

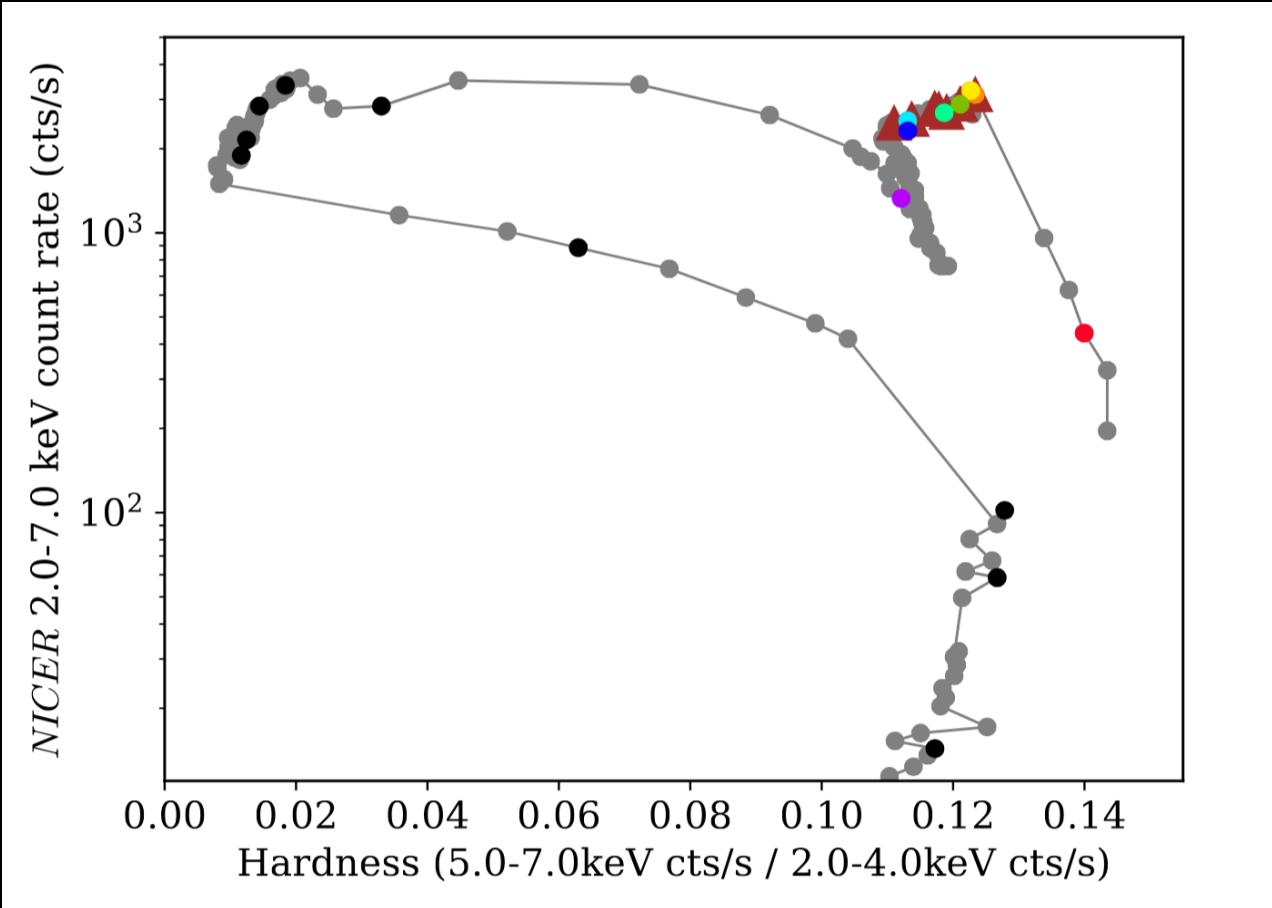
Bayesian sampling best practices in X-ray spectral modelling

Matteo Lucchini

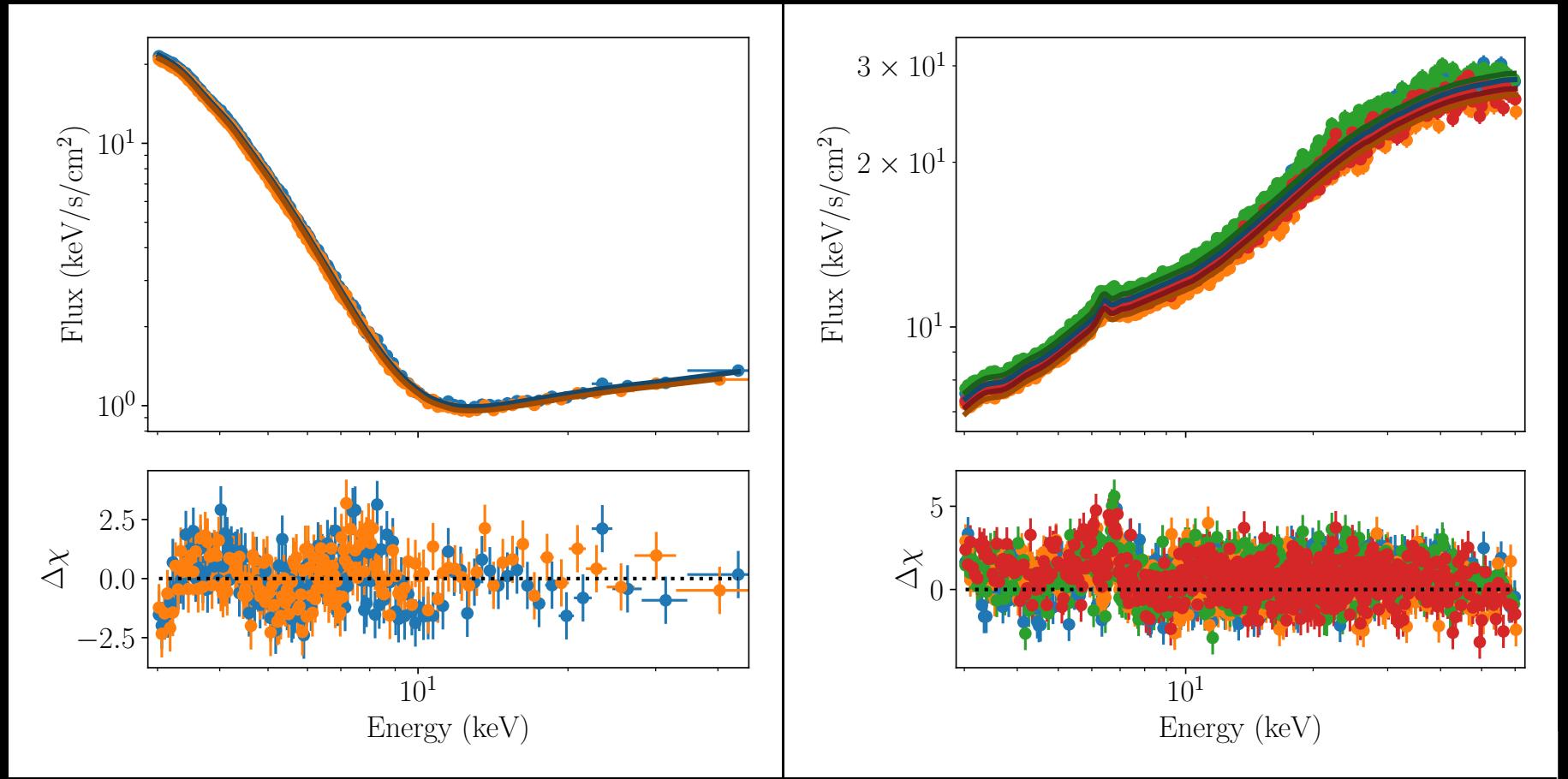
The plots that haunt my dreams

Matteo Lucchini

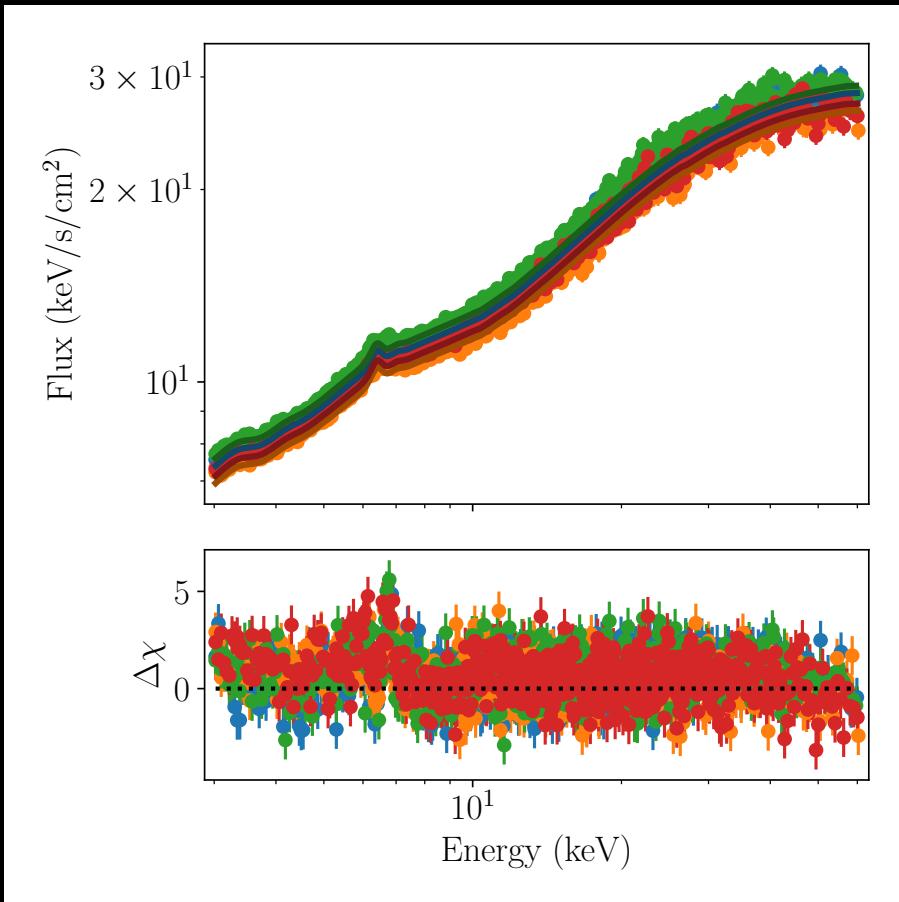
A standard case



Found a decent fit...now what?



Found a decent fit...now what?



How well constrained are my parameters?
Can I measure spin here?

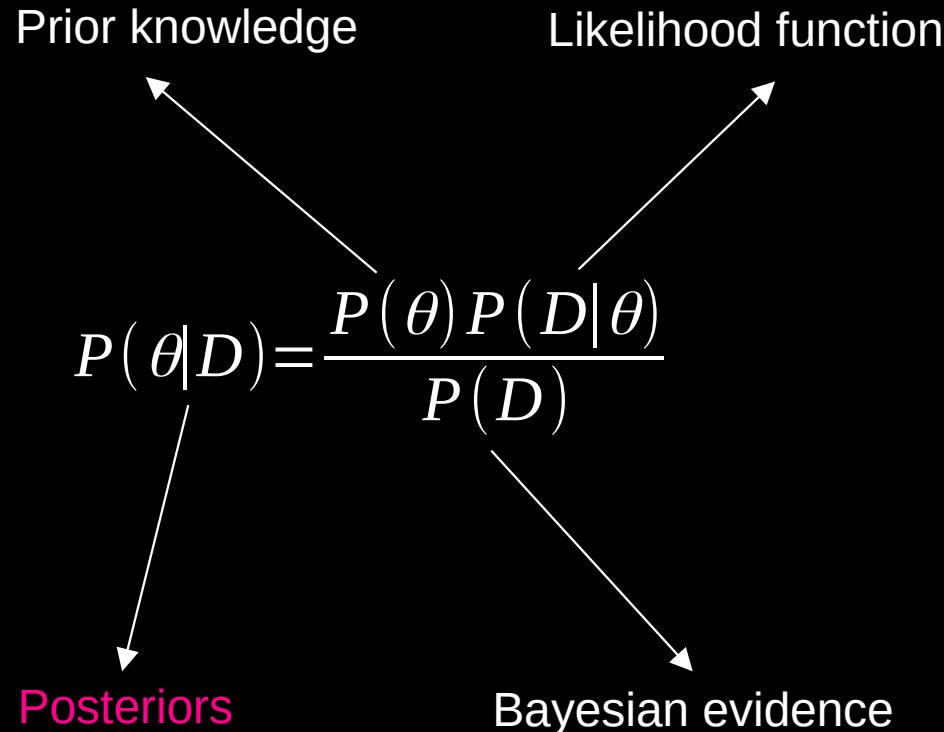


“When all I have is a hammer every problem looks like a nail”



Time for MCMC!

A recap of Bayesian inference



A recap of Bayesian inference

Prior knowledge

Likelihood function

$$P(\theta|D) = \frac{P(\theta) P(D|\theta)}{P(D)}$$

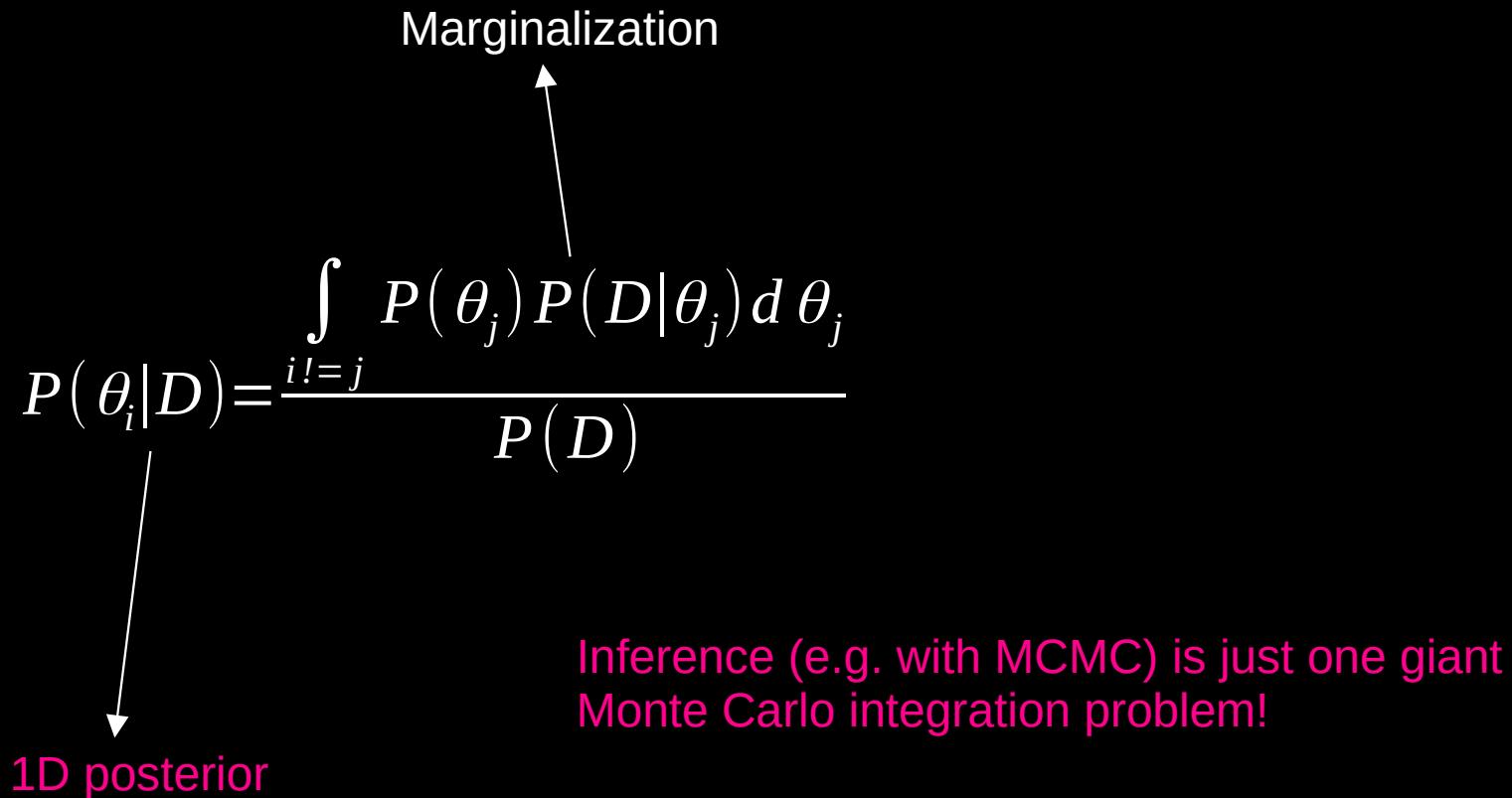
Posteriors

Bayesian evidence

We need to find clever ways to find the ``right'' θ , then calculate $P(\theta)$ and $P(D|\theta)$

$P(D)$ is for model comparison (and a bit difficult to calculate)

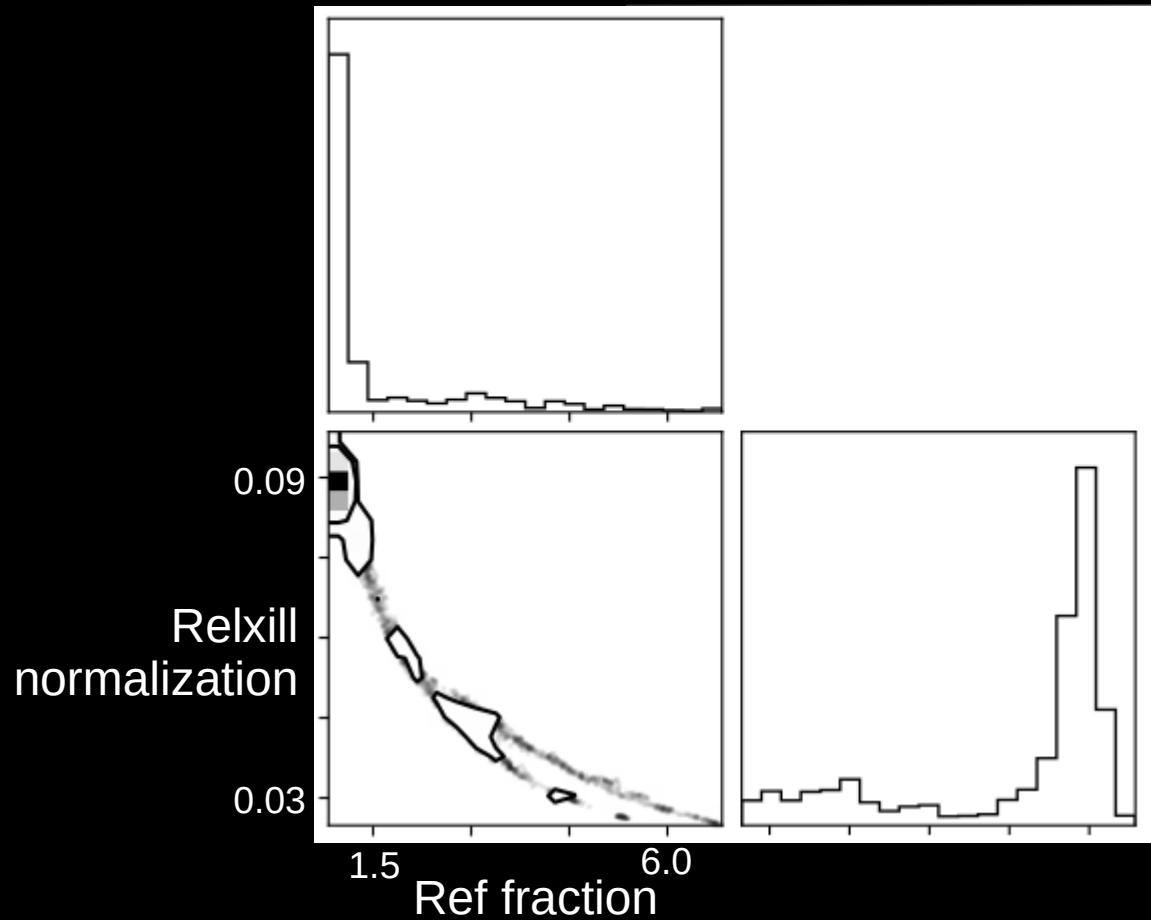
A recap of Bayesian inference



**SEVERAL
DAYS
LATER**



Found a decent fit...now what?



Did it even converge???

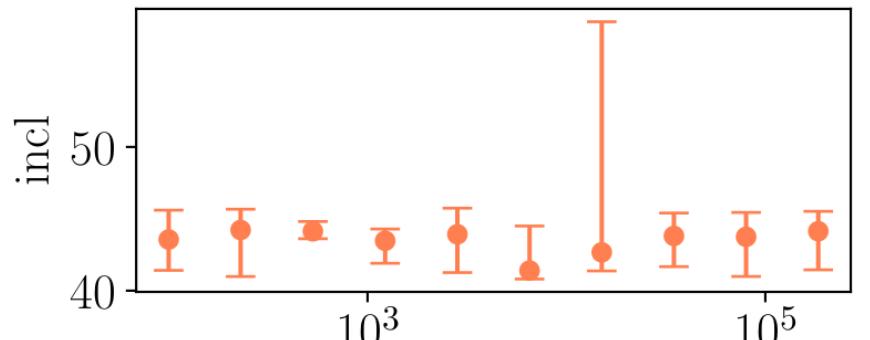
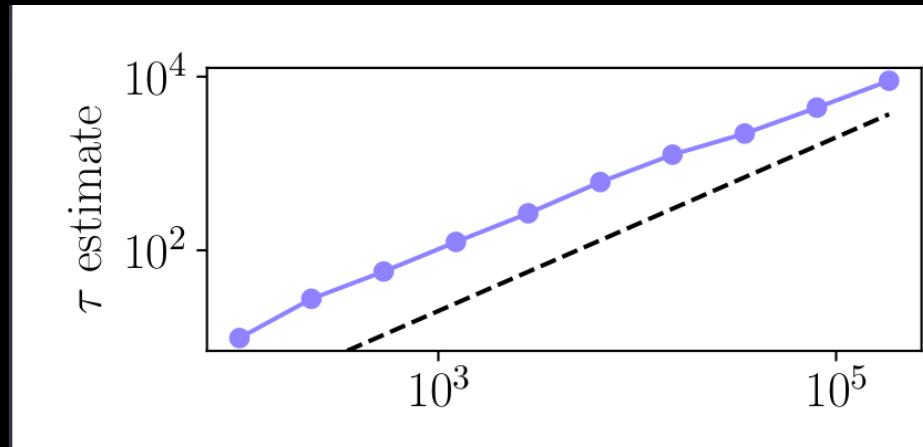
From <https://emcee.readthedocs.io/en/stable/tutorials/autocorr/>:

“With emcee, we follow Goodman & Weare (2010) and recommend using the integrated autocorrelation time τ to quantify the effects of sampling error on your results. The basic idea is that the samples in your chain are not independent and you must estimate the effective number of independent samples”

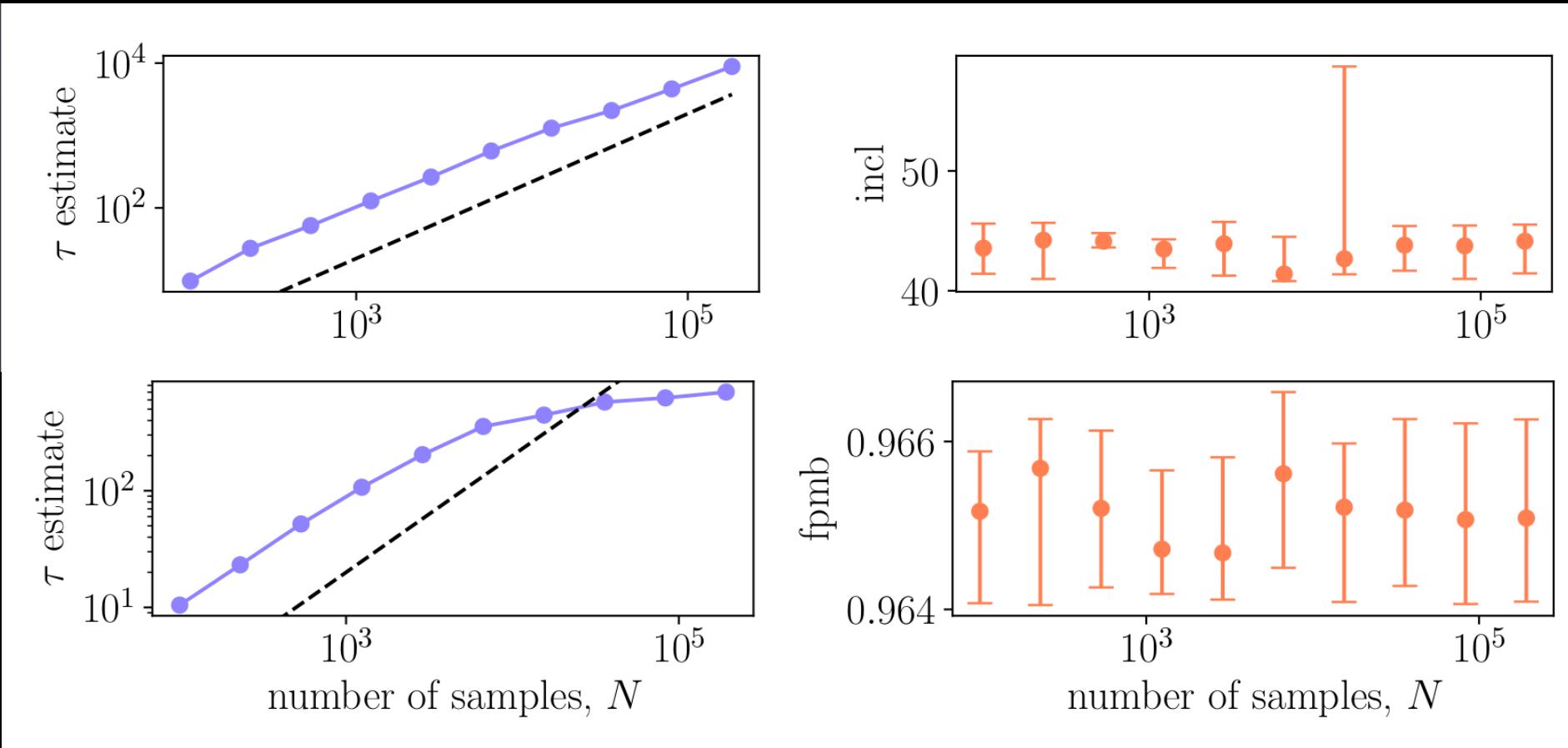
“With emcee we can use the parallel chains to reduce the variance and we've found that chains longer than about 50τ are often sufficient.”

Convergence is not burn-in, it takes much much longer!
BURN IN IS $\sim \tau$, NOT $\sim 50\tau$

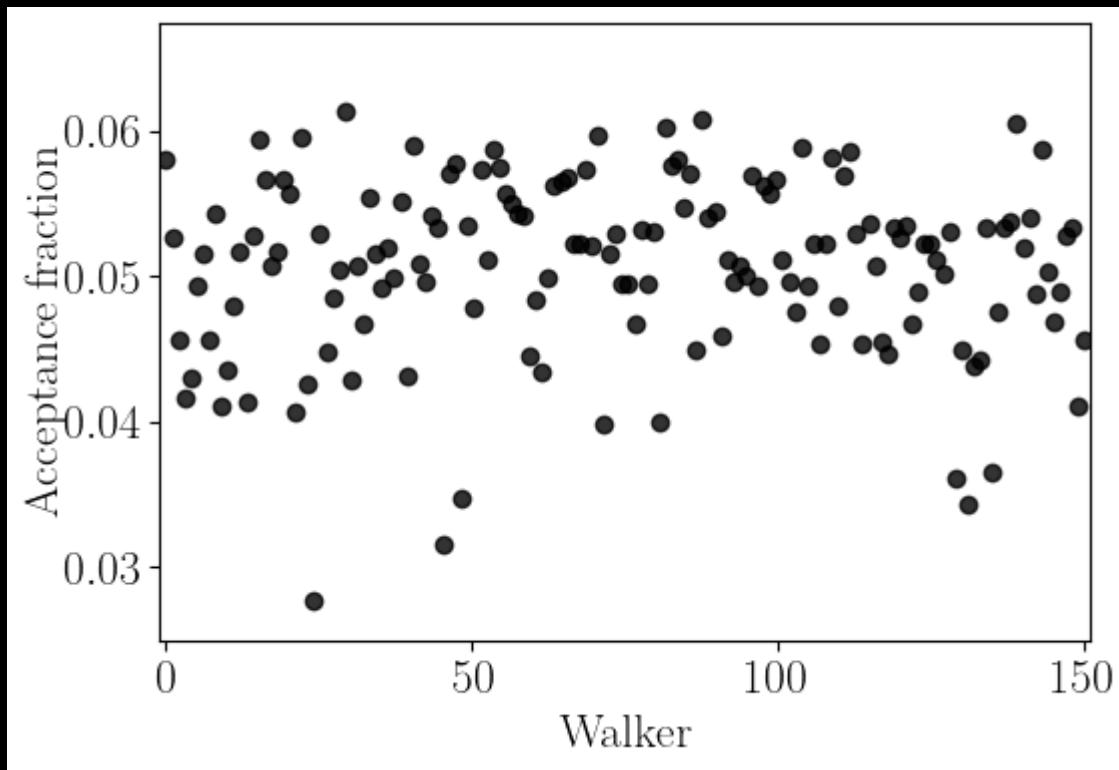
Did it even converge???



Did it even converge???

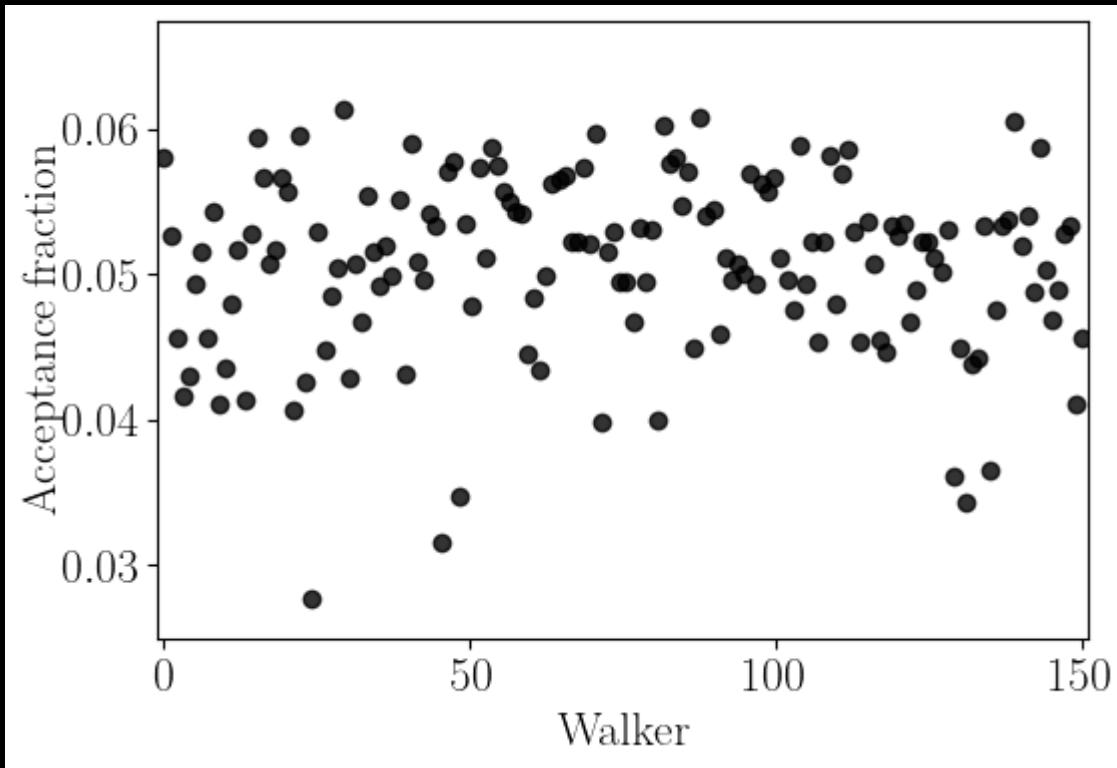


Why is it taking so long?



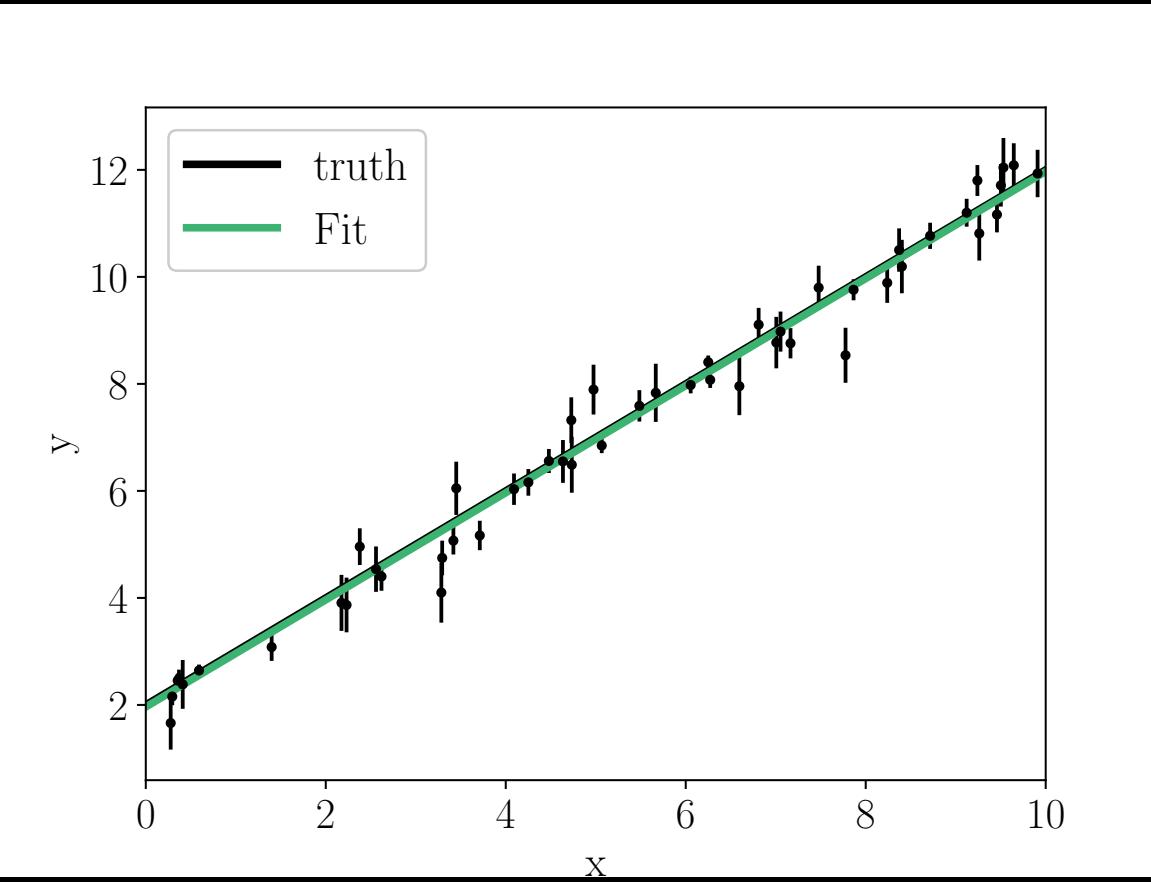
“If the acceptance fraction is getting very low, something is going very wrong.”

Why is it taking so long?

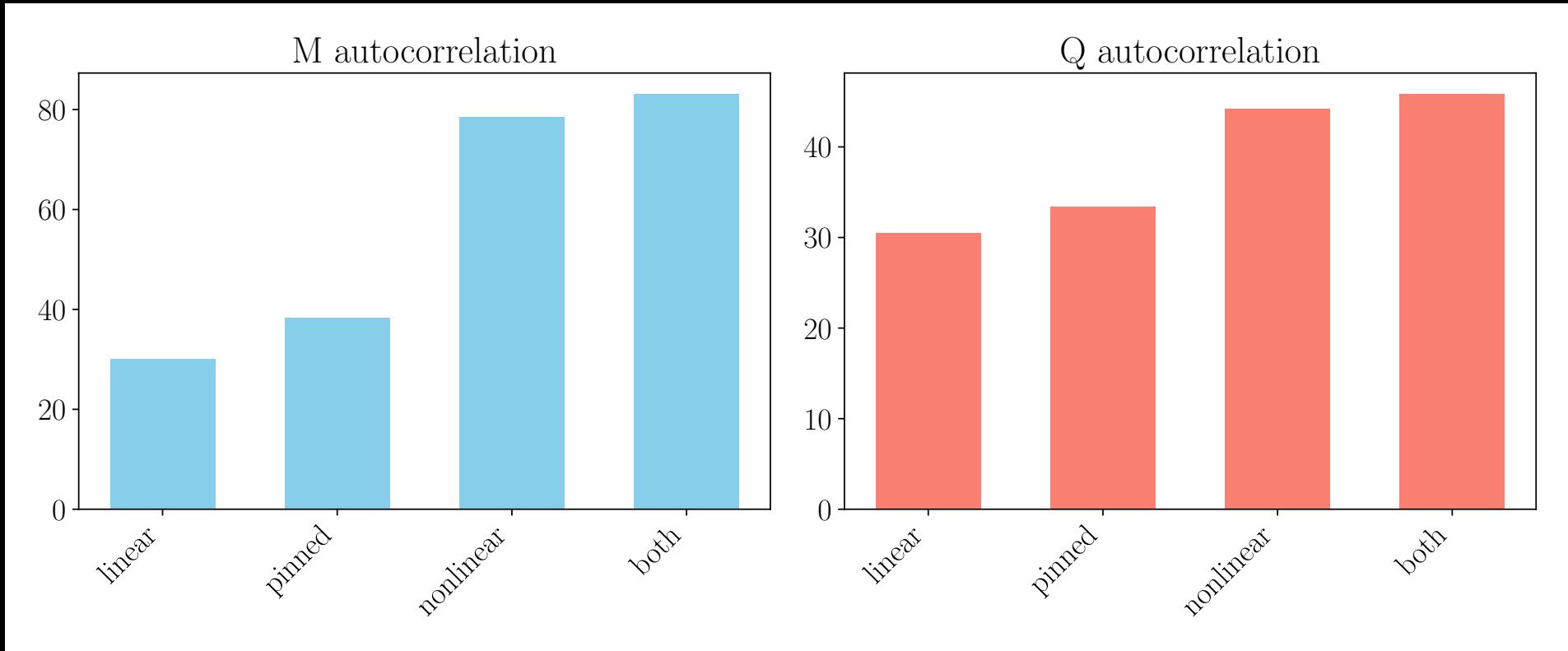


“Typically a low acceptance fraction means that the posterior probability is multi-modal, with the modes separated by wide, low probability “valleys.””

A quick experiment

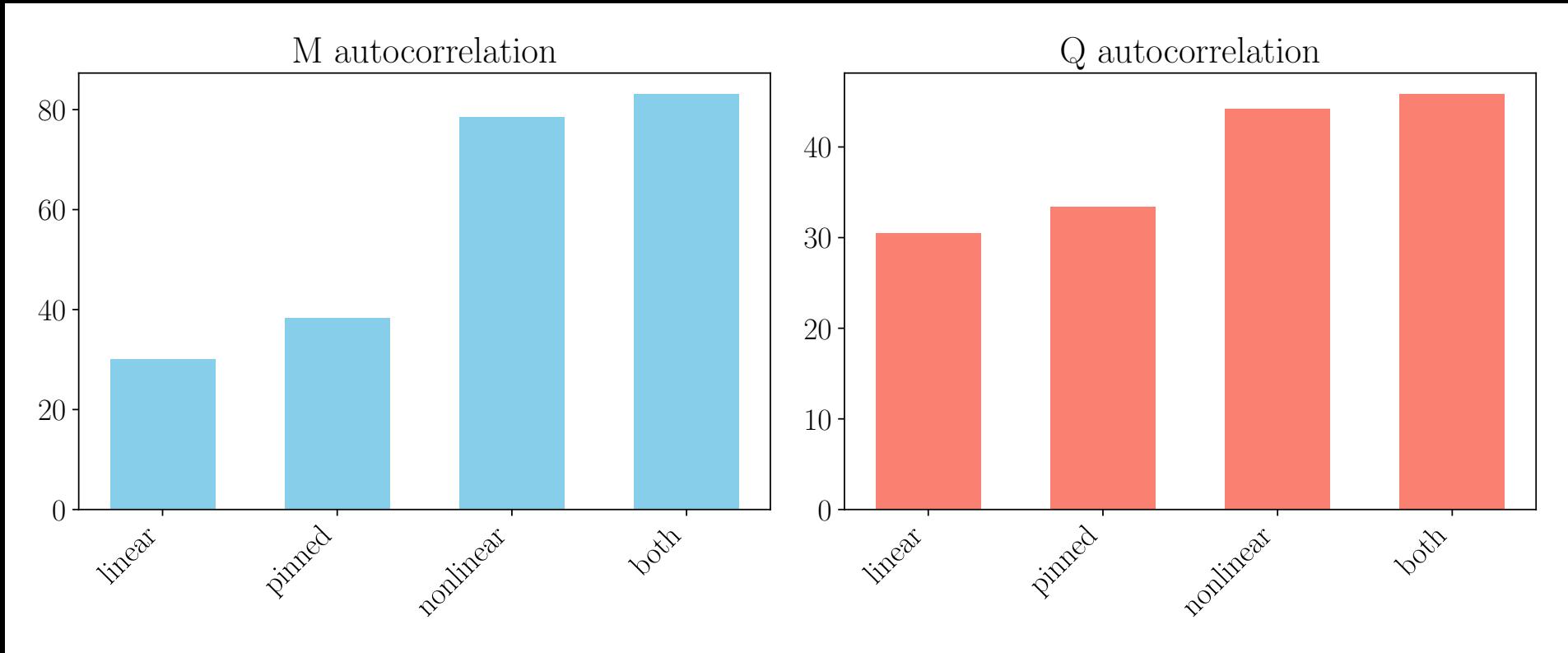


A quick experiment



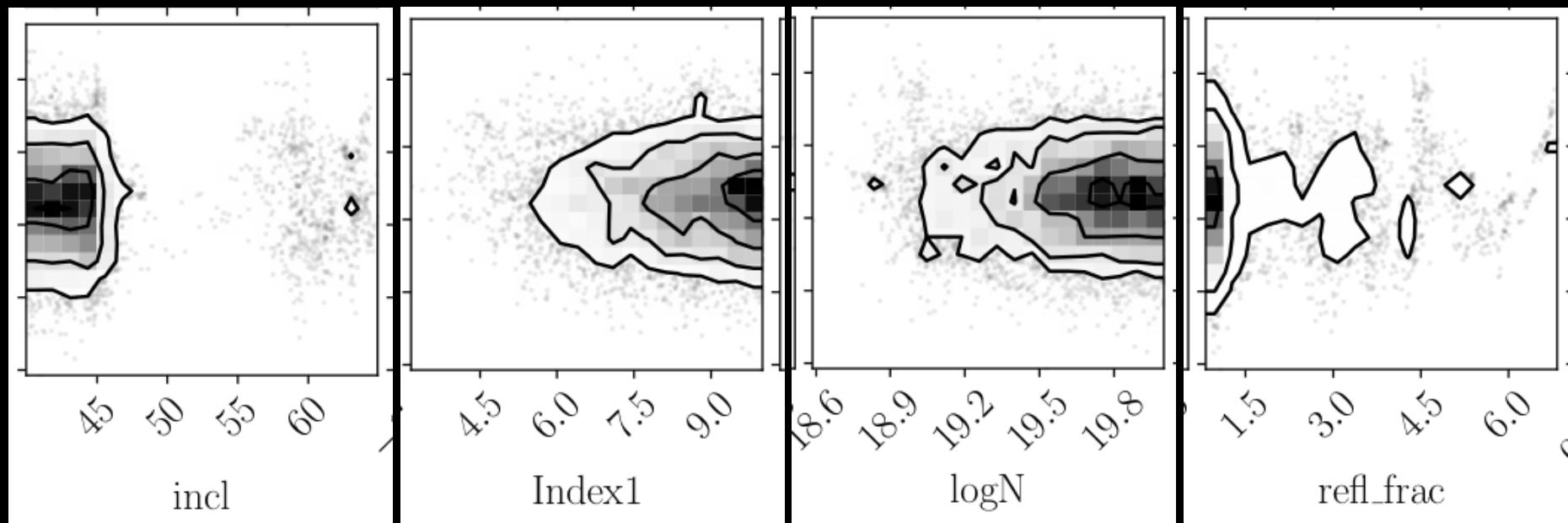
If a parameter is “badly behaved” the chain takes much longer

A quick experiment



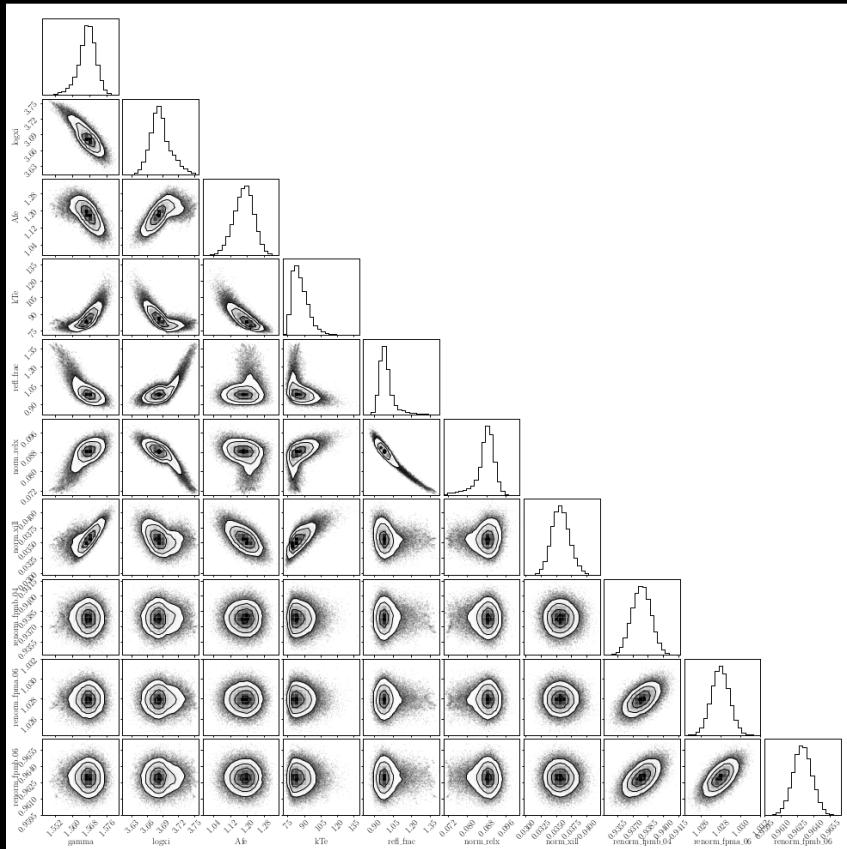
On a side note – these chains took 5000-10000 steps to converge for a straight line!

Back to my chain – fix the culprits



What happens to the rest of the chain?

Back to my chain – fix the culprits

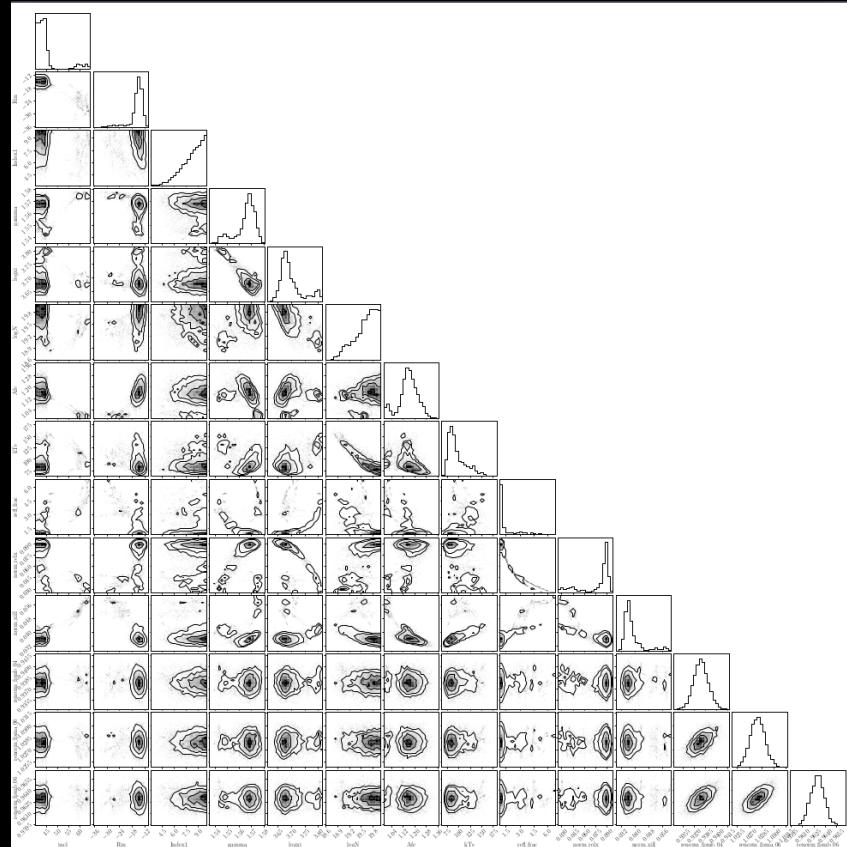
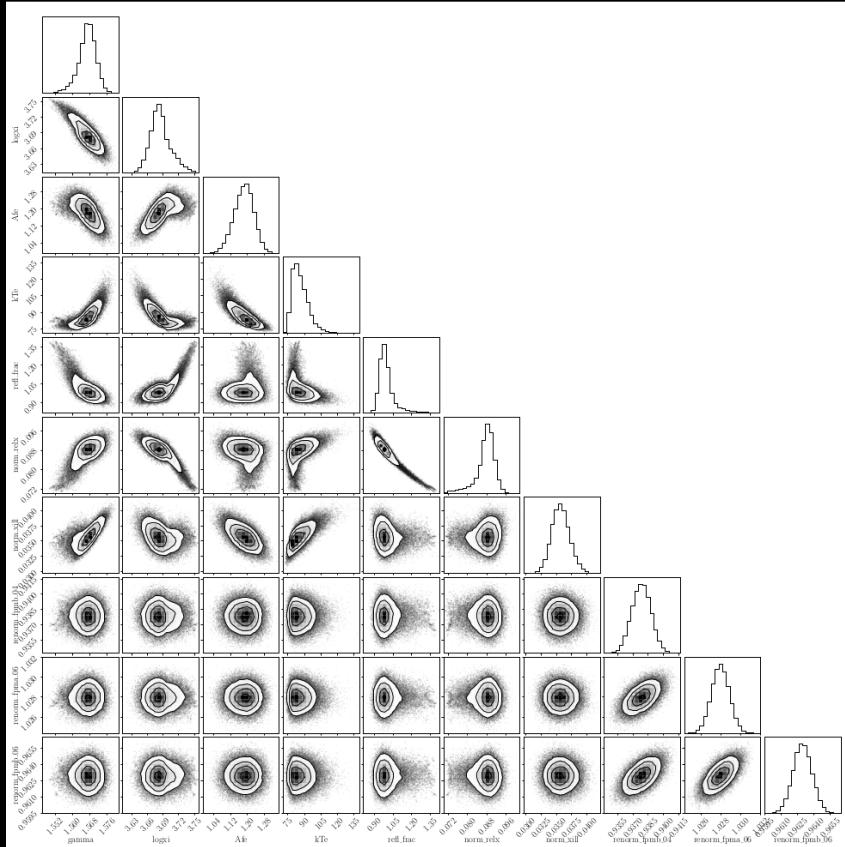


This chain converged perfectly well in one day!

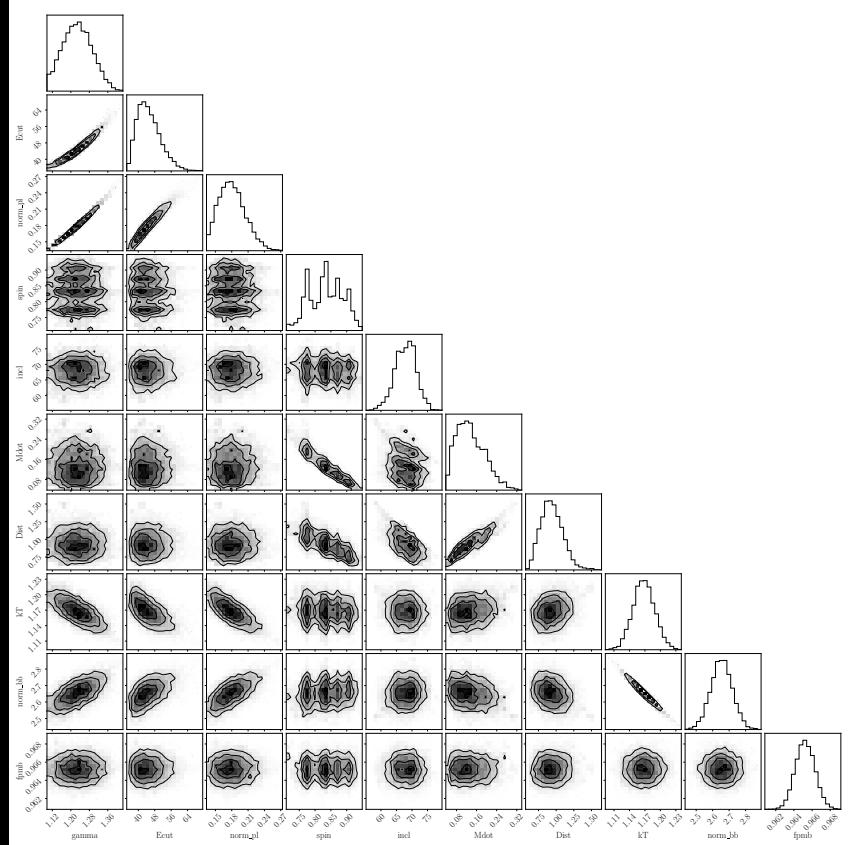
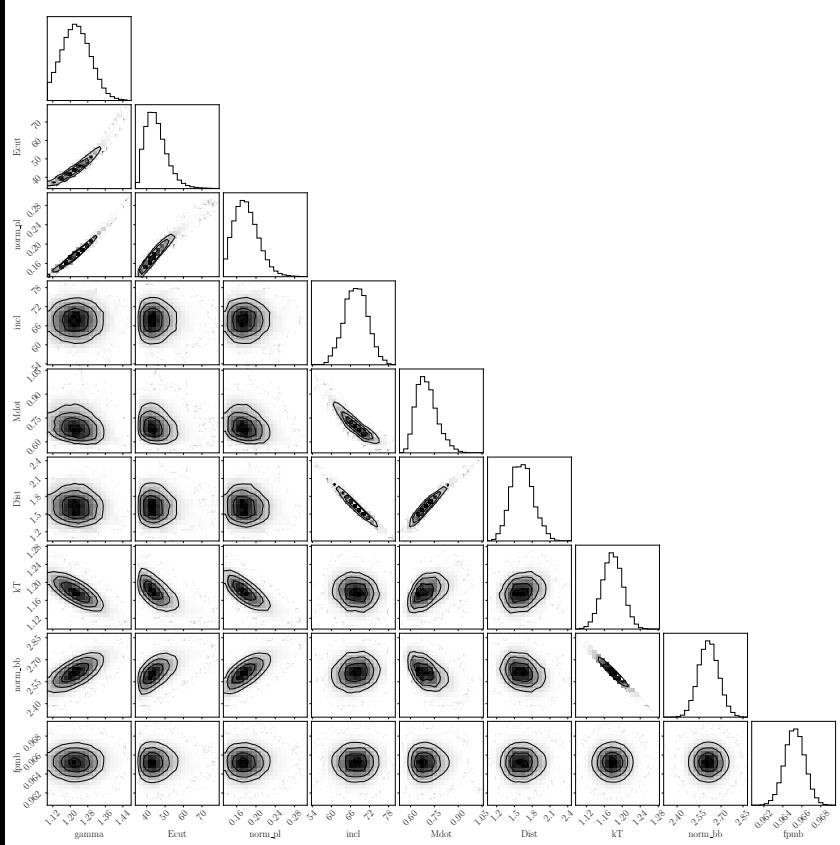
Between ~45 and 300 independent samples depending on the parameter

~100 walkers for 10 parameters, ~80k steps

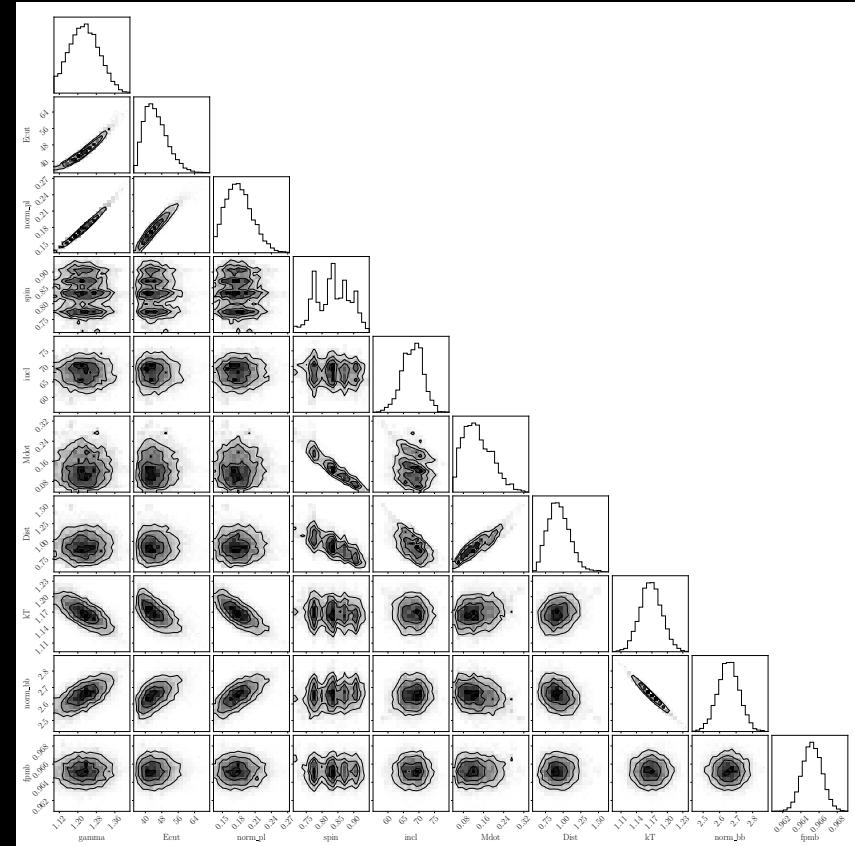
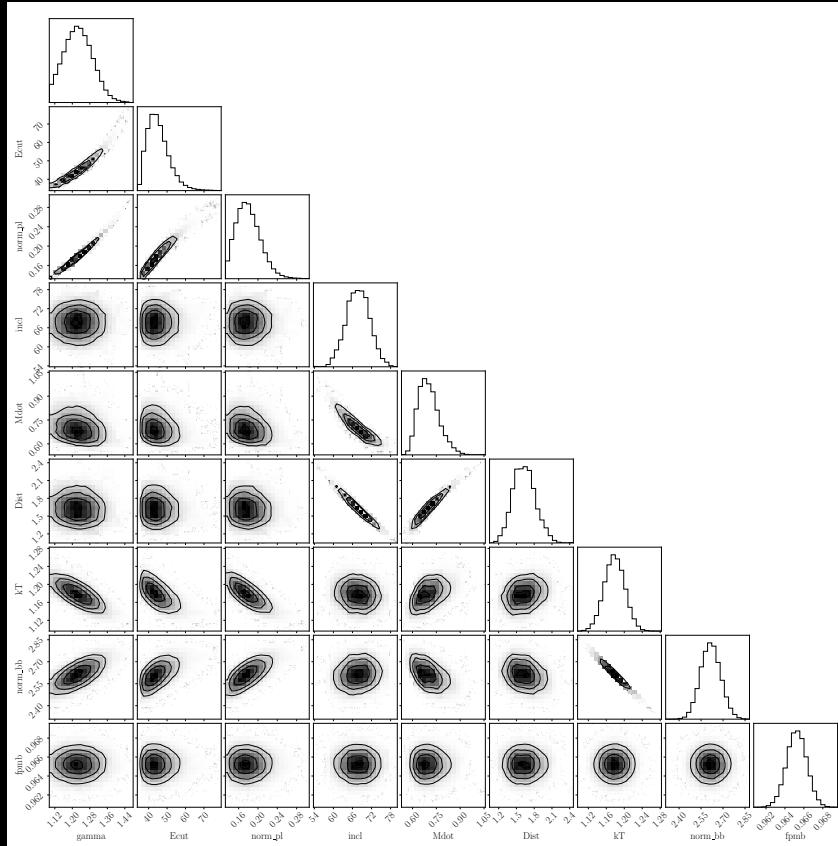
Back to my chain – fix the culprits



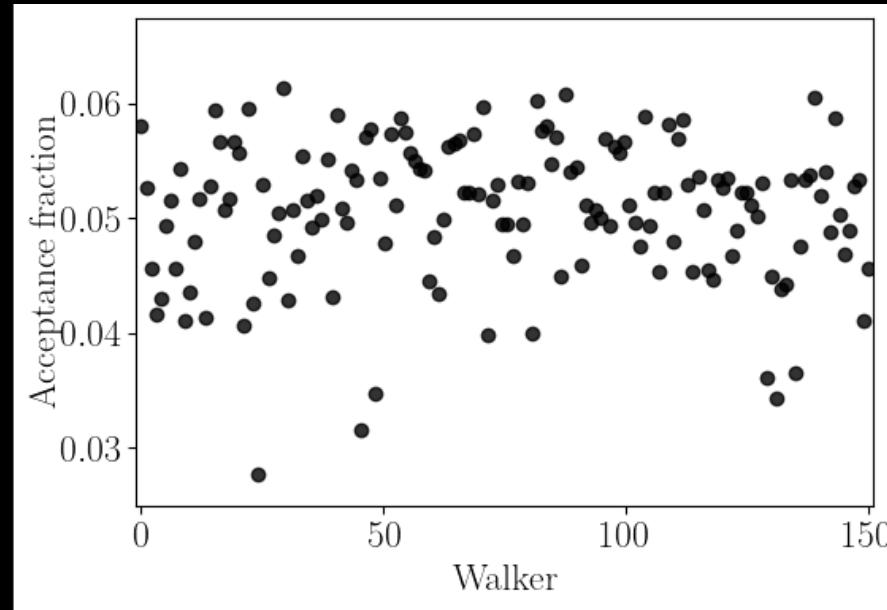
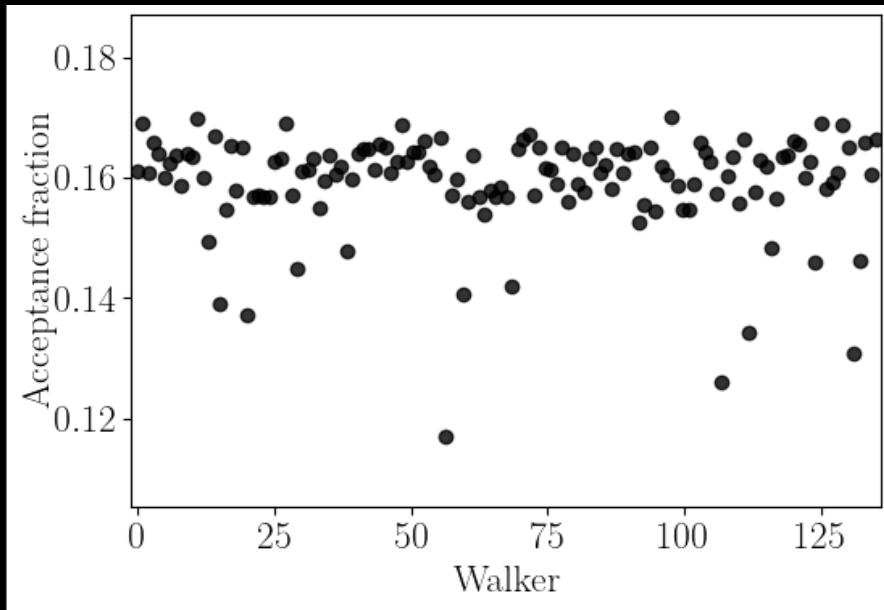
An example from Kerrbb



The only difference is the spin...

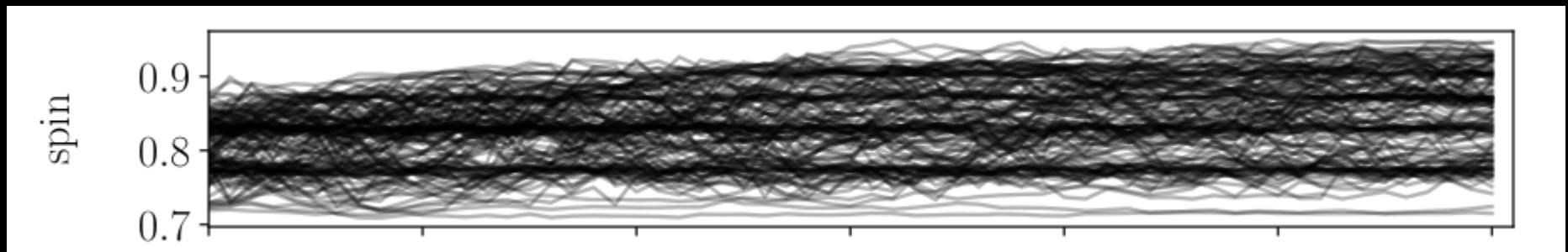


The only difference is the spin...



A single badly behaved parameter can **completely** ruin a chain

The only difference is the spin...



A single badly behaved parameter can **completely** ruin a chain

It was me, interpolation, all along!

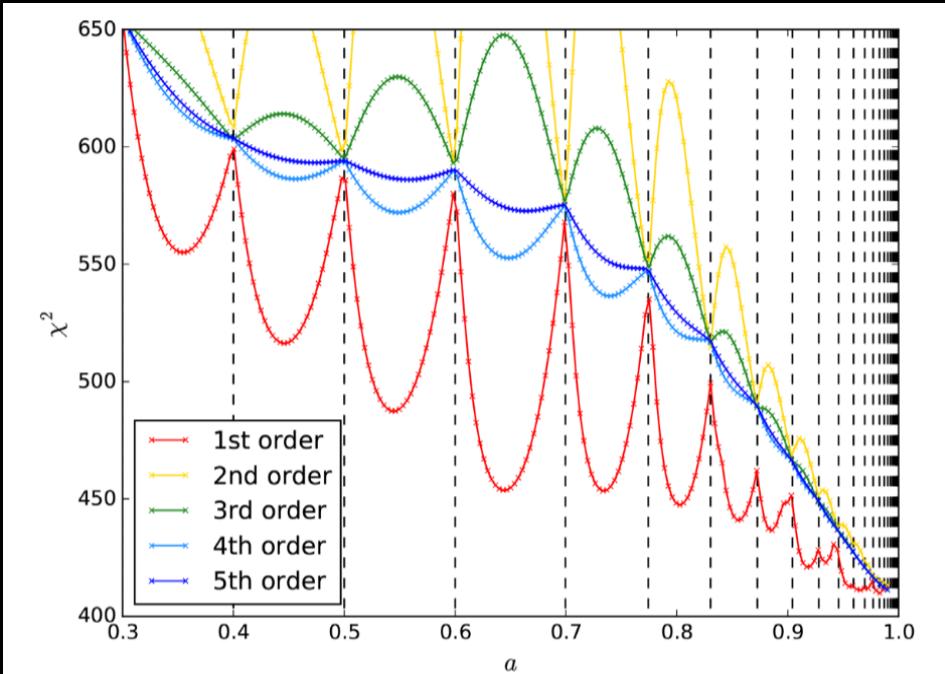
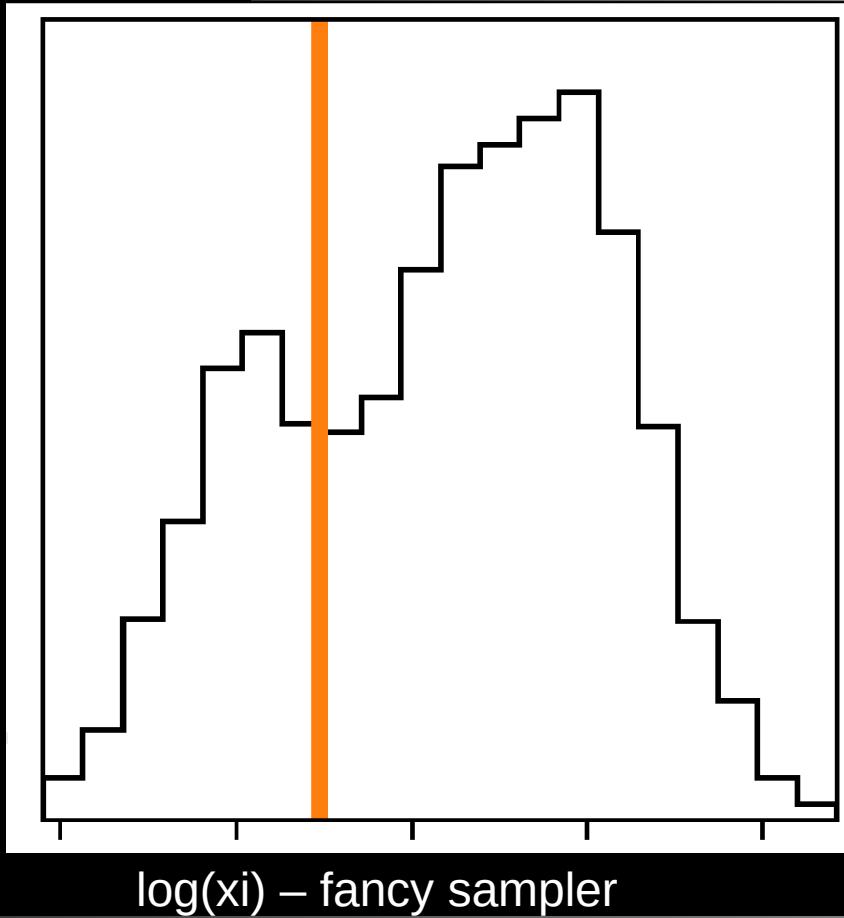
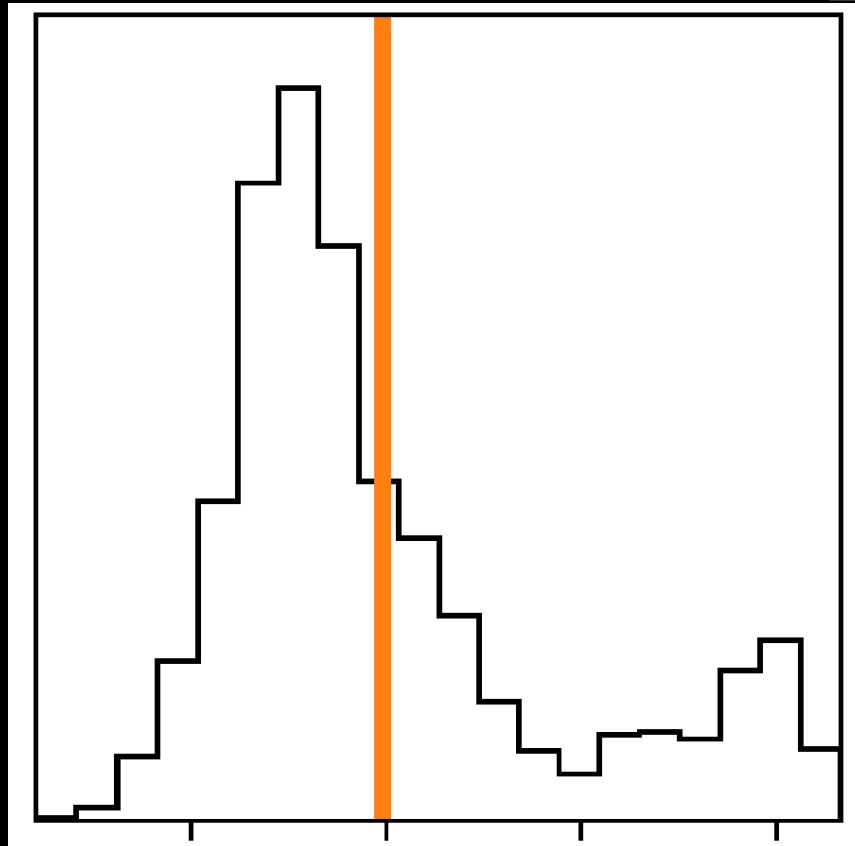


Figure A1. χ^2 contour for simulated POLYKERRBB spectra ($a = 0.998$, $i = 40^\circ$, $M = 10 M_\odot$, $D = 3\text{kpc}$). Crosses indicate where the fit is evaluated, and vertical dashed lines show the values of spin used in KERRBB. The lowest points (red) are for the 1st order interpolation, and correspond to standard KERRBB.

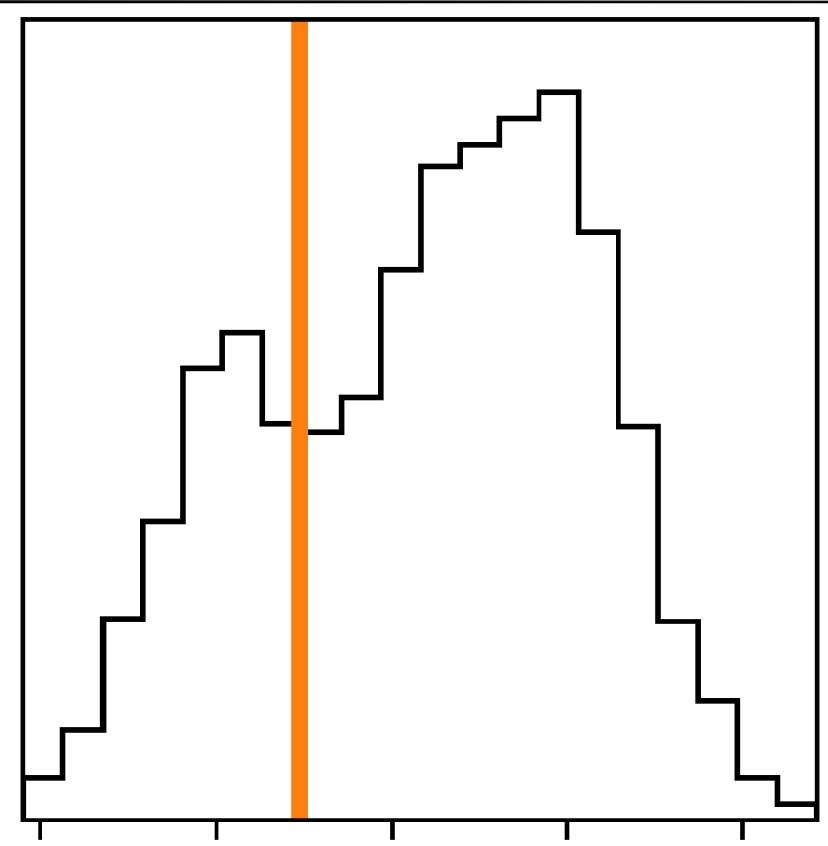
It was me, interpolation, all along!



It was me, interpolation, all along!

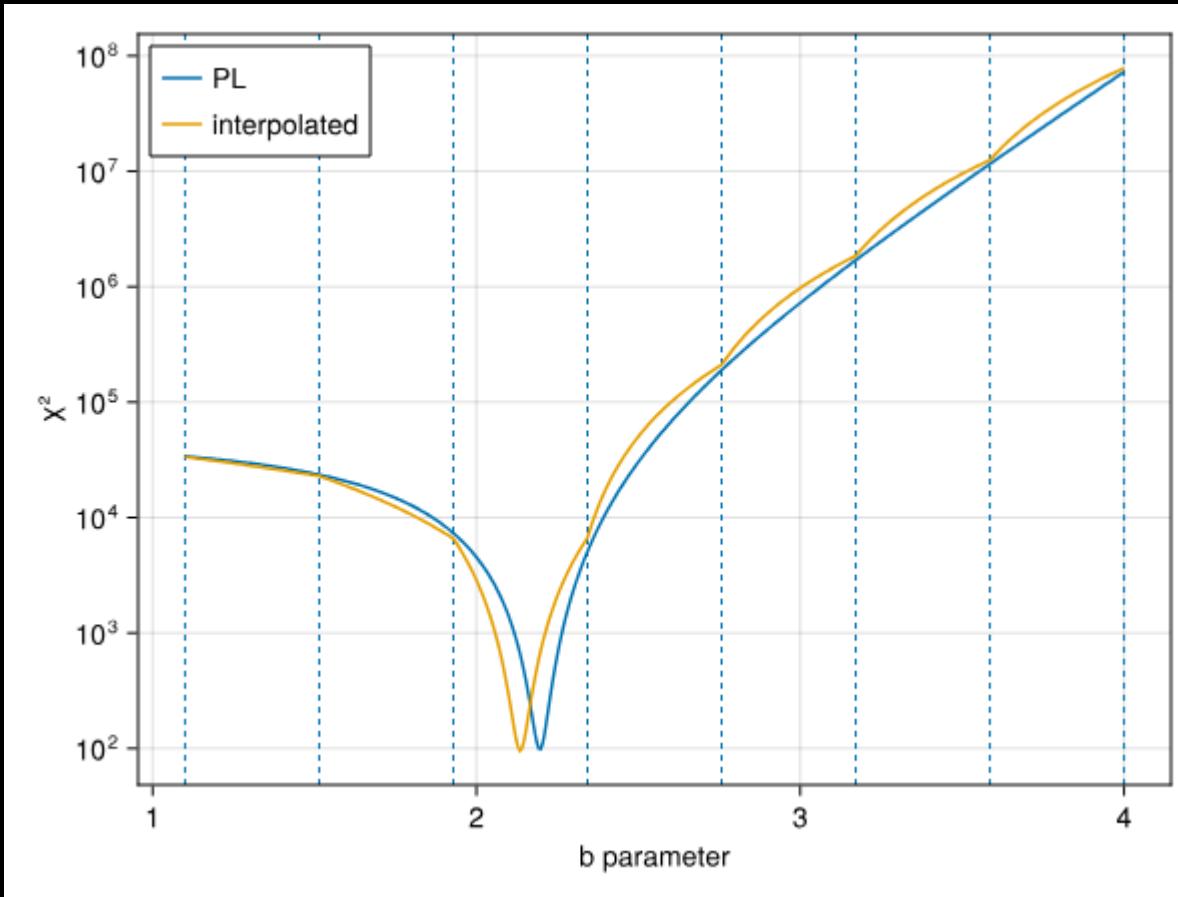


log(ξ) – emcee (failed)



log(ξ) – fancy sampler

It was me, interpolation, all along!



Recommendations for all of us

- 1) Take well known data, calculate likelihood as a function of a single parameter, find if there are discontinuities
- 2) If there are, think if you can rework the code to remove them

Recommendations for all of us

- 1) Take well known data, calculate likelihood as a function of a single parameter, find if there are discontinuities
- 2) If there are, think if you can rework the code to remove them
- 3) If you can't, accept that you have to run much, much longer chains to explore your parameter space

Recommendations for all of us

- 1) Take well known data, calculate likelihood as a function of a single parameter, find if there are discontinuities
- 2) If there are, think if you can rework the code to remove them
- 3) If you can't, accept that you have to run much, much longer chains to explore your parameter space
- 4) If you can't, try different Bayesian samplers. Emcee is not the only one that exists.

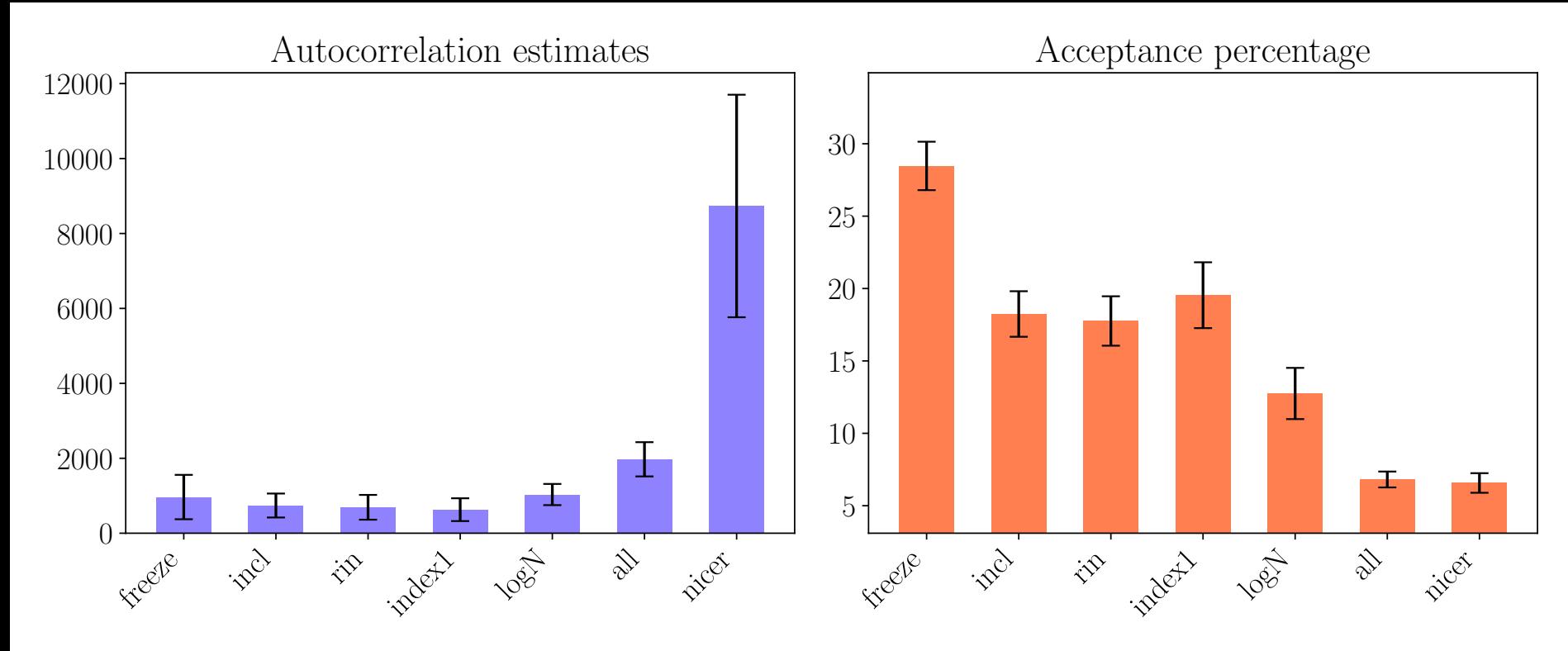
Recommendations for all of us

- 1) Take well known data, calculate likelihood as a function of a single parameter, find if there are discontinuities
- 2) If there are, think if you can rework the code to remove them
- 3) If you can't, accept that you have to run much, much longer chains to explore your parameter space
- 4) If you can't, try different Bayesian samplers. Emcee is not the only one that exists.
- 5) If you can't, report that your chains are not converged and you are not certain about your posteriors

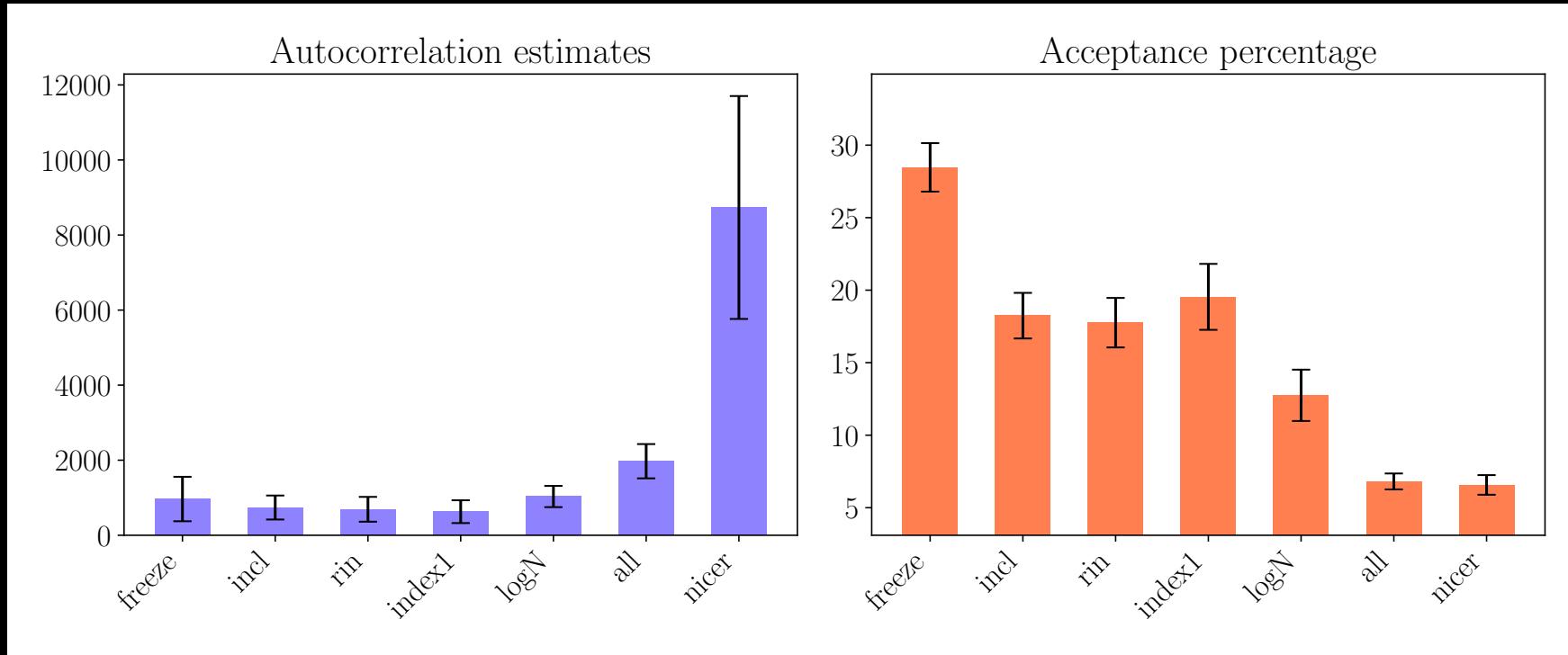
Recommendations for all of us

- 1) Take well known data, calculate likelihood as a function of a single parameter, find if there are discontinuities.
- 2) If there are, think if you can rework the code to remove them!
- 3) If you can't, accept that you have to run much, much longer chains to explore your parameter space
- 4) If you can't, try different Bayesian samplers. Emcee is not the only one that exists.
- 5) If you can't, report that your chains are not converged and you are not certain about your posteriors
- 6) Frequentist likelihood optimization is affected just as badly (if not more...) as Bayesian samplers

Smarter priors or better data?

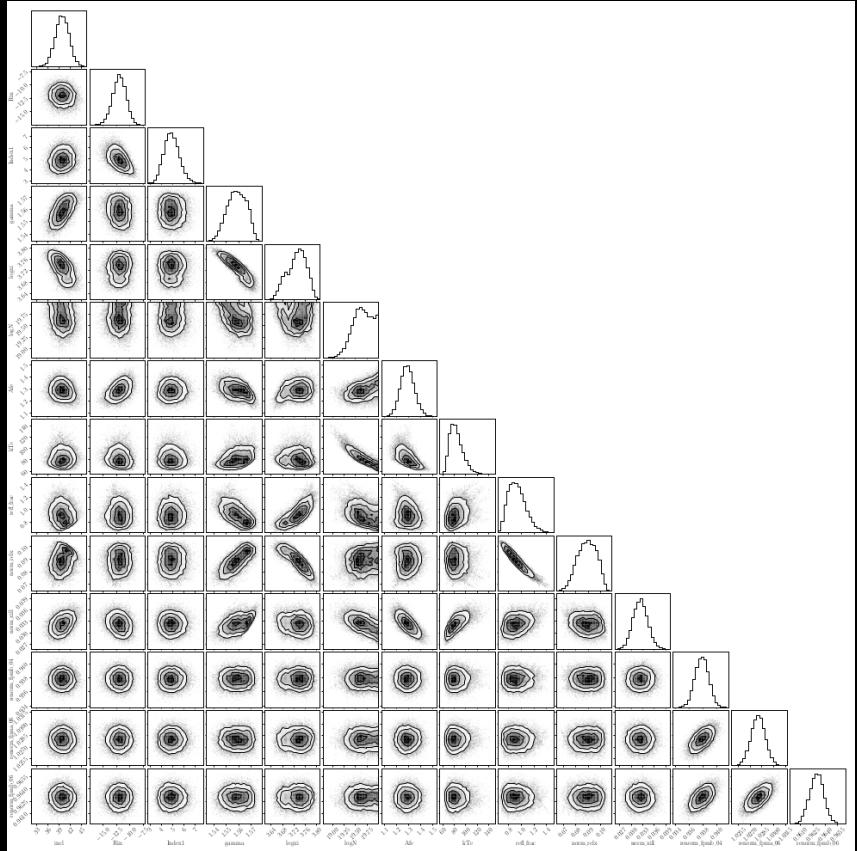


Smarter priors or better data?



- 1) Curse of dimensionality: freeing any parameter makes things worse
- 2) Soft X-rays don't help (it's worse because I need diskbb+tbabs)

Do the posteriors get better?



Yes! This chain at least has a chance of converging...eventually...possibly?

Would require ~1e8-1e9 model evaluations, or weeks on my laptop

Why is this still happening?