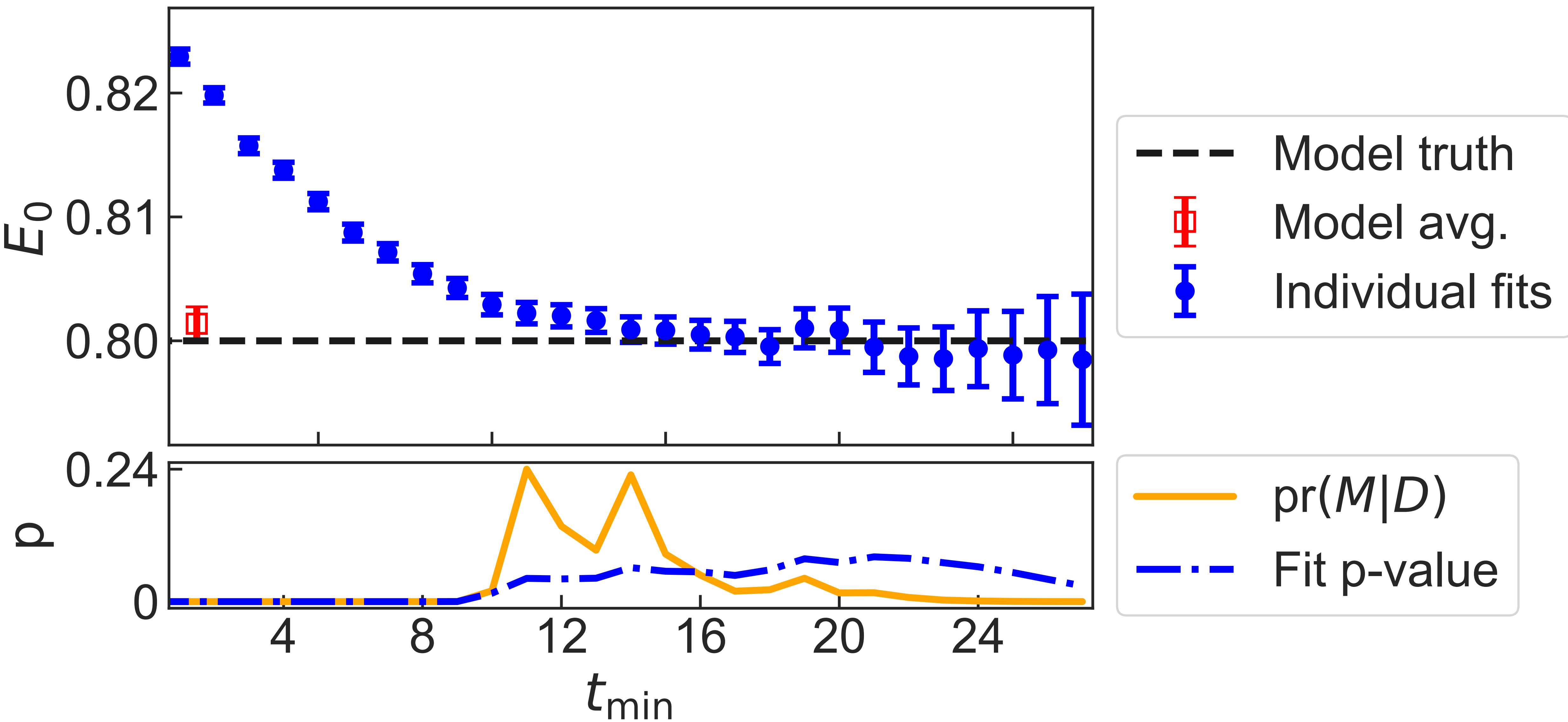
(with Will Jay, arXiv:2008.01069, Phys.Rev.D 103 (2021) 114502)
(also with Will Jay and Jake Sitison, in progress)



Bayesian Model Averaging for Lattice Field Theory

Ethan T. Neil (Colorado)

Lattice 2021 @ MIT/Zoom
07/27/21

Overview and Key Ideas

- Modeling is an important source of systematic error in lattice calculations. (Which model? How many terms do I keep?) Estimate error by trying variations and seeing how outputs fluctuate.

- Bayesian model averaging puts this on firm statistical ground. Obtain any expectation value as a weighted average:

$$\langle O \rangle = \sum_M \langle O \rangle_M \ \mathrm{pr}(M|D)$$

- Asymptotically correct model weights are given by the Akaike information criterion (AIC). Cutting data can also be treated within this framework, resulting in an extra 'penalty term' for removed points. Our main result is:

$$\mathrm{pr}(M|D) \approx \exp\left[-\frac{1}{2}\left(\chi^2_{\mathrm{aug}} + 2k + 2N_{cut}\right)\right]$$

Outline

- Sketch of the model weight derivation
- Data cuts (subset selection) as model selection
- Example results for two-point functions
- Looking forward: improved ICs

- Start with the general *model probability* or "*Bayes factor*":

$$\mathrm{pr}(M|D) = \int d\mathbf{a} \frac{\mathrm{pr}(D|\mathbf{a}, M)\mathrm{pr}(\mathbf{a}|M)\mathrm{pr}(M)}{\mathrm{pr}(D)}$$

model prior (usually: ignore)

likelihood function ~ $e^{-\chi^2}$

prior probability

- Expanding χ^2 about the best-fit parameters a* reduces this to a Gaussian integral, which we can do! (Discard sub-leading terms that can distort results with improper/wide priors; see paper for details on that.)

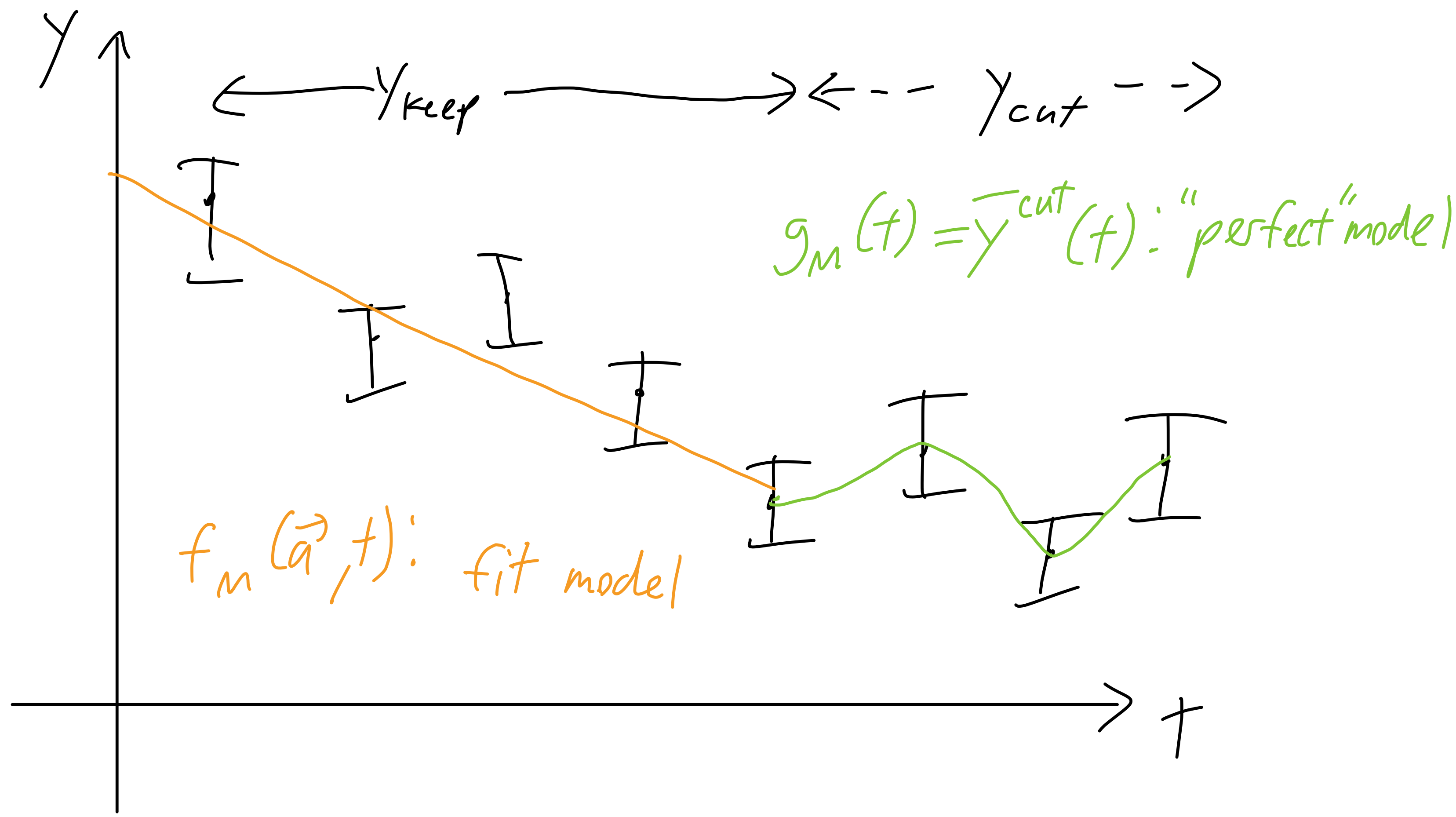
- Asymptotic bias: sample likelihood tends to overestimate true likelihood, effect *does not vanish* as number of samples n—>∞. Add a bias correction term which depends on the number of parameters, k:

$$\mathrm{pr}(M|D)_{BC} = \exp\left(-\mathrm{tr}[J^{-1}I]\right)\mathrm{pr}(M|D) \approx e^{-k}\mathrm{pr}(M|D)$$

- Thus, we find (at large sample size, assuming our model is specified correctly):

$$-2\log \mathrm{pr}(M|D)_{BC} \xrightarrow{N\to\infty} \chi^2 + 2k = AIC_M$$

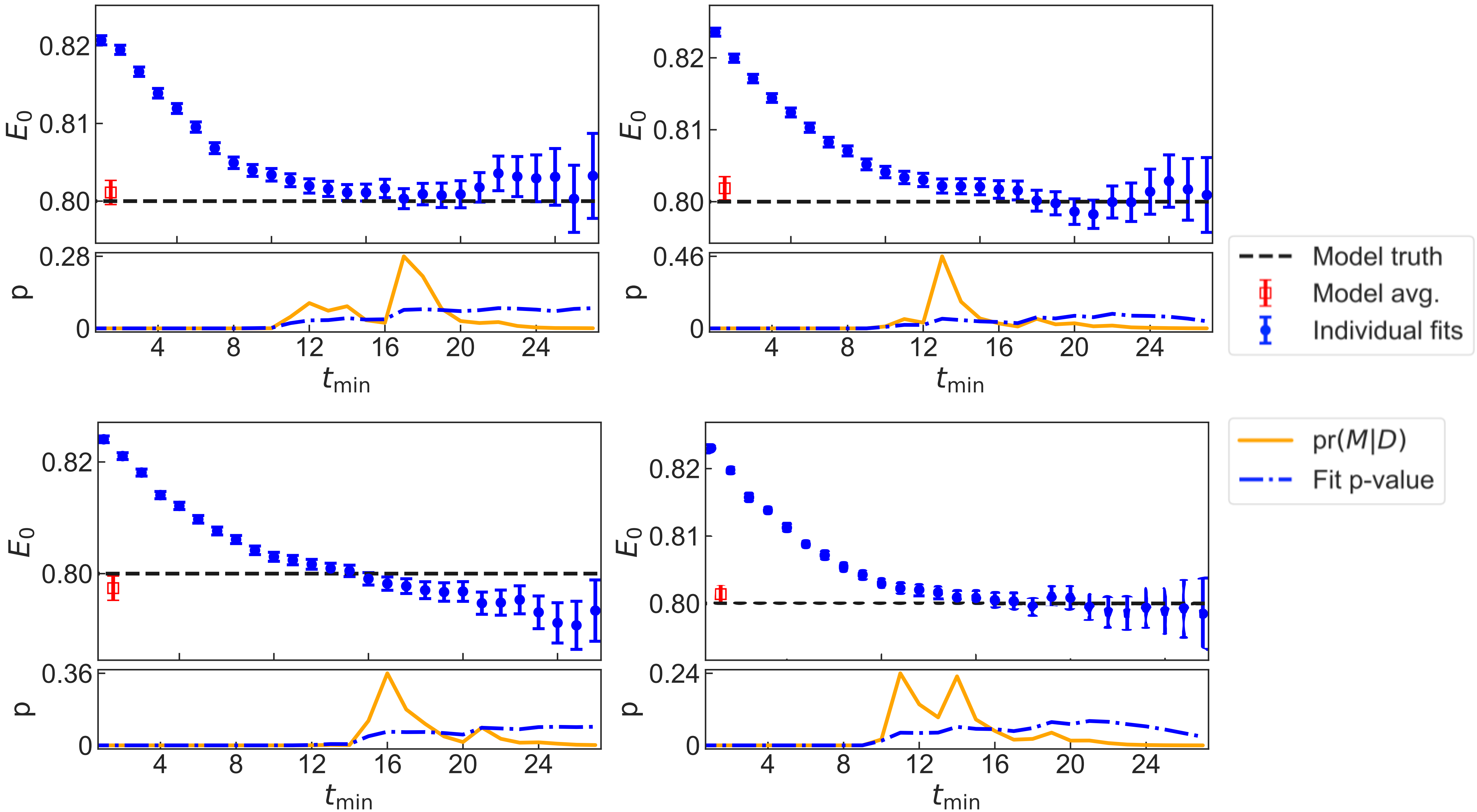
Data subset selection



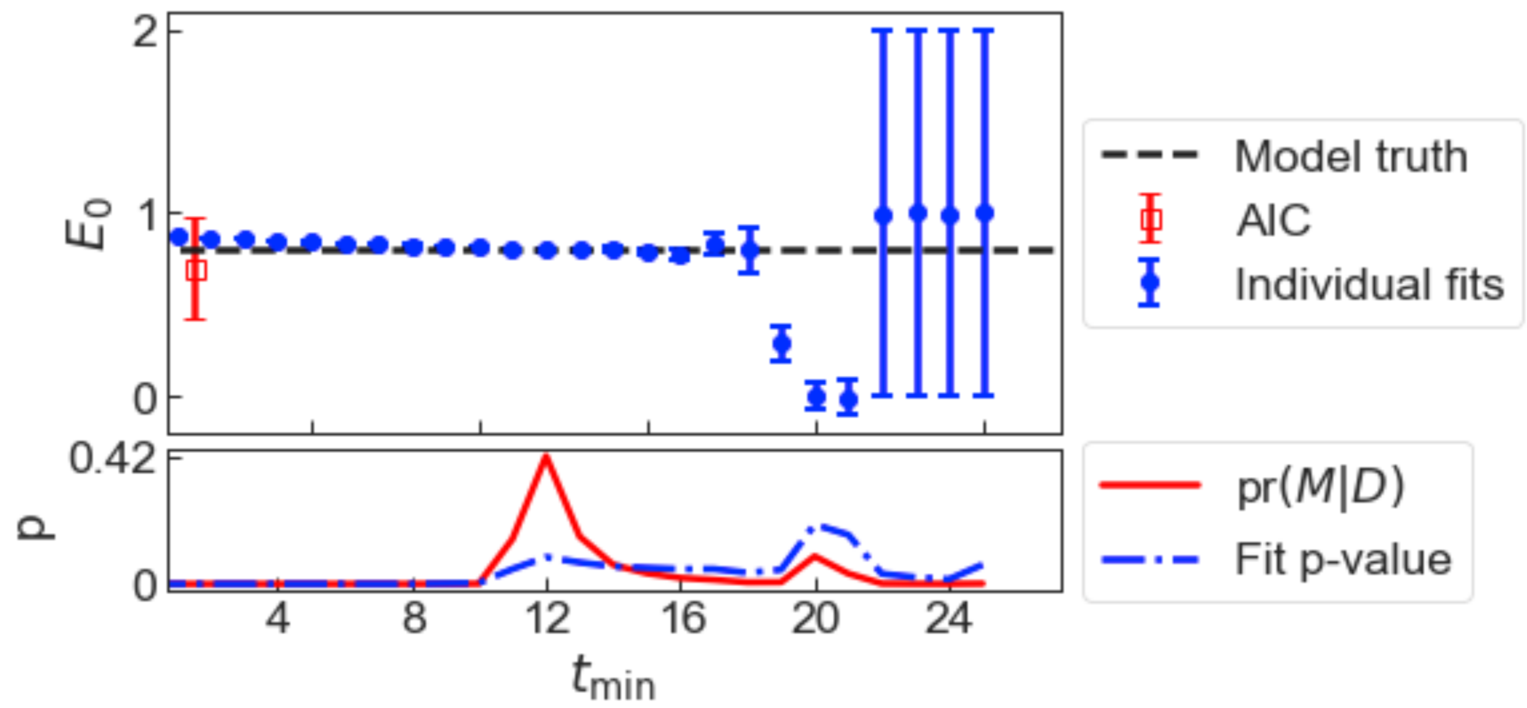
- **Idea:** Subdivide data into fit region (y_{keep}) and cut region (y_{cut}). "Perfect model" (e.g. high-order polynomial) is used in the cut region, so $\chi_{cut}^2 = 0$ identically. (*Don't need to actually do this in practice!*)
- Cut region still contributes a bias-correction term to AIC (the individual samples still fluctuate around the mean):

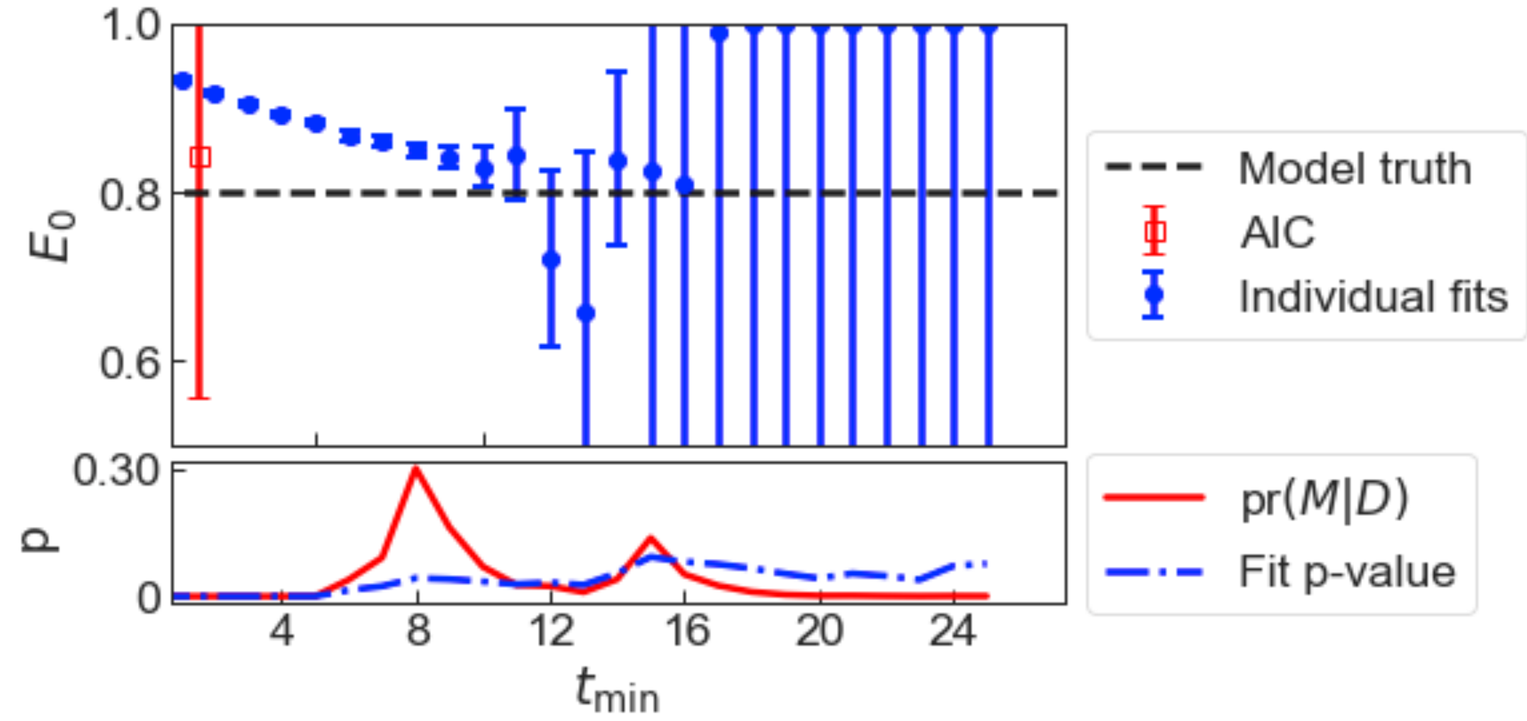
$$AIC_{M,N_{\text{cut}}} = \chi^2 + 2k + 2N_{\text{cut}}$$

Example: fitting two-state exponential mock data to a one-state exponential.

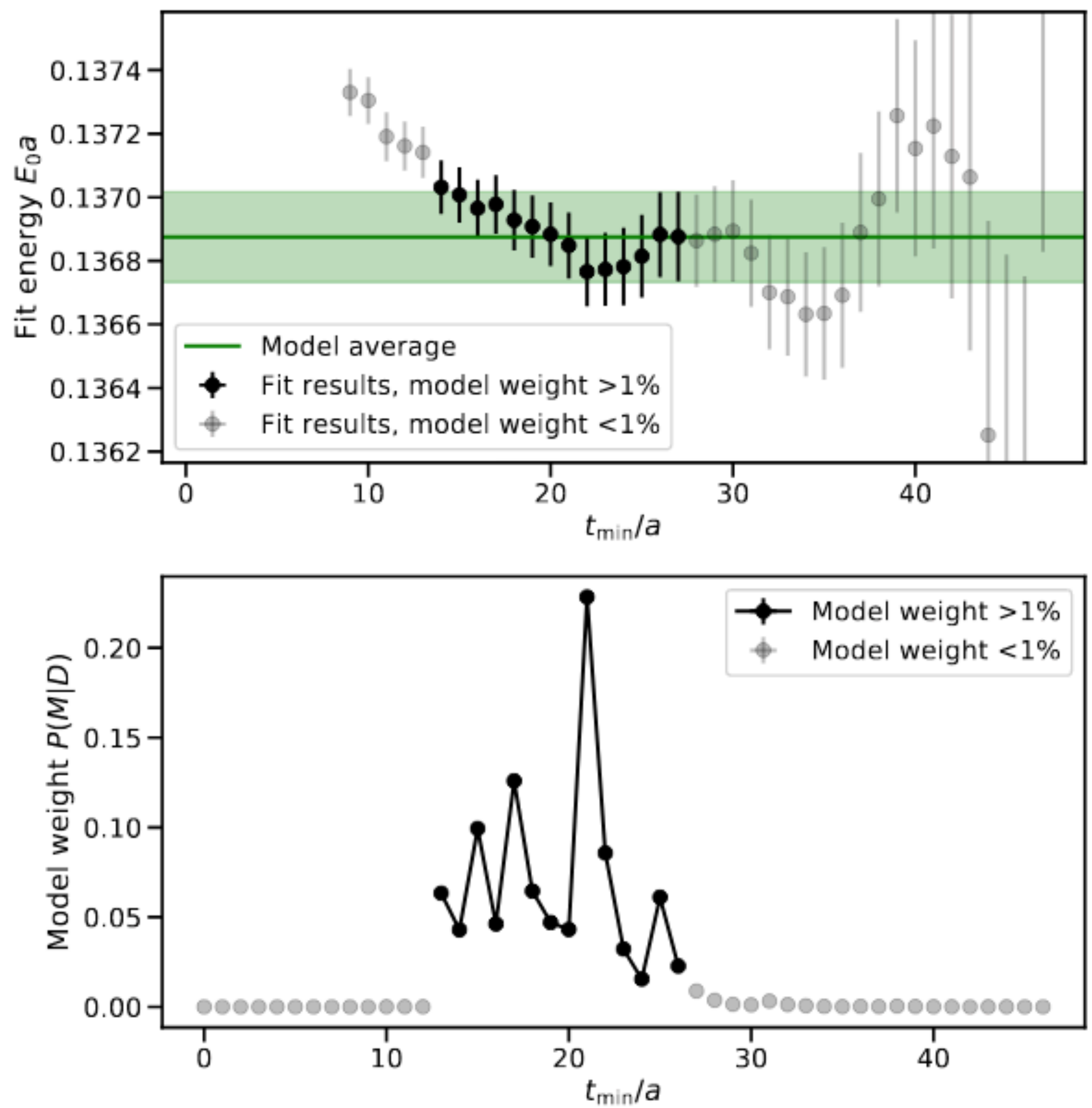


(public code repo at https://github.com/etneil/model_average_paper/)

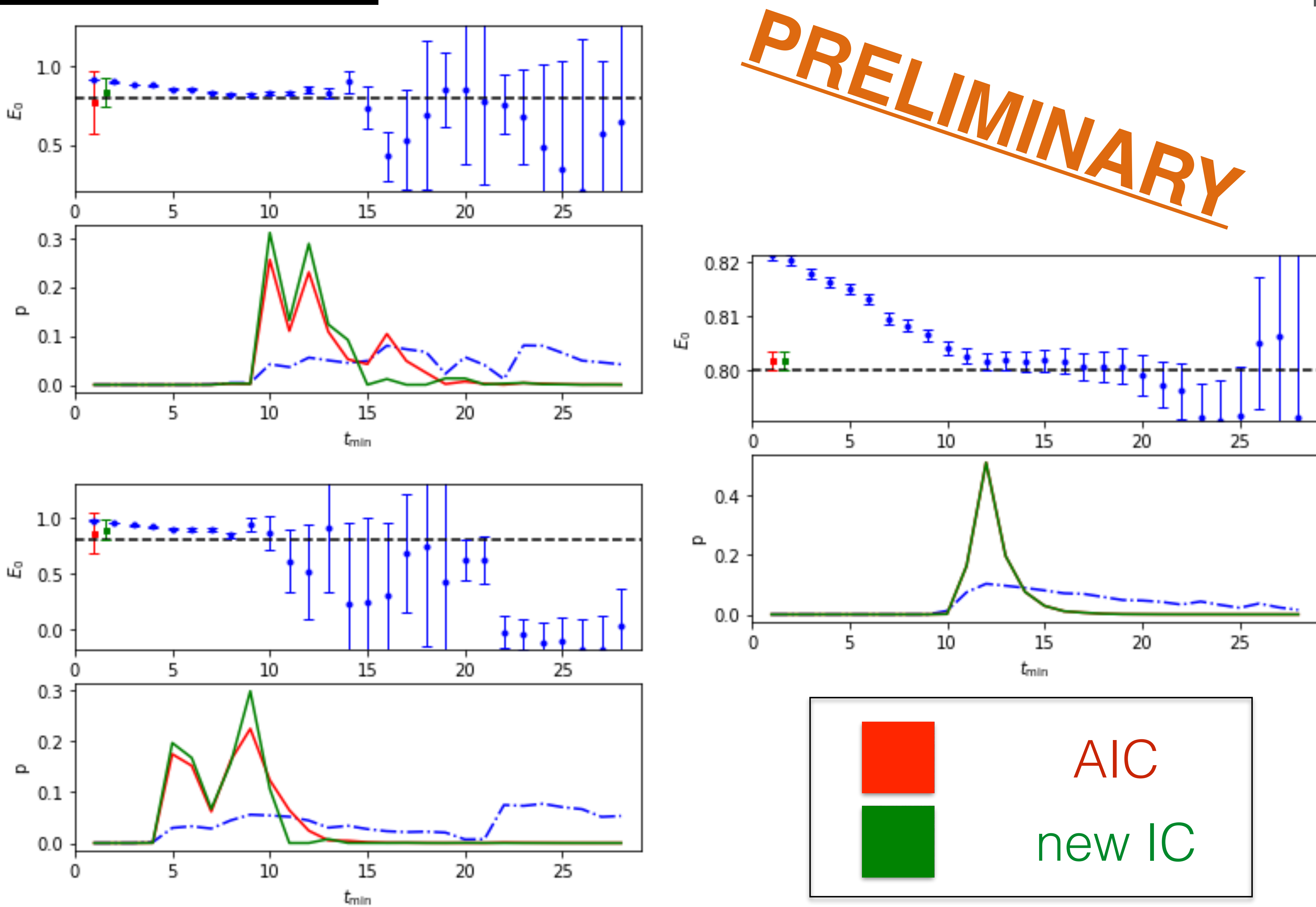




- Cautionary tales: model averaging does not help if you average in bad fits (top) or over extremely noisy data (bottom)! (It is robust against *some* noise - see previous slide.)



- Real lattice QCD correlator example - pion two-point function (courtesy of Fermilab/MILC collaboration.) Model average result for pion E_0 agrees within 1σ with original "by hand" analysis.



- Experimenting with alternative ICs - all are asymptotically equivalent to AIC with enough data, but some look very promising at smaller sample size!

Conclusions

- Model averaging is a statistical approach to handling model choices, without being overly conservative (but, beware of averaging in pure noise…)

- Readily applied to common problems in lattice calculations (how many states? which t_{min}/t_{max} to use? what mass range should I fit to?). Works on expectation values, so compatible with your favorite statistical methods (e.g. bootstrap/jackknife)

- Easy to compute within existing least-squares fitting analysis! AIC with modification for cutting data:

$$\mathrm{pr}(M|D) \approx \exp\left[-\frac{1}{2}\left(\chi^2_{\mathrm{aug}} + 2k + 2N_{cut}\right)\right]$$

- Stay tuned for work on improved estimators!