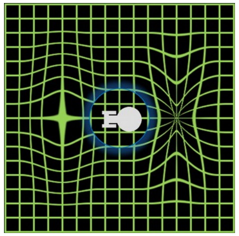
$k(E,E') = \delta^2 \exp\left(-\frac{1}{2}\frac{(E-E')^2}{\lambda^2}\right)$

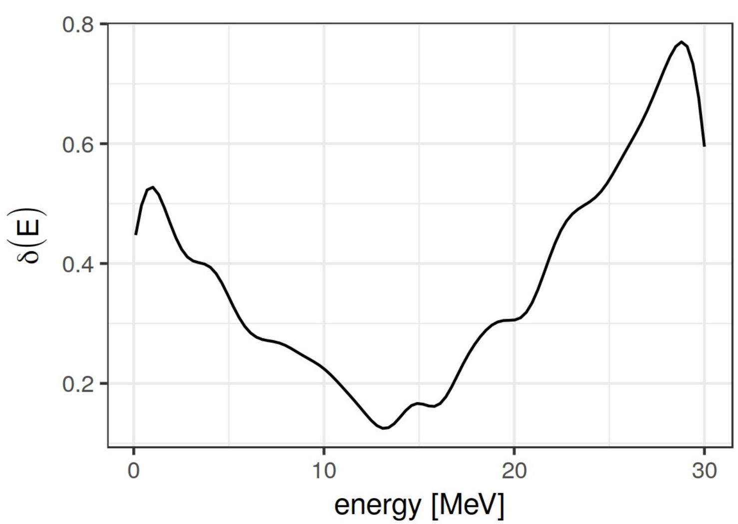
Dynamic time warping GP

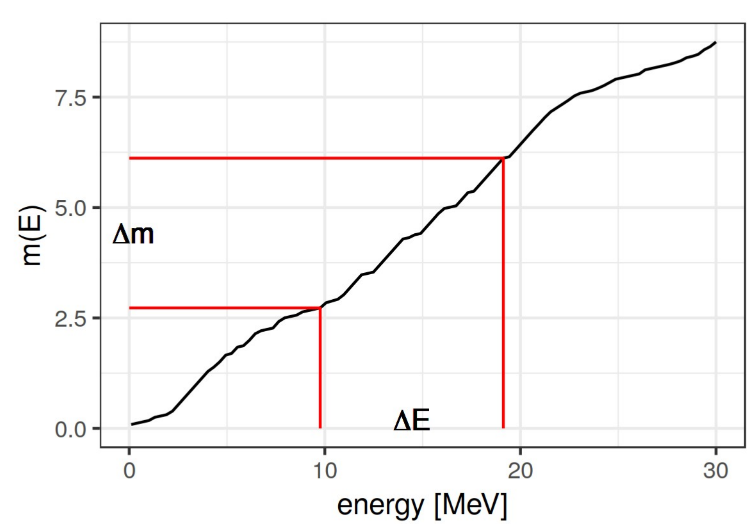
$$k(E,E') = \delta(E)\delta(E') \exp\left(-\frac{1}{2}\Big(m(E) - m(E')\Big)^2\right)$$

$$\delta(E) = \sum_{i=1}^{99}\left(\frac{E_{i+1}-E}{E_{i+1}-E_i}\mathbf{y_i} + \frac{E-E_i}{E_{i+1}-E_i}\mathbf{y_{i+1}}\right)\mathcal{I}_{(E_i \le E < E_{i+1})}(E)$$

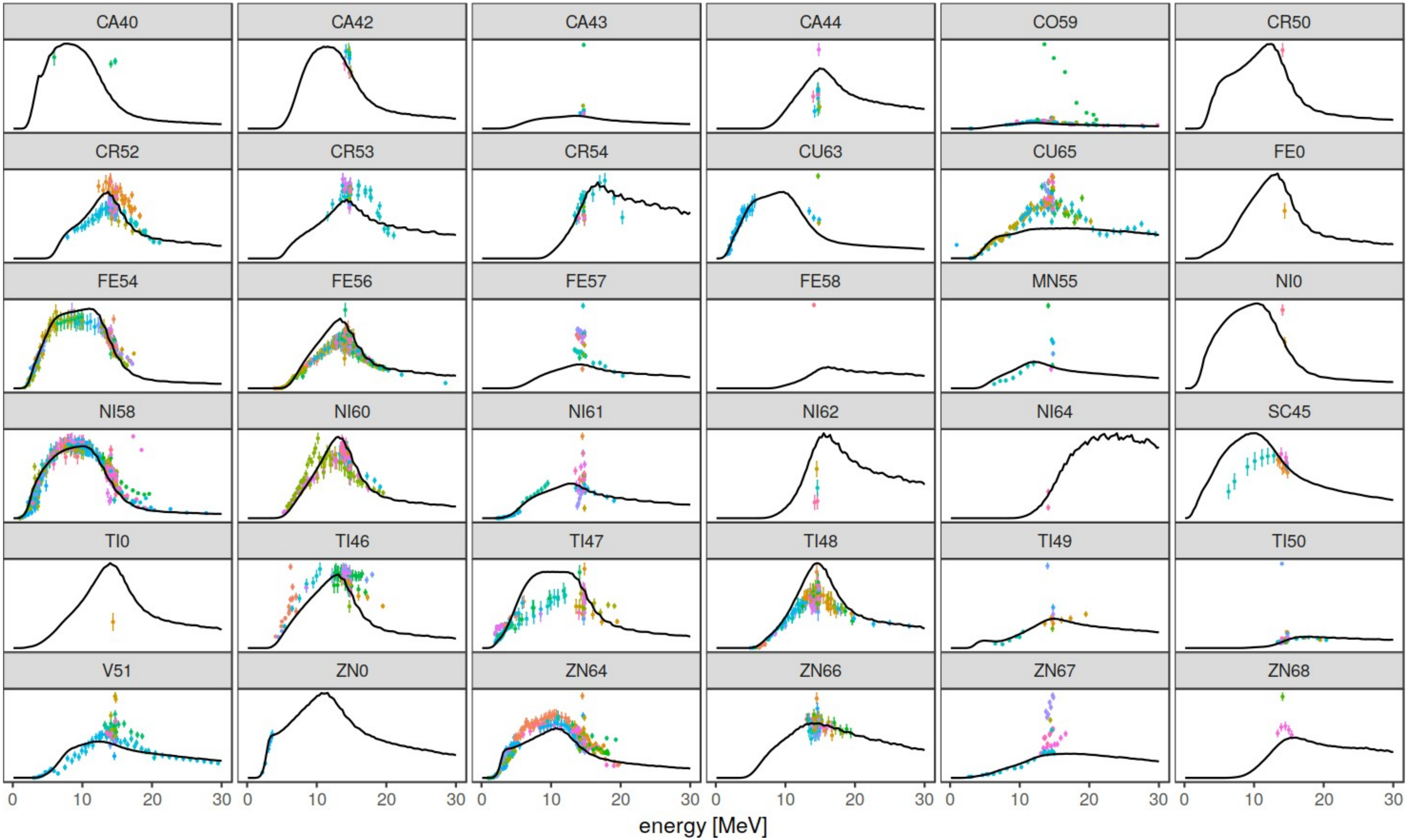
$$m(E) = \sum_{i=1}^{99}\left(\frac{E_{i+1}-E}{E_{i+1}-E_i}\mathbf{z_i} + \frac{E-E_i}{E_{i+1}-E_i}\mathbf{z_{i+1}}\right)\mathcal{I}_{(E_i \le E < E_{i+1})}(E)$$



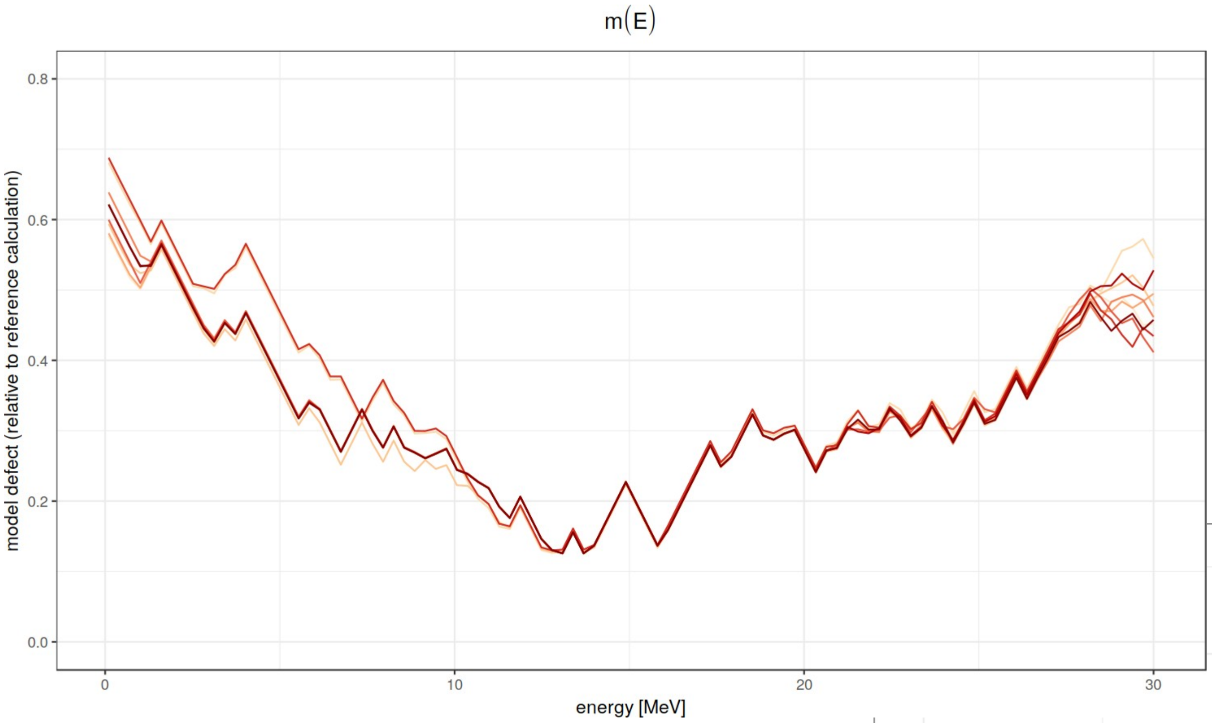




(n,p) reactions as an example

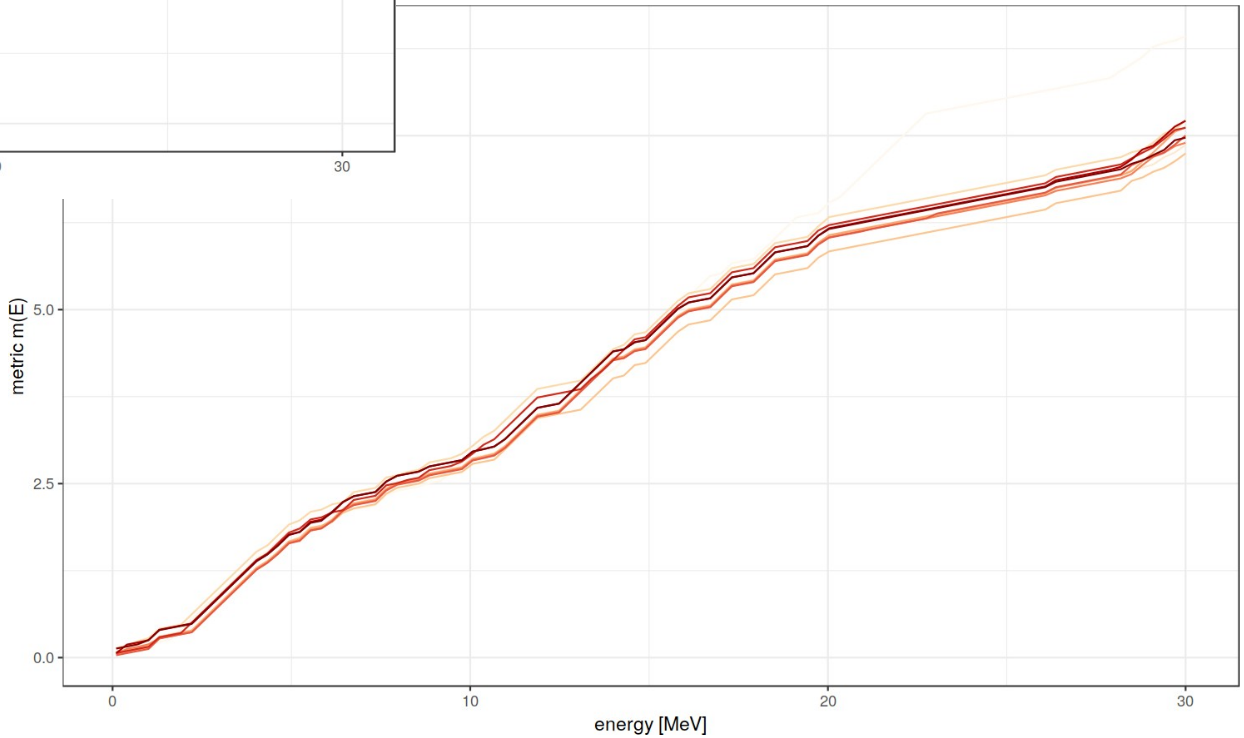


Marlike maximization with (n,p) data



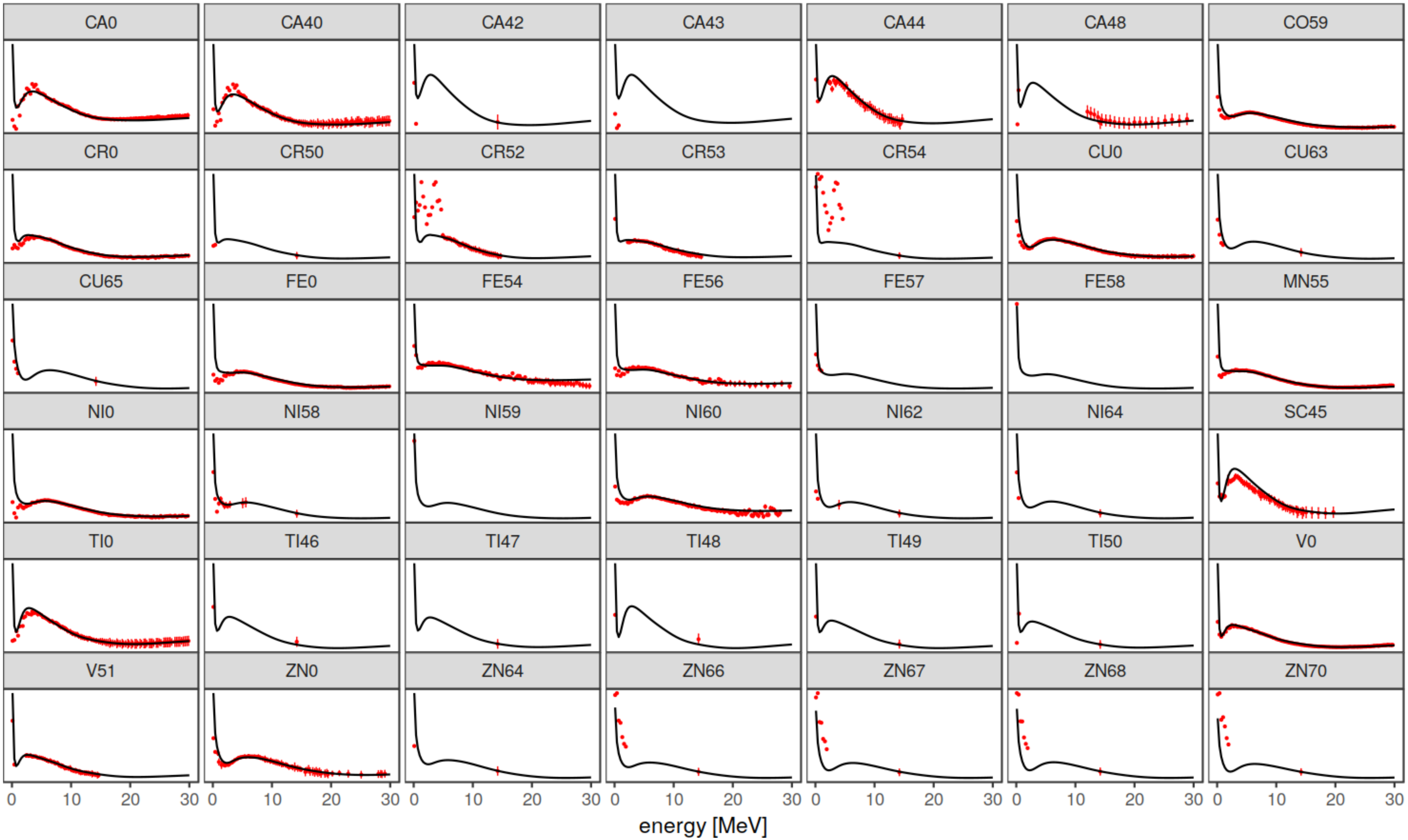
Link to animation [MP4]

Link to animation [GIF]



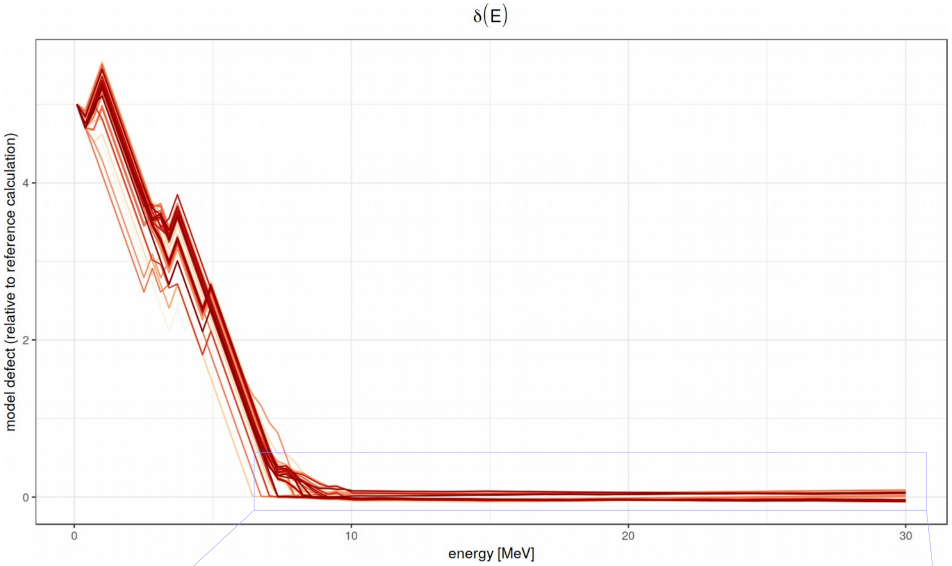
Remark
Reasonable constraints are important for a successful optimization (e.g., lower bound for length-scale, upper bound for maximal local difference of amplitude, etc.)

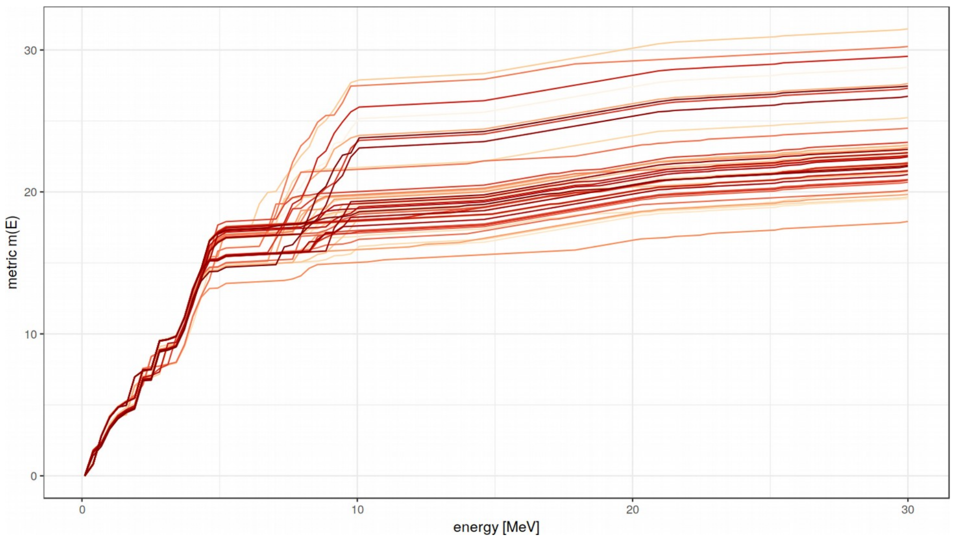
Another example (n,tot)

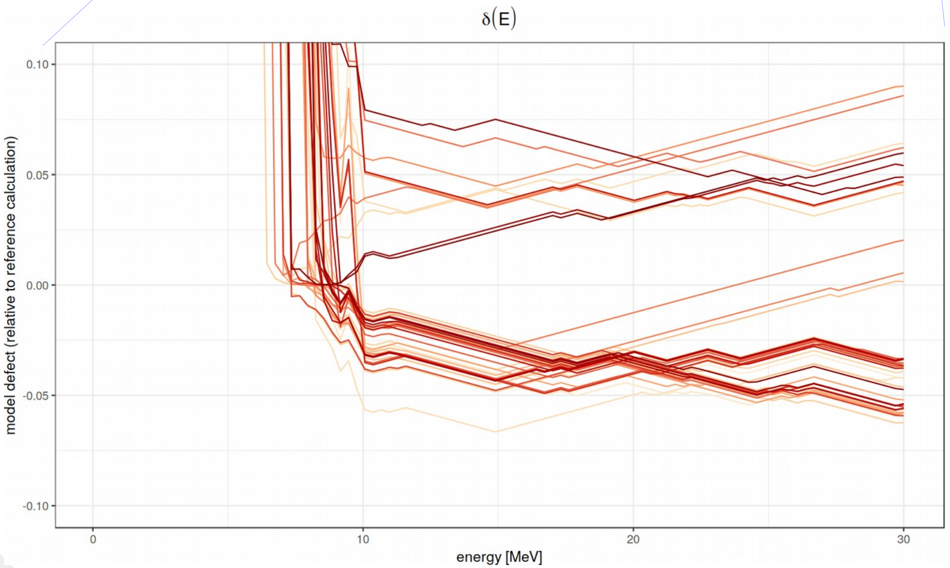


(0,3 MeV resolution)

Marlike maxim with (n,tot) data





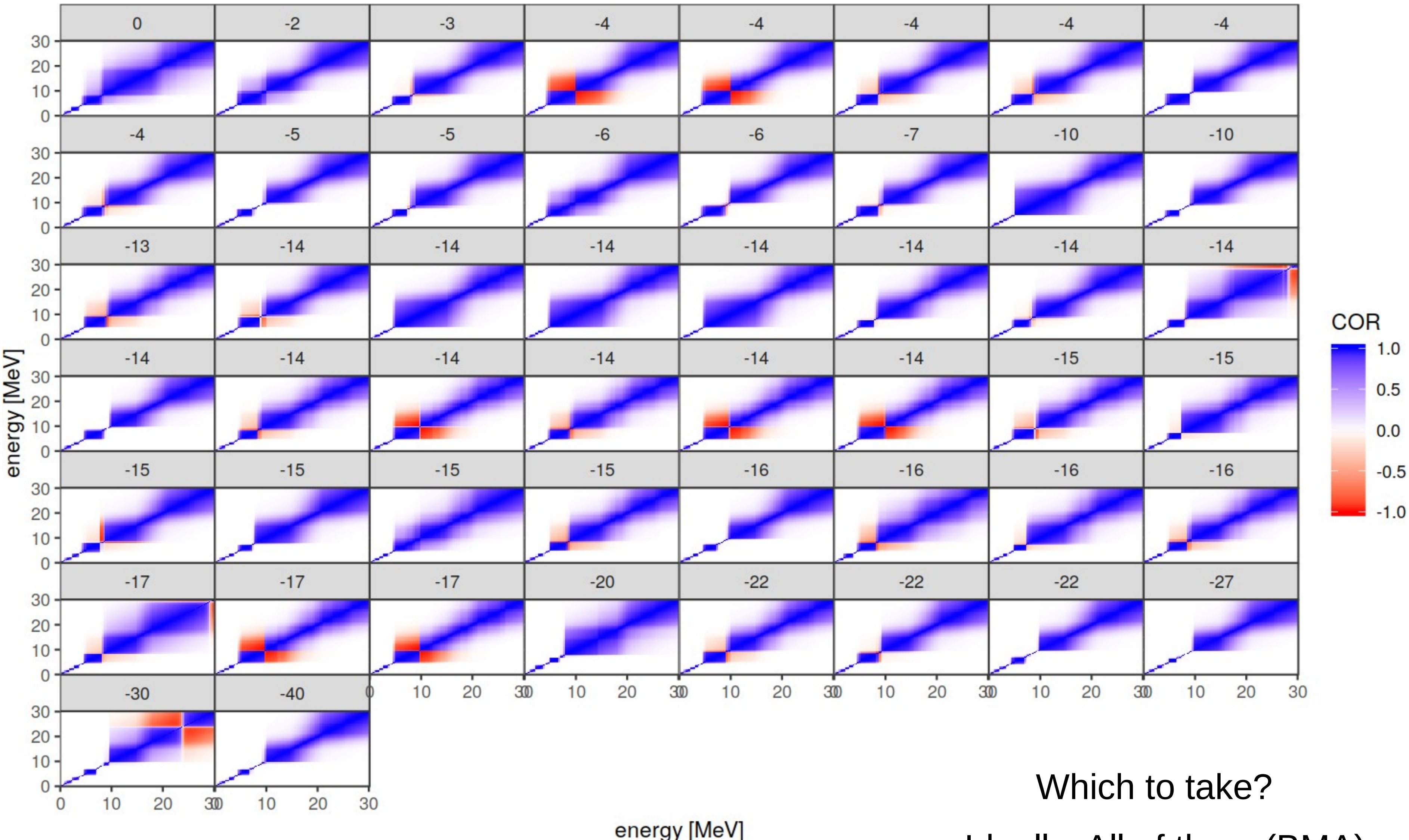


Remark
Optimization was guided by allowing more flexibility at lower than at higher energies.

Link to animation [GIF]

Link to animation [MP4]

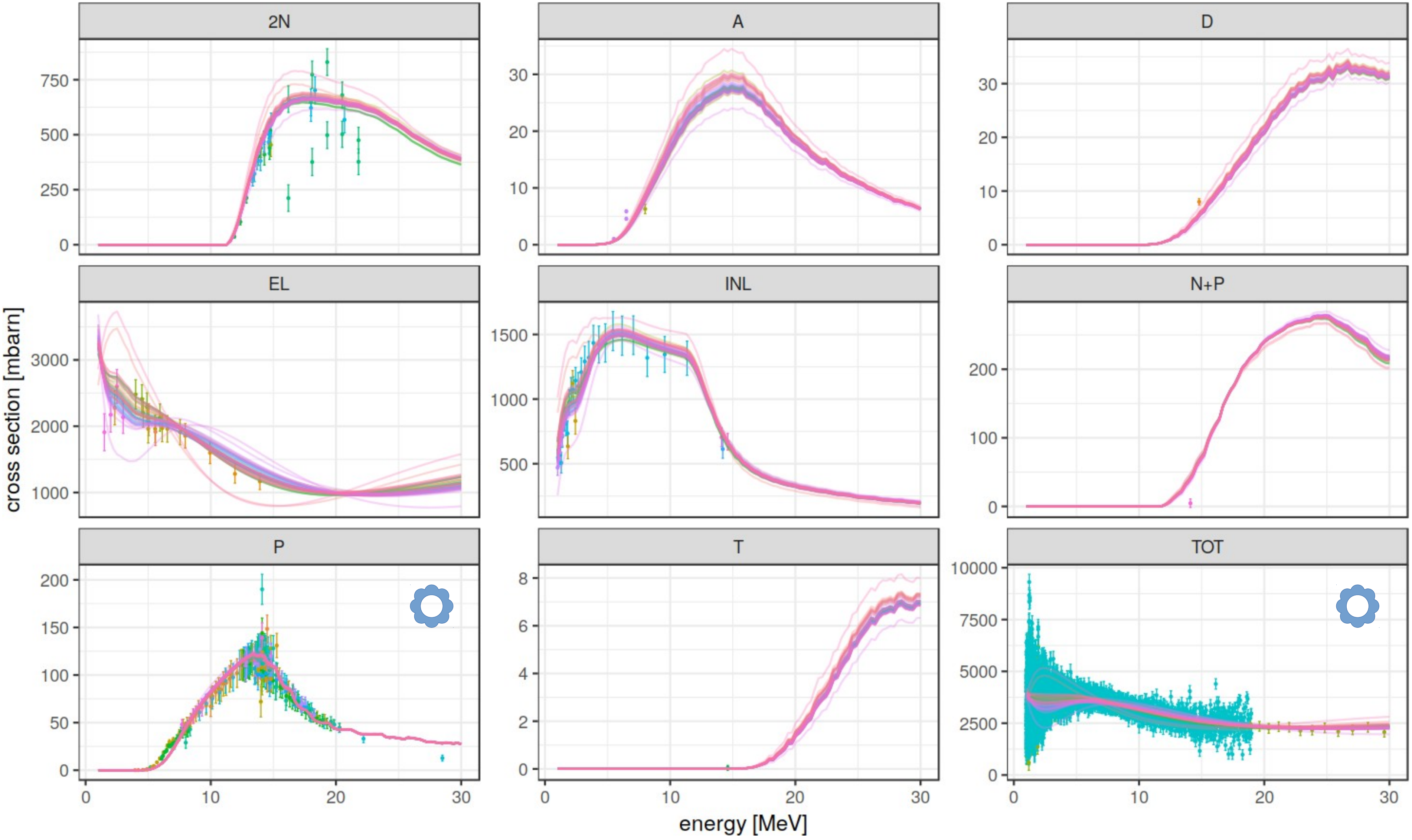
Correlation matrices of defect (n,tot)



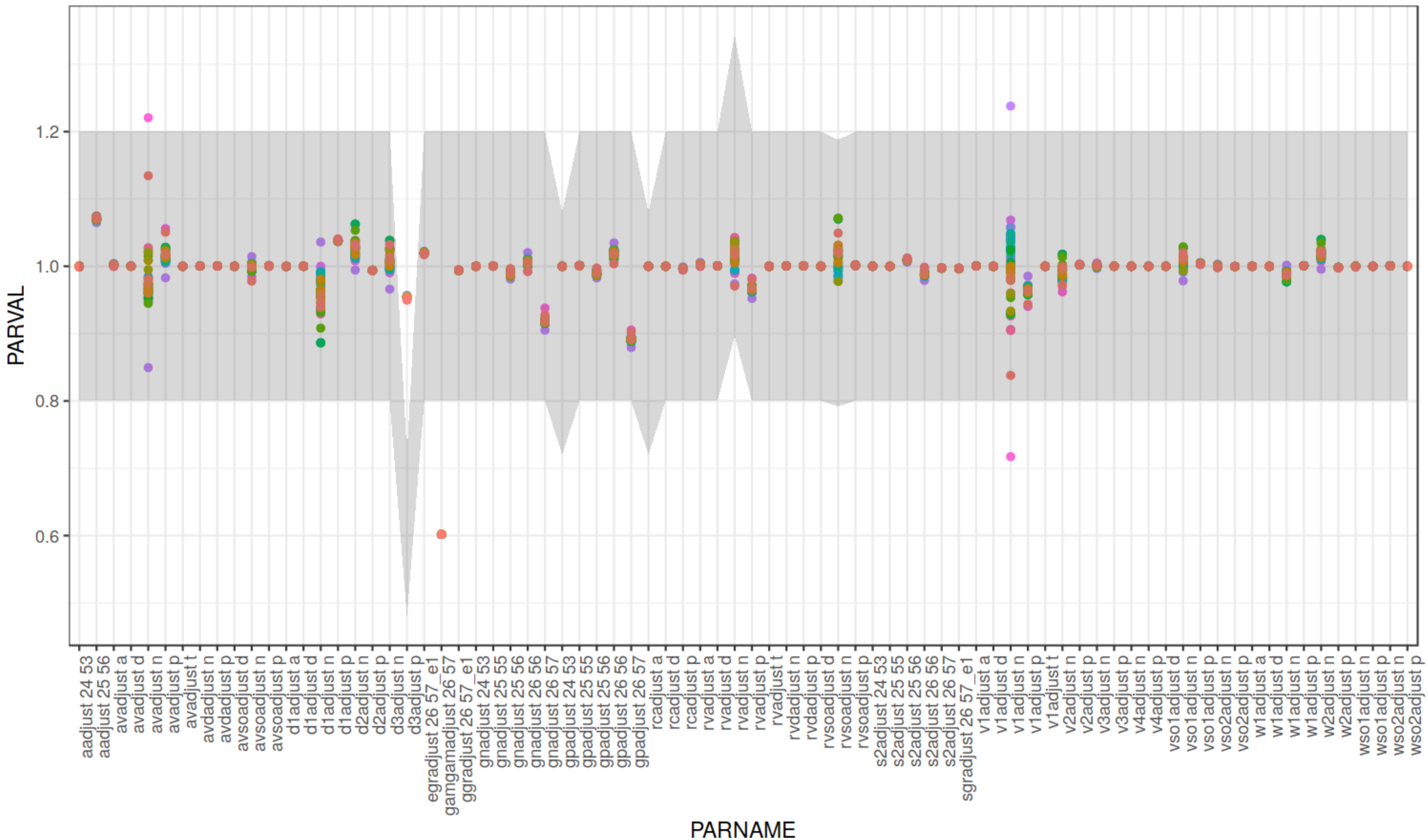
Which to take?

Ideally: All of them (BMA)

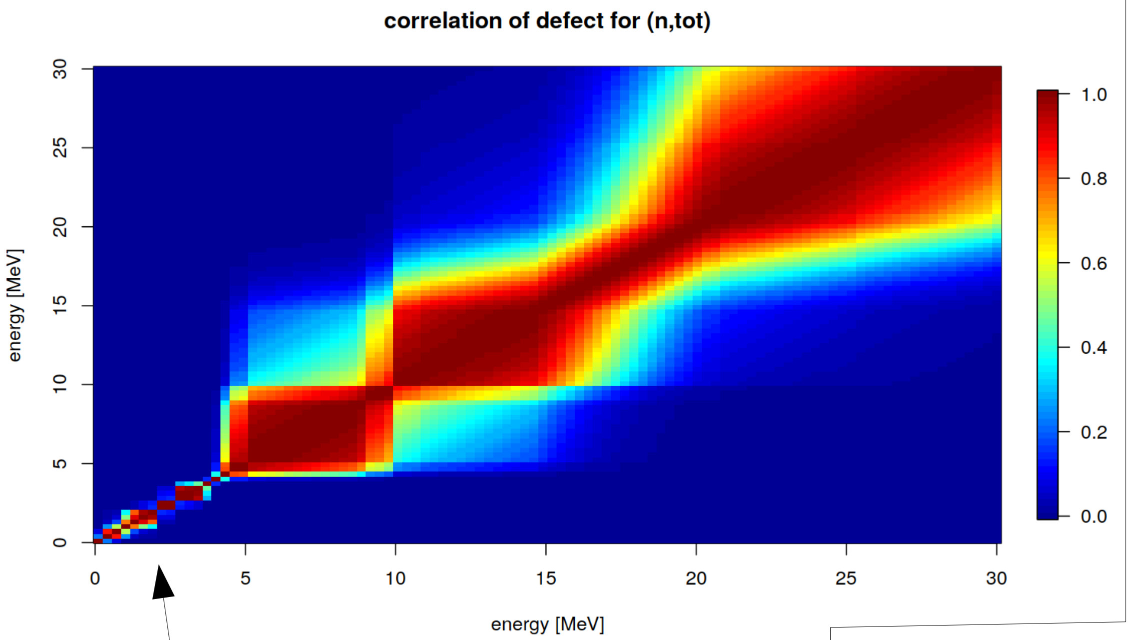
Poor man's BMA [again ^{56}Fe update (n,p) and (n,tot)]

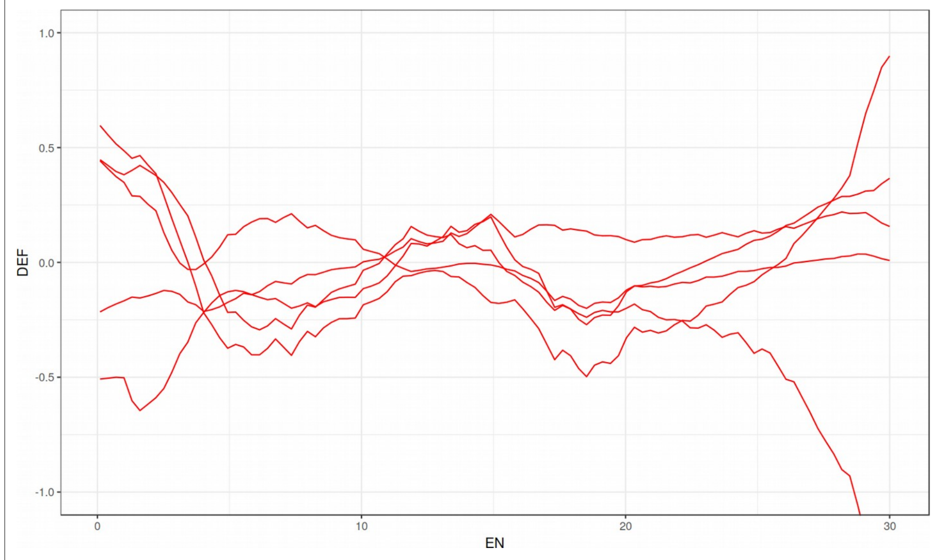


Poor man's BMA
[again ^{56}Fe update (n,p) and (n,tot)]

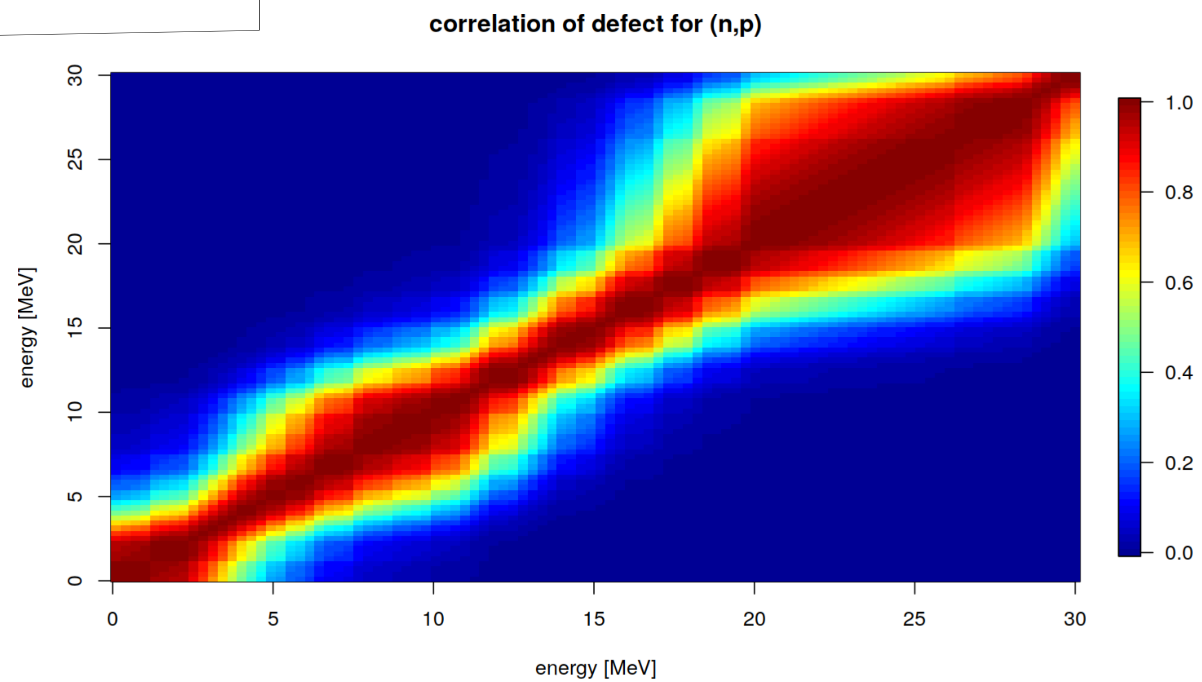


Correlation structure





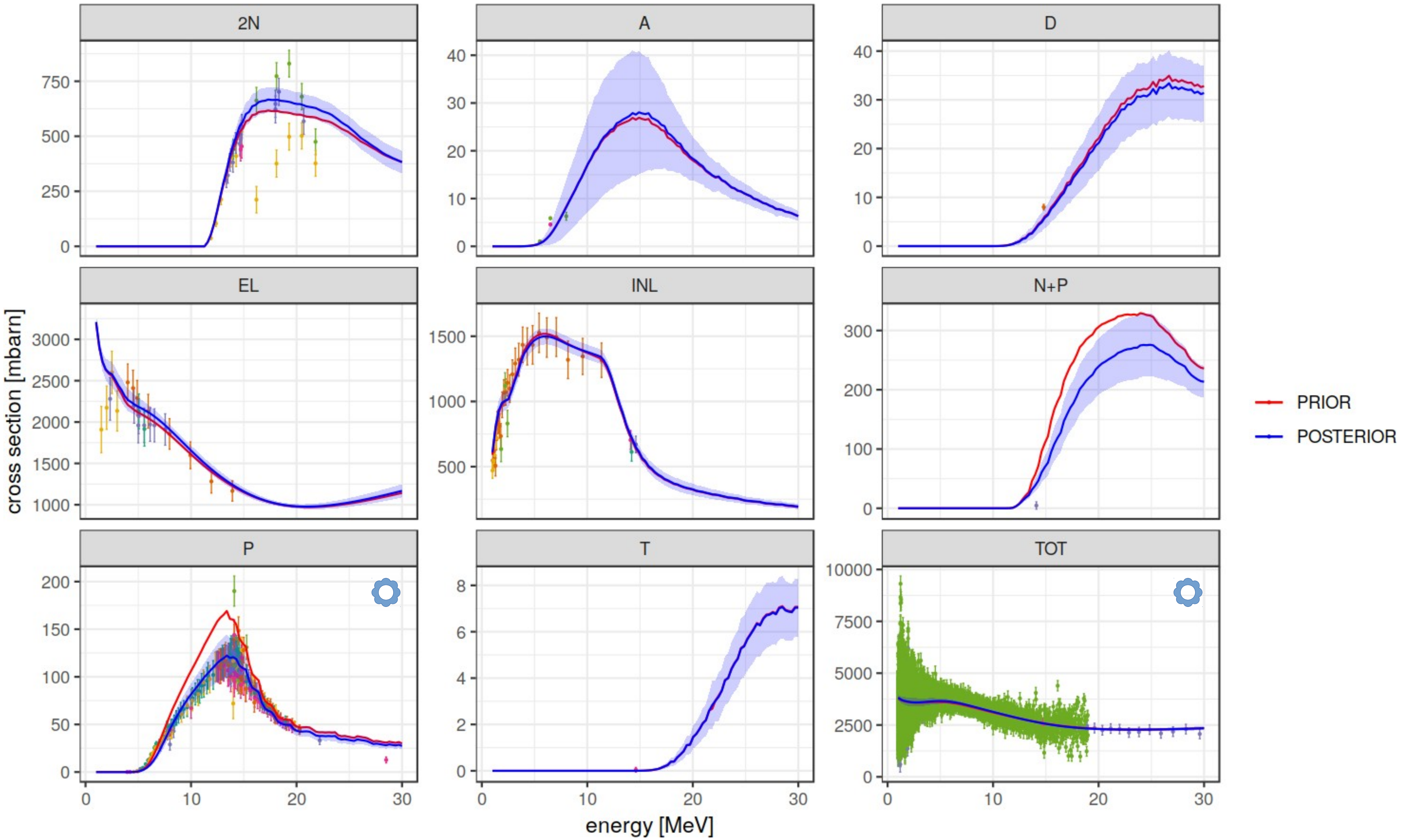
resonance region



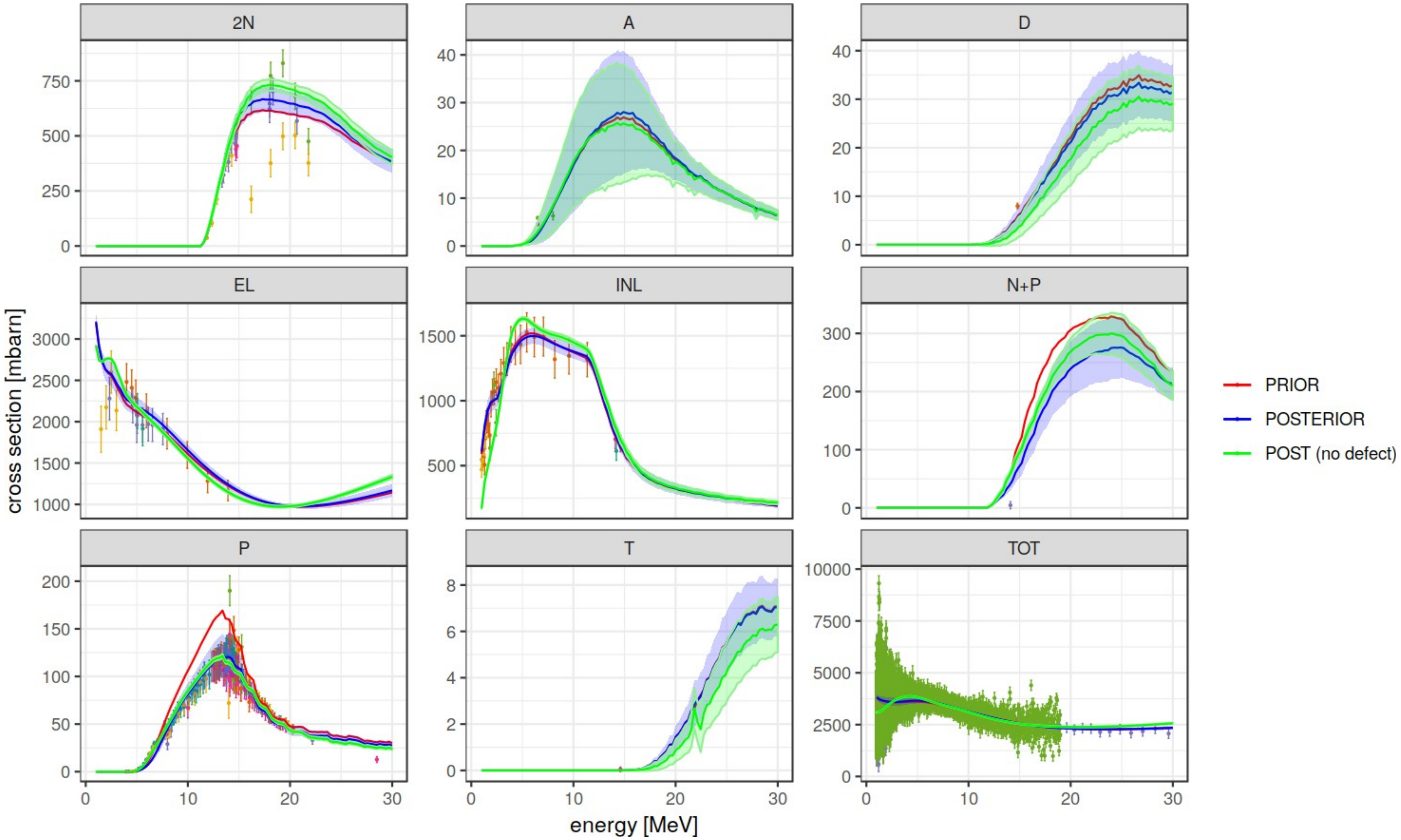
Link to animation [GIF]

Link to animation [MP4]

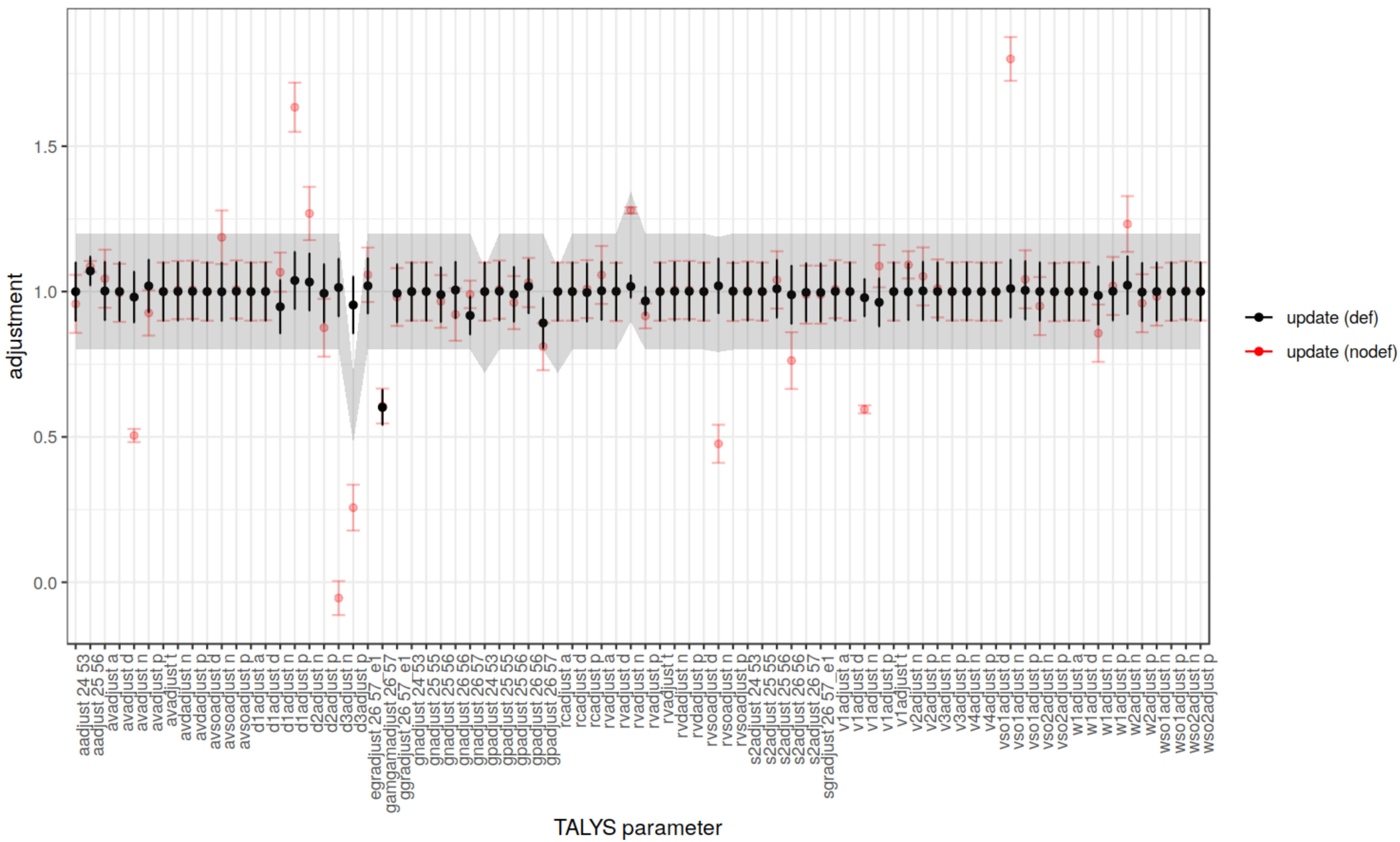
Updating including defect prior



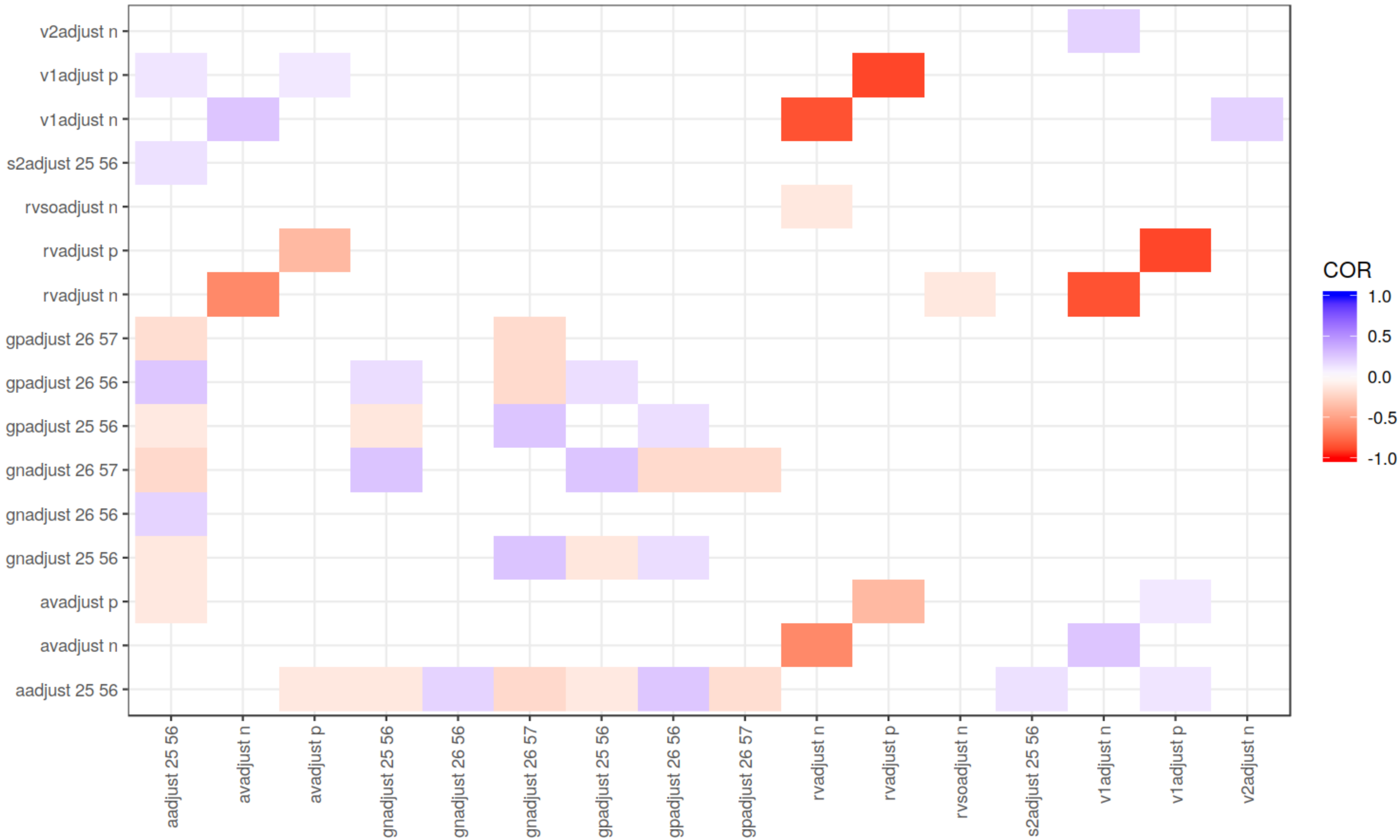
Comparison update (def/nodef)



(Most) Parameters abide by the law



Updated correlations (with def)



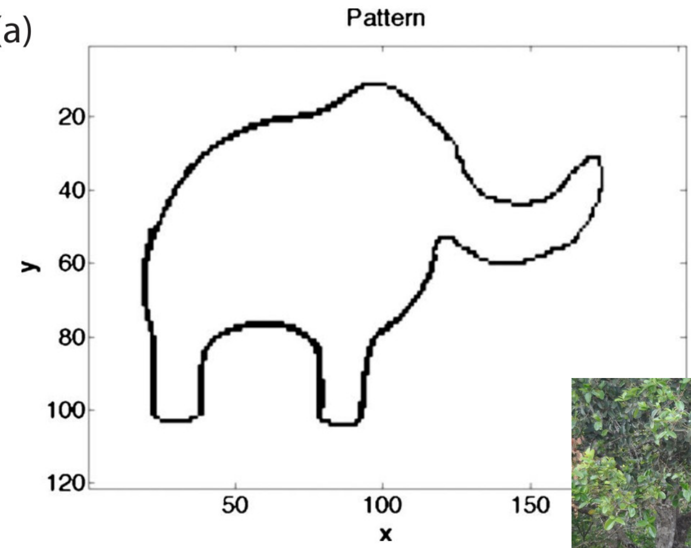
Summary

- Fitting without using model defects is problematic
- Data-driven approach to learn on a per-energy basis about the performance of the model using a “dynamic time warping GP”
- We can use this information within the Bayesian framework in a principled way to get more reasonable parameter adjustments, uncertainty estimates, etc.
- We can also equip existing methods with this extra information, e.g., GLS, UMC-B, UMC-G, BMC, BFMC, EMPIRE-Kalmann, …
- Of course, a lot of work still ahead (smoothness, BMA, etc.)
- One promising route: Combine it with the idea of energy-dependent TALYS parameters (*)

(*) Helgesson, P., Sjöstrand, H., 2018. Treating model defects by fitting smoothly varying model parameters: Energy dependence in nuclear data evaluation. Annals of Nuclear Energy 120, 35–47. https://doi.org/10.1016/j.anucene.2018.05.026

Thank you!

(a)



georg.schnabel@physics.uu.se