Simulation Based Inference

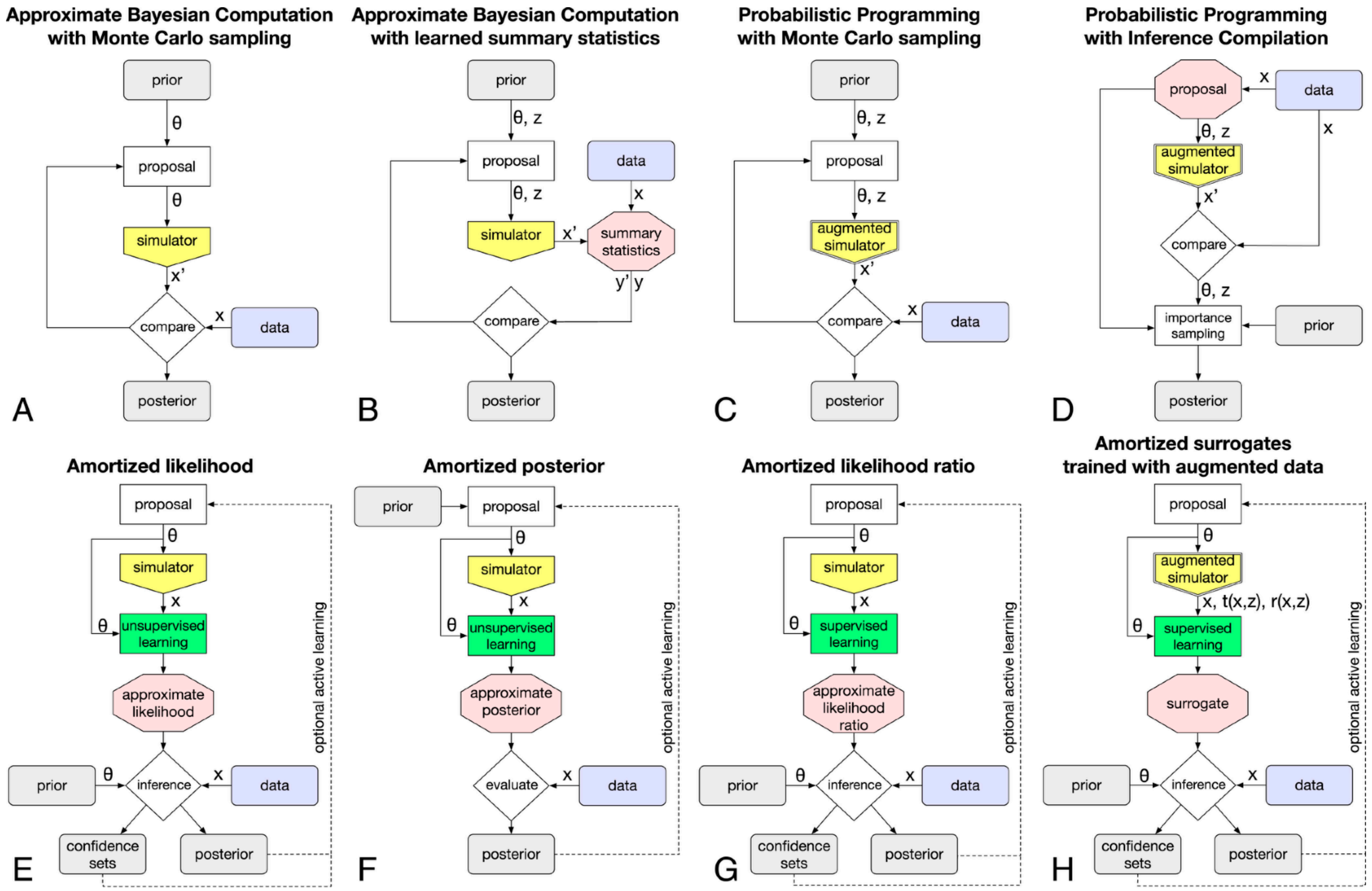
Astro Data Science Discussion group

David Yallup - 28/02/24

Building a mindmap of SBI

Rather than analyse parameters (of models), analyse mock data (from models)

- NLE
- NPE
- NRE



Either using **classifiers** to distinguish between two piles of mock data, or using invertible **density estimators** to learn distributions over piles of mock data

[https://github.com/sbi-dev/sbi]

SBI Package `sbi`

Seems to be generally touted as one stop shop for all SBI needs

Neural Posterior Estimation: amortized (NPE) and sequential (SNPE)

- `SNPE_A` (including amortized single-round `NPE`) from Papamakarios G and Murray I *Fast ε-free Inference of Simulation Models with Bayesian Conditional Density Estimation* (NeurIPS 2016).
- `SNPE_C` or `APT` from Greenberg D, Nonnenmacher M, and Macke J *Automatic Posterior Transformation for likelihood-free inference* (ICML 2019).
- `TSNPE` from Deistler M, Goncalves P, and Macke J *Truncated proposals for scalable and hassle-free simulation-based inference* (NeurIPS 2022).

Neural Likelihood Estimation: amortized (NLE) and sequential (SNLE)

- `SNLE_A` or just `SNL` from Papamakarios G, Sterrat DC and Murray I *Sequential Neural Likelihood* (AISTATS 2019).

Neural Ratio Estimation: amortized (NRE) and sequential (SNRE)

- `(S)NRE_A` or `AALR` from Hermans J, Begy V, and Louppe G. *Likelihood-free Inference with Amortized Approximate Likelihood Ratios* (ICML 2020).
- `(S)NRE_B` or `SRE` from Durkan C, Murray I, and Papamakarios G. *On Contrastive Learning for Likelihood-free Inference* (ICML 2020).
- `BNRE` from Delaunoy A, Hermans J, Rozet F, Wehenkel A, and Louppe G. *Towards Reliable Simulation-Based Inference with Balanced Neural Ratio Estimation* (NeurIPS 2022).
- `(S)NRE_C` or `NRE-C` from Miller BK, Weniger C, Forré P. *Contrastive Neural Ratio Estimation* (NeurIPS 2022).

These slides: bit.ly/ras-sbi-disc
Q&A: bit.ly/ras-sbi-qs

Structure

- Why SBI?
- What are the limits?
- (When) Will SBI replace Likelihood-based?

https://ras.ac.uk/events-and-meetings/ras-meetings/simulation-based-inference-astrophysics

These slides: bit.ly/ras-sbi-disc
Q&A: bit.ly/ras-sbi-qs

What are the limits?

- When not to use it?
- What is still missing from SBI frameworks?
- Is it only a matter of building trust in SBI techniques?
- Simulations are key, but what about compression?
- Are the emerging SBI methods multi-purpose enough that we now only need to worry about how we perform our simulations?

Thanks to organisers of RAS SBI meeting:
Alessio Spurio Mancini
Ian Harrison
Will Hartley

Watch out for community document summarising meeting and SBI outlook.

I know no other discipline where half of the principle equation is so widely ignored

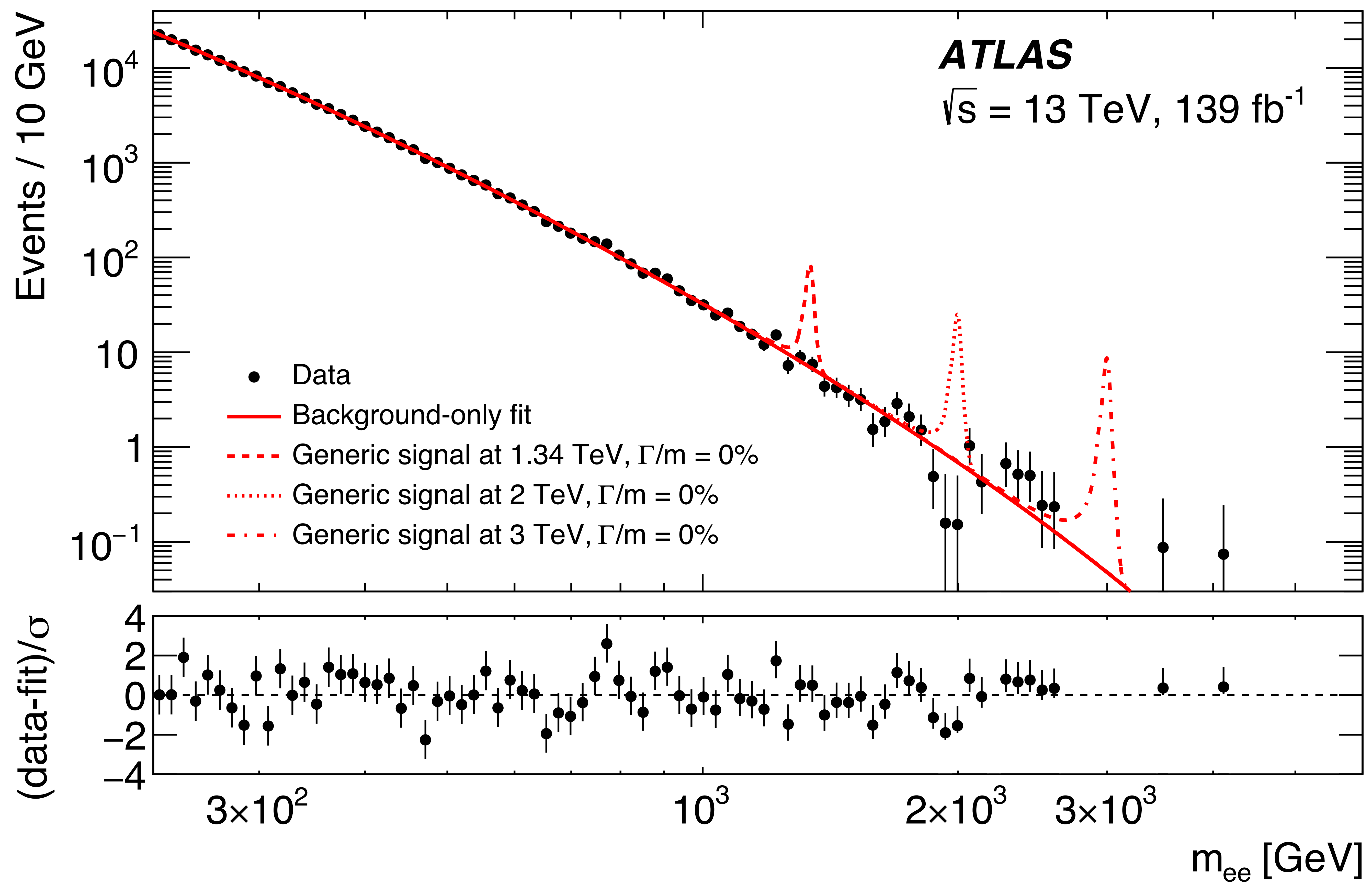
John Skilling

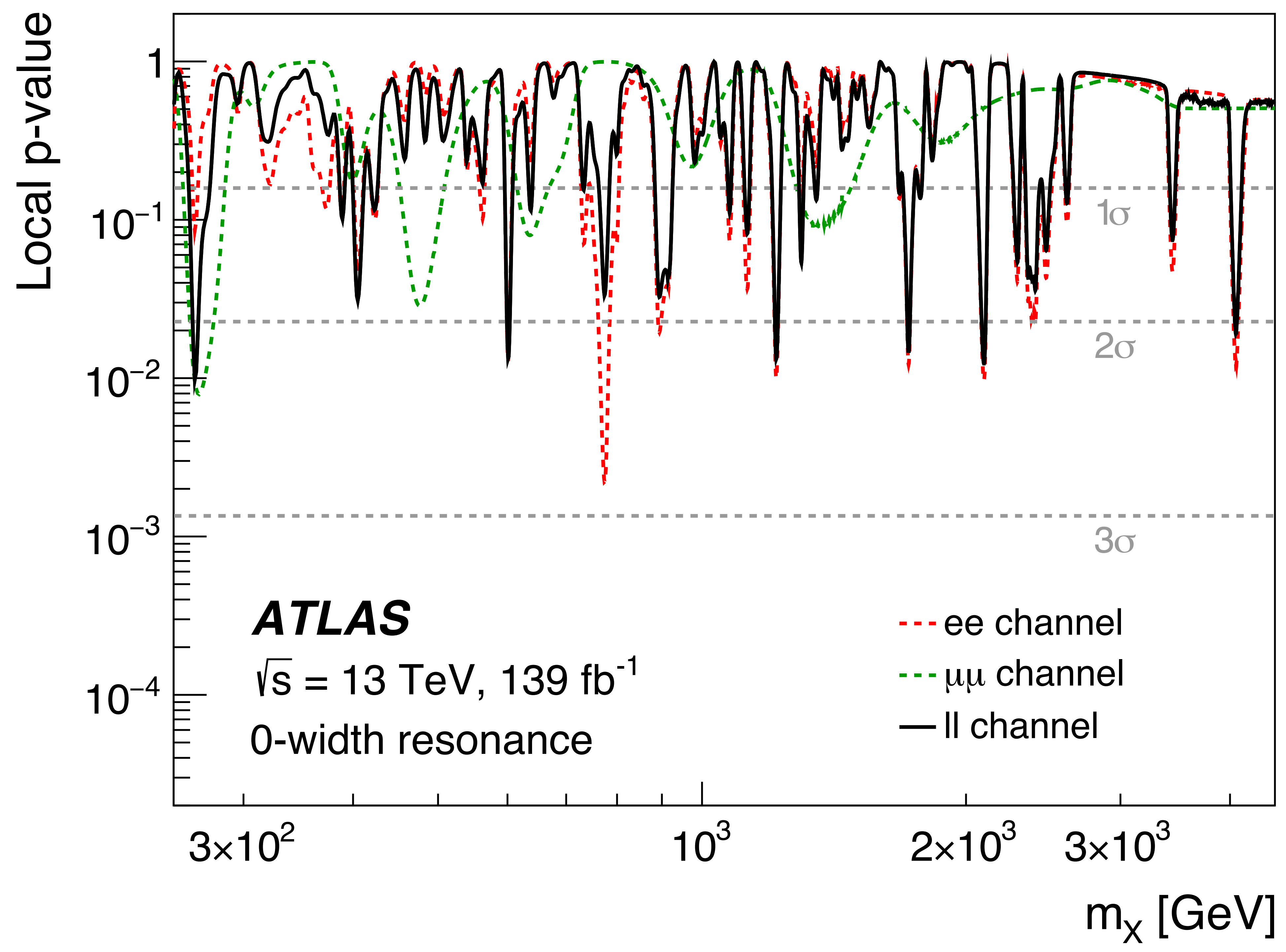
$$\mathscr{L} \times \Pi = Z \times P$$

Likelihood x Prior = Evidence x Posterior

"Think of "Bayesian Inference" as generation of the Evidence, with the Posterior following if needed.
Evidence is primary."

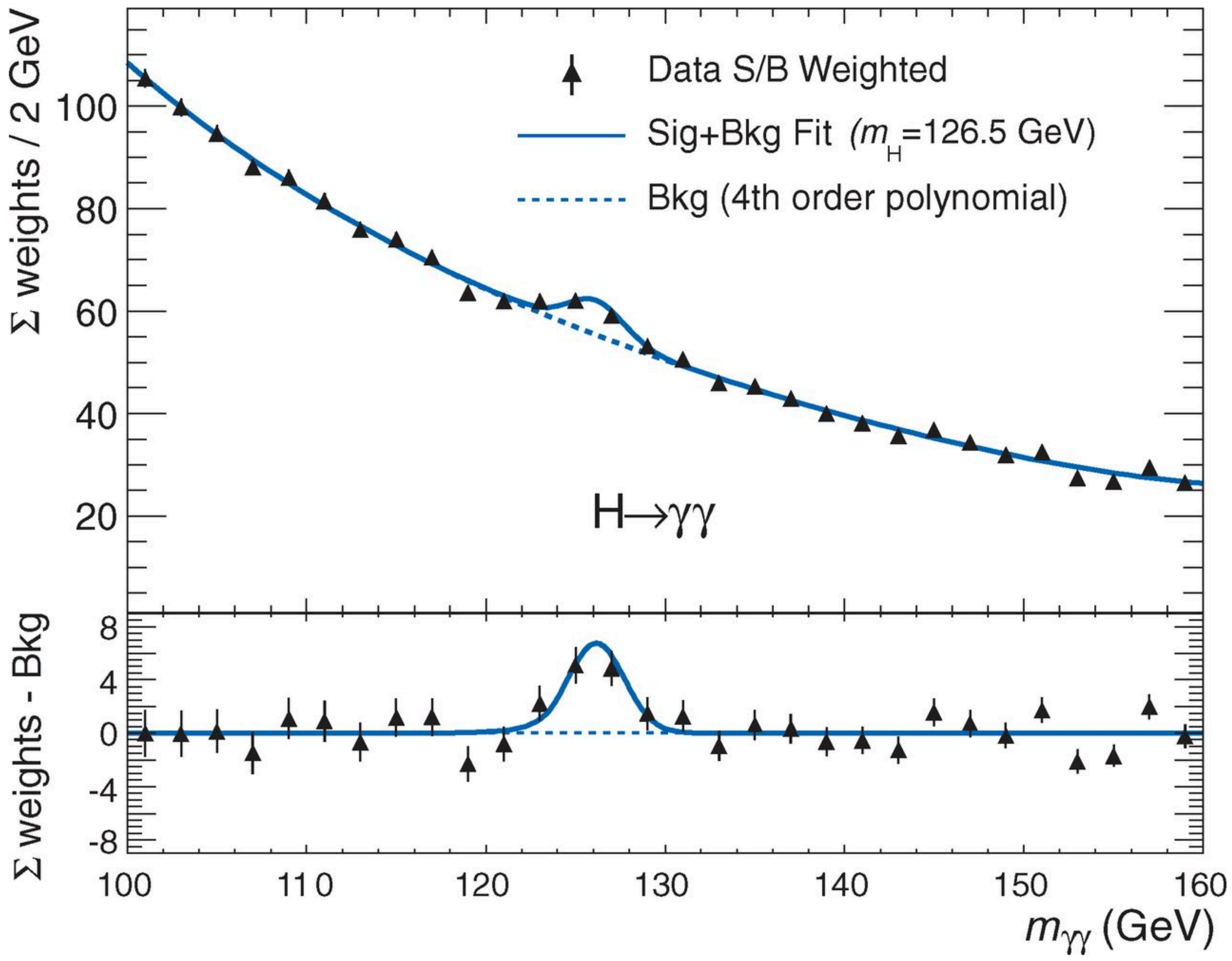
Evidence is all you need

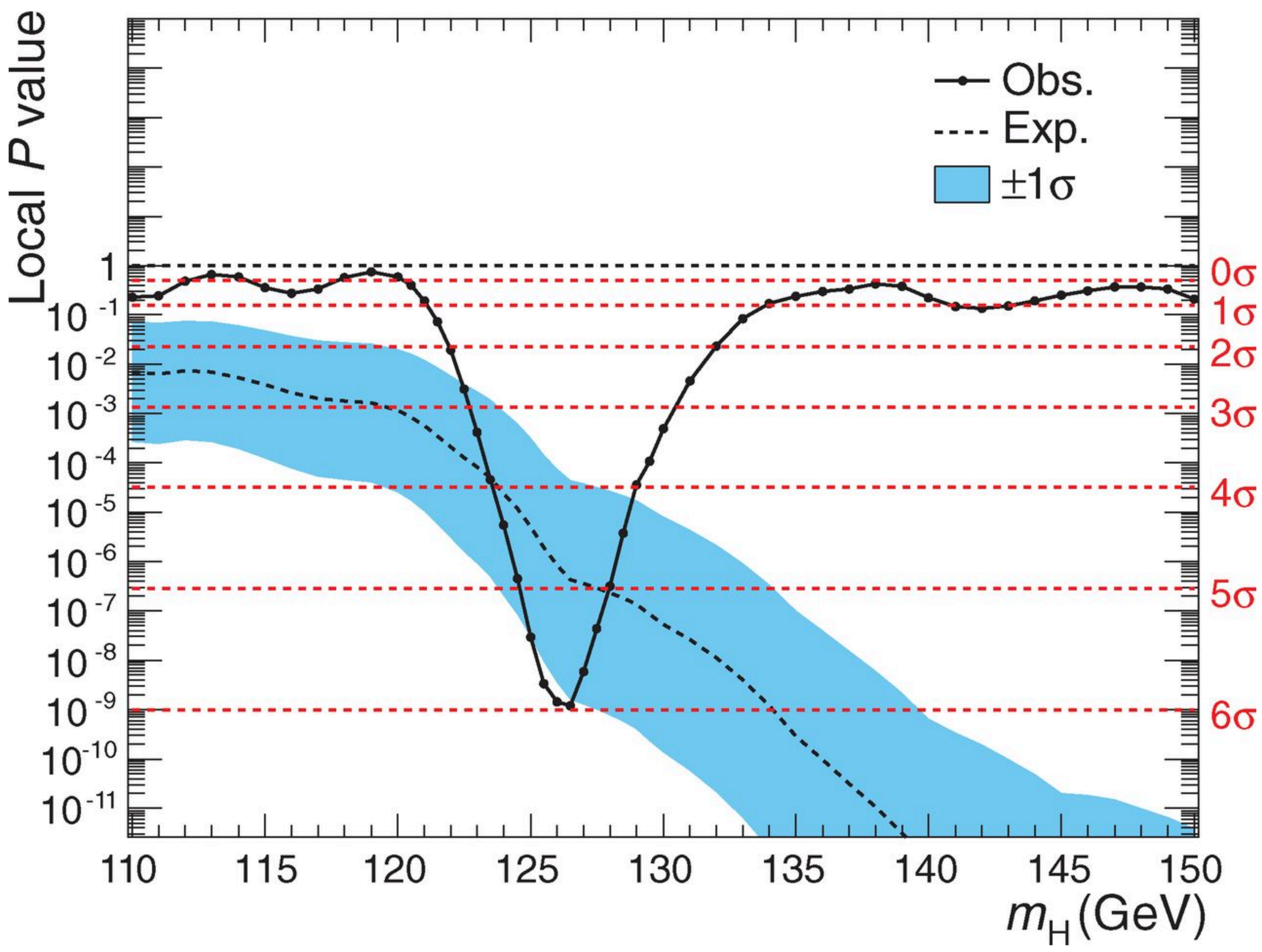


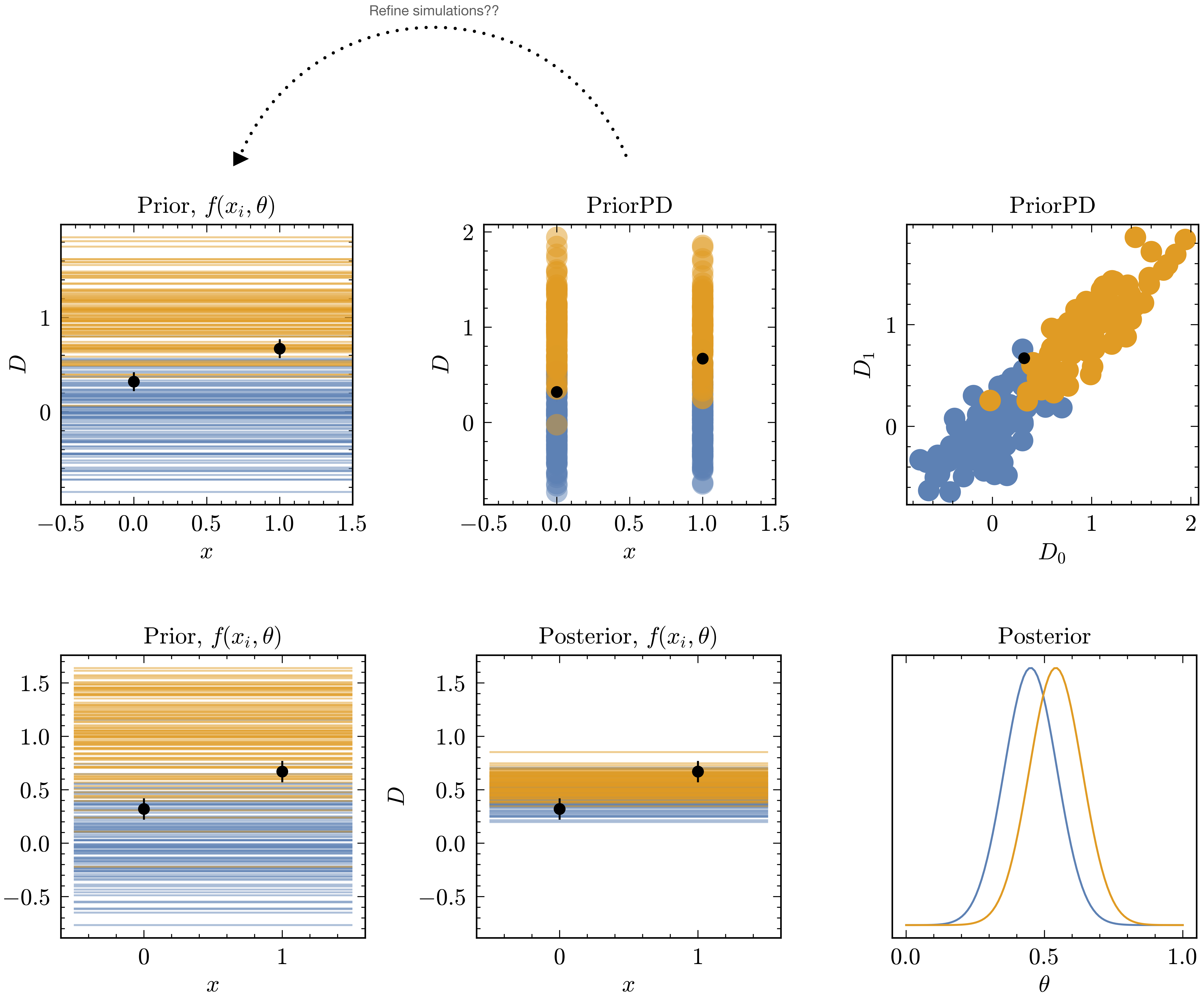


No significant excess is observed. The largest deviations from the background-only hypothesis in the dielectron, dimuon and combined dilepton channels are observed at masses of 774 GeV, 267 GeV and 264 GeV for zero-width signals with a local p0 of 2.9 σ, 2.4 σ and 2.3 σ and a **global significance of 0.1 σ, 0.3 σ, and zero**, respectively

Evidence is all you need (sort of)







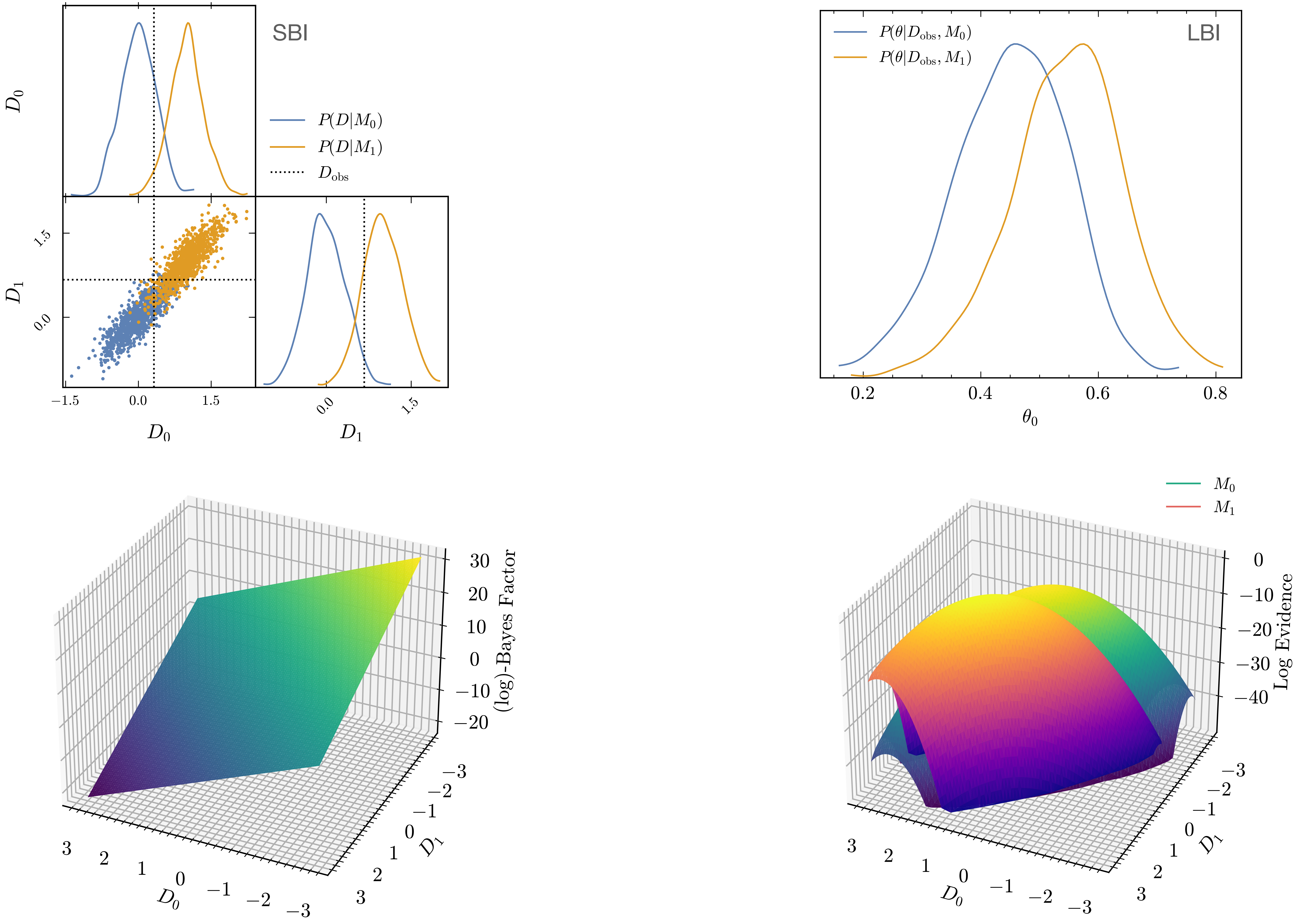
Ingredients:
Model for your data: $f(x_i, \theta)$
Noise model for your data e.g.: $n_i \leftarrow \mathcal{N}(0,\sigma)$

$$D_i = f(x_i, \theta) + n_i$$

SBI

$$L(D\,|\,\theta, M) = \prod_i \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(f(x_i,\theta) - D_i^{obs})^2}{2\sigma^2}}$$

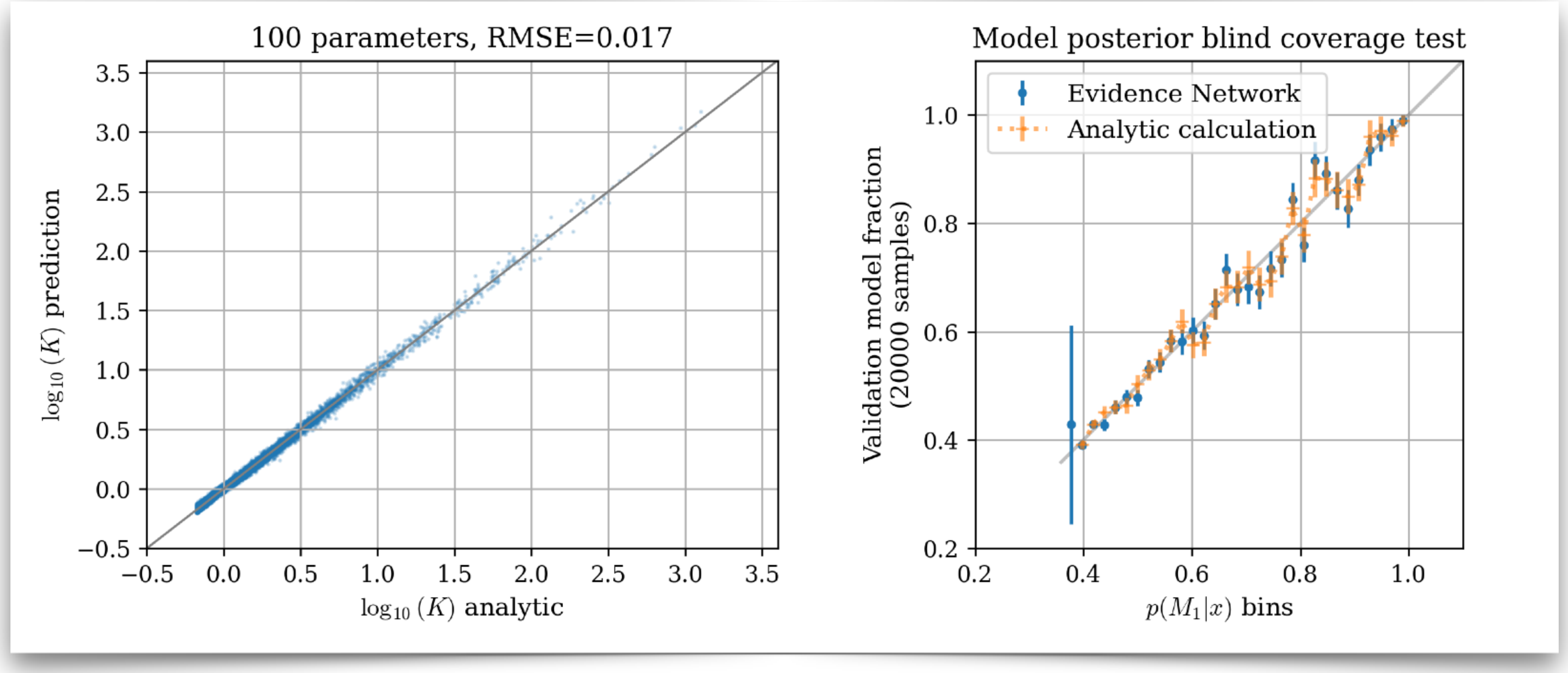
LBI



SBI
$P(D|M_0)$
$P(D|M_1)$
$D_{\rm obs}$
D_0
D_1
LBI
$P(\theta|D_{\rm obs}, M_0)$
$P(\theta|D_{\rm obs}, M_1)$
θ_0
(log)-Bayes Factor
M_0
M_1
Log Evidence

The likelihood ratio trick appears to imply model comparison is “easier” than parameter estimation in SBI.

This is what I find most interesting about SBI!



[2305.11241]

Evidence Networks:

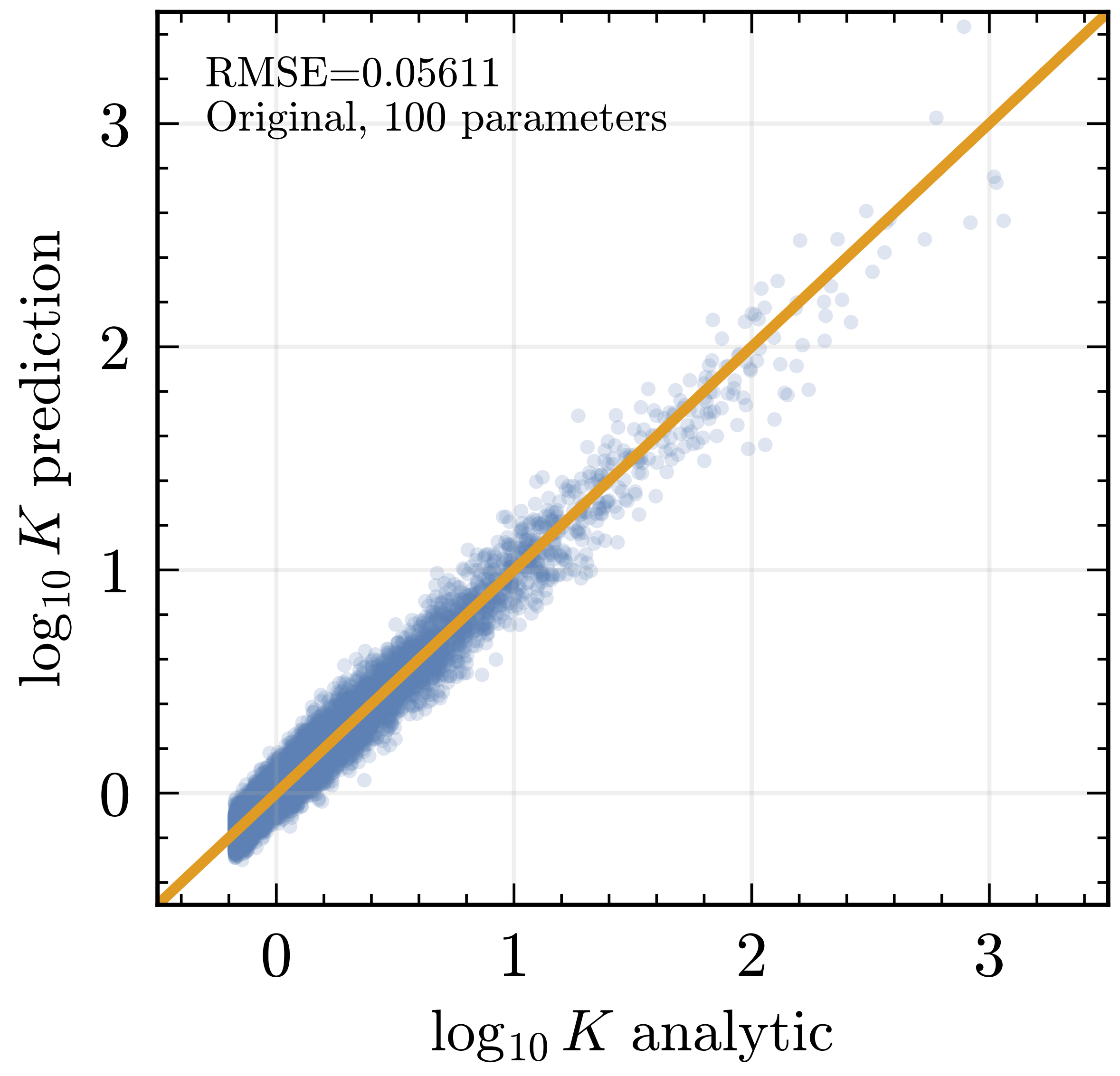
Take the "Likelihood ratio trick" seriously with generation of data from a prior distribution.

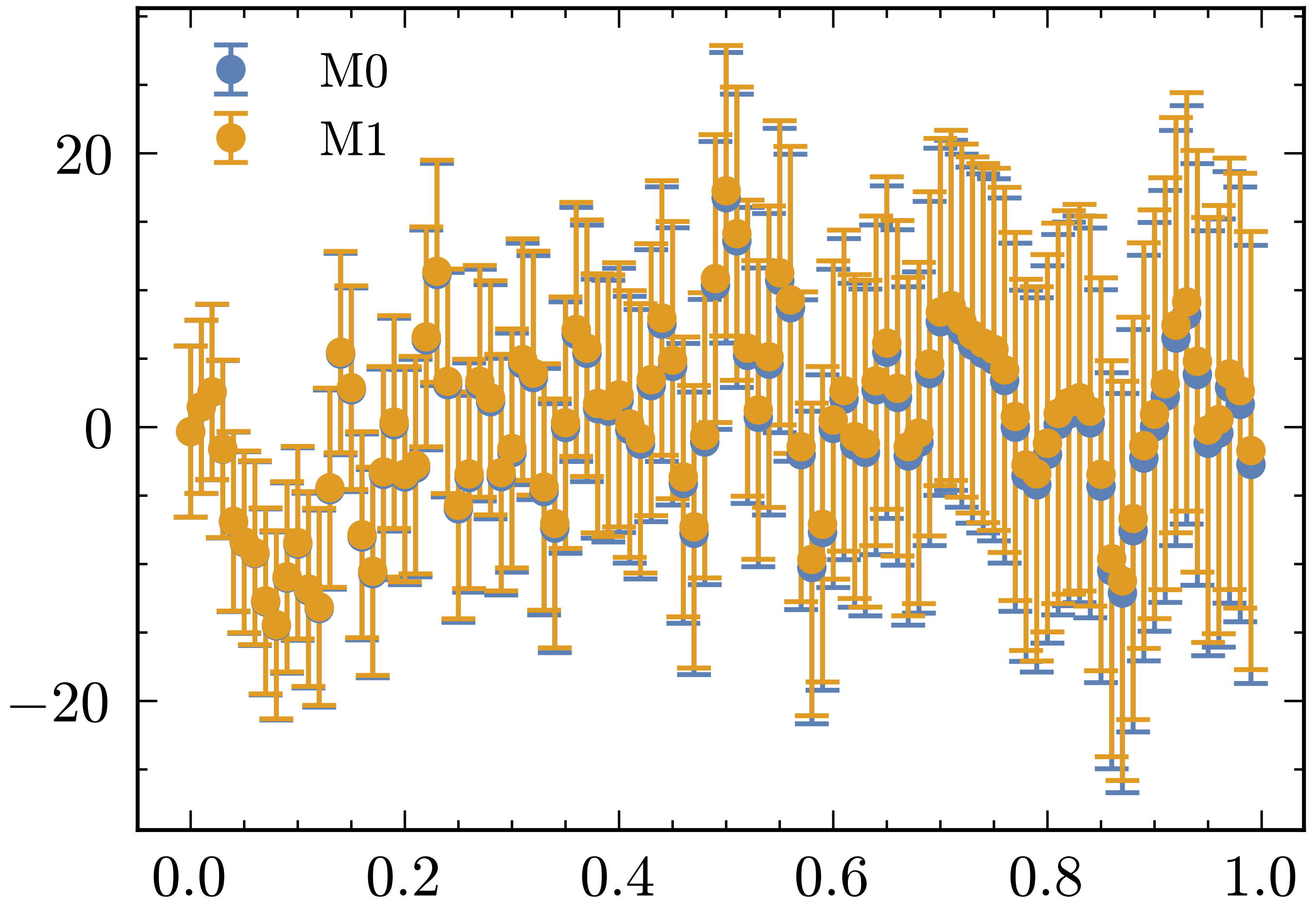
[1805.12244]

The likelihood ratio trick (LRT). A surrogate model for the likelihood ratio $\hat{r}(x|\theta_0,\theta_1)$ can be defined by training a probabilistic classifier to discriminate between two equal-sized samples $\{x_i\} \sim p(x|\theta_0)$ and $\{x_i\} \sim p(x|\theta_1)$. The binary cross-entropy loss

$$L_{\mathrm{XE}} = -\mathbb{E}[\mathbb{1}(\theta=\theta_1)\log\hat{s}(x|\theta_0,\theta_1) + \mathbb{1}(\theta=\theta_0)\log(1-\hat{s}(x|\theta_0,\theta_1))] \tag{1}$$

is minimized by the optimal decision function $s(x|\theta_0,\theta_1) = p(x|\theta_1)/(p(x|\theta_0)+p(x|\theta_1))$. Inverting this relation, the likelihood ratio can be estimated from the classifier decision function $\hat{s}(x)$ as $\hat{r}(x|\theta_0,\theta_1) = (1-\hat{s}(x|\theta_0,\theta_1))/\hat{s}(x|\theta_0,\theta_1)$. This "likelihood ratio trick" is widely appreciated [5–



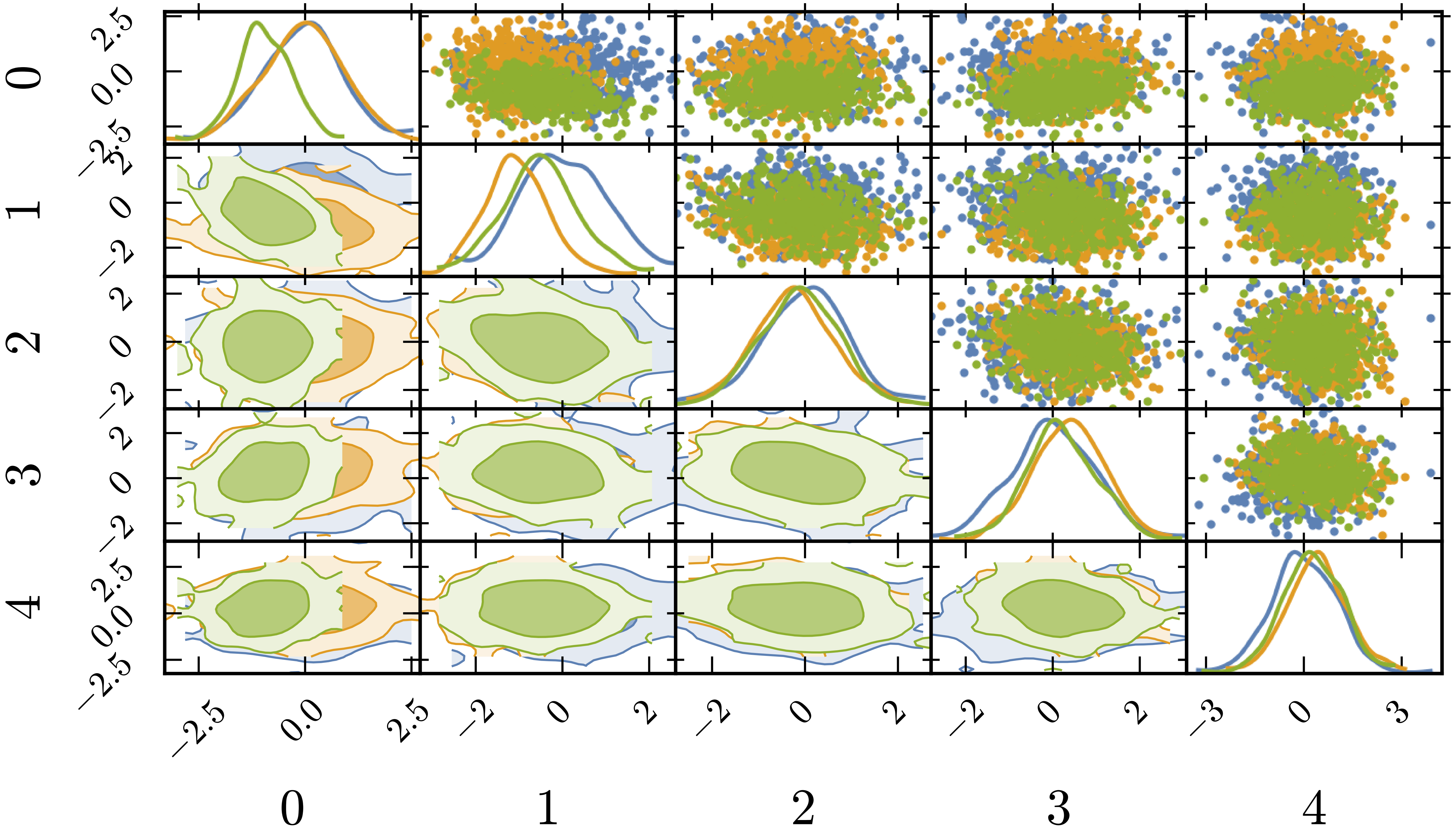


$$A_{i0} = 2x_i\,,$$
$$A_{ij} = \cos(j-1/2)x_i\,,$$
$$n_i \leftarrow \mathcal{N}(0,\{2,\ldots,4.5\})\,,$$
$$\theta_i \leftarrow \mathcal{N}(0,1)\,.$$
$$D_i = A_{ij}\theta_j + n_i$$

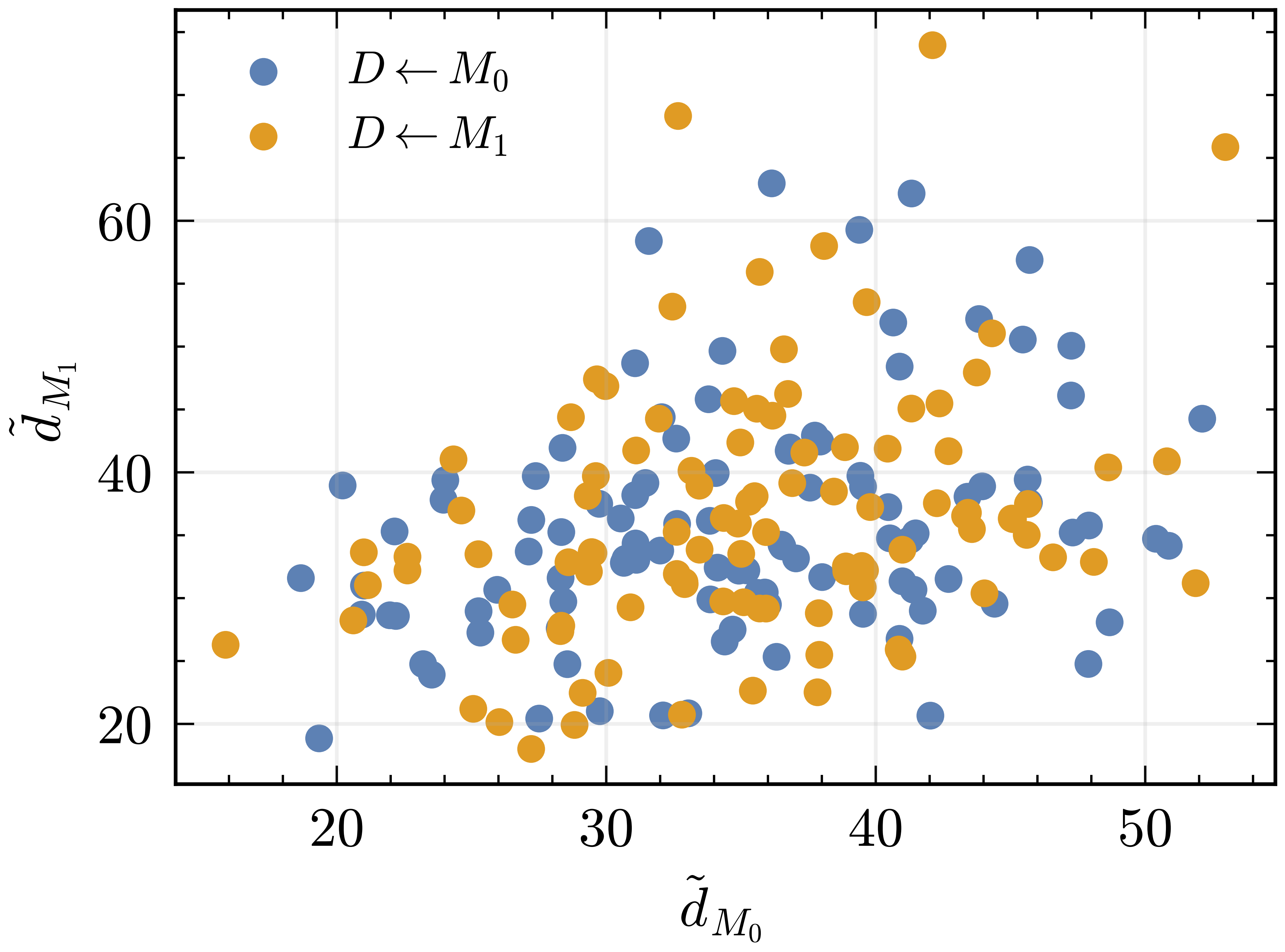
[1903.06682]

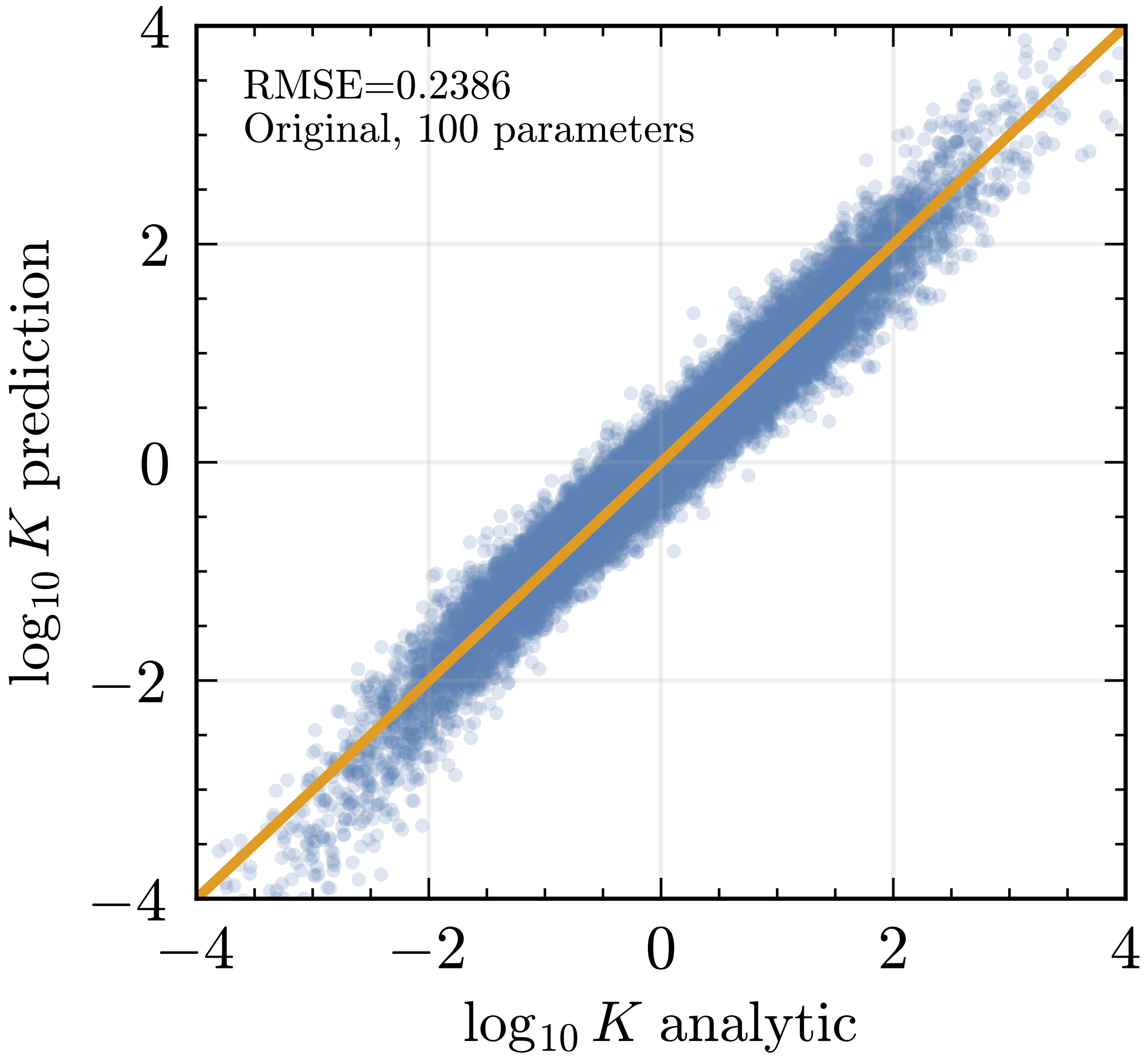
$$\frac{\tilde{d}}{2} = \langle(\ln \mathscr{L})^2\rangle_P - \langle \ln \mathscr{L}\rangle_P^2$$

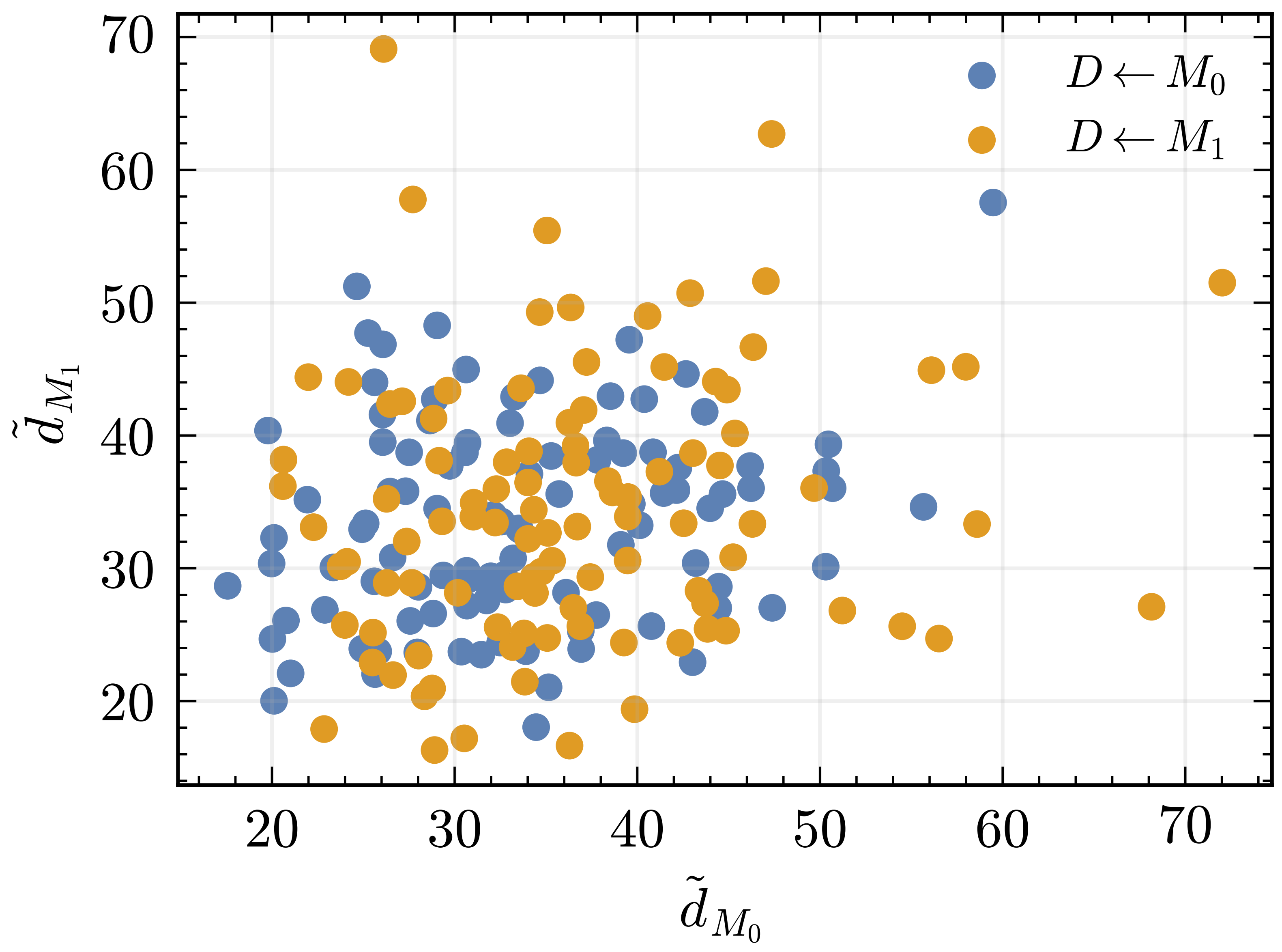
Blue (Prior), Yellow/Green (Posterior) first 5 params



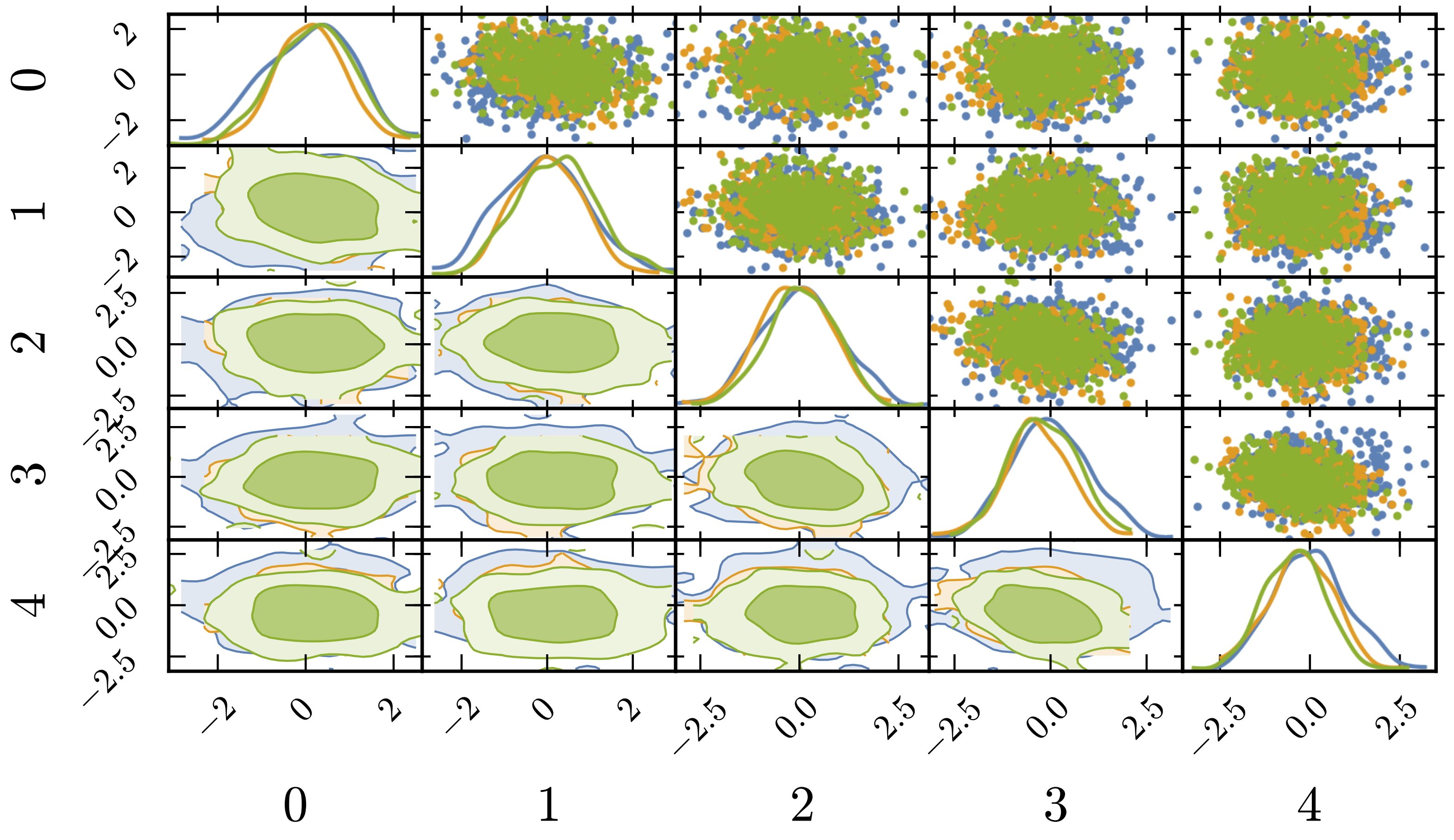
Bayesian Model Dimensionality - on a Gaussian gives the effective DoF

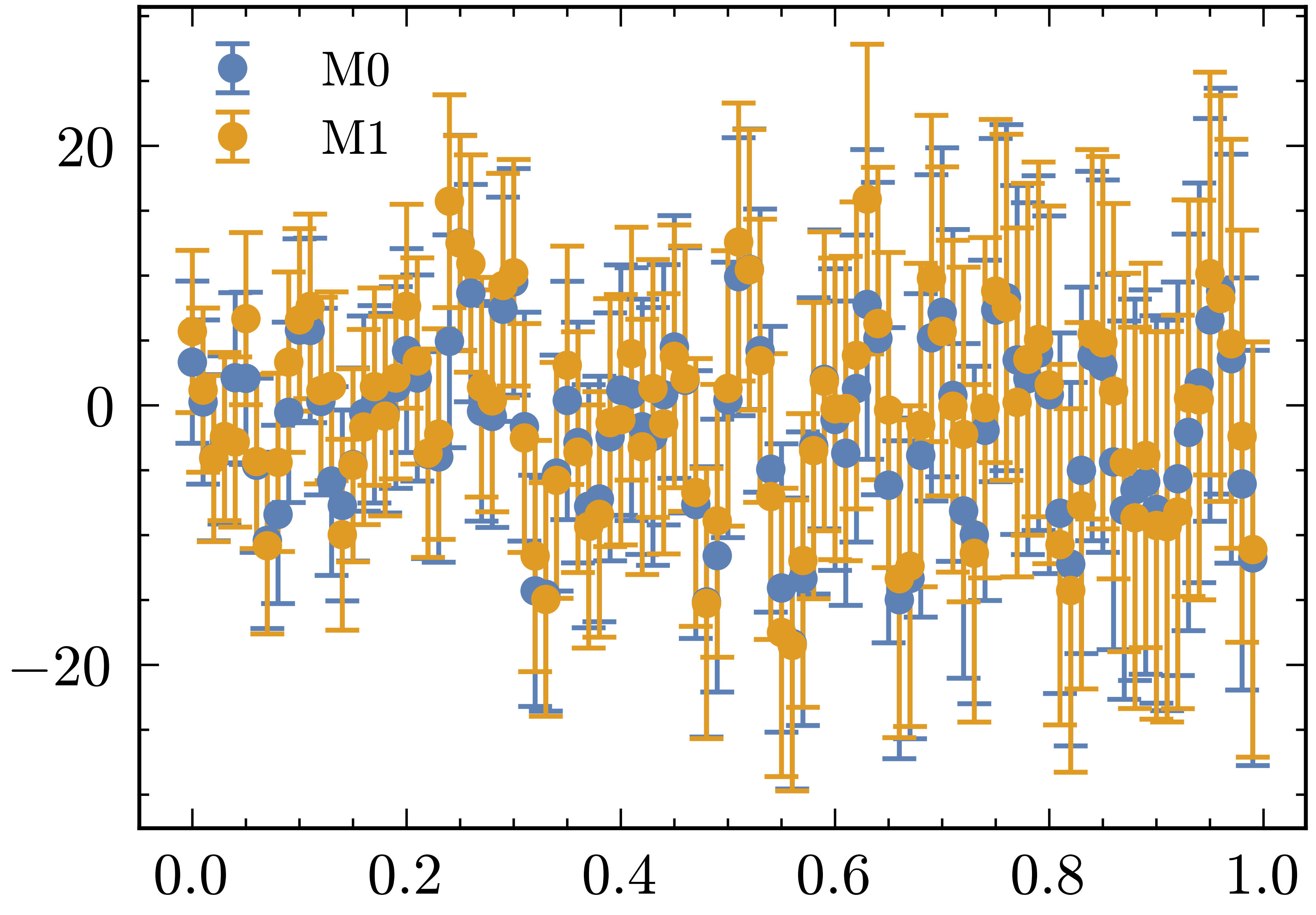


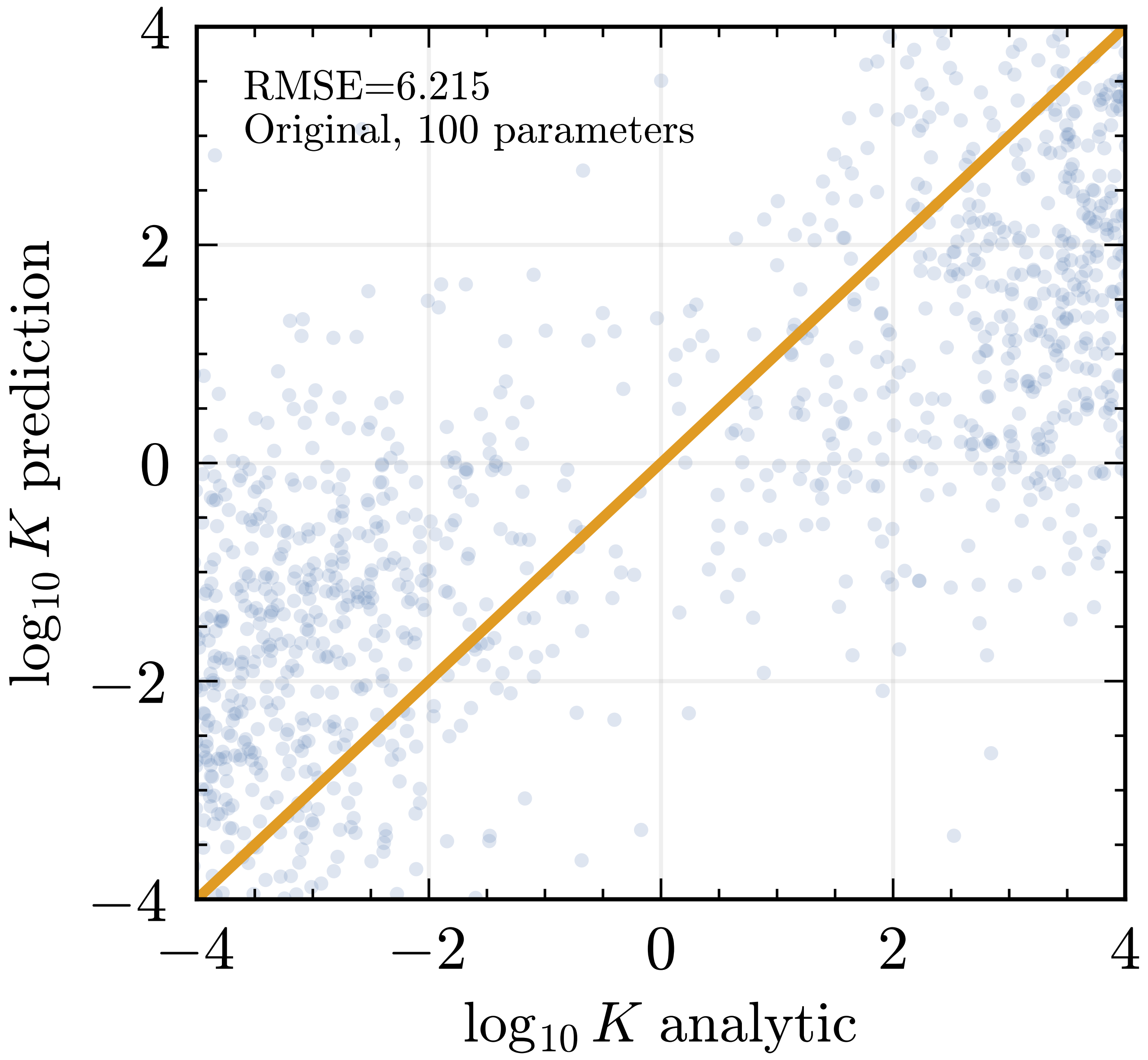


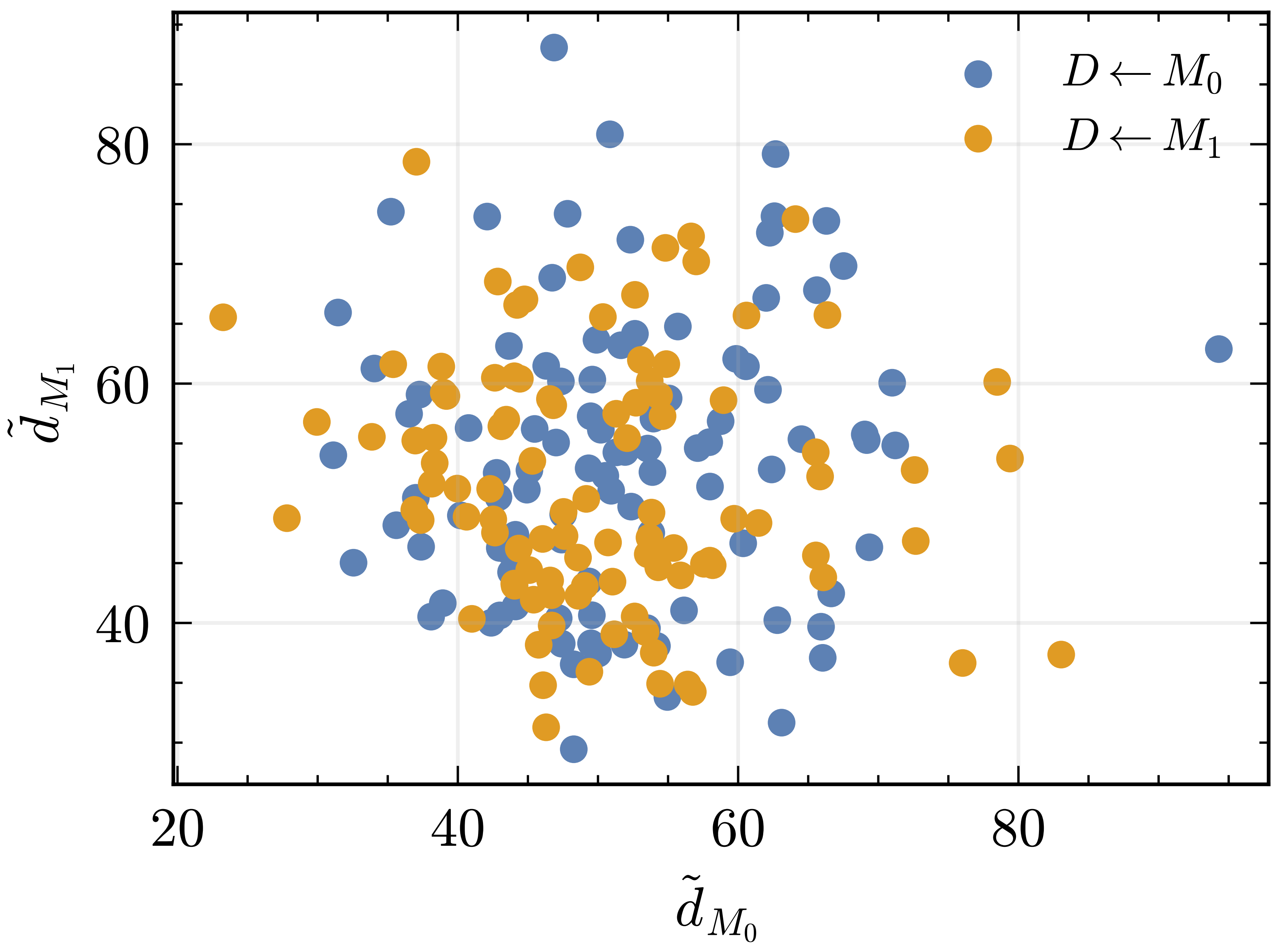


$$A^0_{ij} = \cos(j - 1/2)x_i\,,$$
$$A^1_{ij} = A^0_{ij} \cdot R$$
$$R \leftarrow O(N, \epsilon)$$

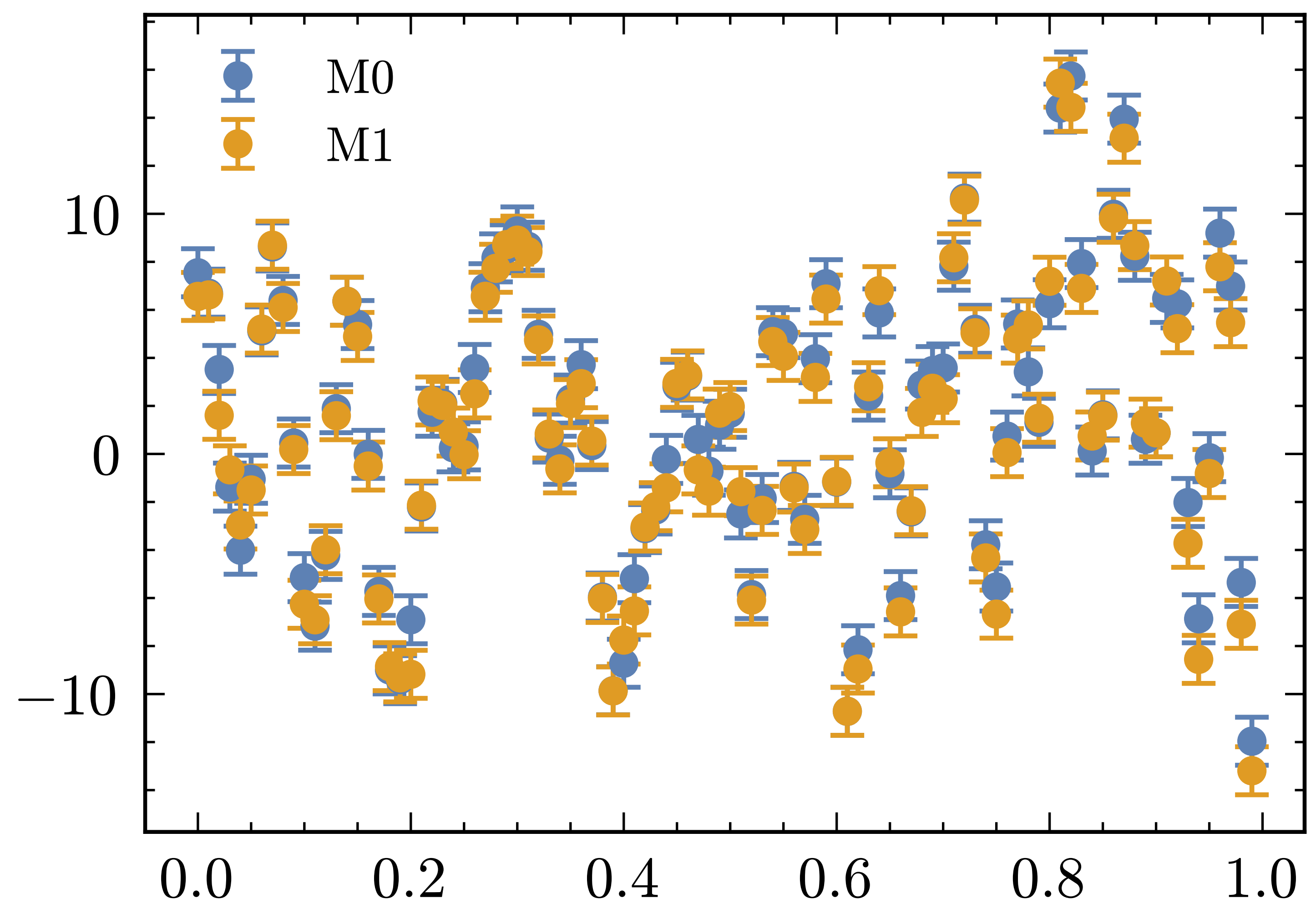


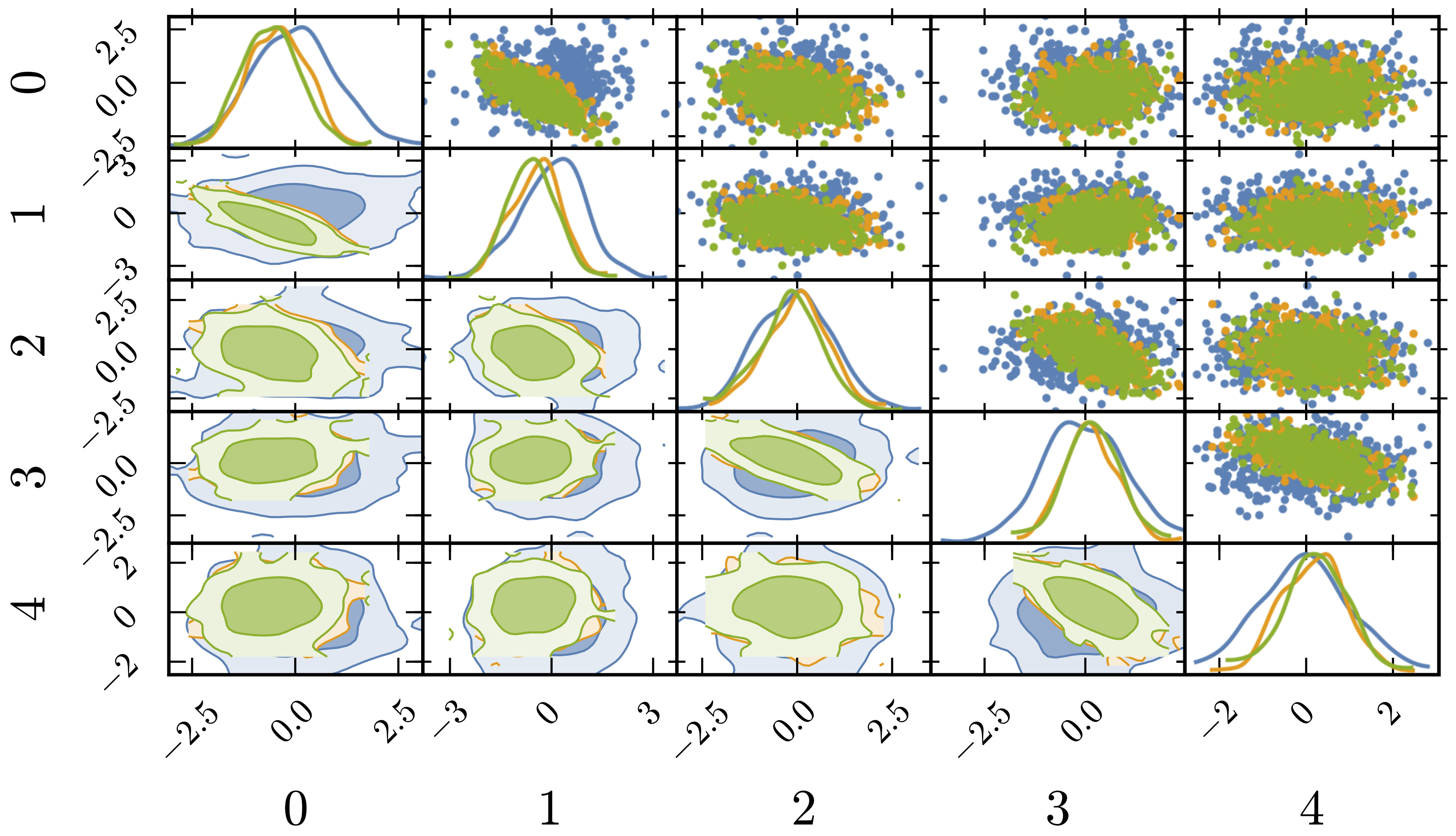






$$A^0_{ij} = \cos(j - 1/2)x_i\,,$$
$$A^1_{ij} = A^0_{ij} \cdot R$$
$$R \leftarrow O(N, \epsilon)$$
$$n_i \leftarrow \mathcal{N}(0,1)$$





How good do your simulations have to be?

- Model misspecification — *"All models are wrong some are useful"*, expect data we analyse will *never* be drawn from the model we are proposing
    - How well can neural techniques extrapolate

- Relevance of simulation data to observed data
    - A lot of the really juicy parts of SBI are gaining from being apparently vastly more economical here, amortization over all data?
    - **How good do your {simulations} have to be?**

Assorted thoughts

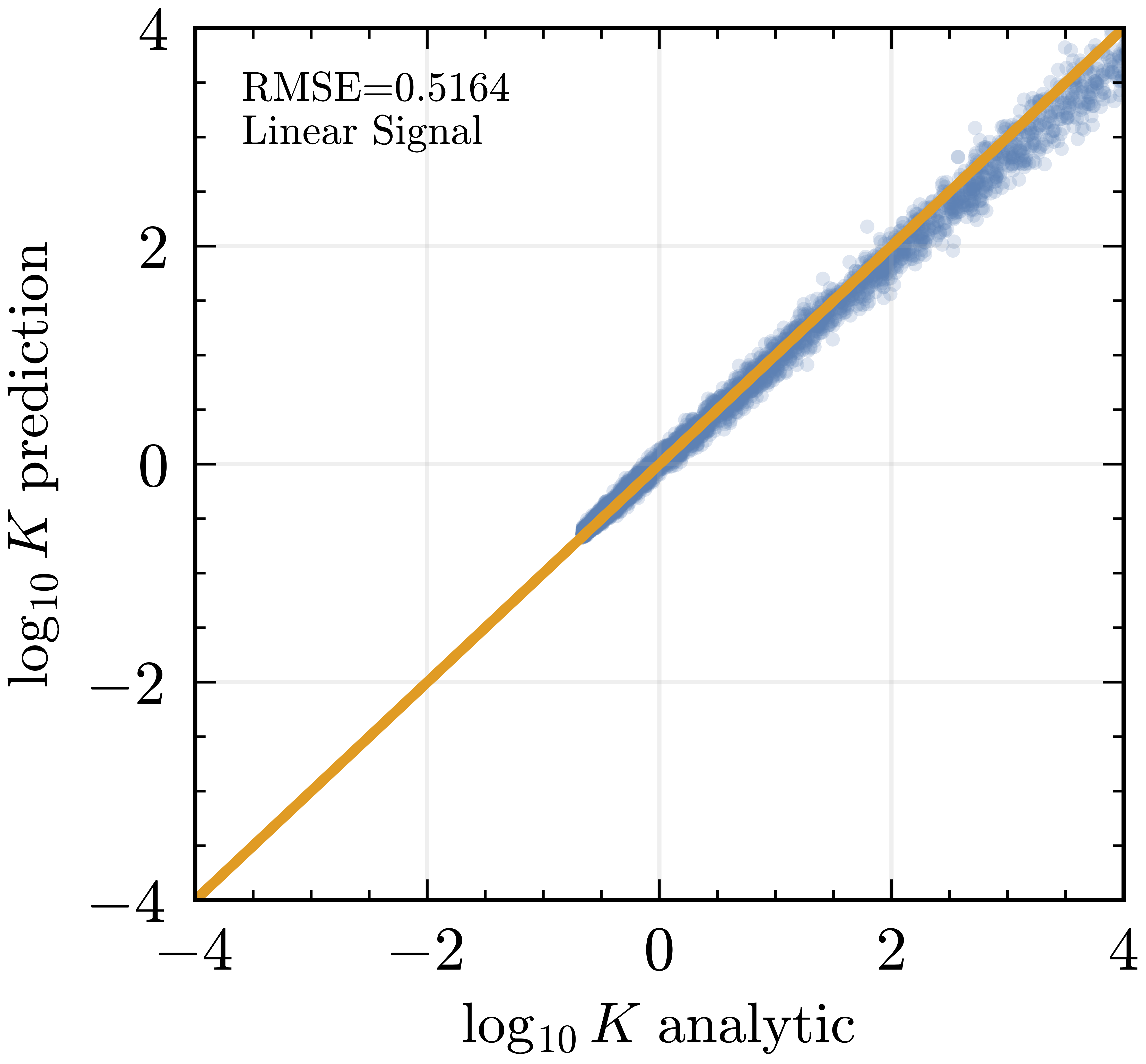
- How different is using a GP acquisition to get a new set of sims for my amortized SBI pipeline from using a GP surrogate in parameter space?
- What regime are we in: $$\mathcal{O}(D) \ll \mathcal{O}(\theta),\quad \mathcal{O}(D) = \mathcal{O}(\theta),\quad \mathcal{O}(D) \gg \mathcal{O}(\theta)$$
- It's very hard to think of something analytic, non-linear, representative of real problems! Writing good test problems is a lot harder than it seems

Conclusions

- Ratios are interesting, potential cancellation that simplifies problems
- How we [sequential] or [amortize] + [active learning??] is for my money the open problem
- The above can be mapped onto traditional methods?
- SBI is here to stay

If your experiment needs statistics, you ought to have done a better experiment.

Ernest Rutherford



$$A_{i0} = 2x_i\,,$$

$$A_{ij} = \sin\frac{\pi}{10}(j+1/2)x_i\,,$$

$$n_i \leftarrow \mathcal{N}(0,1)\,,$$

$$\theta_i \leftarrow \mathcal{N}(0,1)\,.$$

$$D_i = A_{ij}\theta_j + n_i$$