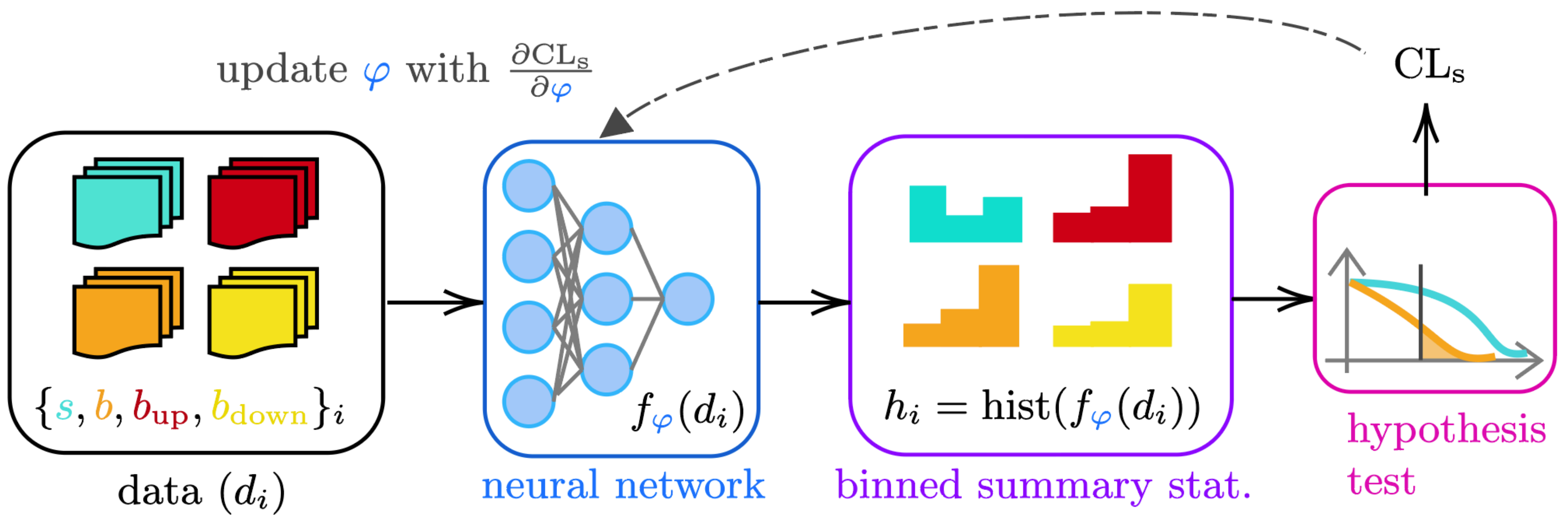
What could go wrong?

Frederic Renner, 05.03.2026

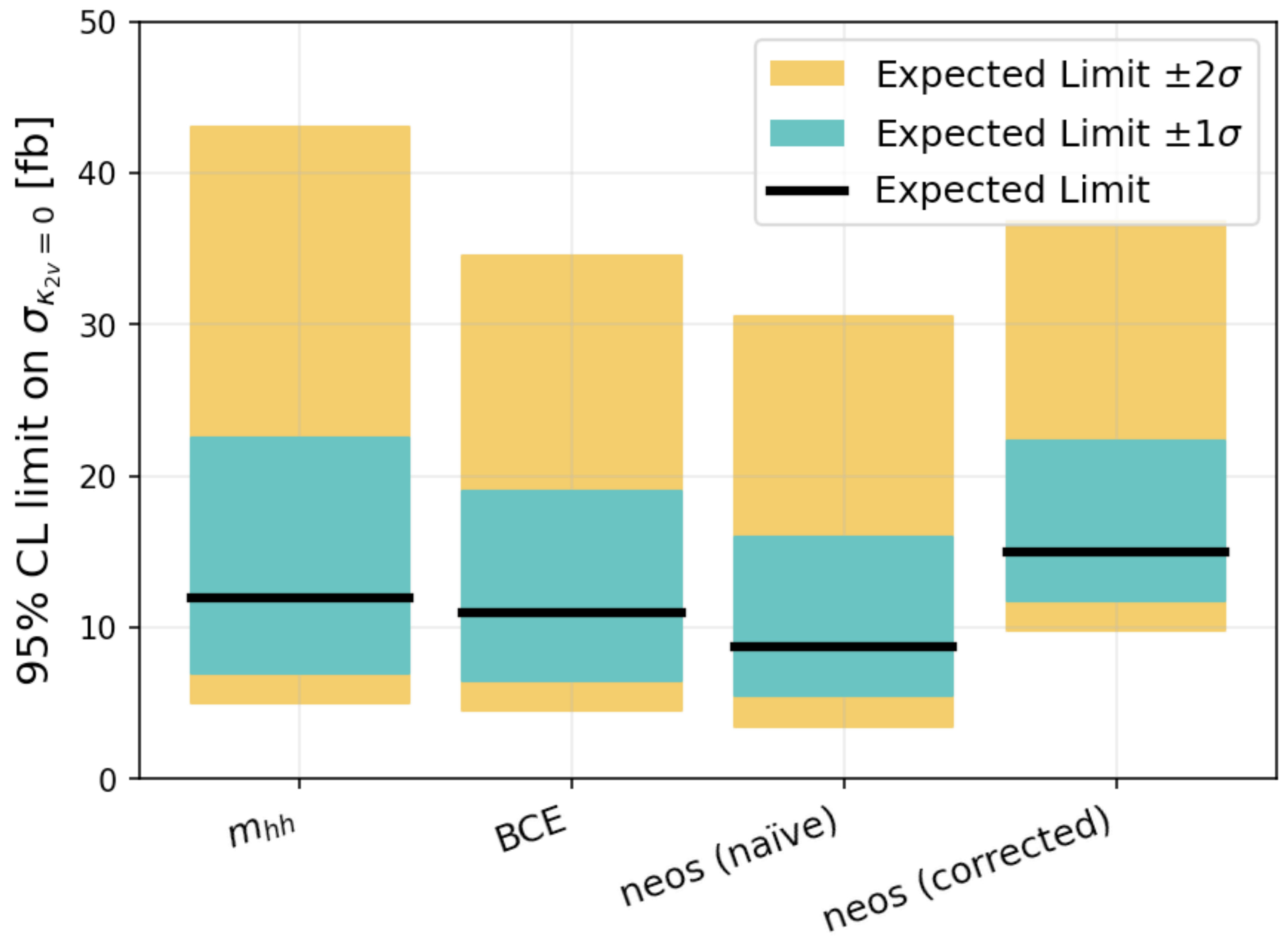
Setup

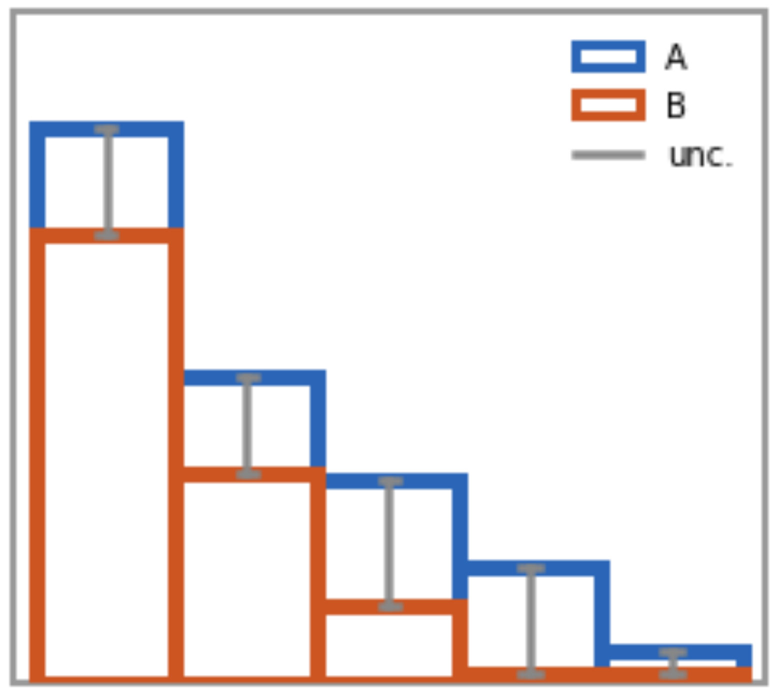


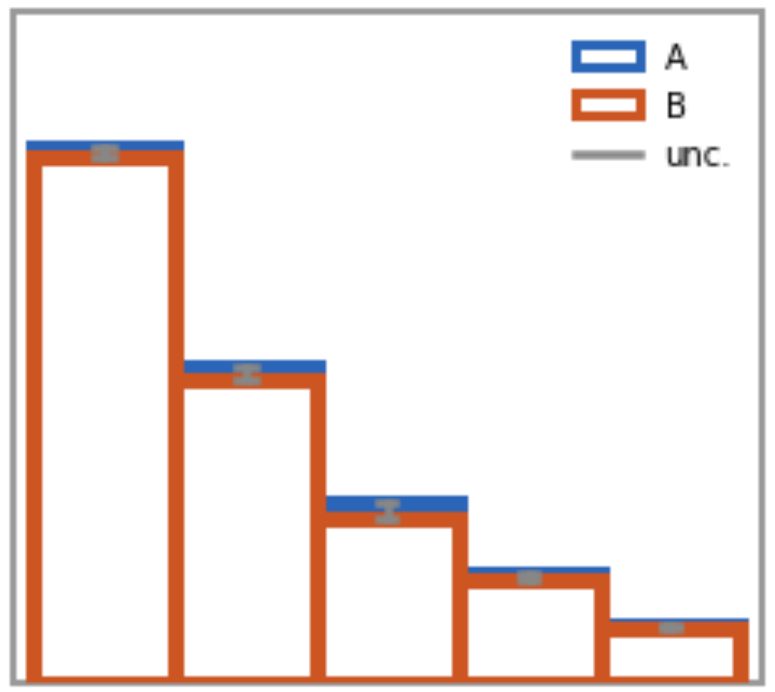
- 1.5 years crunch time
- no agents
- no idea of jax
- Based on NEOS (autodiff hypothesis test)
- Thesis: Automated Optimization of an ATLAS Search for Higgs Boson Pair Production at the LHC
- showcase optimizations: NN, kinematic variable, cuts, bins
- Framework: TOMATOS - auTOMATed Optimization of Sensitivity
- Goal: working example

First Attempt

- translate search into NEOS
- input $\mathscr{D}$ —> NEOS($\mathscr{D}, \varphi$) —> p-value
- often in fit config
    - measurements from control regions:
    - unc = hist_A/hist_B
    - You want this unbiased
- first lesson:
- —> anything will be exploited
- —> separate clearly what you want to be optimized and what not







NEOS —>

Framework talk

- Debugging in jax general hard
- pyhf fix: error points to NEOS but actually phyf
- Out of Memory: aggressive manual jax JIT cache clearing
- How to preprocess?
    - Min-Max scaling, train/test splitting, k-folding, resampling
    - Reframe problem: input —> ML box —> output
    - change input —> counteract with scale factors; e.g. inputs/2 —> histogram*2 —>p-value
- Need extensive diagnostics

My JAX pitfalls

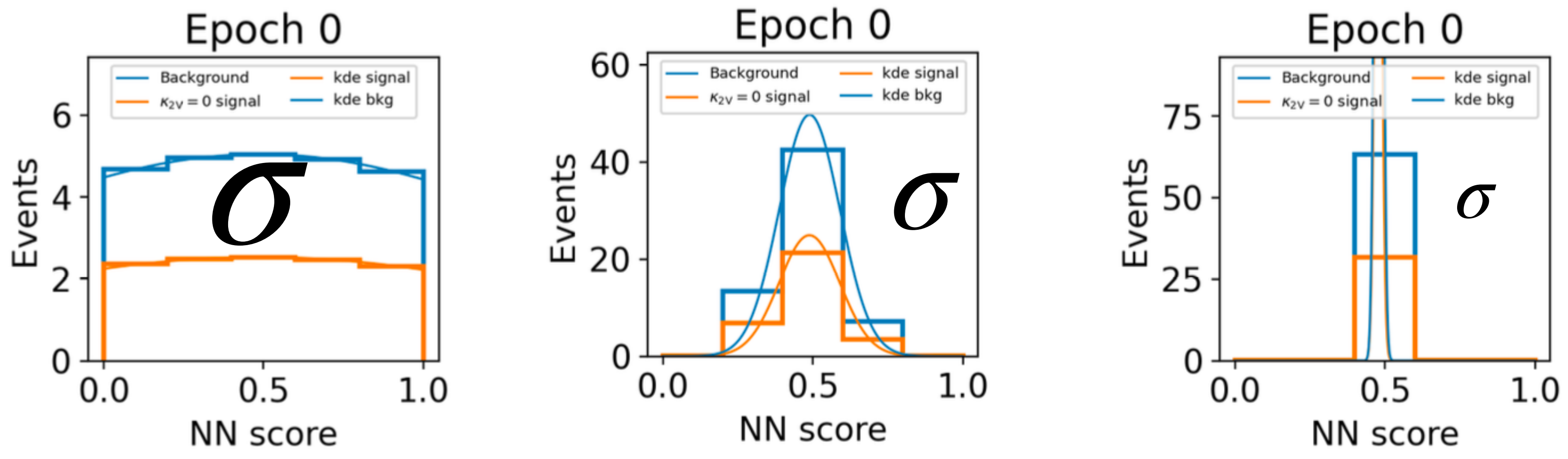
(even though i read the jax docs 100 times)

- array.at[i].set(val) is NOT in-place! updated array is returned as a new array:
  - updated_array = array.at[i].set(val)
- NaNs are your worst enemy, one NaN everything NaN
  - need many protections: flooring histograms +1e-3, zero divison, …
- optimization parameters are just parameters, bin edges can cross, go out of bounds
  - bins = [0,opt_pars[“bins”],1]
  - opt_pars["bins"] = jnp.clip(jnp.sort(jnp.abs(opt_pars[“bins"])), 1e-6, 1 - 1e-6)

Guardrail gradient descent

- opt. parameters
  - min-max scaled
  - limit gradient updates
- Crucial: Dynamic bandwidth σ:
  - start high, finalize low
  - penalize for out of boundary conditions, e.g. N_min per bin
  - Full Exploration of possibilities
  - Finalize with high accuracy histogram approx.

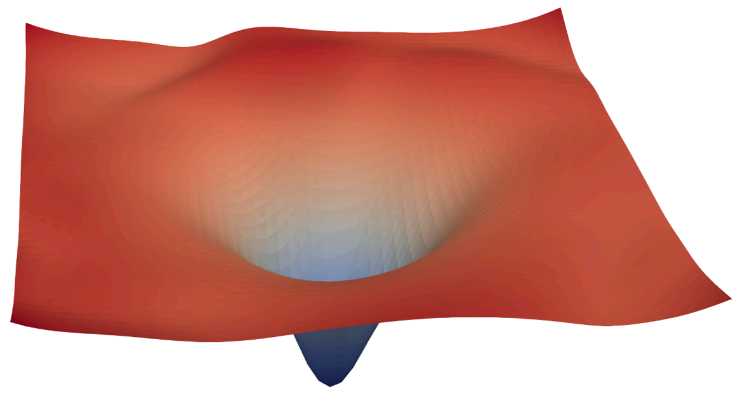
Initial setting for



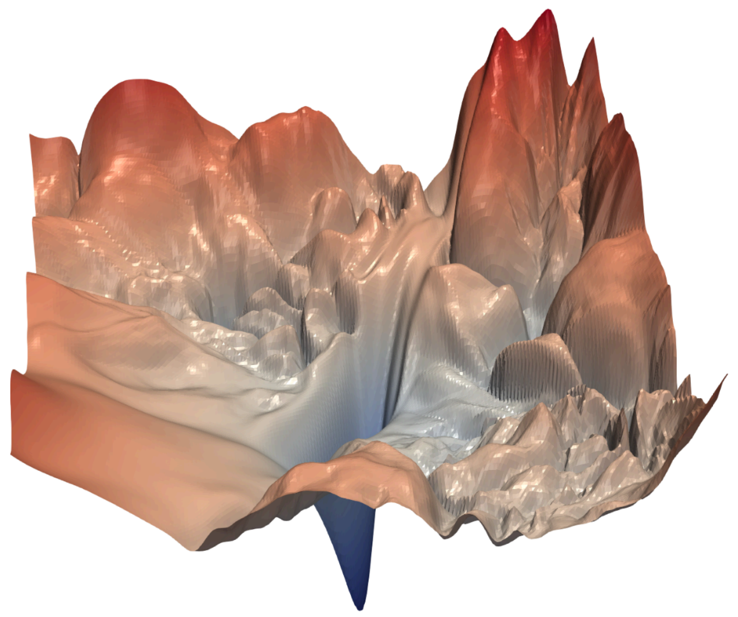
During training

if $(N_{\mathrm{bkg}} < N_{\mathrm{min}})$

start σ → end σ

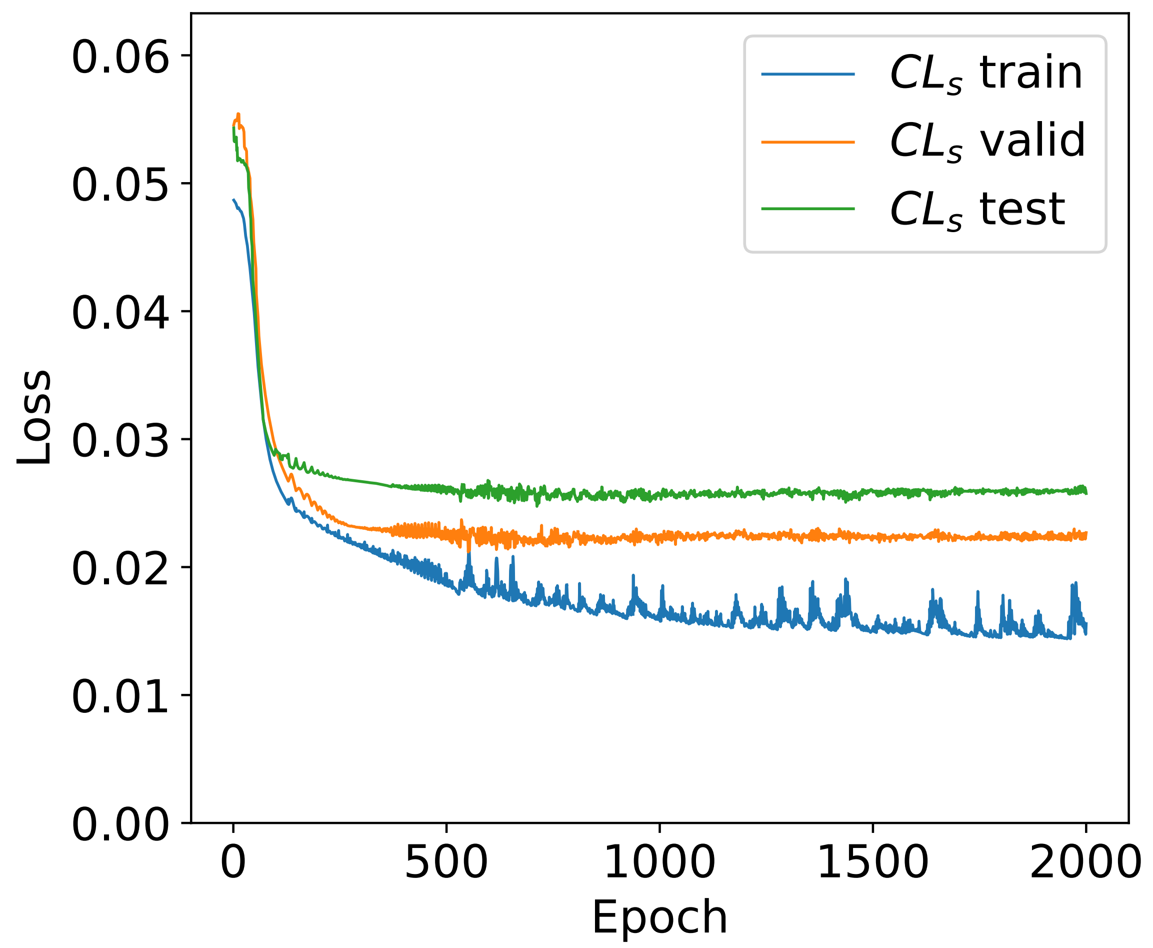


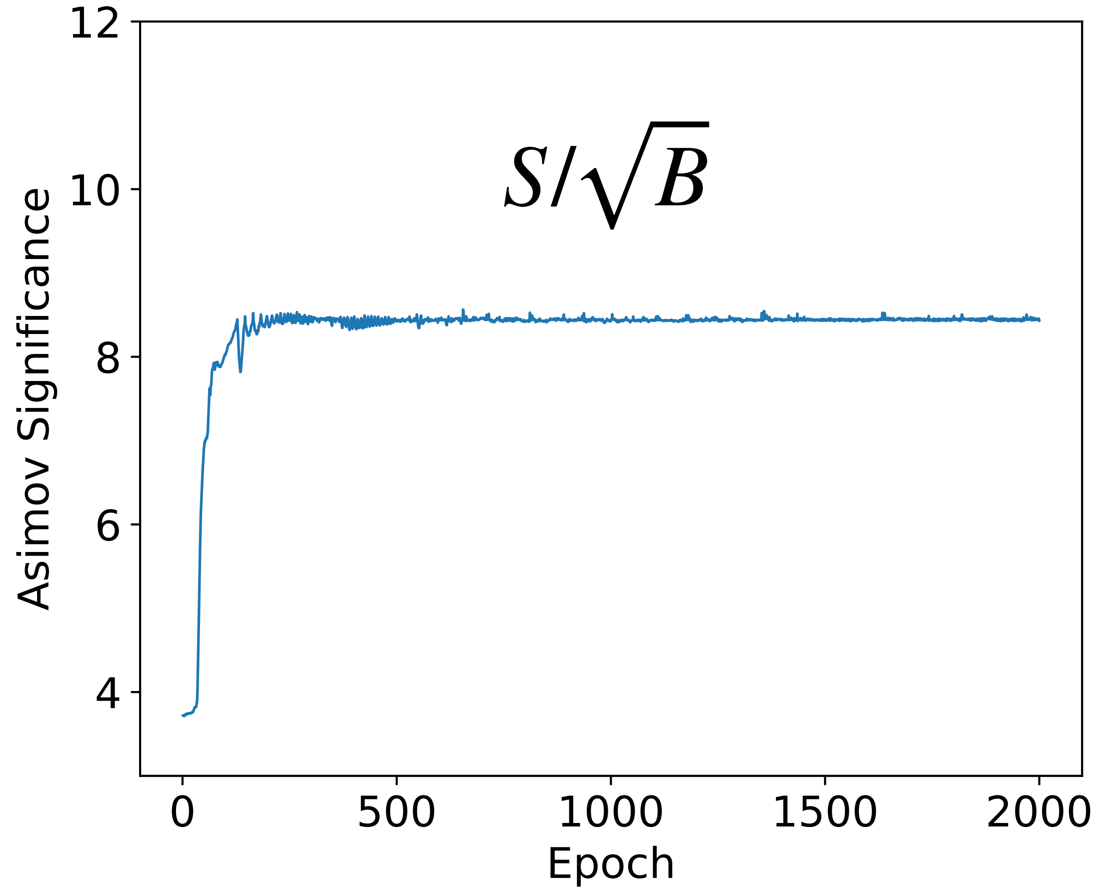
$\mathrm{CL_s}(\varphi_1, \varphi_2)$

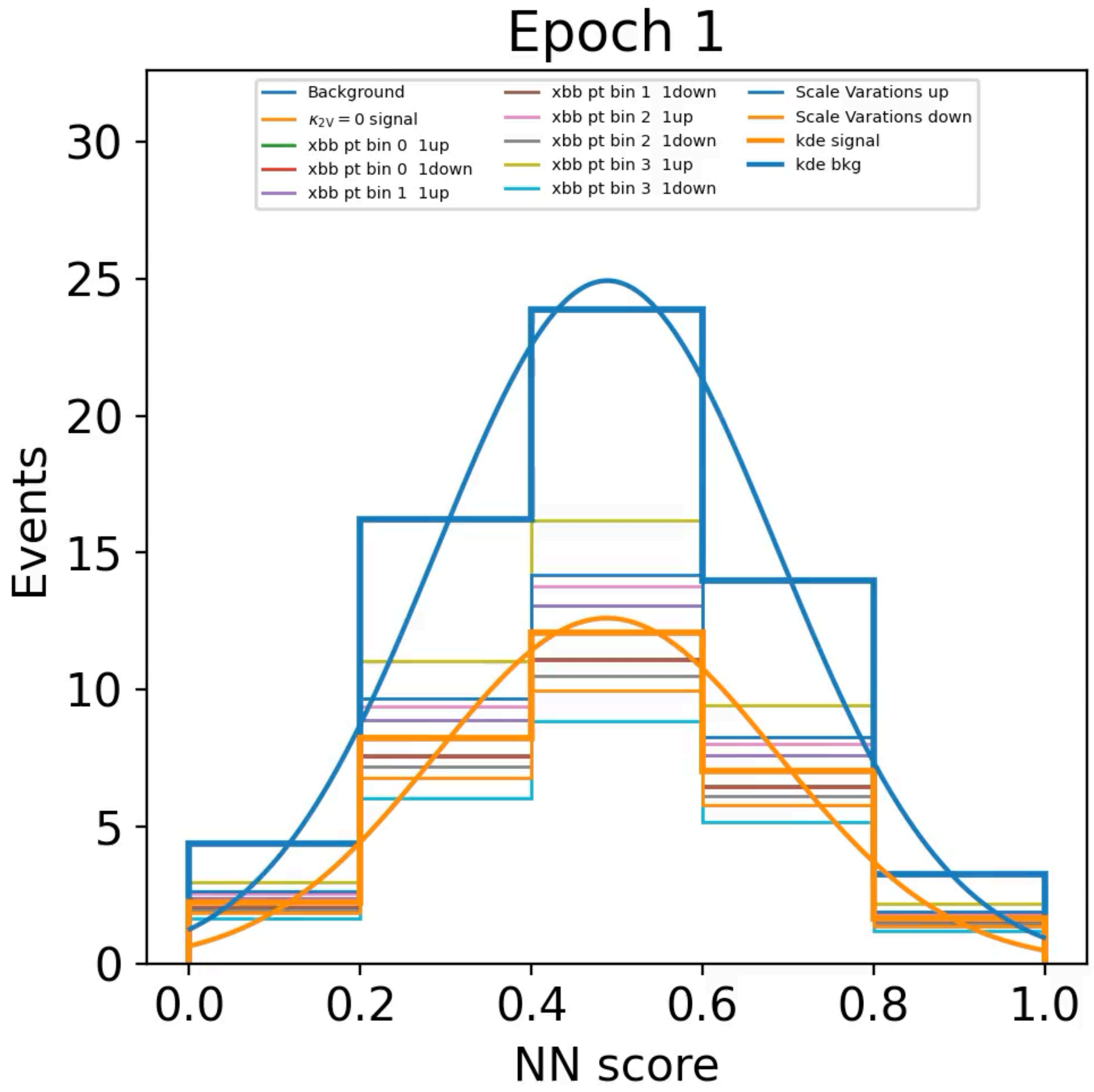


Automated NN Optimization

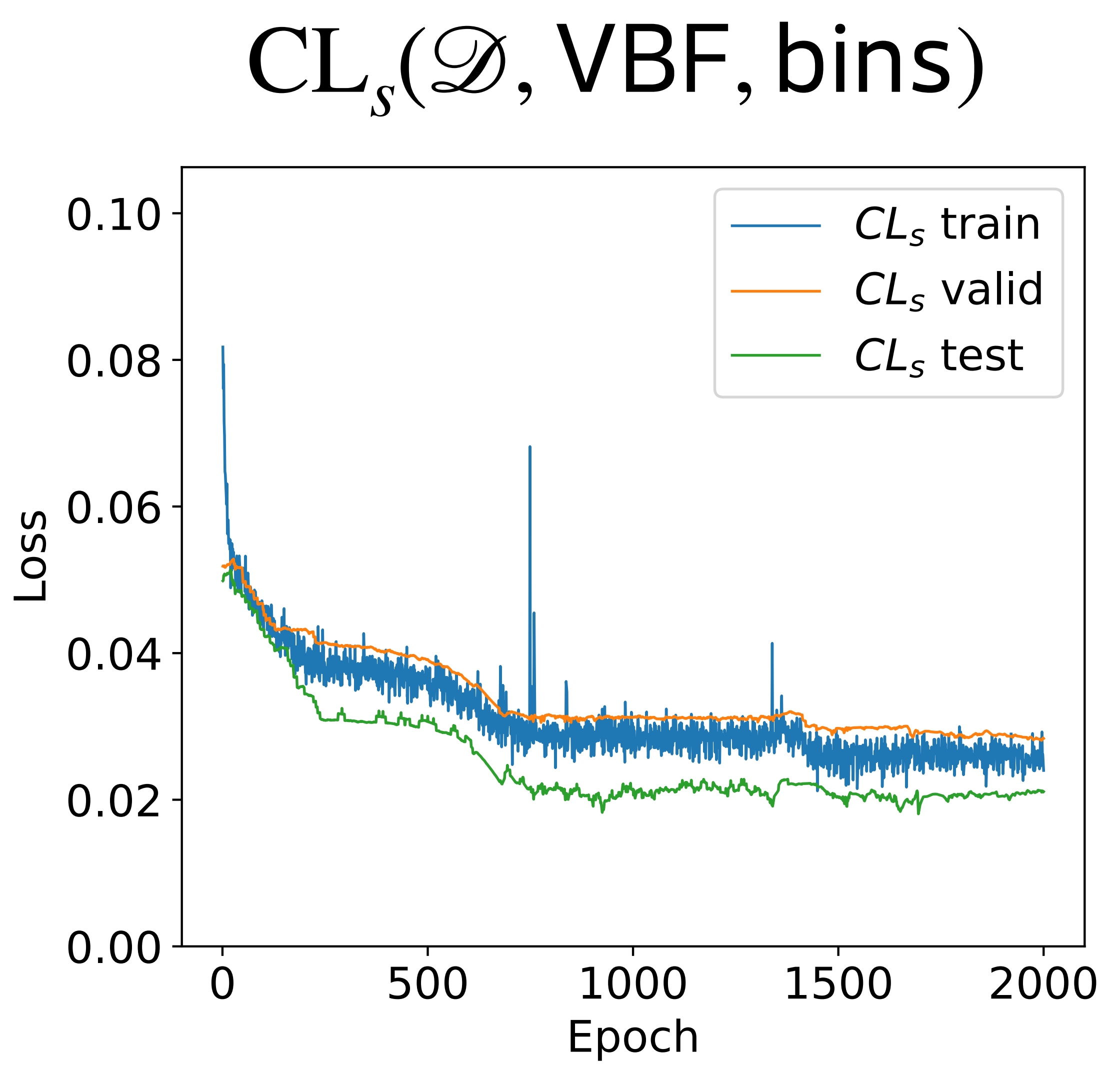
$\mathrm{CL}_s(\mathscr{D}, \mathrm{VBF}, \mathrm{NN})$

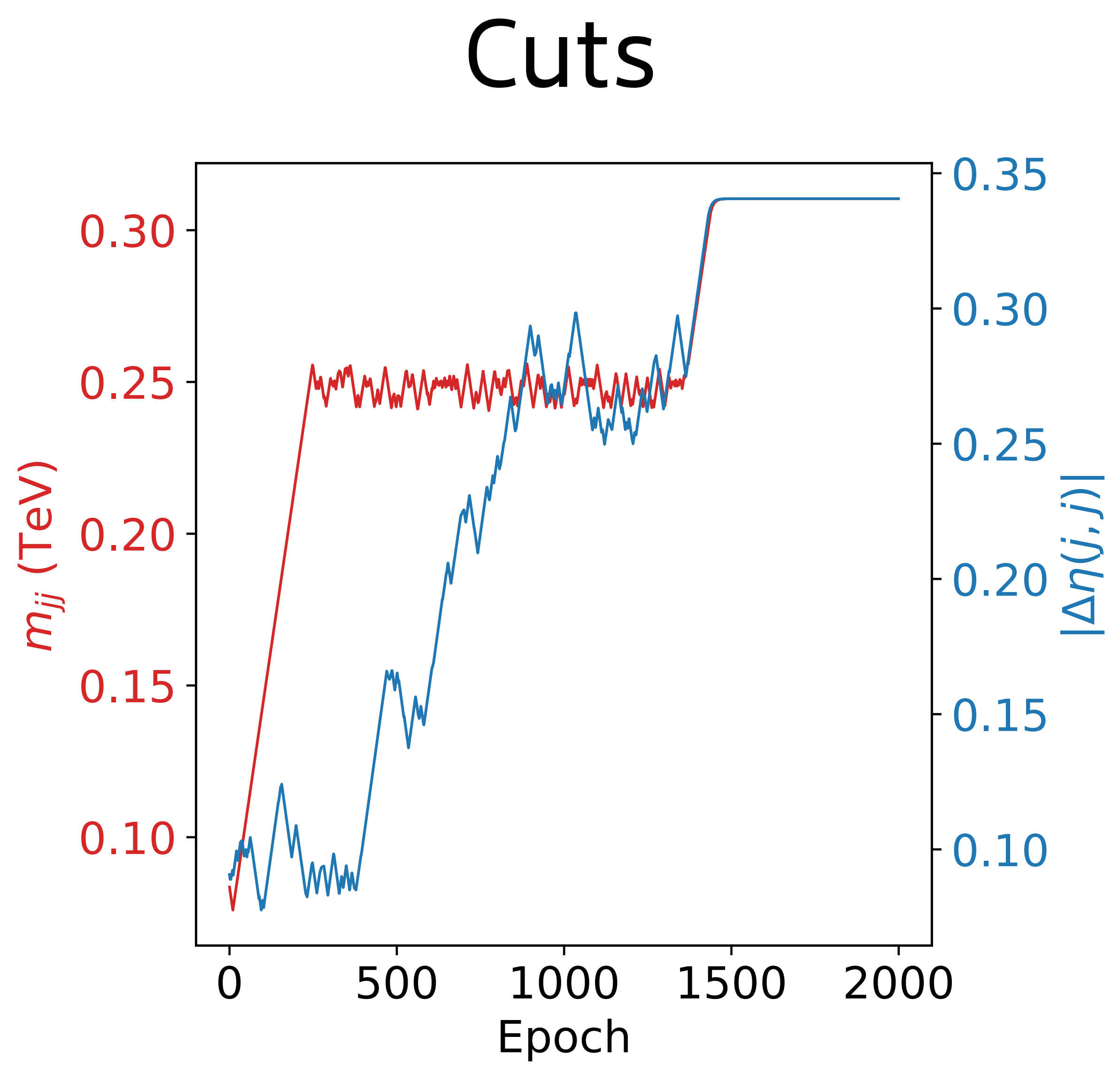


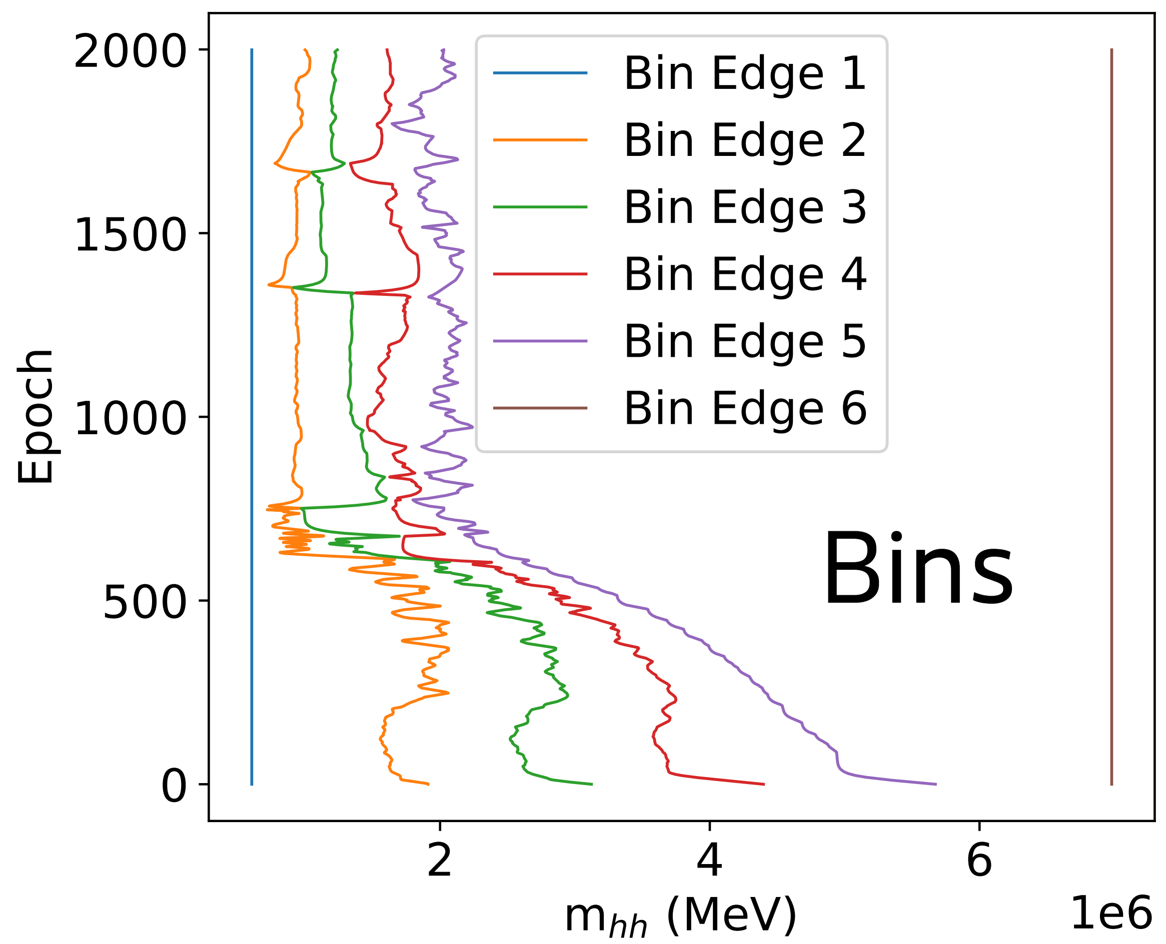


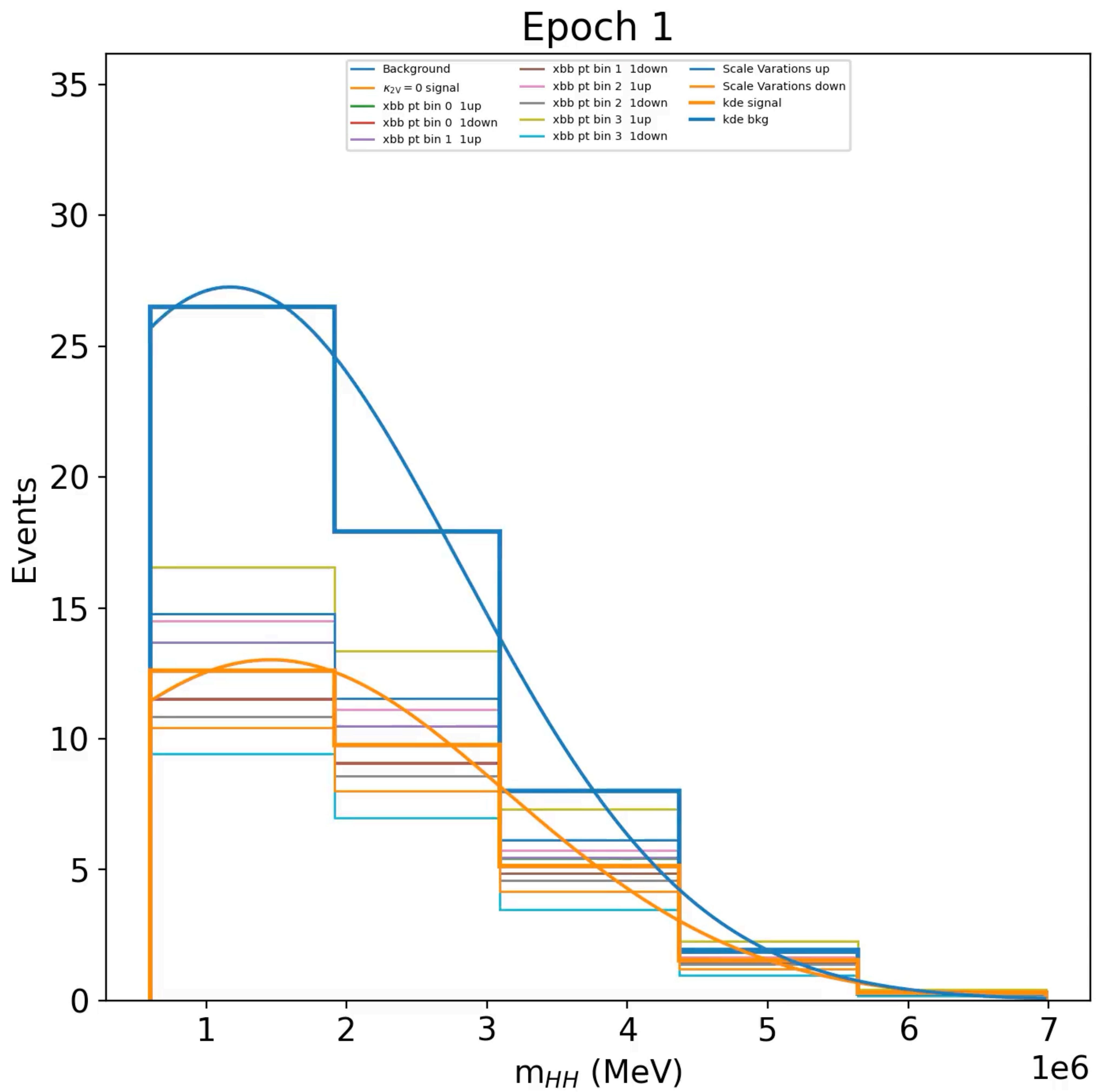


Automated m_{HH} Optimization









Outlook

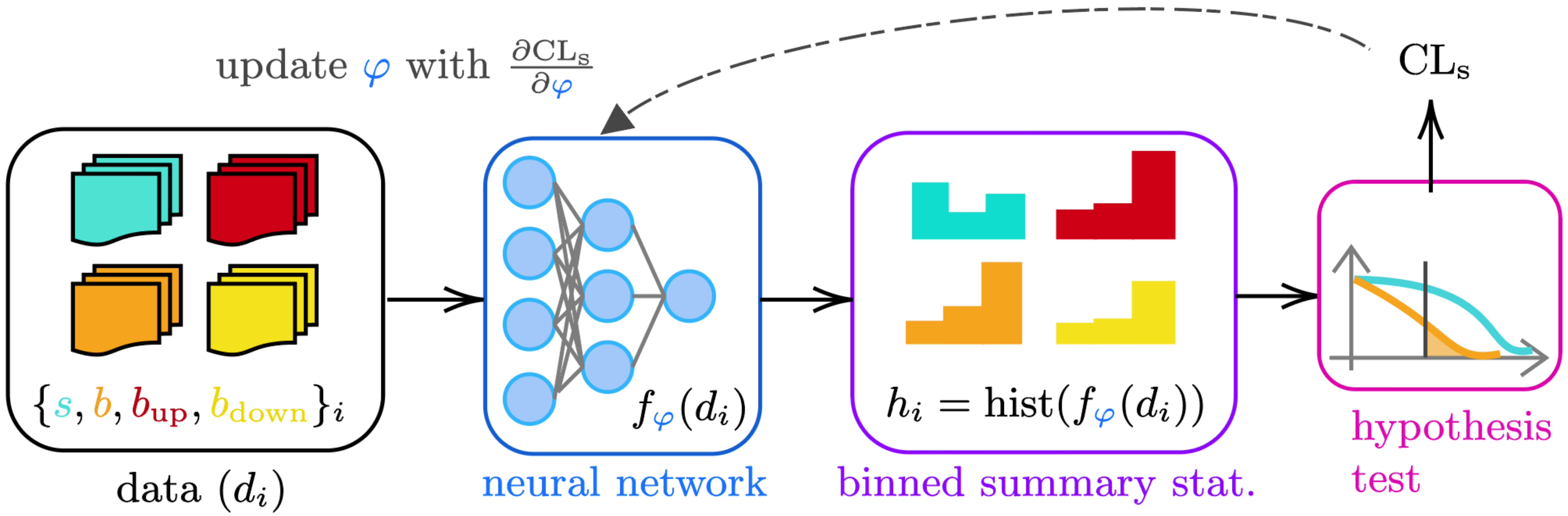
- Auto-optimized not necessarily better —> thesis
- Largest issue is adoption
    - learning jax
    - consider many things simultaneously
- by far not everything that can go wrong :)

Backup

NEOS: neural end-to-end-optimized summary statistics

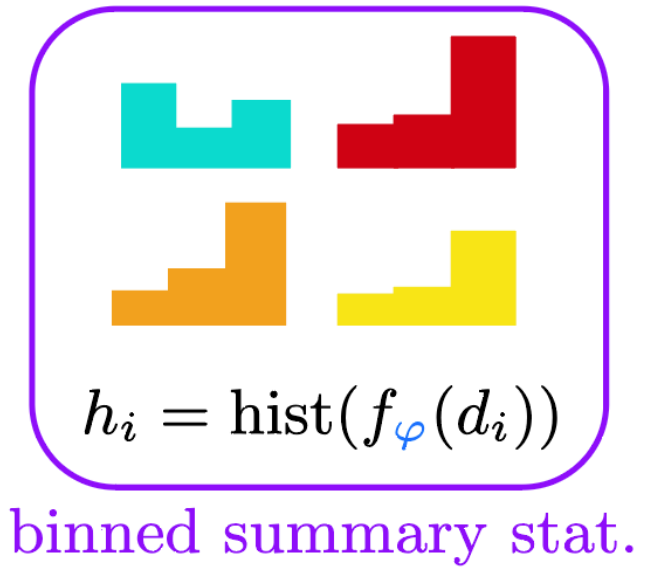
N. Simpson, L. Heinrich
arXiv:2203.05570

- $$\frac{\partial CL_s}{\partial \varphi} = \frac{\partial f_{sensitivity}}{\partial f_{teststat}} \times \ldots \times \frac{\partial f_{observable}}{\partial \varphi}$$
- How? —> Automatic Differentiation (JAX)
- Challenges: rewrite hypo test, Histogram?
- But did not answer:
  - How does it scale?
  - Uncertainties by Differentiability?

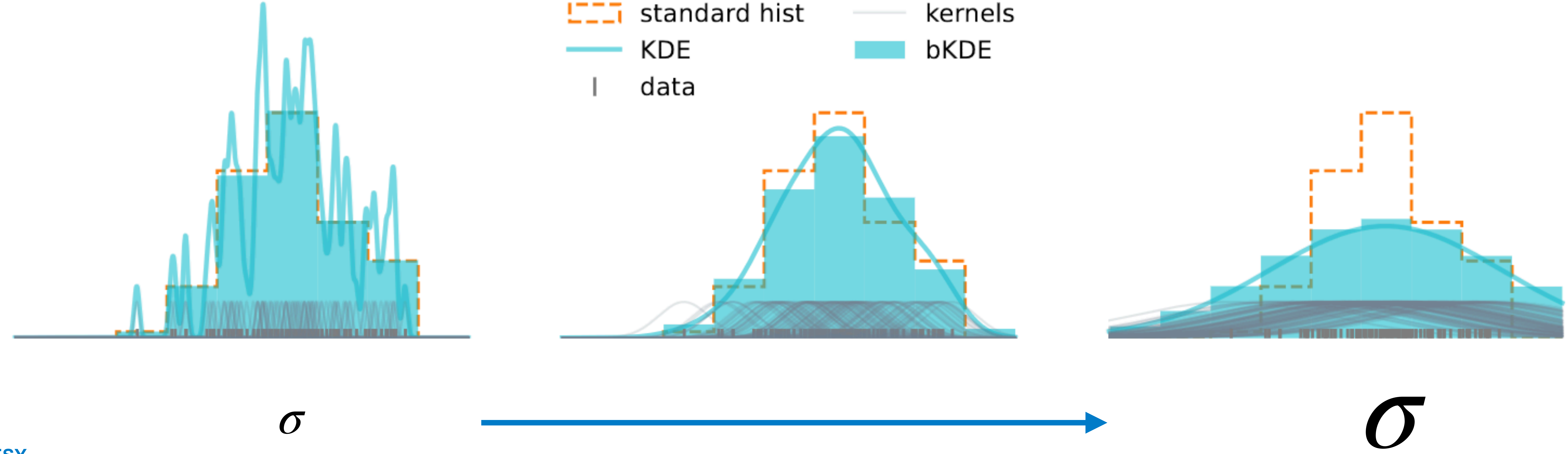


TOMATOS - auTOMATed Optimization of Sensitivity

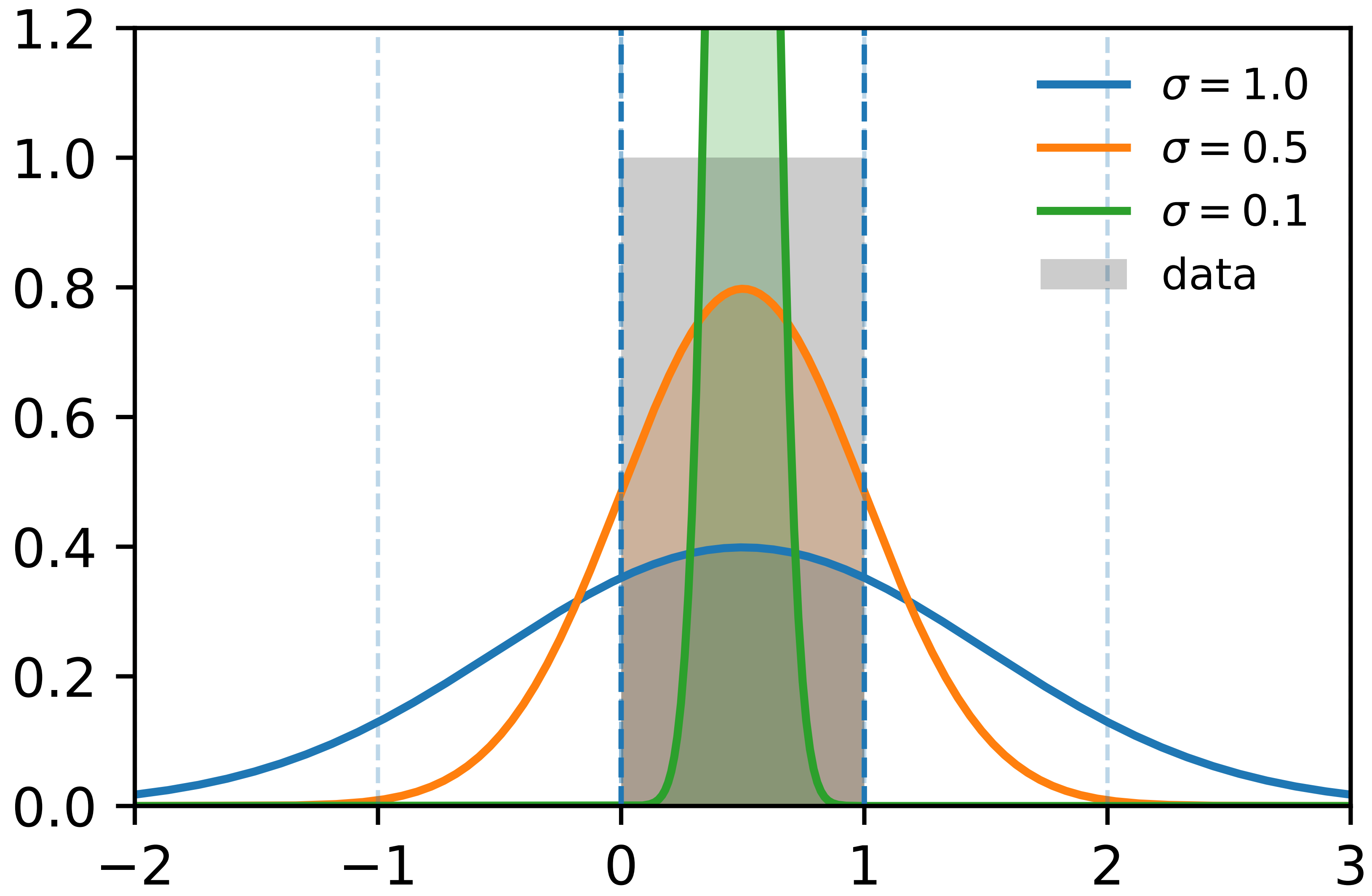
Differentiable Histograms



- Histogram: approximate datapoints with Gaussians
- $$\text{KDE}(\mathbf{x}, \sigma) = \sum_i \text{Gauss}(x', \mu = x_i, \sigma = \sigma)$$
- σ: Tradeoff Gradient vs Accuracy
- Whats the right σ?



Dynamic Bandwidth σ

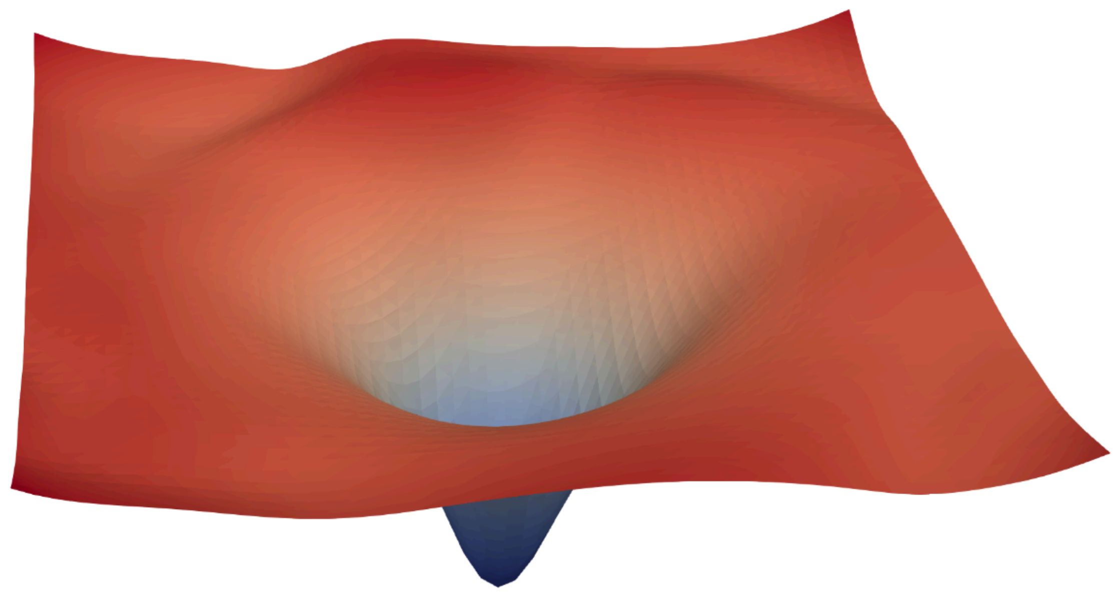


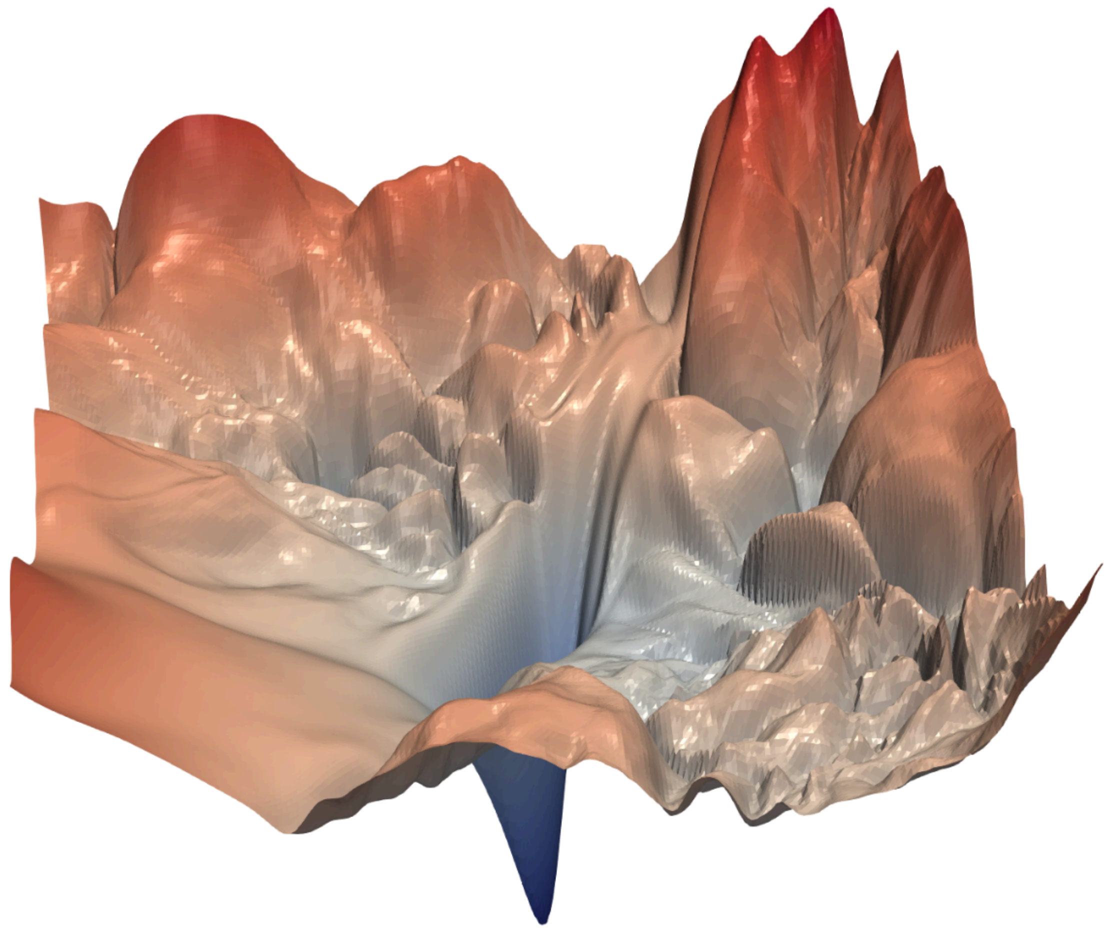
- Crucial:
- Full Exploration of Histograms
- Full Automation (Regularization)
- Finalize with high accuracy: <1% deviation per bin

if $(N_{\text{bkg}} < N_{\text{min}})$

start σ → end σ

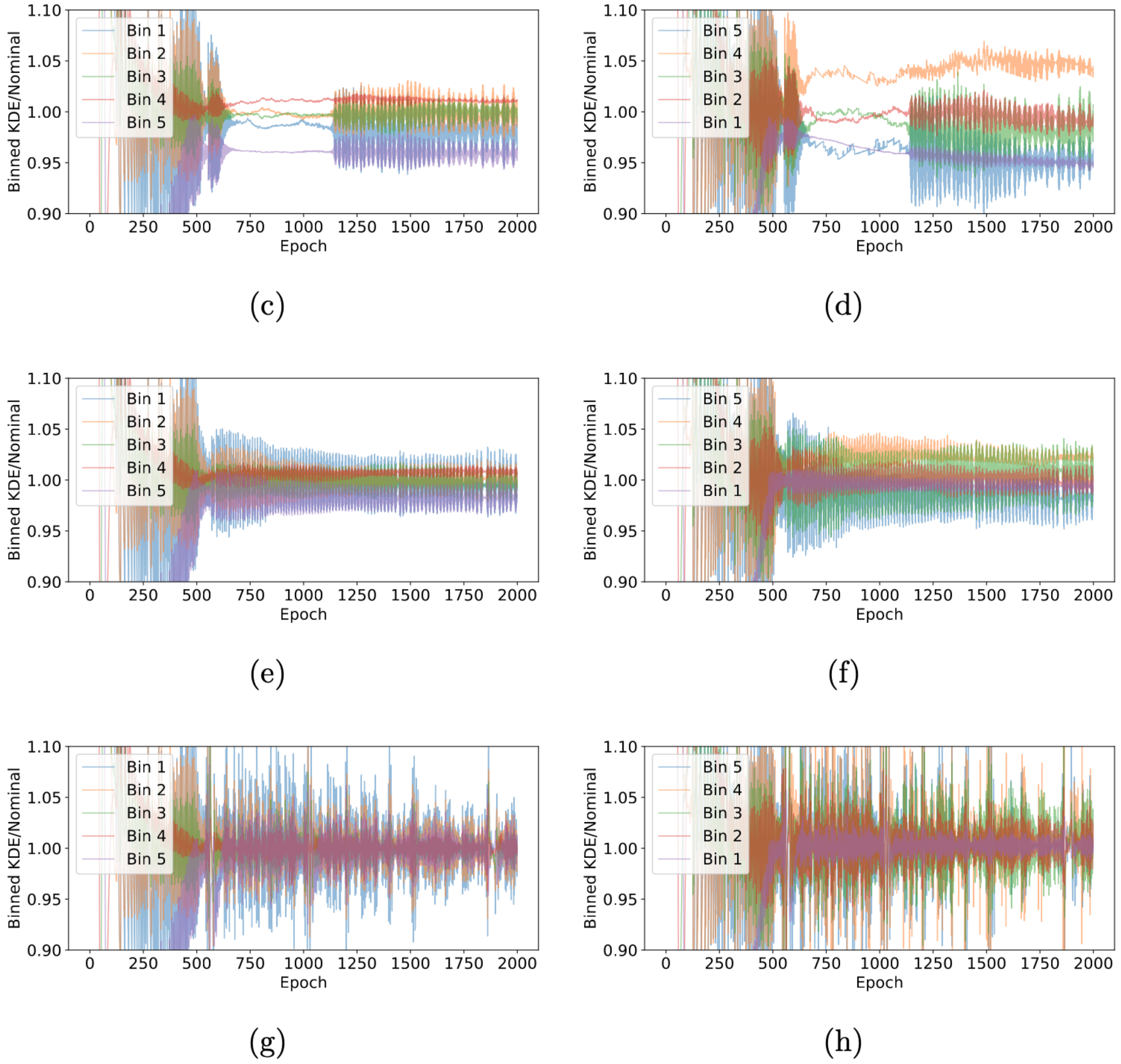
$\mathrm{CL_s}(\varphi_1, \varphi_2)$





Diffable methods accuracy

see thesis appendix



as illustrated in figure B.12. Thus for input features scaled between [0, 1], features are turned on and off over a cut value range of ~ 0.0003. This translates to the cuts here as $\Delta m_{jj} \approx 20\,\mathrm{GeV}$ and $\Delta|\Delta\eta(j,j)| \approx 0.003$. As a sufficient approximation $m = 20,000$ is assumed as a stable training value.

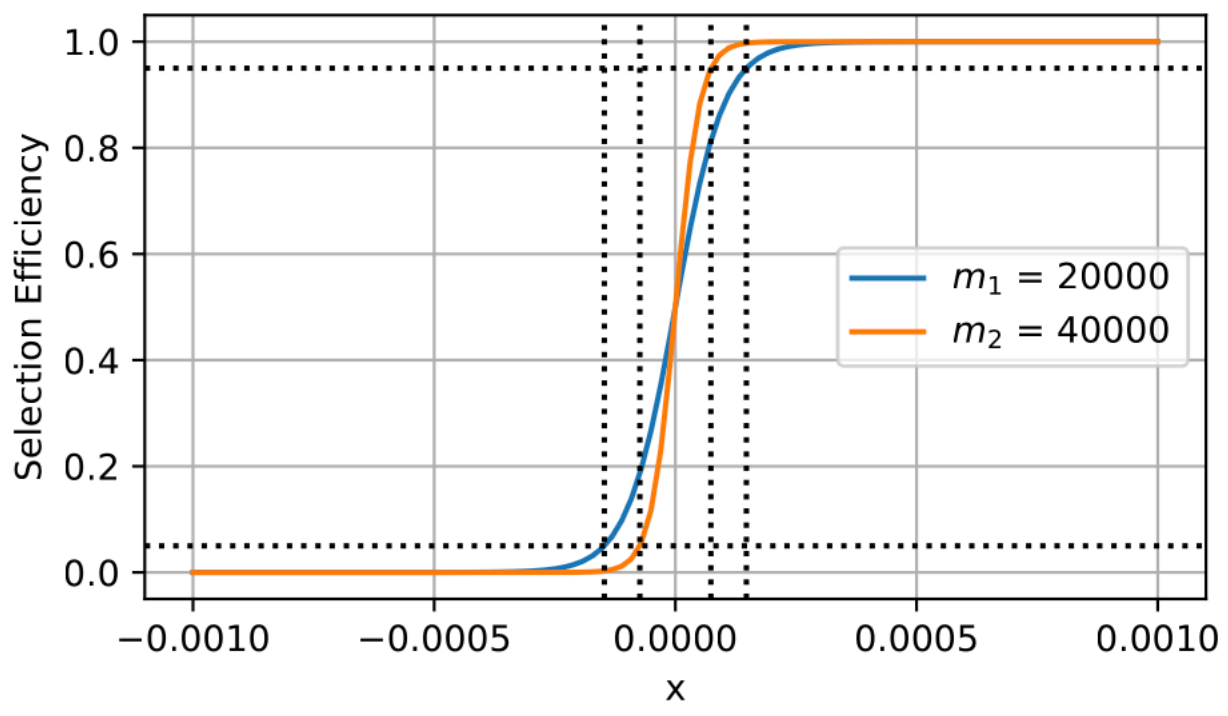
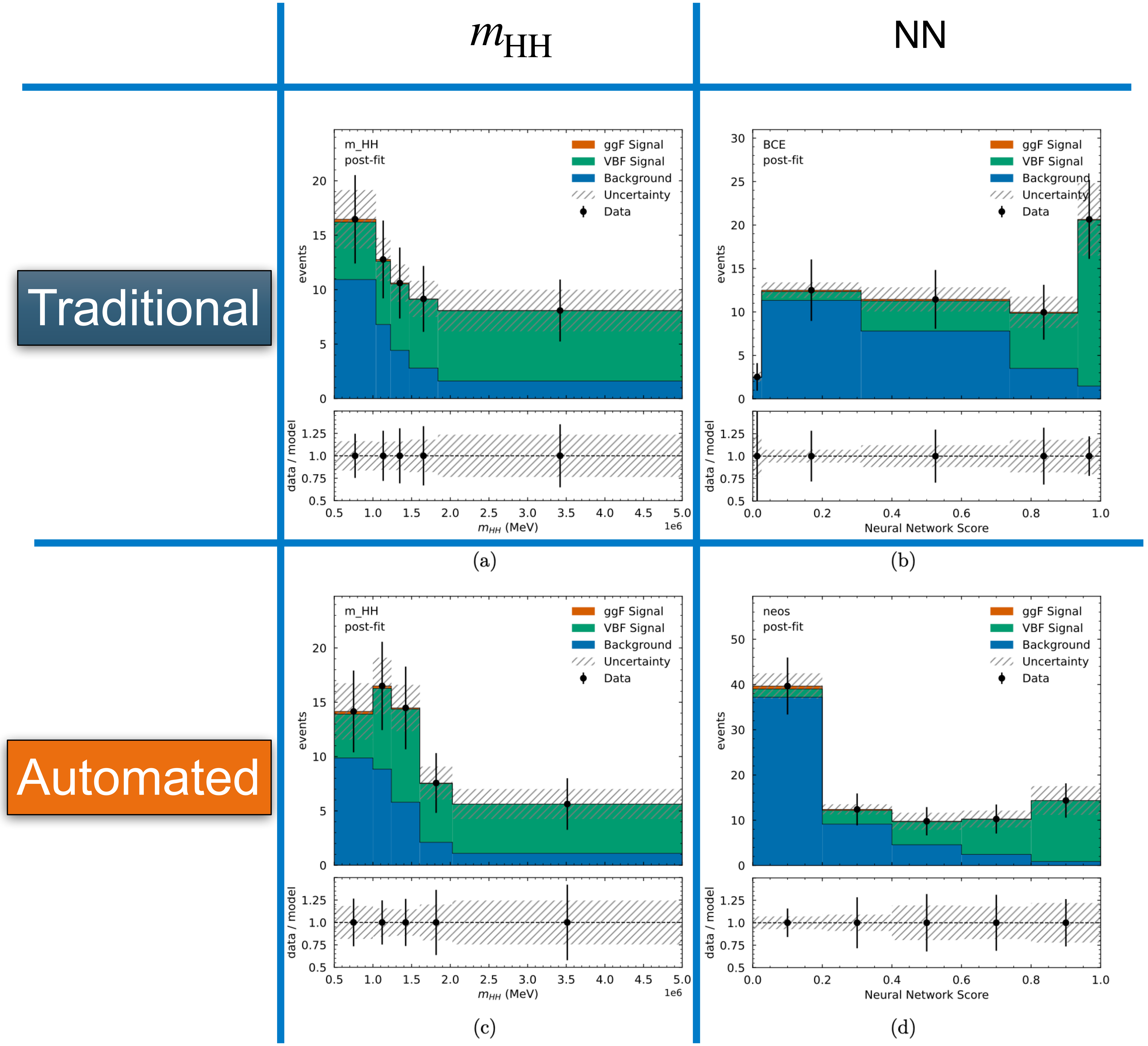
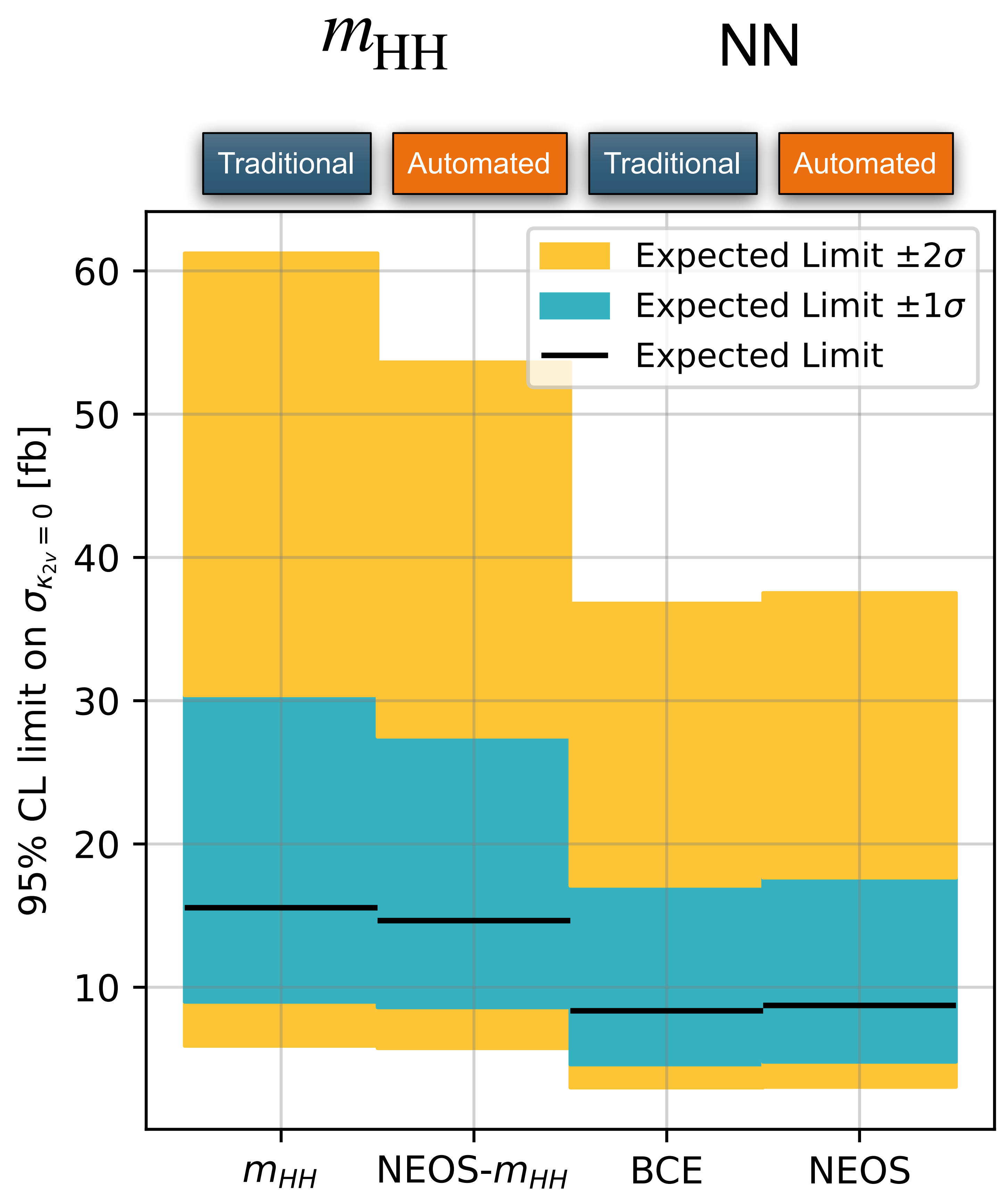


Figure B.12: Sigmoids for slopes 20000 and 40000 and guiding lines for 5% and 95% selection efficiency.

Performance Comparison





- Automated finds similar distributions + limits
- Automated Optimization feasible in real Analysis!

TOMATOS Preprocessing

- autoload data incl. uncertainties into predefined array (jax likes that)
  - $\mathscr{D}$ = (N_sample_systematics, N_events, N_variables)
- caveat: small dataset <1gb, 2 samples: signal, background, 10 uncertainties