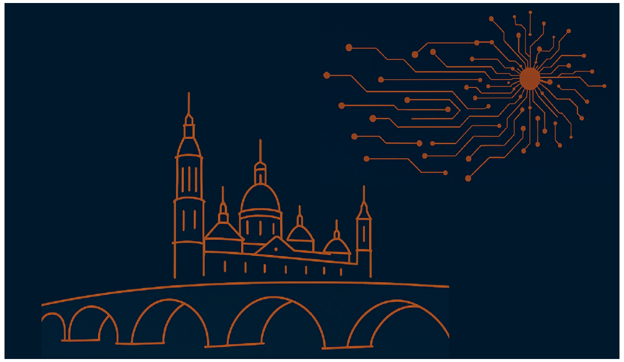


ML for HEP
Neural Simulation Based Inference

David Rousseau

IJCLab-Orsay

david.rousseau@in2p3.fr **@dhpmrou**

COMCHA, Zaragoza, April 2026

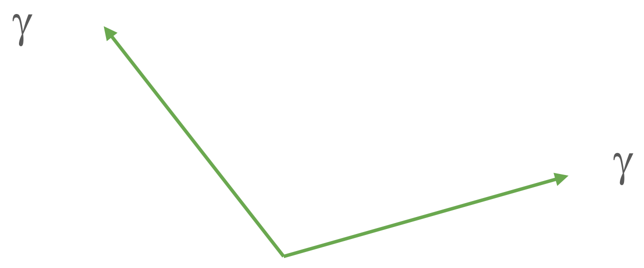
NUCLÉAIRE & PARTICULES

université PARIS-SACLAY

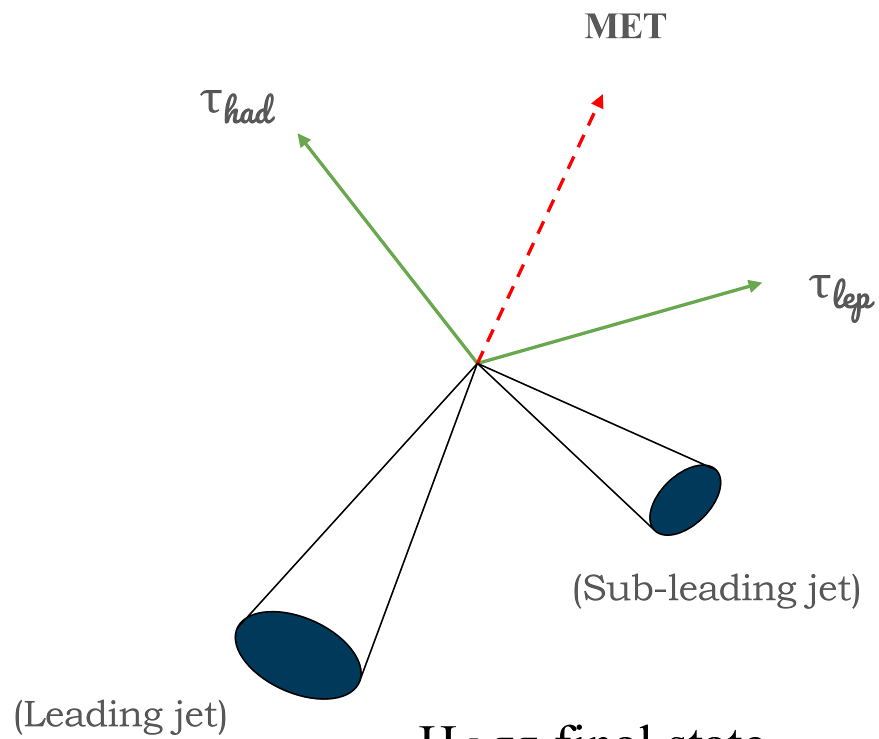
INSTITUT DATAIA Science des données, Intelligence & Société

IJCLab Irène Joliot-Curie Laboratoire de Physique des 2 Infinis

Introduction - ML in HEP



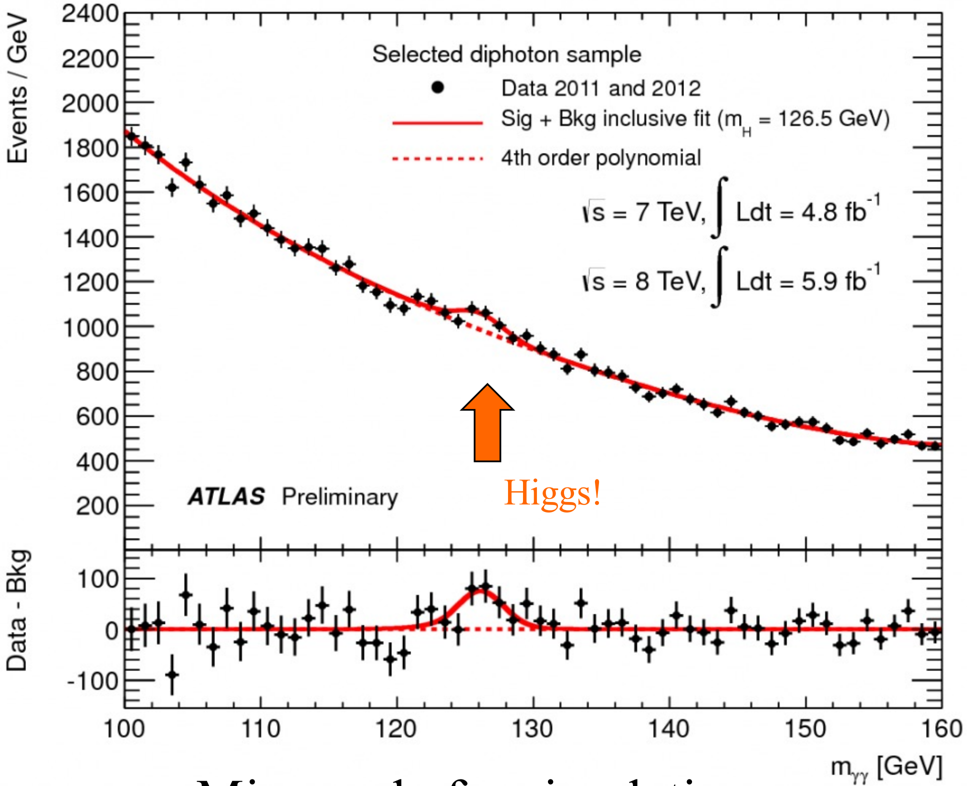
H→γγ final state



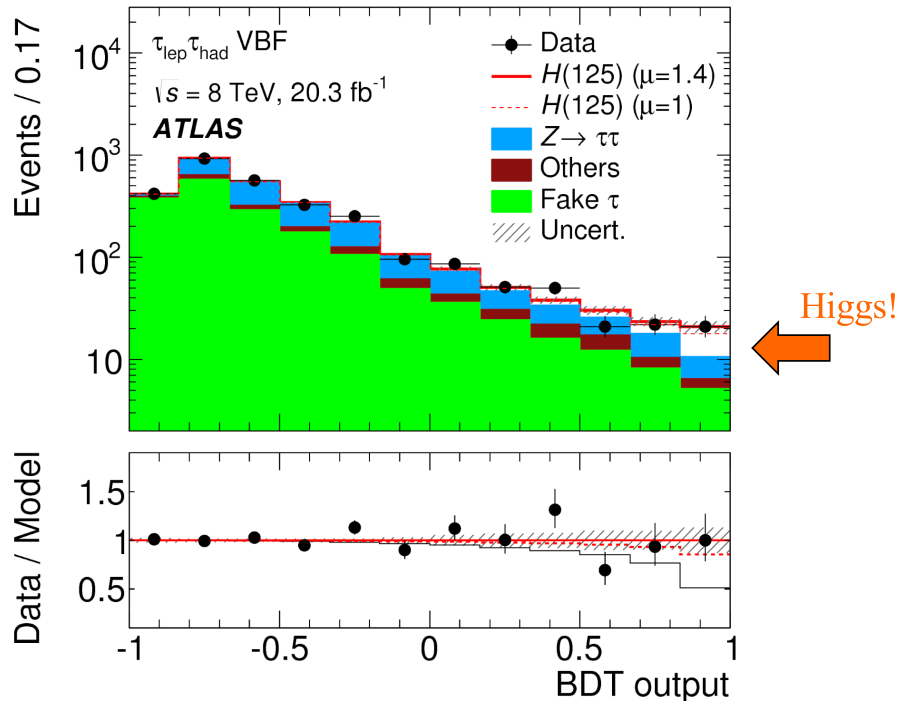
H→ττ final state

Features: 4-momenta of particles,
and derived features (invariance mass etc…)

Higgs boson evidence (classical)

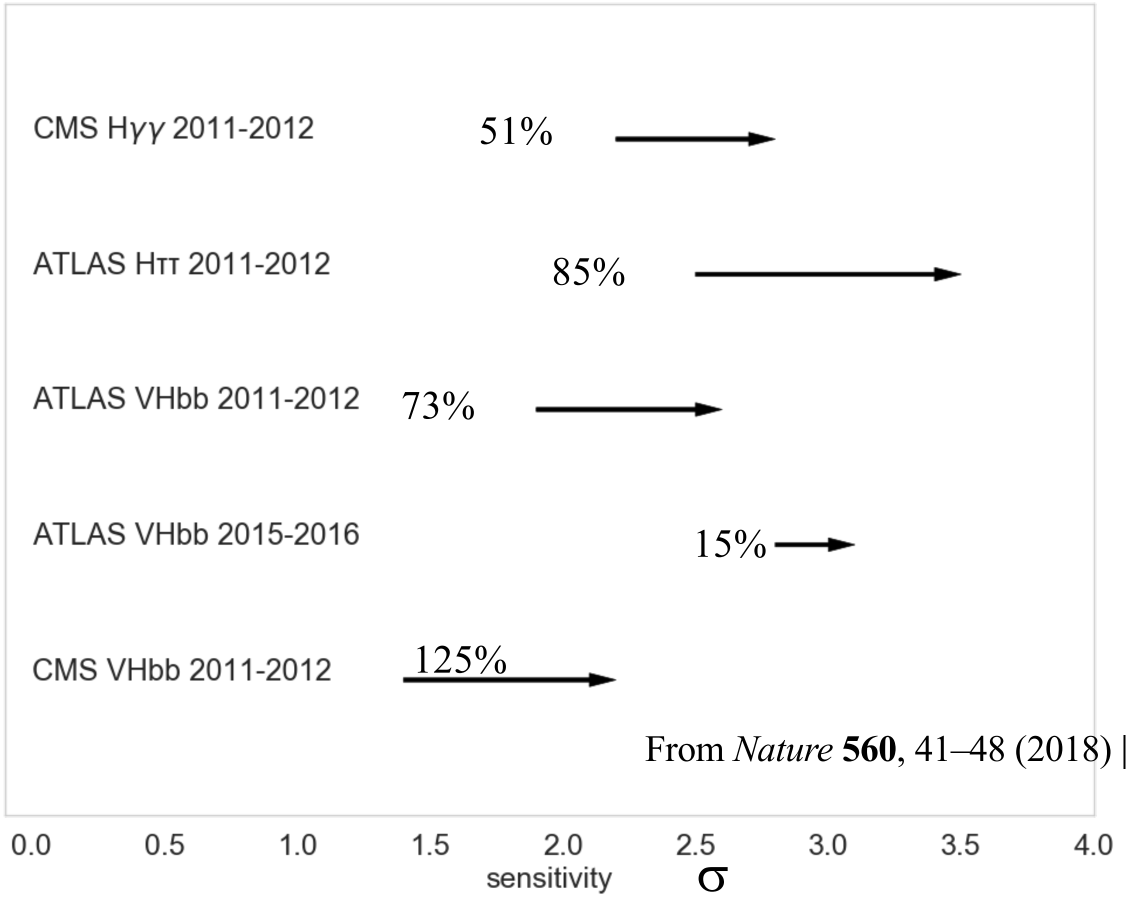


Minor role for simulation



Major role for simulation

Impact of ML on Higgs Physics



➔~ 50% gain on LHC data!

Discovery of the positron

[« The positive electron », Phys Rev 43, 491 (1933)]

Carl Anderson
Nobel 1936

No real statistics analysis,
but evaluation of probability of a fake for each event

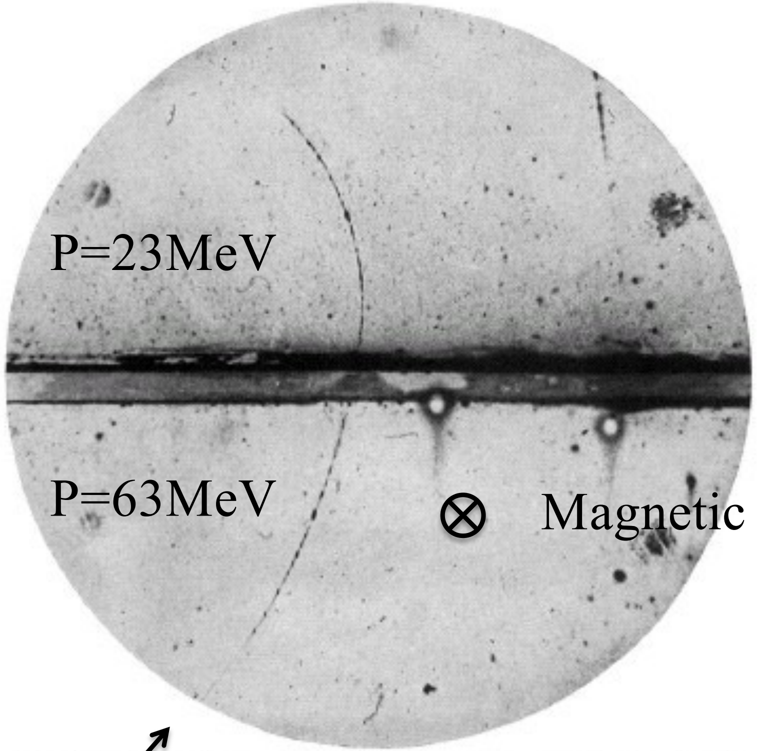
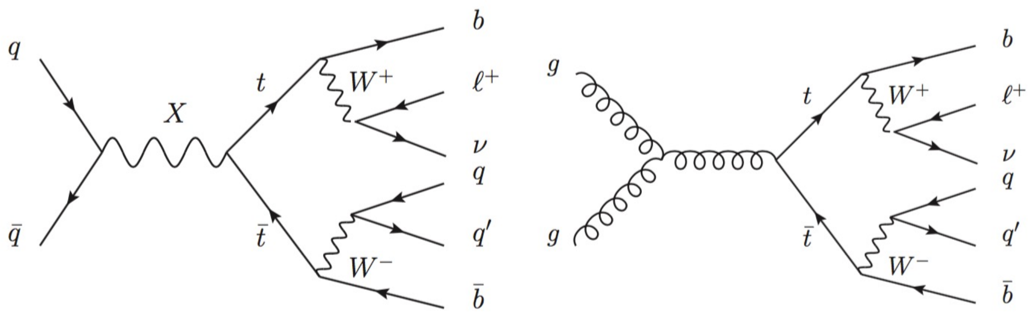


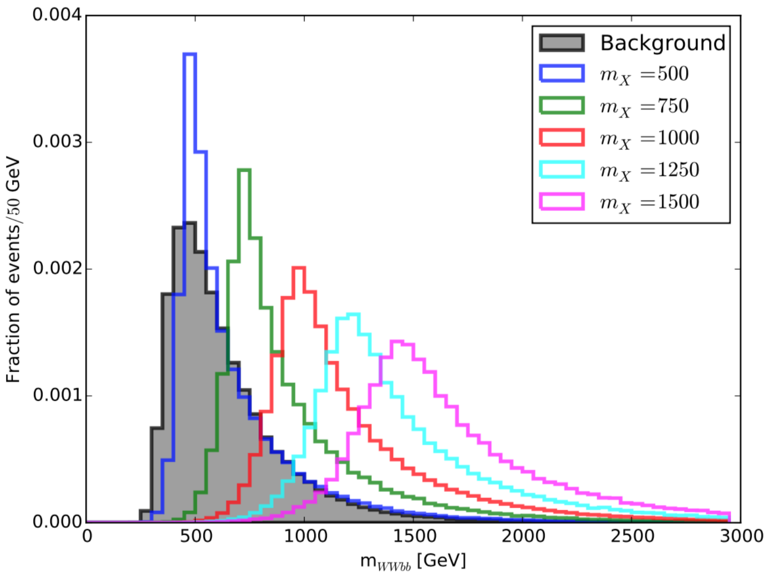
FIG. 1. A 63 million volt positron ($H\rho=2.1\times10^5$ gauss-cm) passing through a 6 mm lead plate and emerging as a 23 million volt positron ($H\rho=7.5\times10^4$ gauss-cm). The length of this latter path is at least ten times greater than the possible length of a proton path of this curvature.

digression on Parameterised Learning

Parameterised learning

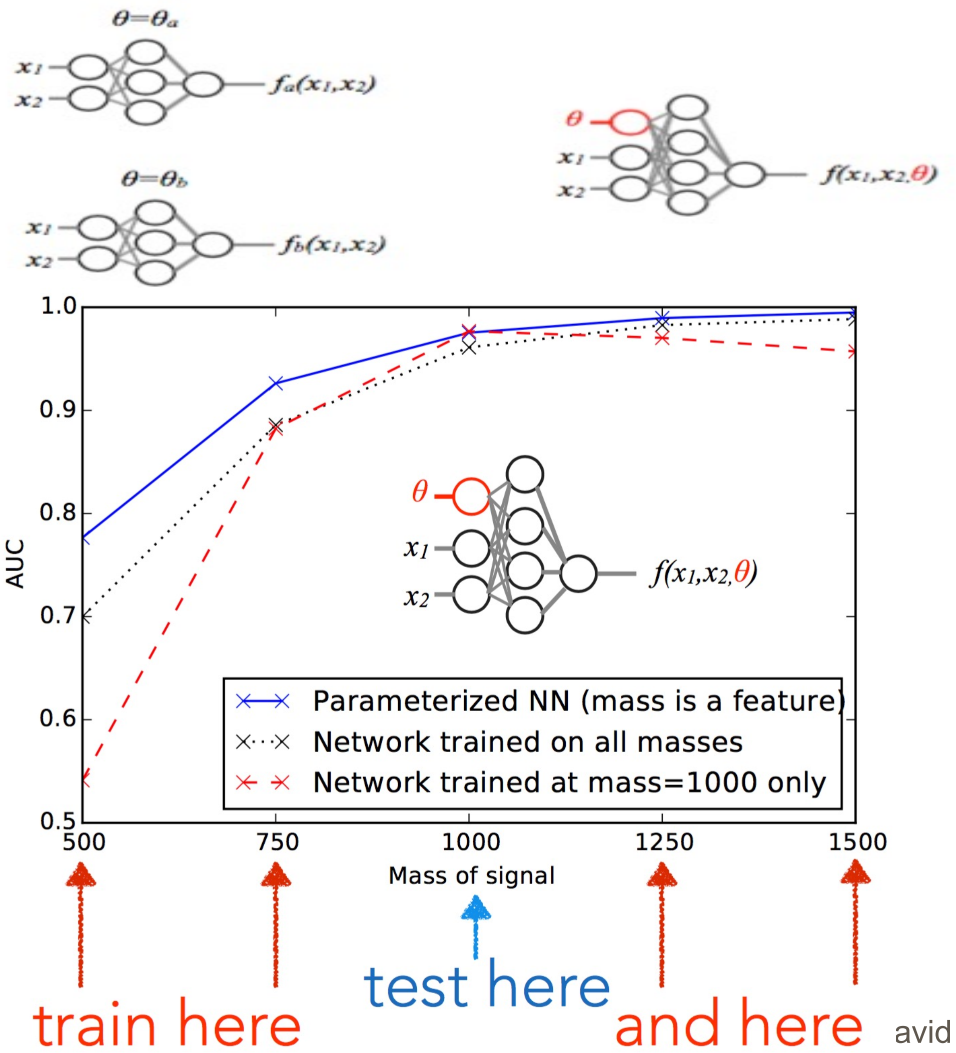
1601.07913 Baldi, Cranmer, Faucett, Sadowksi, Whiteson



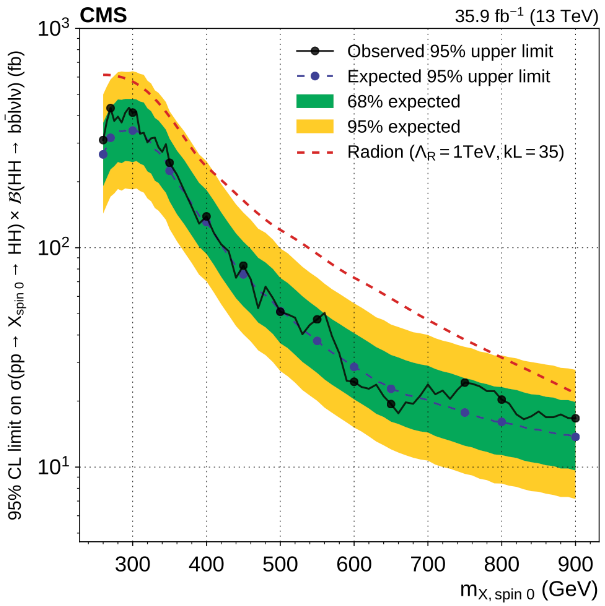


- Typical case: looking for a particle of unknown mass
- E.g. here tt decay

Parameterised learning (2)



- Train on 28 features plus true mass
- NN is conditioned on the true mass
  - ➔ at inference time, need a hypothesised mass for evaluation
- Parameterised NN as good as single mass training
- ➔clean interpolation
- Used for real e.g. arXiv:1708.04188, resonant HH



Neural Simulation Based Inference

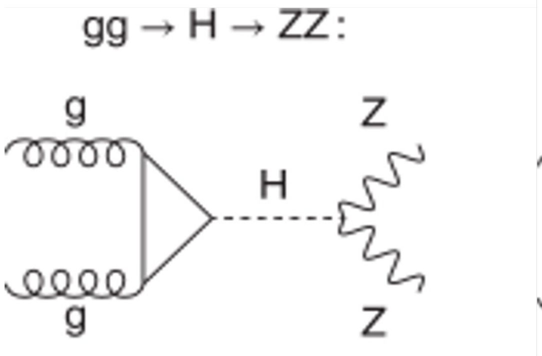
focus on off-shell Higgs cross-section
on behalf of the ATLAS Collaboration

G. Aad et al. Rept.Prog.Phys. 88 (2025) 6, 067801, https://arxiv.org/abs/2412.01600

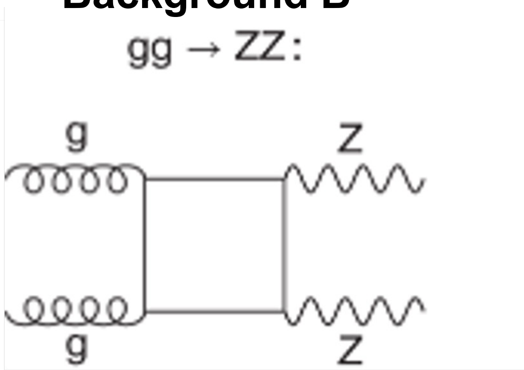
Interference in Higgs Boson cross-section

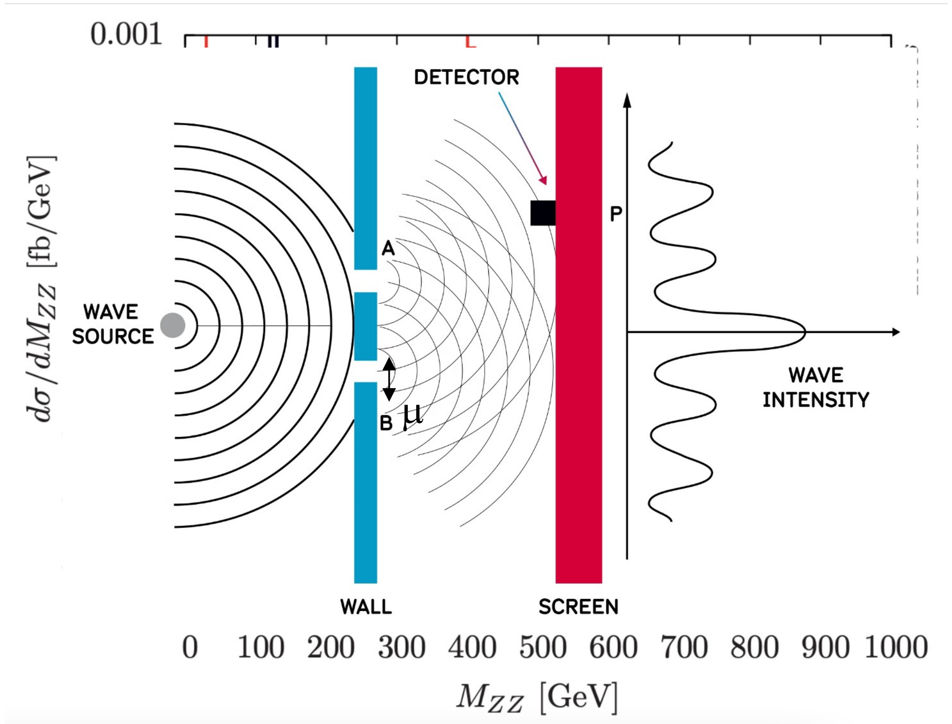
μ : signal strength

Signal S



Interfering Background B





$$|\sqrt{\mu}\, M_S + M_B|^2 = \underbrace{\mu |M_S|^2}_{signal\,\propto\,\mu} + \overbrace{|M_B|^2}^{background} + \underbrace{\overbrace{2\sqrt{\mu}\,\Re\,[M_S M_B^*]}^{<\,0}}_{interference\;\propto\;\sqrt{\mu}}$$

Yield : Number of events counted, for a given number of proton collision @ LHC

$$\nu_{ggF}(\mu) = (\mu - \sqrt{\mu})\,\nu_S + \sqrt{\mu}\,\nu_{SBI} + (1 - \sqrt{\mu})\,\nu_B$$

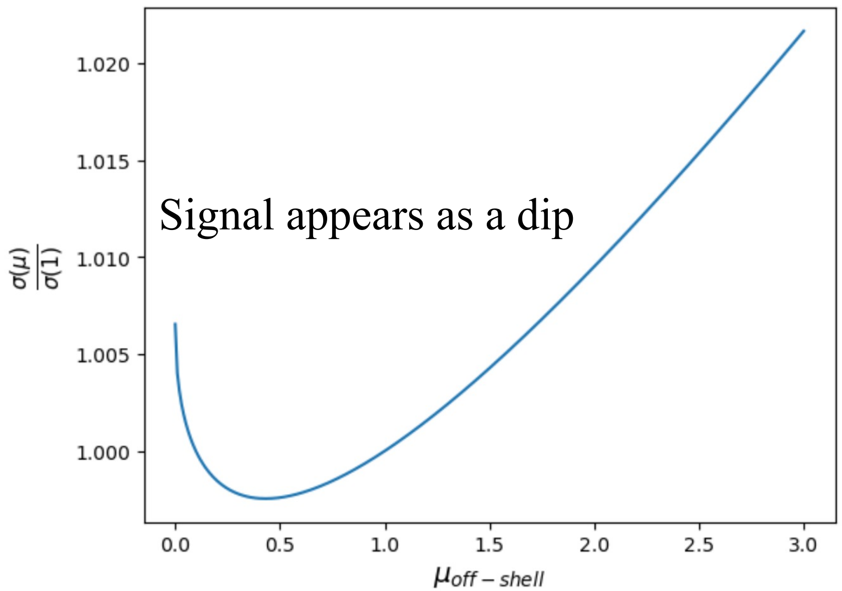
Off-shell Higgs Boson cross-section

Signal Strength

$$\mu = \frac{\nu_{obs}}{\nu_{exp}}$$

In Standard Model

$$\mu = 1$$



$$|\sqrt{\mu}\,M_S + M_B|^2 = \underbrace{\mu|M_S|^2}_{\textit{signal} \propto \mu} + \overbrace{|M_B|^2}^{\textit{background}} + \underbrace{2\sqrt{\mu}\,\overbrace{\Re\left[M_S M_B^*\right]}^{<0}}_{\textit{interference} \propto \sqrt{\mu}}$$

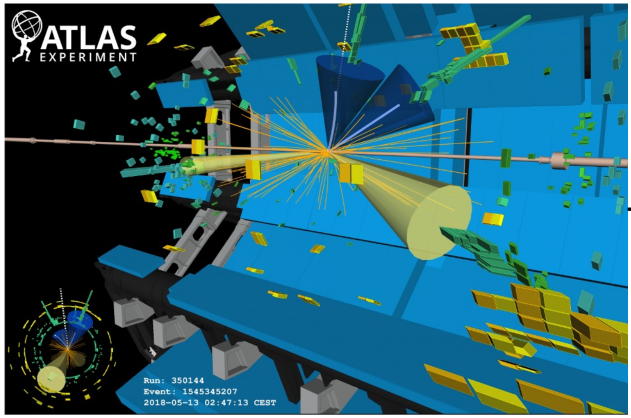
Yield : Number of events counted.

$$\nu_{ggF}(\mu) = (\mu - \sqrt{\mu})\,\nu_S + \sqrt{\mu}\,\nu_{SBI} + (1 - \sqrt{\mu})\,\nu_B$$

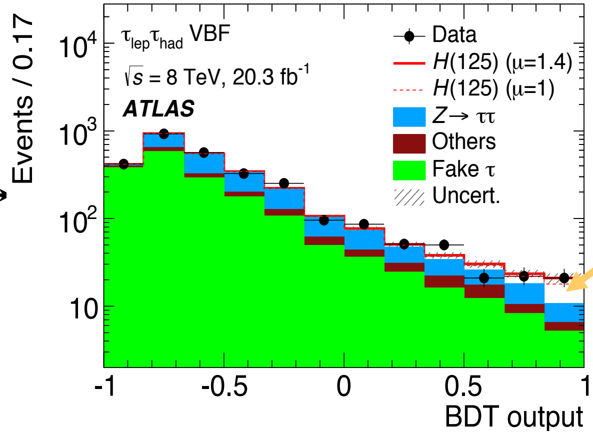
Classical Histogram Method

JHEP 04, 117 (2015) 1501.04943

Boosted Decision Tree using ~dozen of high level features



High dimensional data



Higgs evidence

Summary Histogram
(BDT or NN classifier Score)

Parameter of interest : $\boldsymbol{\mu}$

Statistical Fit

Inference:

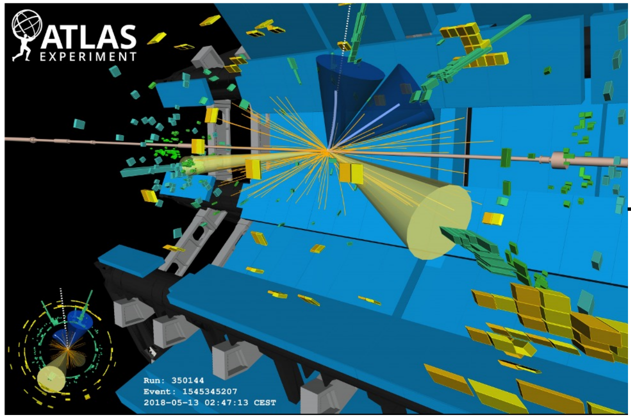
NLL(μ)==-2 log (L(μ))

$$\mathrm{NLL}(\mu) = \sum_{j}^{N_{\mathrm{bins}}} \left[-2\, N_{\mathrm{data}}^{(j)} \left(\log \nu^{(j)}(\mu)\right) + 2\left(\nu^{(j)}(\mu)\right) \right]$$

Poisson

Here the likelihood is computed per **bin**

Unbinned high-dimensional neural inference



High dimensional data

Neural
Networks

Inference:

$$
\begin{aligned}
t_\mu &= -2\log\frac{\mathcal{L}(\mu|\mathcal{D})}{\mathcal{L}(\text{ref}|\mathcal{D})} \\
&= -2\left(N_{\text{data}}\log(\nu(\mu)) - \nu(\mu)\right) - 2\sum_{i=1}^{n} w_i \log\left(\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)}\right)
\end{aligned}
$$

Here the likelihood is computed per **event**

Mixture model

Cranmer 2016

Let $p(x_i|\mu)$ be the probability of finding any event x_i for a given parameter μ, and C the set of processes producing the desired final state

Event rates known from simulation

Known analytically from theory

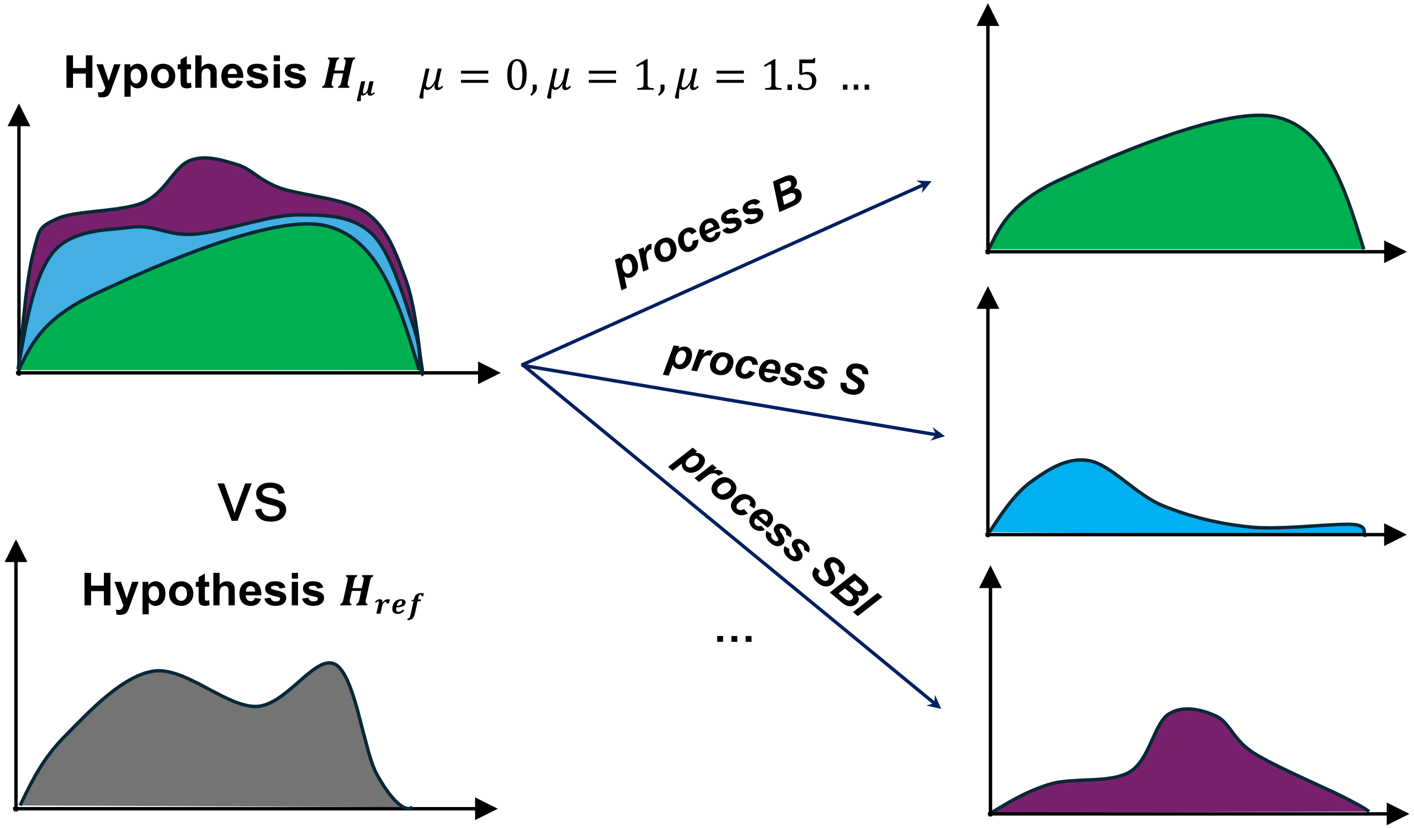
$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_{X}^{C} f_X(\mu) \cdot \nu_X \cdot p_X(x_i)$$

Density ratio estimated using an ML Classifier. (independent of μ!)

$$\frac{p(x_i|\mu)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu)} \sum_{X}^{C} f_X(\mu) \cdot \nu_X \cdot \frac{p_X(x_i)}{p_{ref}(x_i)}$$

SBI == Unbinned Maximum Likelihood fit with NN estimated event likelihood!

NSBI for Off-shell Higgs Boson



d

$$f_B(\mu) = 1 - \sqrt{\mu}$$

$$f_S(\mu) = \mu - \sqrt{\mu}$$

$$f_{SBI}(\mu) = \sqrt{\mu}$$

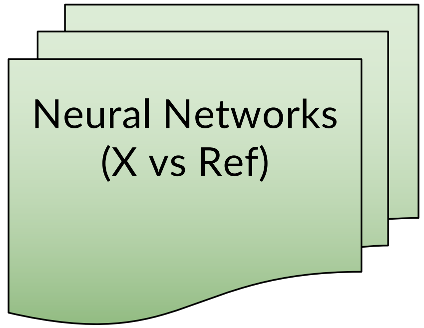
Non Interfering Background

$$\frac{p(x_i|\mu)}{p_{ref}} = \frac{1}{\nu(\mu)}\left[(\mu - \sqrt{\mu})\nu_S \underbrace{\frac{p_S}{p_{ref}}}_{\text{NN}}(x_i) + \sqrt{\mu}\nu_{SBI}\underbrace{\frac{p_{SBI}}{p_{ref}}}_{\text{NN}}(x_i) + (1 - \sqrt{\mu})\nu_B \underbrace{\frac{p_B}{p_{ref}}}_{\text{NN}}(x_i) + \nu_{NIB}\underbrace{\frac{p_{NIB}}{p_{ref}}}_{\text{NN}}(x_i)\right]$$

Density ratio estimation using NNs

Dataset:

- Tabular datasets with 13 features
- 4ℓ system decay kinematic variables: $\cos(\theta^*)$, $\cos(\theta_1)$, $\cos(\theta_2)$, ϕ_1, ϕ, m_{Z_1} , m_{Z_2}
- Higgs production kinematic variables: $p_{T_{4\ell}}$, $\Upsilon_{4\ell}$
- Jet kinematics variables (if applicable): n_{jets} , m_{jj} , ϕ_{jj} , η_{jj}



$$\hat{s}_{\text{pred}}(x_i) = \frac{p_X(x_i)}{p_X(x_i) + p_{\text{ref}}(x_i)}$$

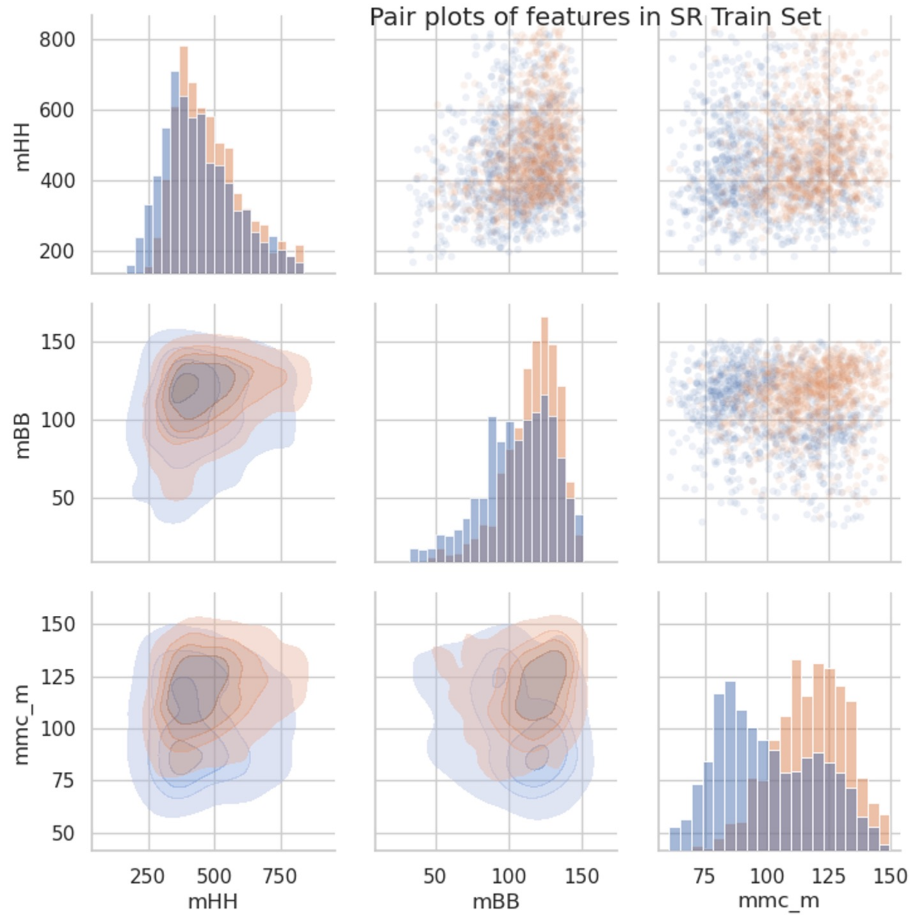
$$L = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right)$$

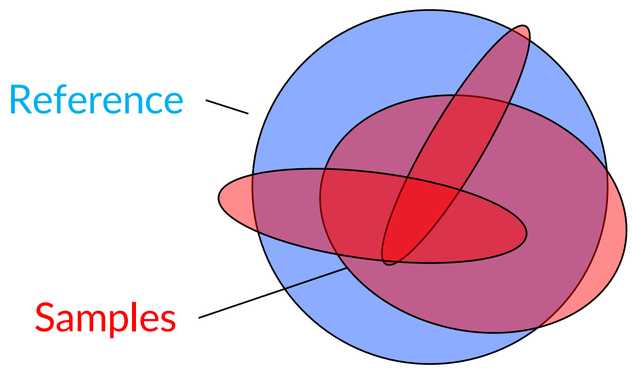
- **Binary Cross Entropy loss**
- Dense NN classifier with 5 hidden layers of 1000 nodes

$$\frac{p_X(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{\hat{s}_{\text{pred}}(x_i)}{1 - \hat{s}_{\text{pred}}(x_i)}$$

Choice of Reference

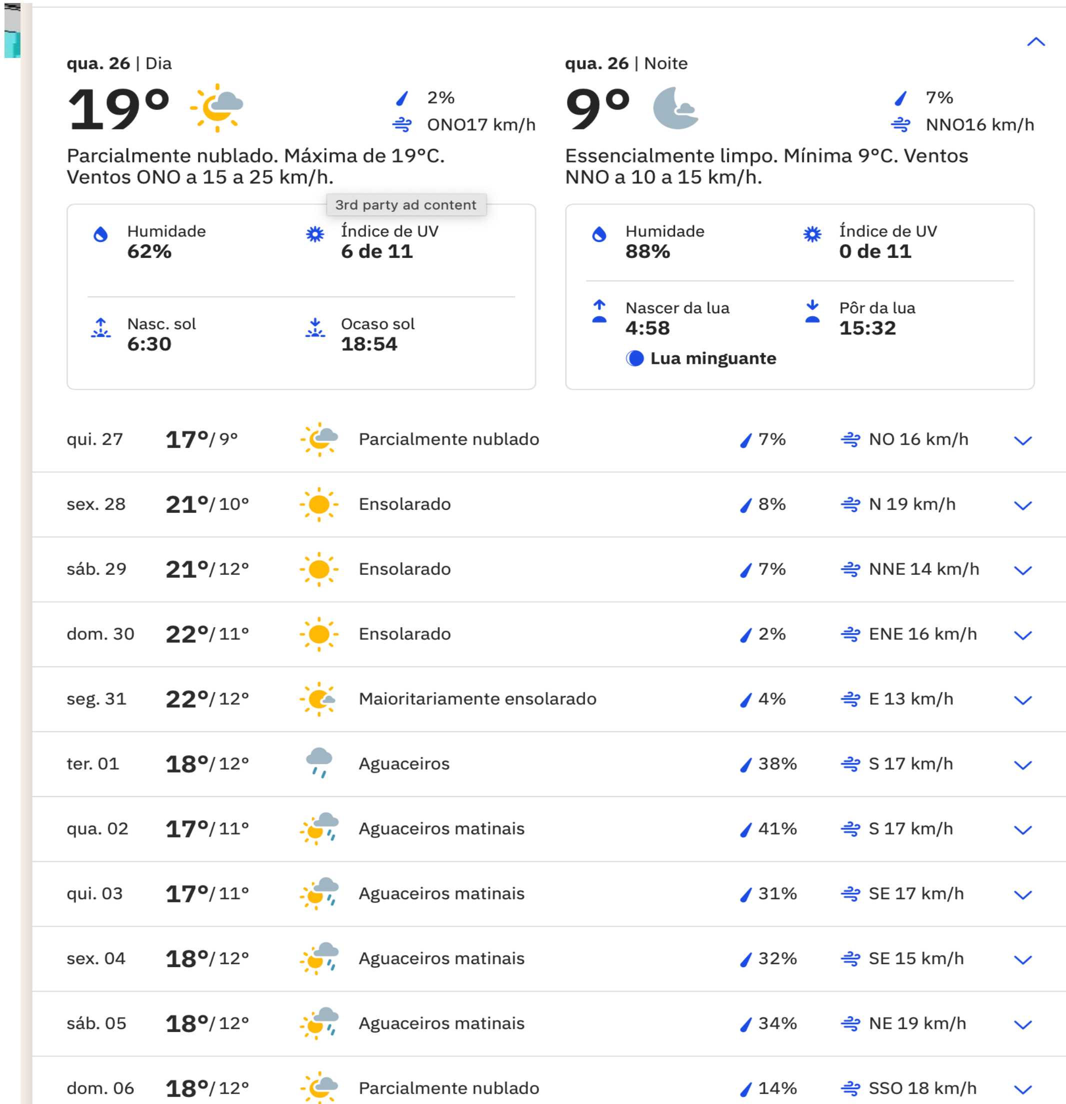
The reference should cover the entire phase-space

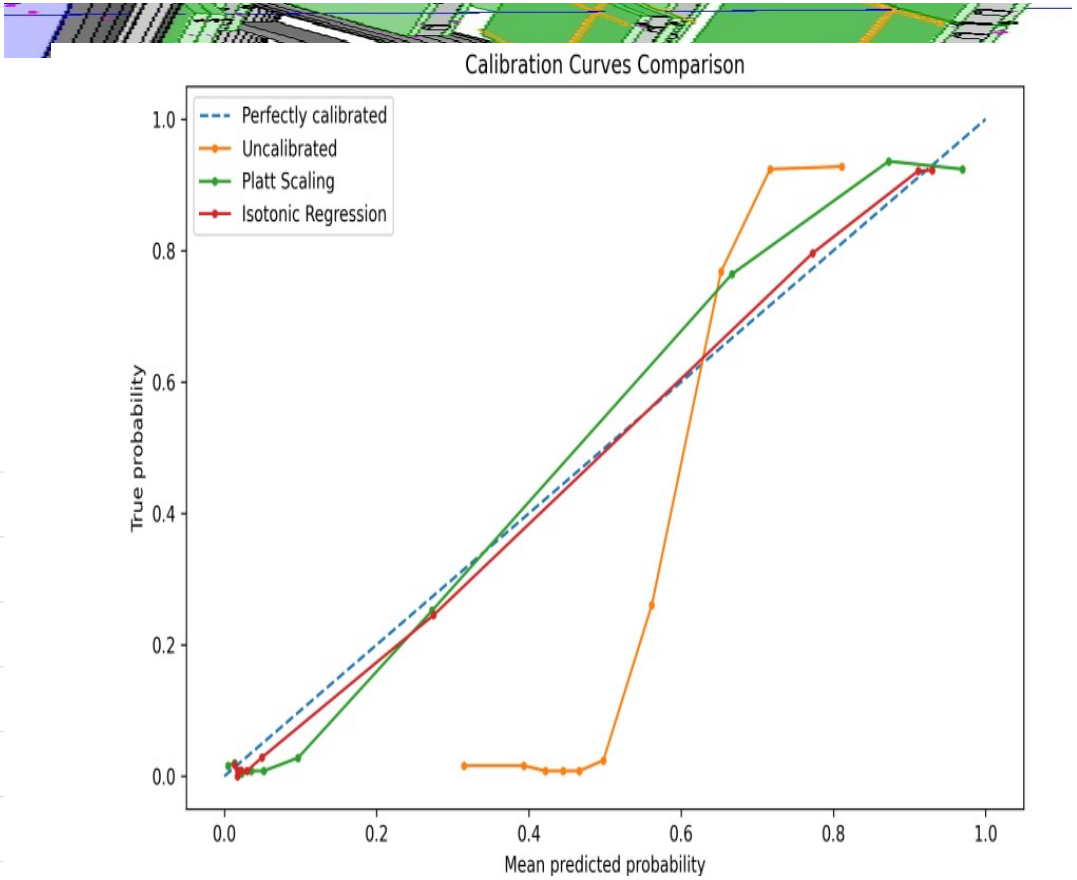




Probability calibration

Probability Calibration

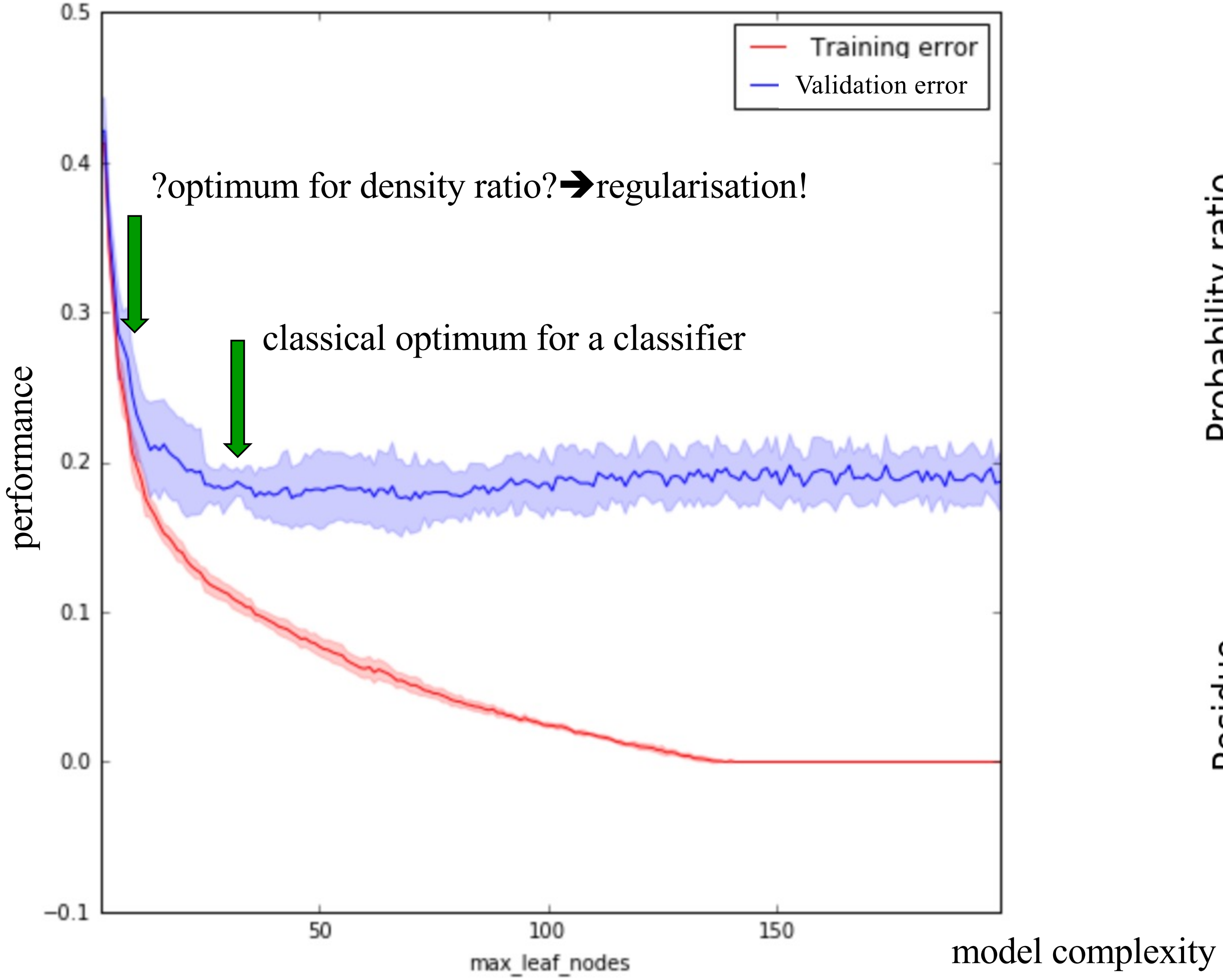


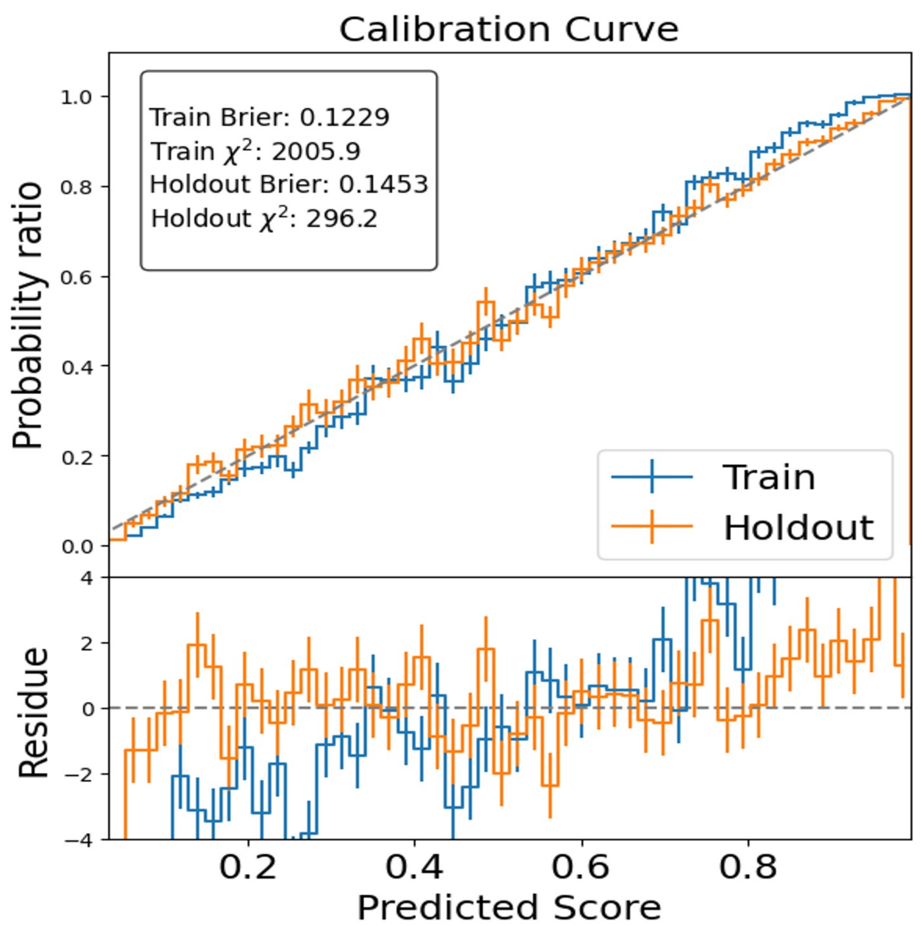


eau,

In practice, many practical difficulties with post-hoc calibration technique.
➔stick to ensembling (but resource hungry)

NN Density ratio : overtraining is fatal!



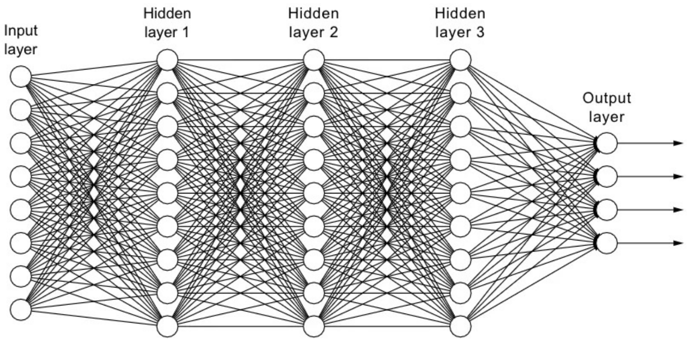


Ensembling

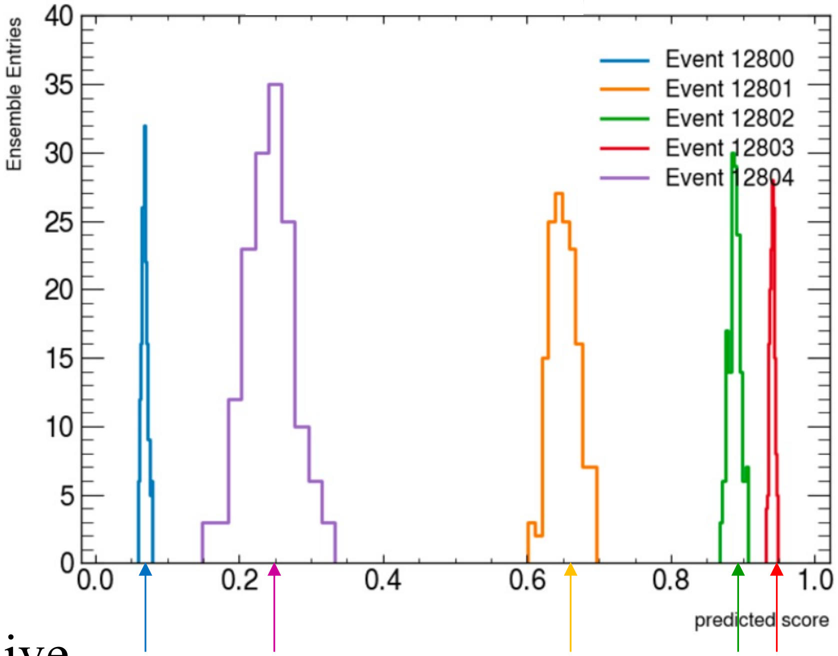
Improve calibration by reducing model complexity and ensembling

Randomize training dataset (train/val split or cross-validation)

Randomize initialization weights



Average score



Ensemble by up to 100! very resource intensive

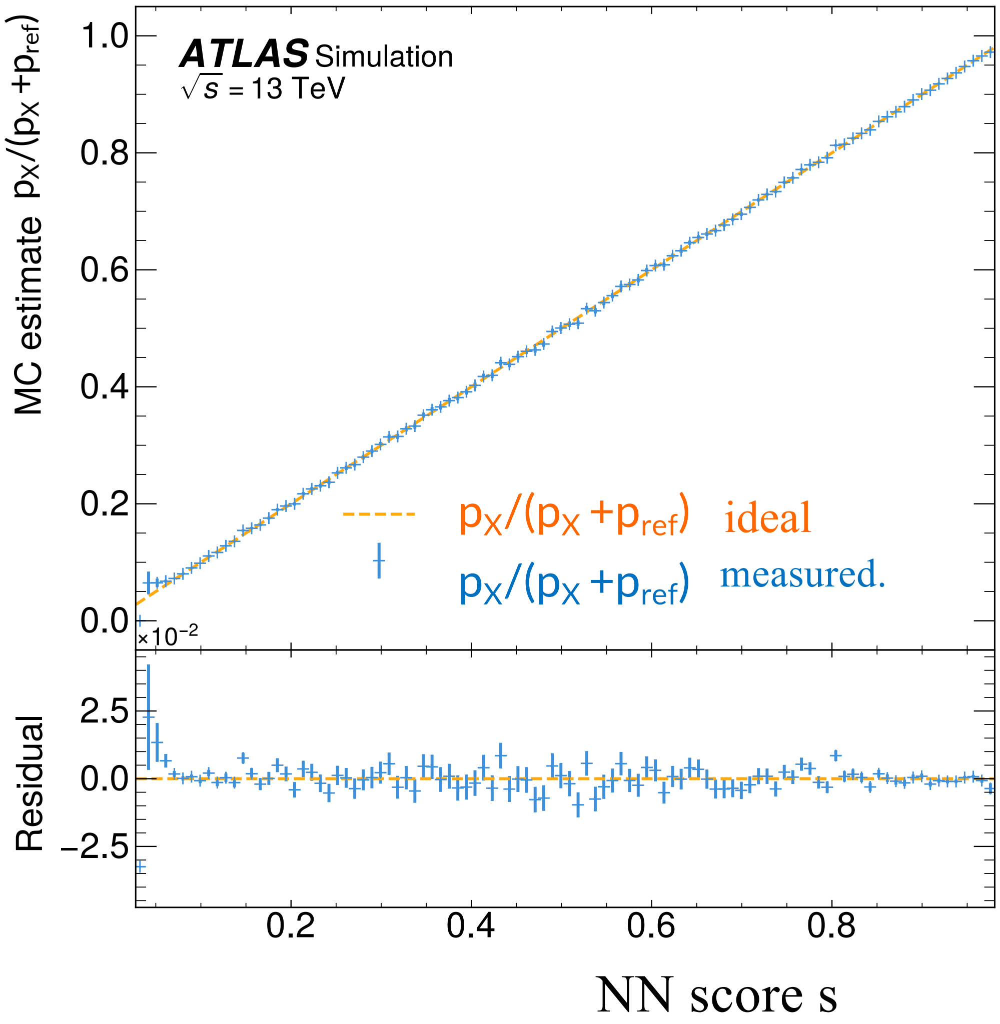
Calibration check 1

With cross-entropy loss and regularisation NN score is really a probability:

$$s(x_i) = \frac{p_X}{p_X + p_{ref}}(x_i)$$

Check it on a holdout dataset !

...not the full story



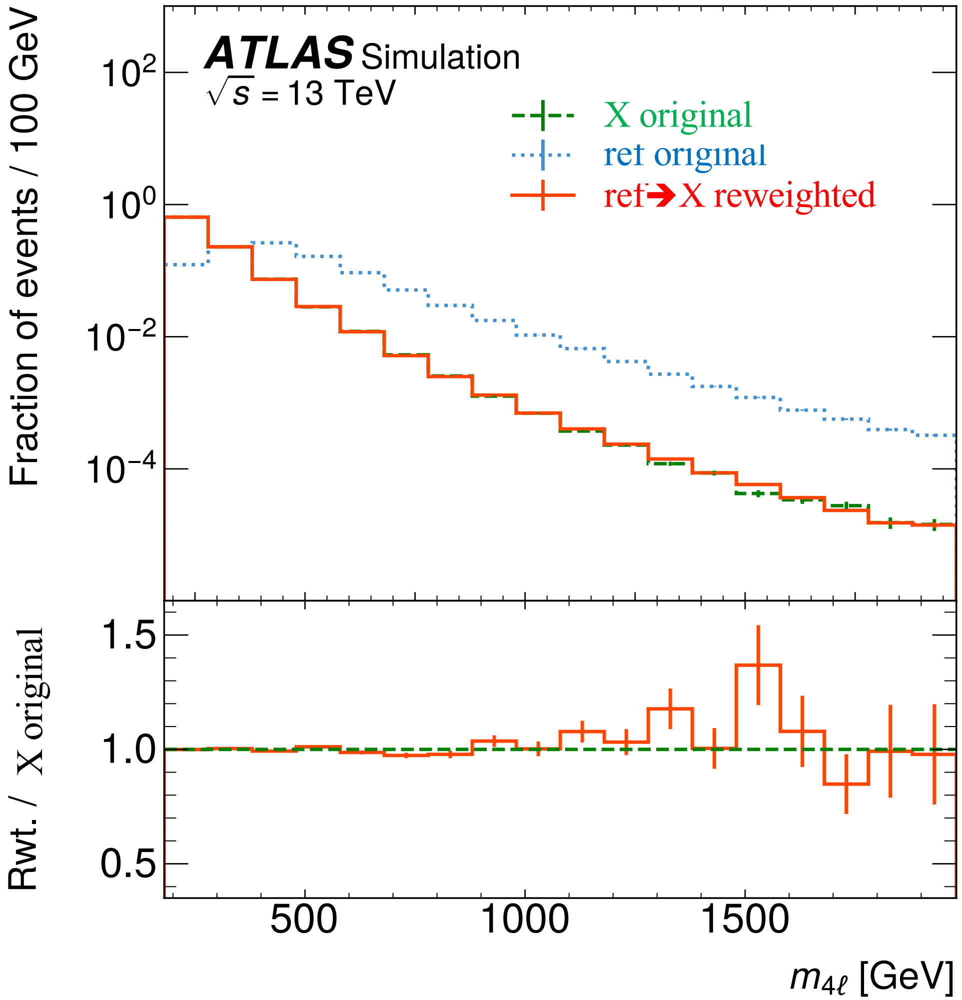
Calibration check 2

$$s(x_i) = \frac{p_X}{p_X + p_{ref}}(x_i)$$

➔

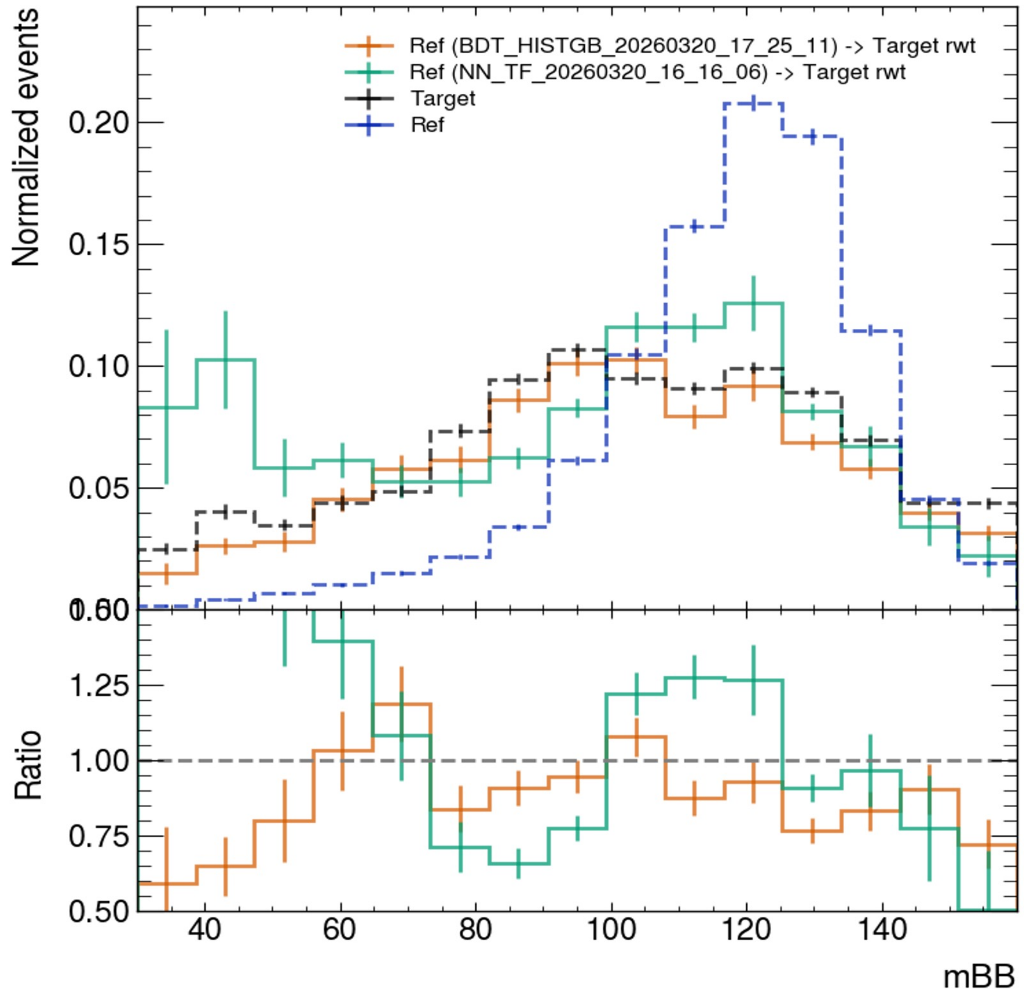
$$\frac{p_X}{p_{ref}}(x_i) = \frac{s}{1-s}(x_i)$$

- Reweighting with p_X/p_{ref}, one **can turn ref dataset into X dataset**
- ➔detailed cross checks across feature space. Here m_{4L} bins.

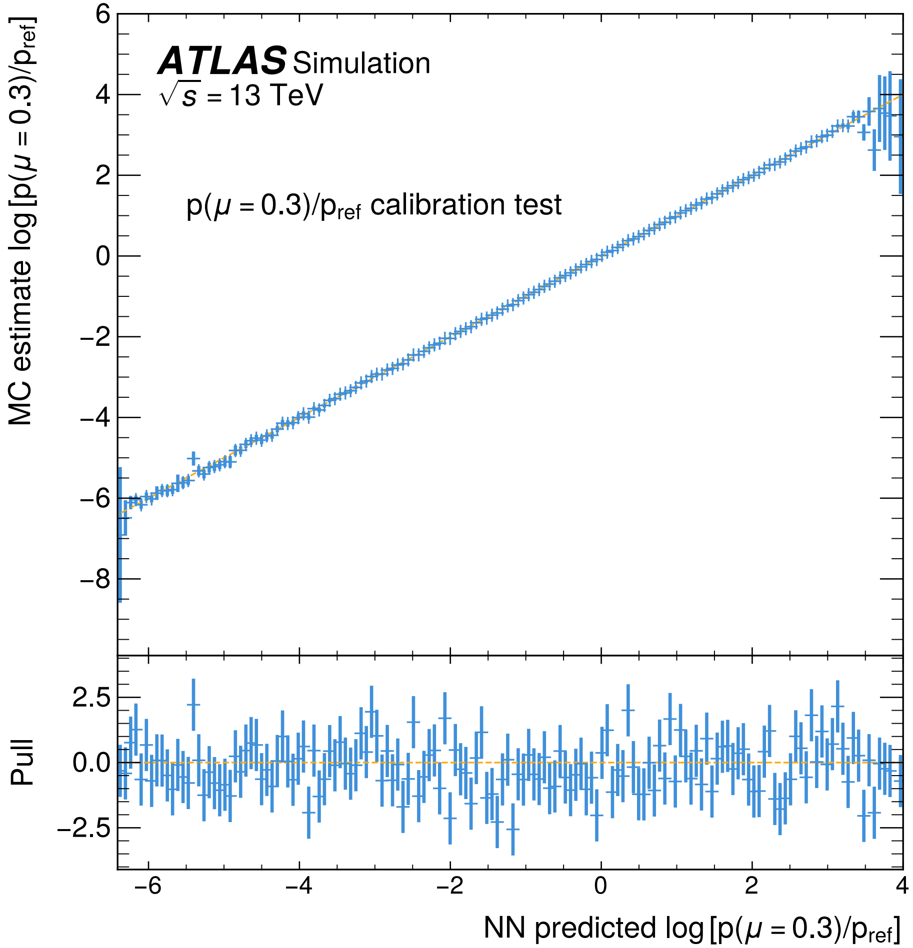


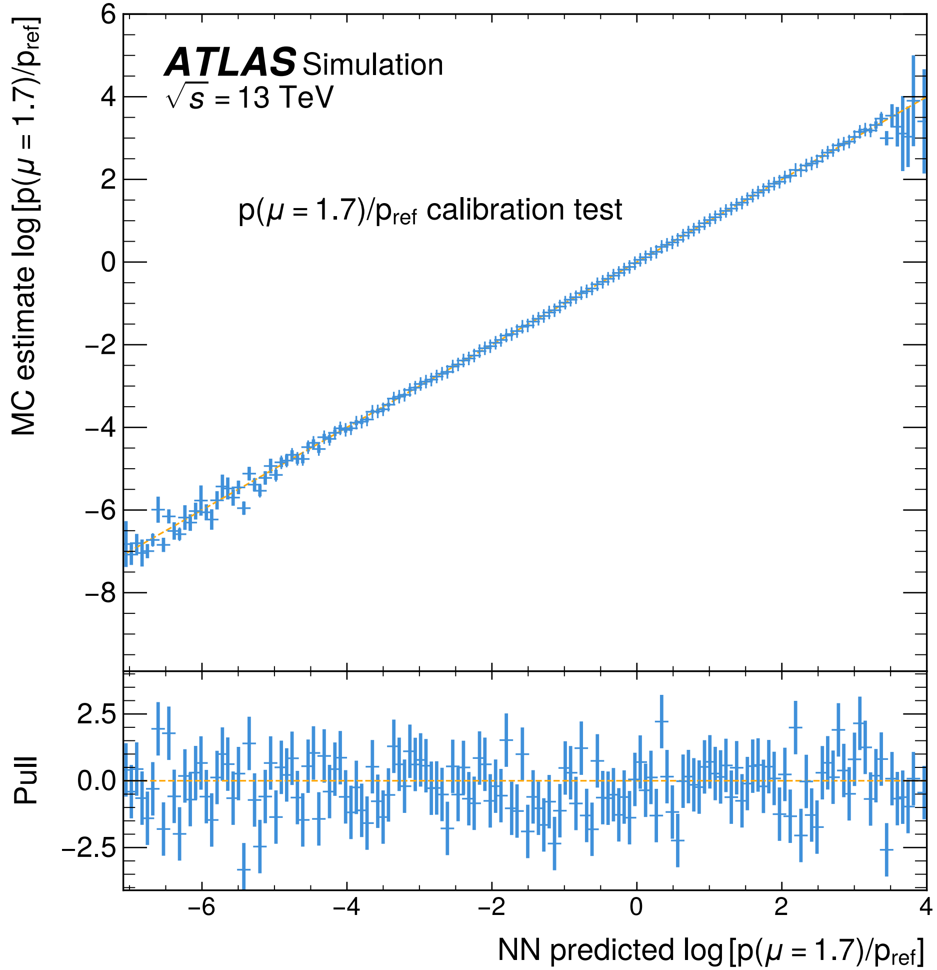
Calibration check 2

It never works on first attempt….



Calibration check 3





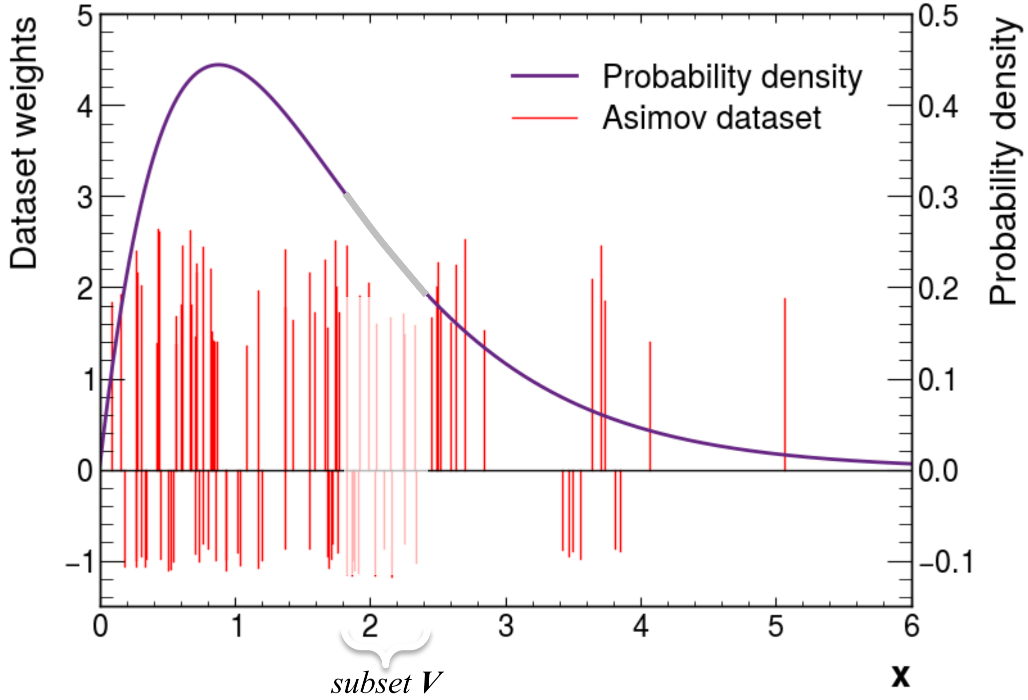
Negative Log Likelihood fit

Asimov dataset

- Common concept in Particle Physics data analysis, frequently used in binned analysis, more tricky in unbinned analysis
- Asimov dataset built from expensive simulator (~10^7 CPU.h)
- Events weighted to approximate the average dataset to be obtained from a specific experiment (with specific value of μ or systematics)
- Sampling from Asimov dataset as a proxy for running the simulator $\mathbb{E}$
- More important (Cowan et al (2010) arXiv:1007.1727) :

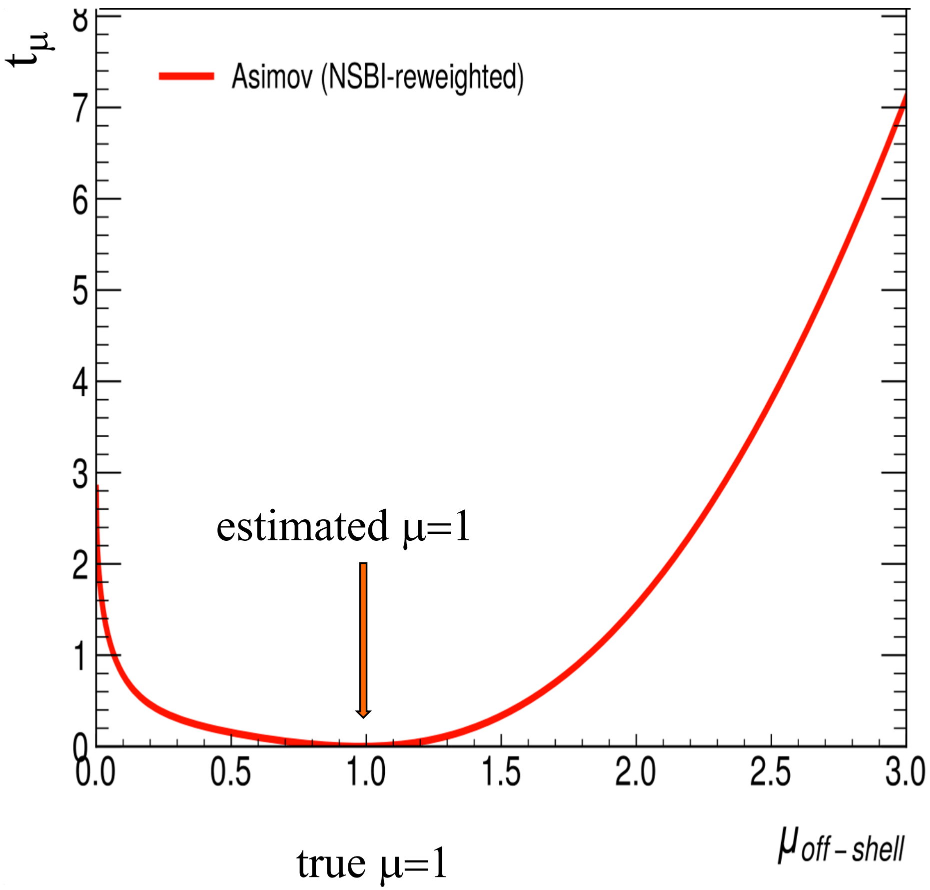
  NLL (Asimov)~ $\mathbb{E}$ (NLL (data))

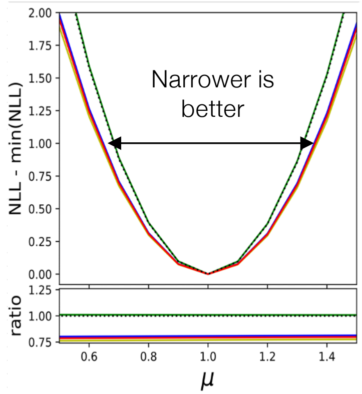
$$\forall V \subset E, \sum_{i \in V} w_i^{\text{Asimov}(\mu)} \sim \int_{x \in V} dx \nu(\mu) p(x|\mu)$$



n dimensions as large as necessary

Negative Log Likelihood fit and Confidence Interval



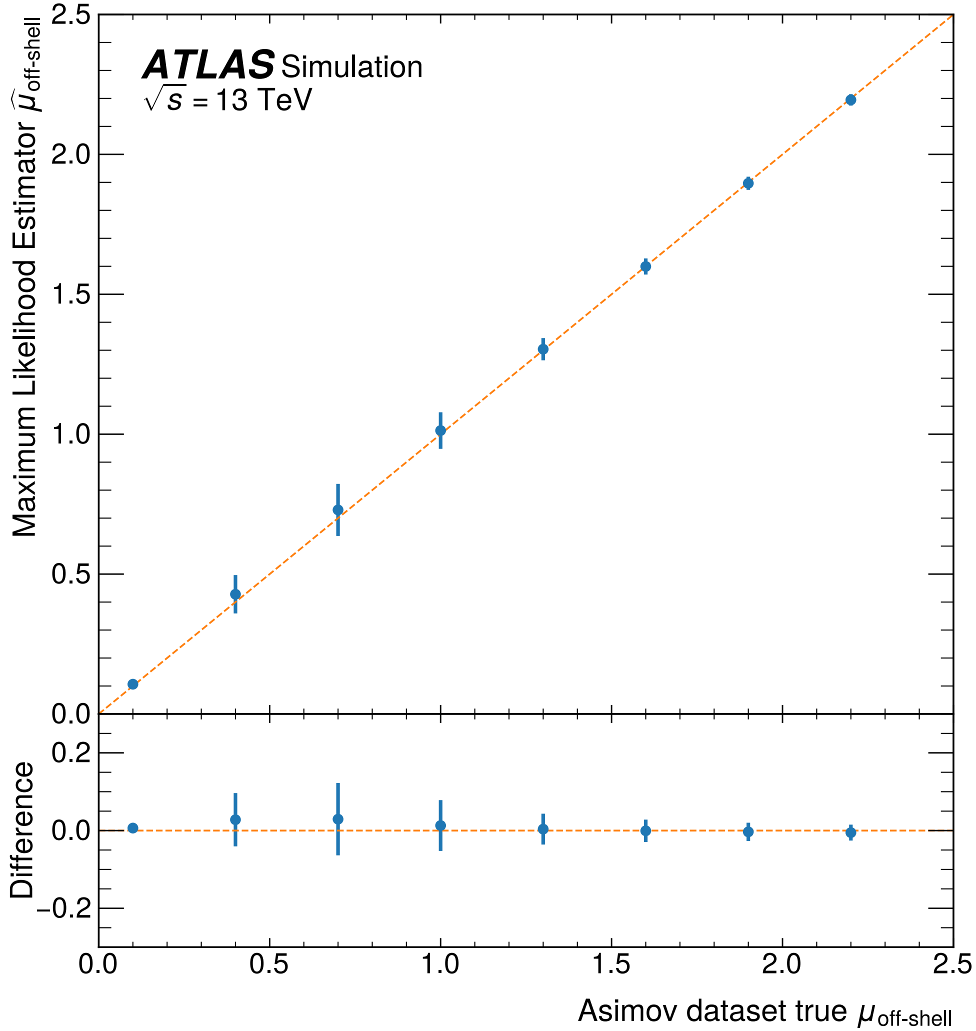


Classical NLL fit

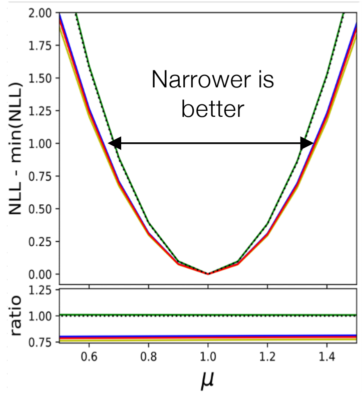
$$t_\mu == NLL(\mu)\text{-}min(NLL)$$

Cross-check 4

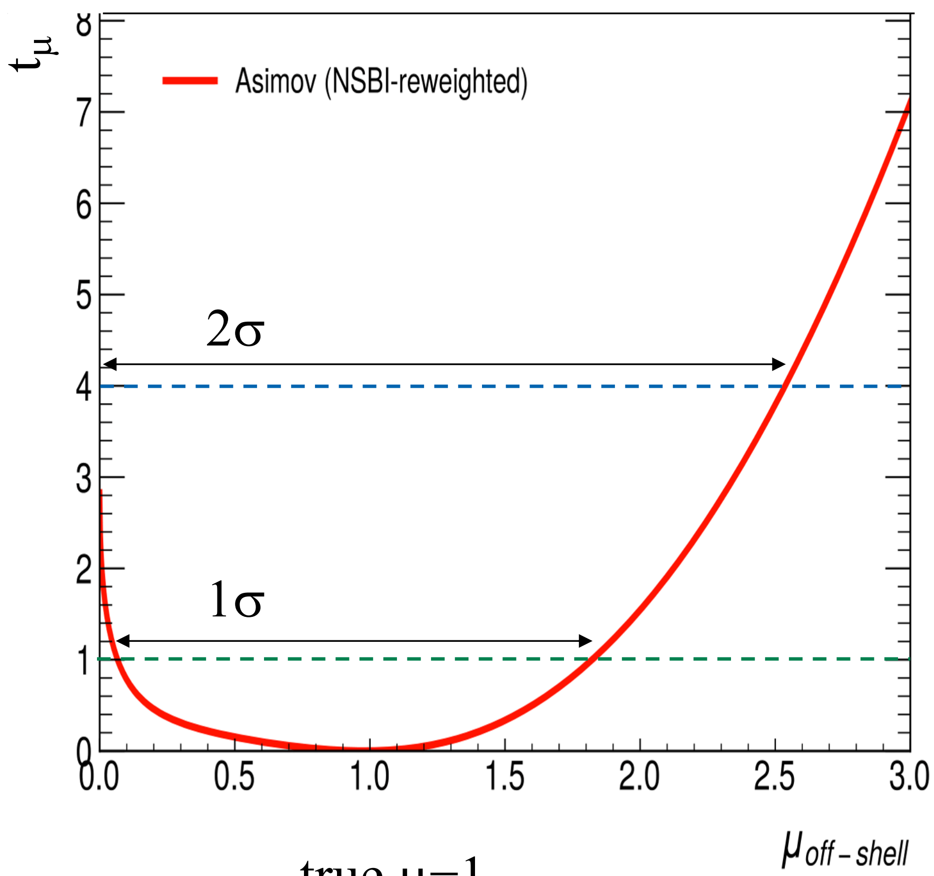
- Compared measured μ to injected μ_{true}



Negative Log Likelihood fit and Confidence Interval



Classical NLL fit



true μ=1

$t_\mu == NLL(\mu)\text{-}min(NLL)$

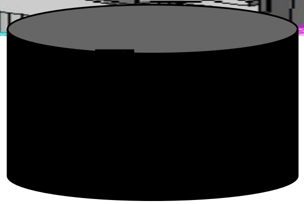
$H^{*}4l$ NLL clearly non-parabolic

Are we at the asymptotic limit of Wilk's theorem ?

1σ and 2σ Confidence Intervals might not be at 1 and 4…

Neyman Construction

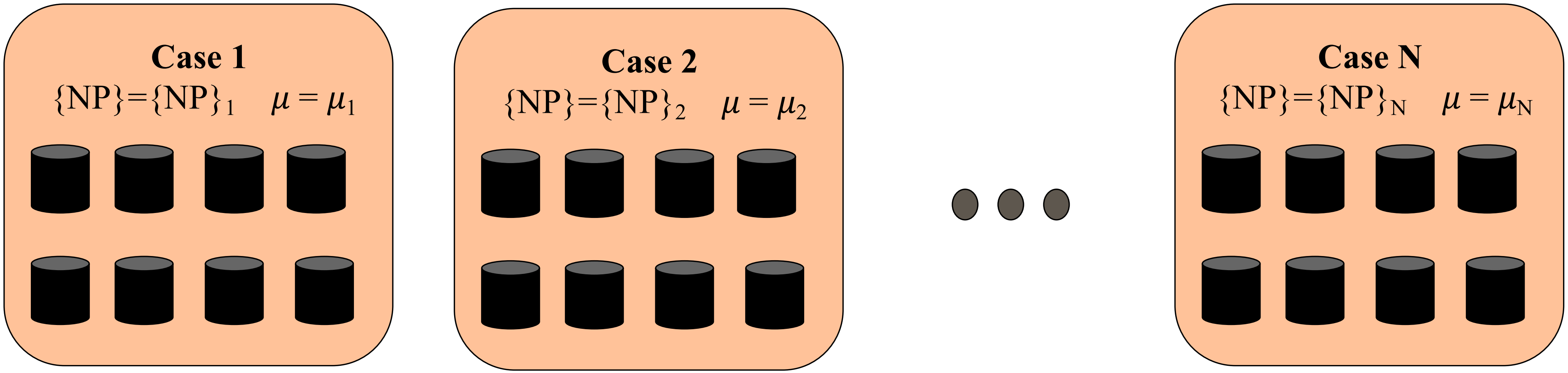
Pseudo-experiments

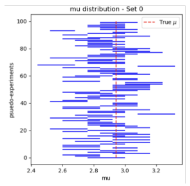


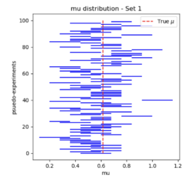
Parent master dataset

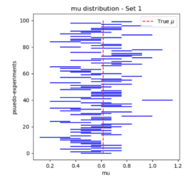
Create pseudo experiments for each test cases

NP = Nuisance Parameter









NLL minimisation for pseudo-experiments

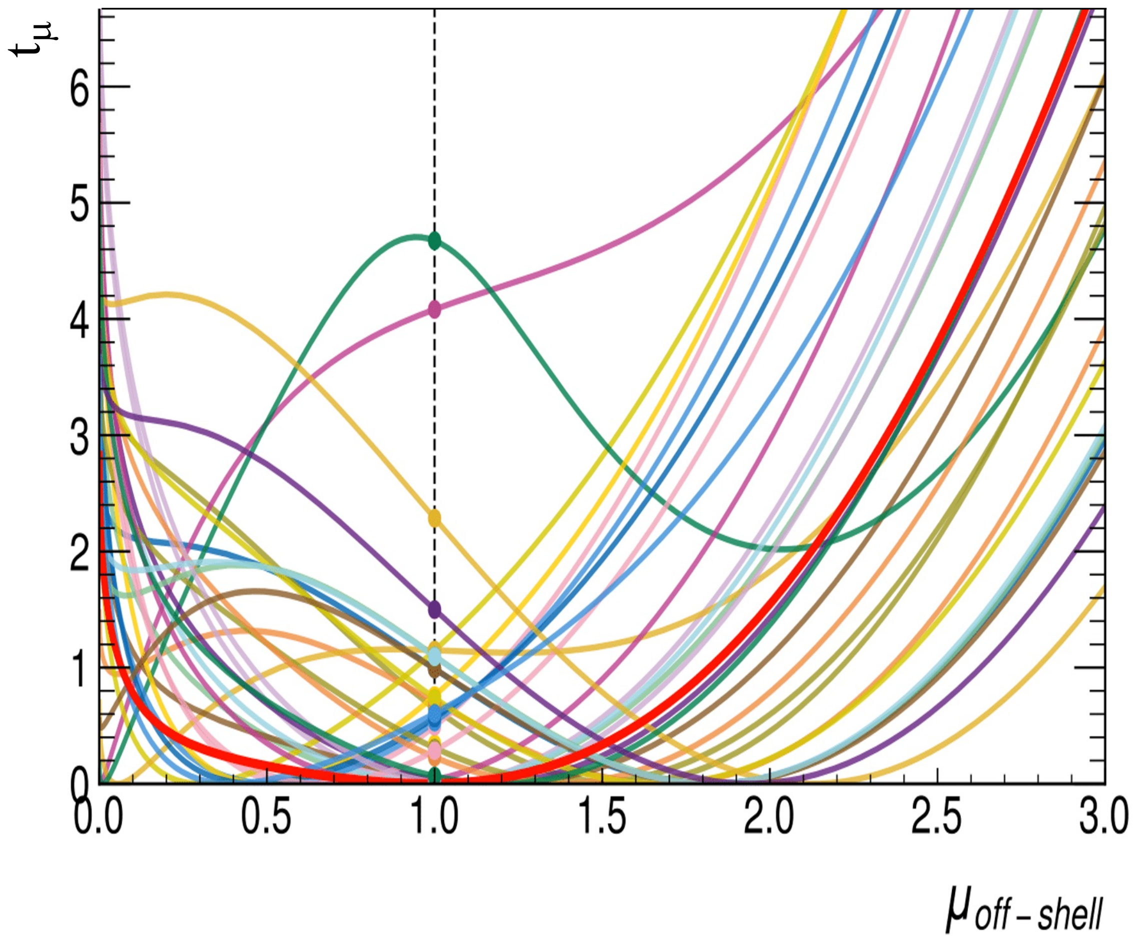
Simulate an actual experiment for a given

μ_{truth} hypothesis:

- ~2000 events sampled **with replacement** from an Asimov dataset of ~10M events
- then fit the NLL as for real data

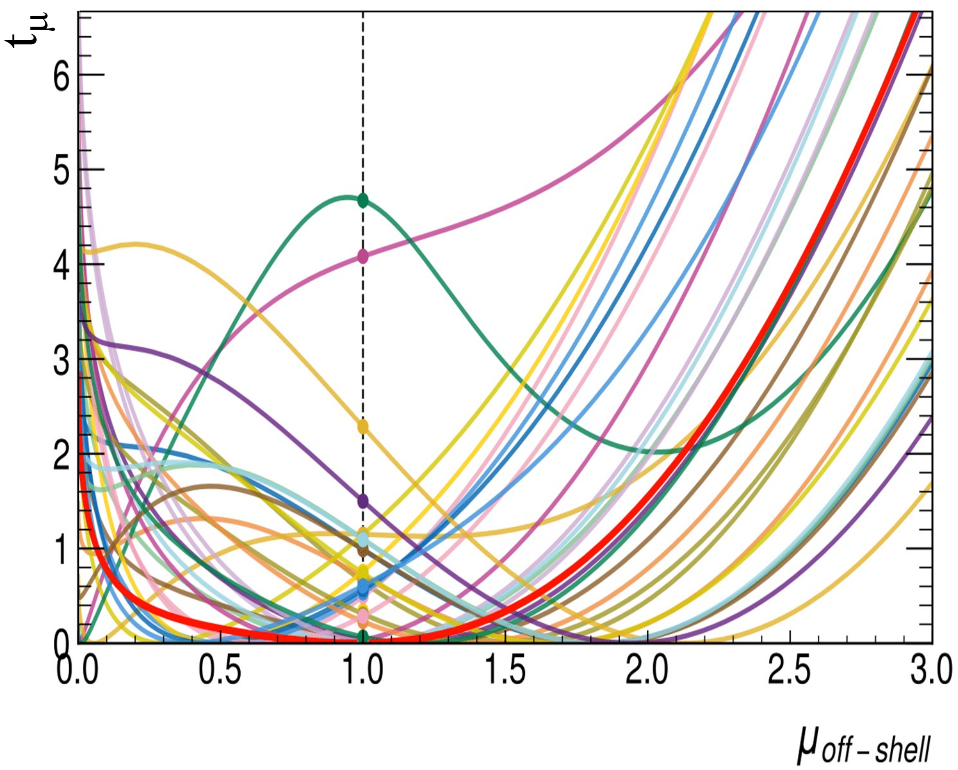
Repeat millions time

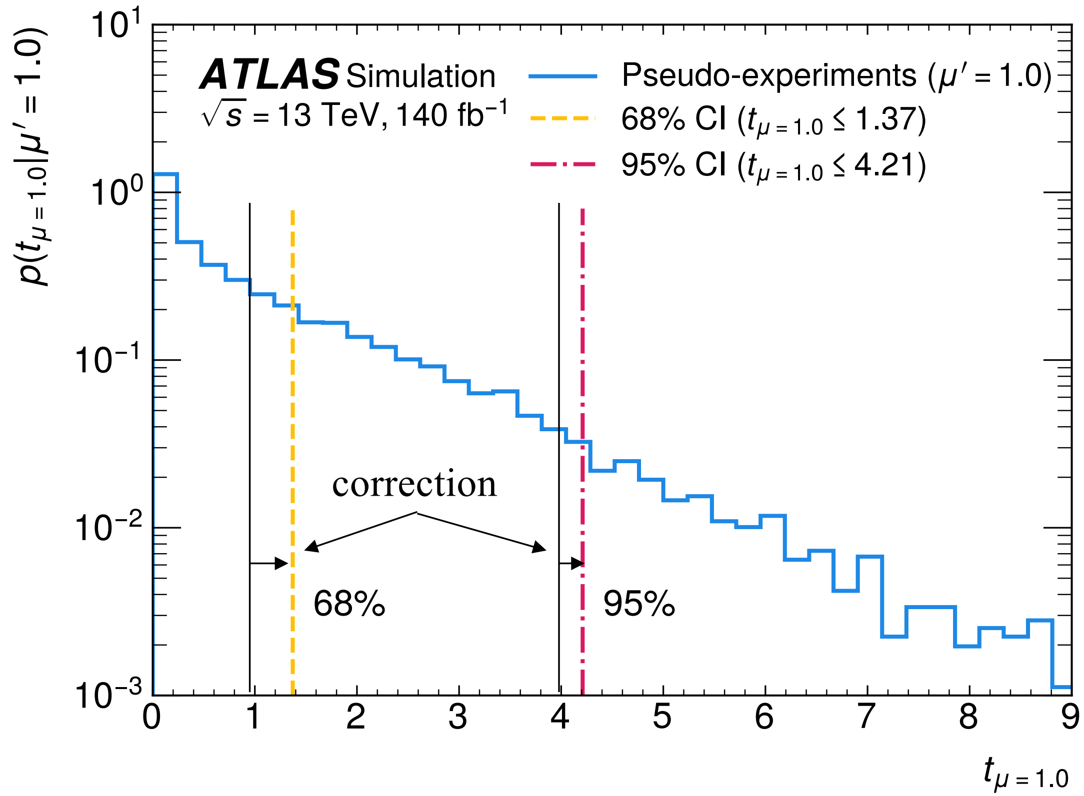
Hypothesis $\mu_{truth} = 1$



Neyman Construction

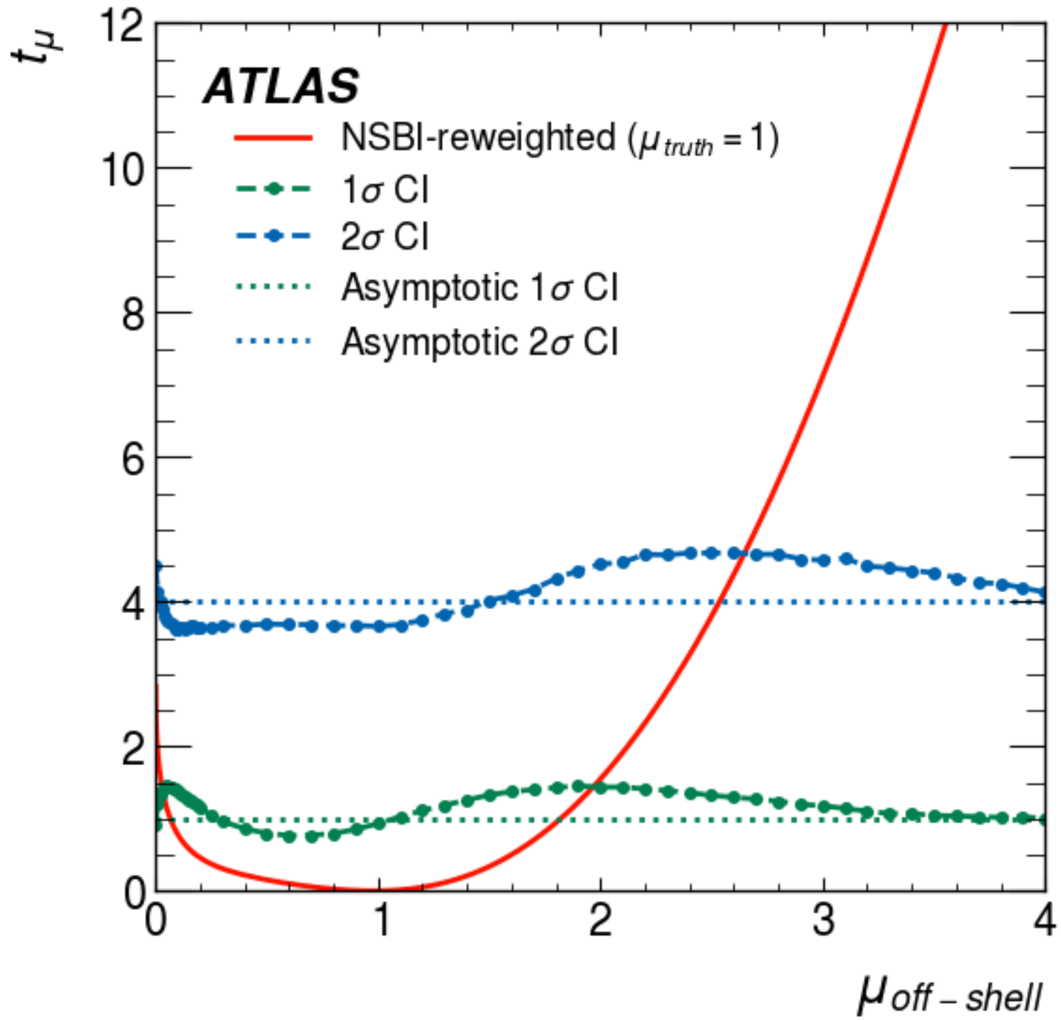
- Compute $t_{\mu=\mu_{truth}} \rightarrow$ the test statistic at $\mu = \mu_{truth}$
- Integrate up to 68.27% (95.45%) to determine the 1σ (2σ) CI for hypothesis μ_{truth}





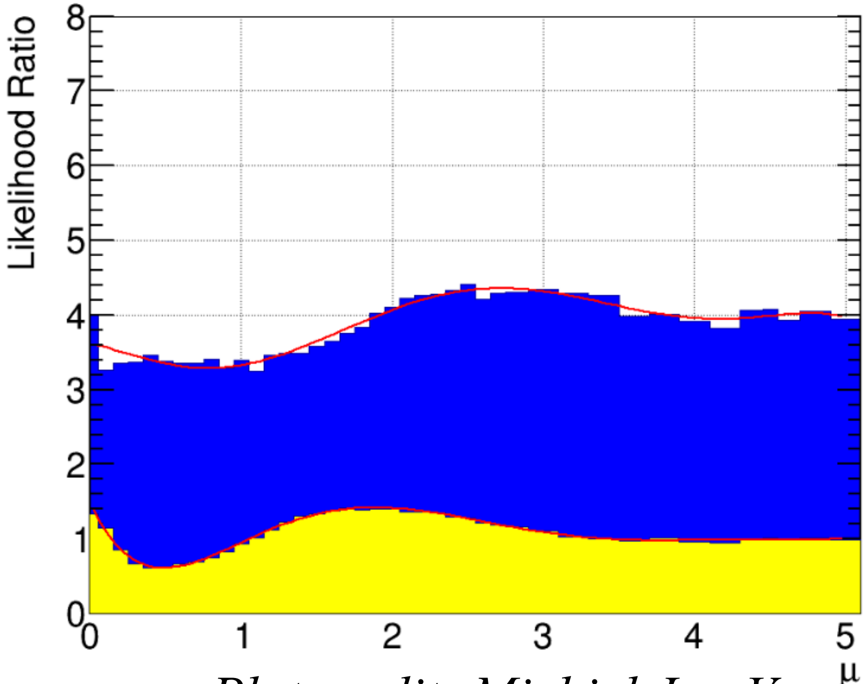
A less resource hungry technique, Quantile Regression, see arxiv:2507.17831

Results of Neyman construction for Stat-only



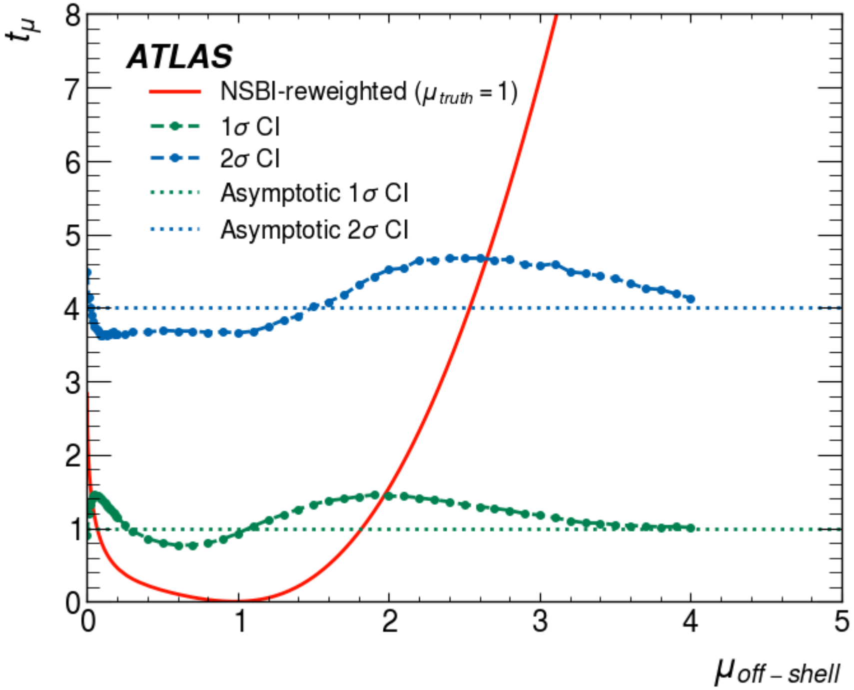
Comparison of Neyman construction

Comparison with histogram-based H*4l analysis ➔ similar 1σ and 2σ CIs !



Plot credit: Michiel Jan Veen

(a) Histogram-based $H^*4\ell$ analysis [126]

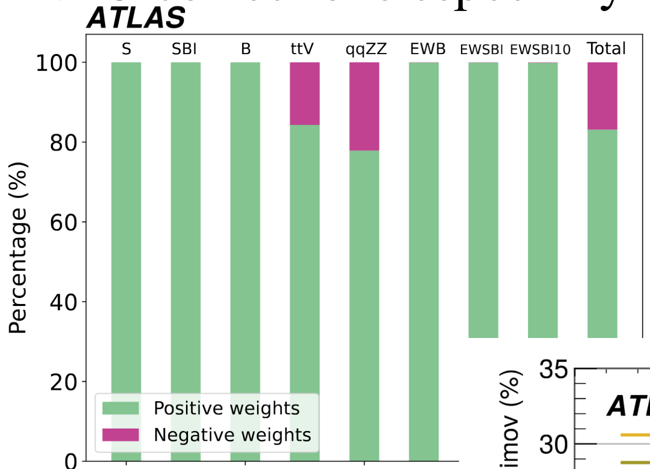


(b) NSBI $H^*4\ell$ analysis

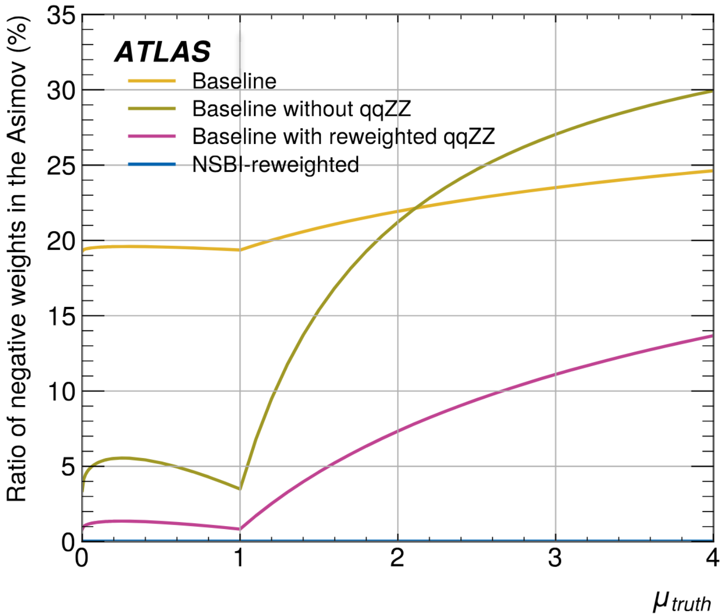
Negative weight handling

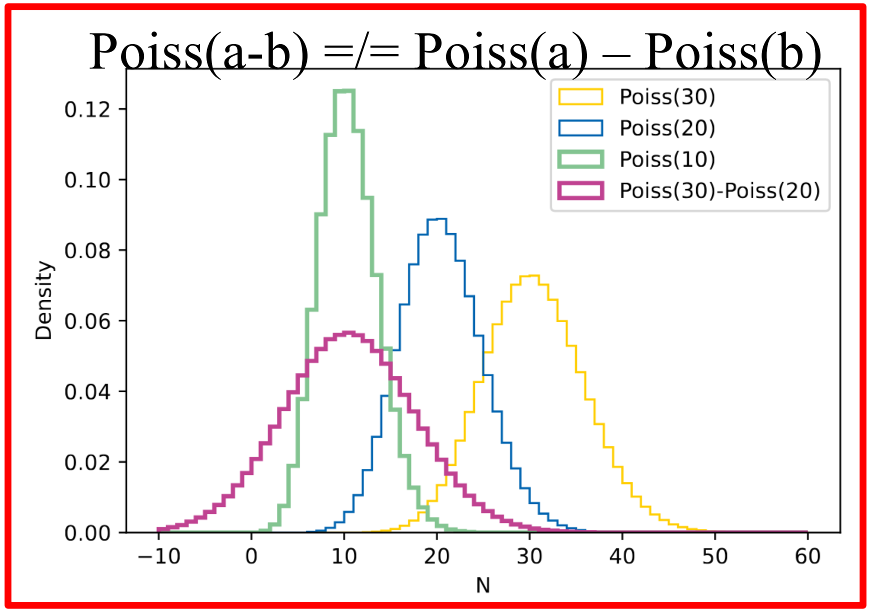
How did we deal with negative weights for NC?

NLO corrections especially for qqZZ



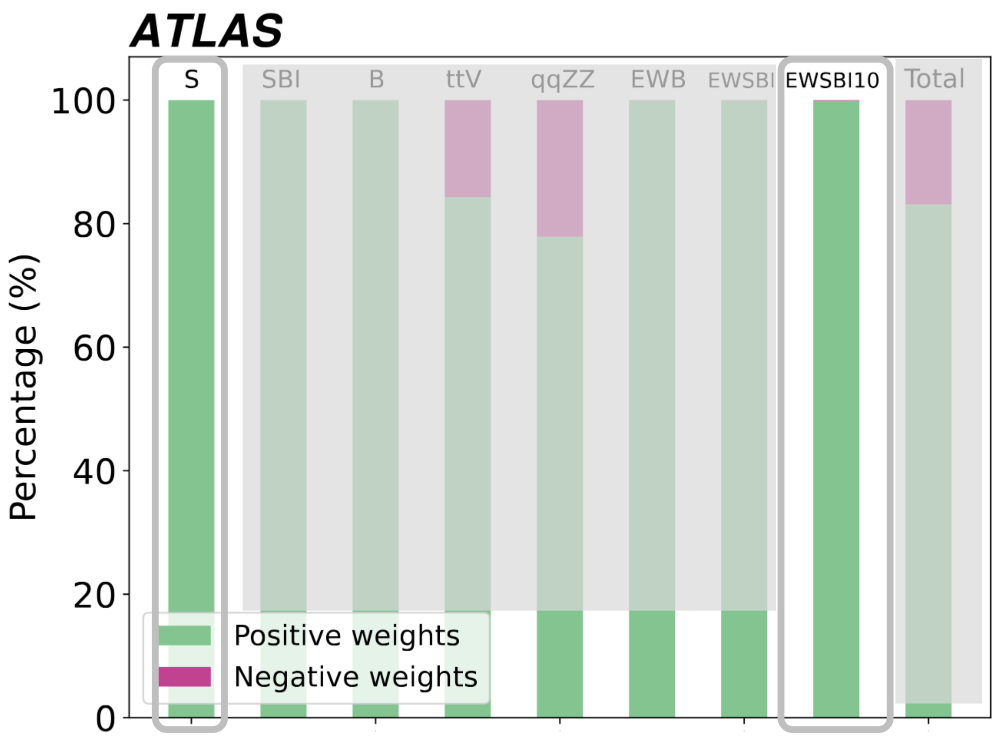
Event type	μ multiplication factor
ggF SBI	$\sqrt{\mu}$
ggF S	$\mu - \sqrt{\mu}$ ← negative if 0<mu<1
ggF B	$1 - \sqrt{\mu}$ ← negative if mu>1
EW SBI	$\frac{1}{\sqrt{10}-10} \cdot (\sqrt{10} \cdot \mu - 10\sqrt{\mu})$
EW B	$\frac{1}{\sqrt{10}-10} \cdot (1 - \sqrt{10} \cdot \mu + 9\sqrt{\mu} - 10 + \sqrt{10})$
EW SBI10	$\frac{1}{\sqrt{10}-10} \cdot (\sqrt{\mu} - \mu)$
$q\bar{q}ZZ$	1
$t\bar{t}V$	1





Producing Asimov with NSBI-reweighting of *Ref*

Reference dataset → defined as ggF S + EW SBI10



- Reference dataset chosen to have no negative weights $\dfrac{p(x_i|\mu)}{p_{\text{ref}}(x_i)}$

- NSBI estimated density ratios →

$$p_{\text{ref}}(x_i) \to p^{\text{Asimov}}_{\text{NSBI-rewgt}}(x_i|\mu) = \frac{\nu_{SR}}{\nu_{\text{Ref}}} \frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} \cdot p_{\text{ref}}(x_i)$$

- Instead of using X (with neg weight), us Ref reweighted to X ➔ same distributions, but no negative weights!

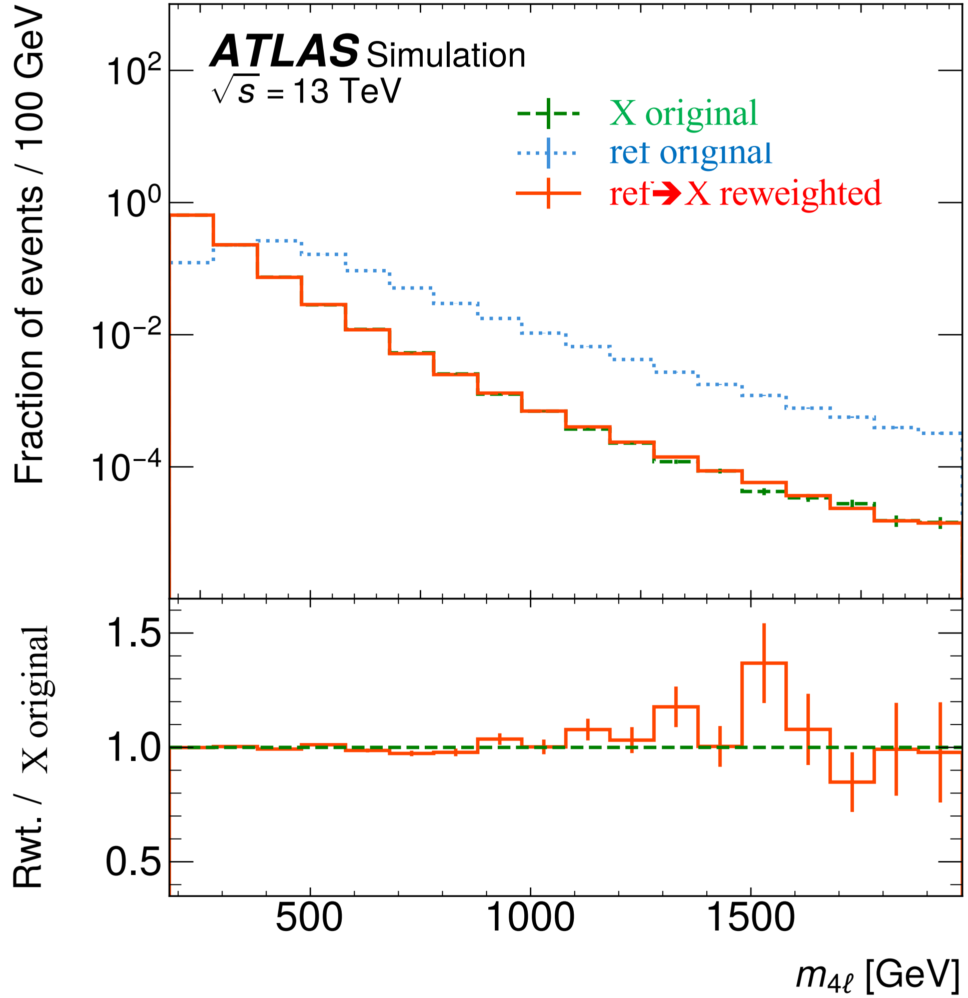
Reminder Calibration check 2

$$s(x_i) = \frac{p_X}{p_X + p_{ref}}(x_i)$$

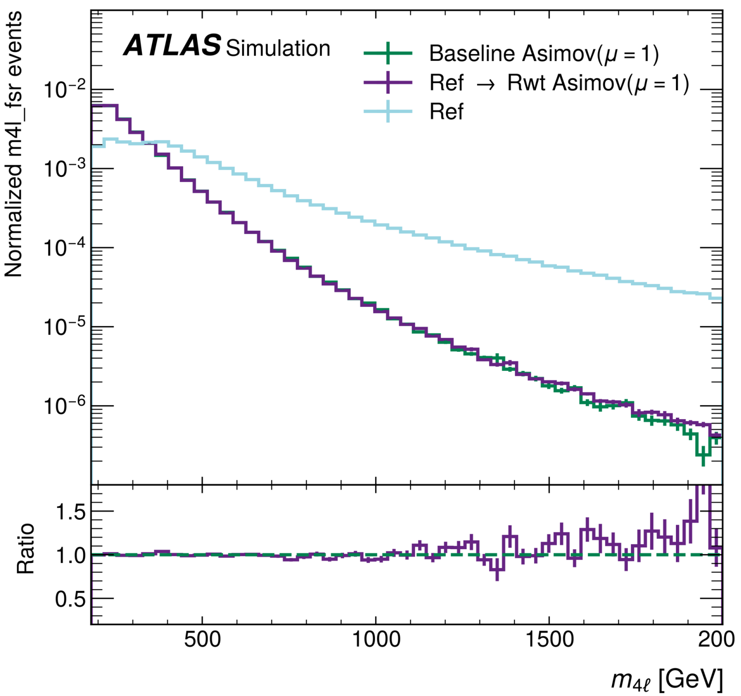
➔

$$\frac{p_X}{p_{ref}}(x_i) = \frac{s}{1-s}(x_i)$$

- Reweighting with p_X/p_{ref}, one **can turn ref dataset into X dataset**
- Ref chosen without neg weight=>ref➔X same distributions as X but without negative weights
- ==> use it for Neyman Construction

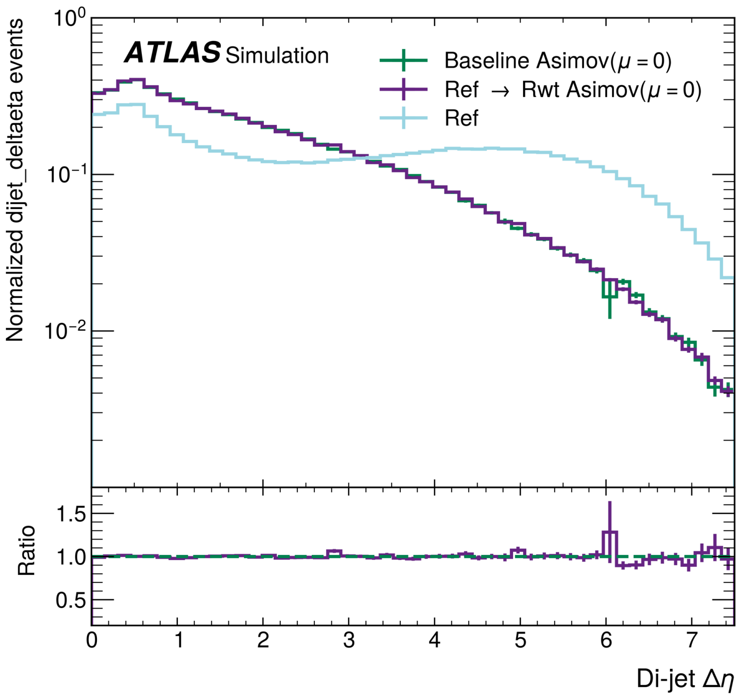


NSBI-reweighted Asimov datasets



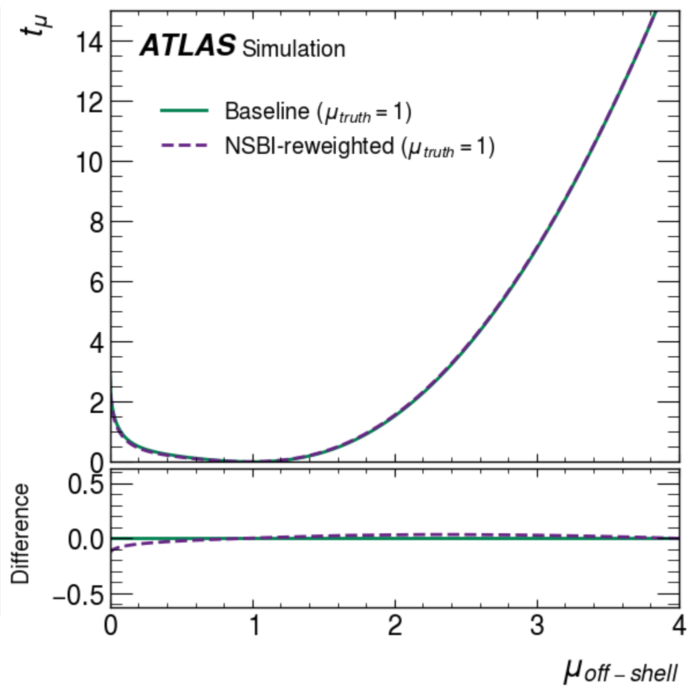
Hypothesis $\mu_{truth} = 1$

Kinematic comparison along m4l
(used for NN training)



Hypothesis $\mu_{truth} = 0$

Kinematic comparison along di-jet $\Delta\eta_{jj}$
(**NOT** used for NN training)



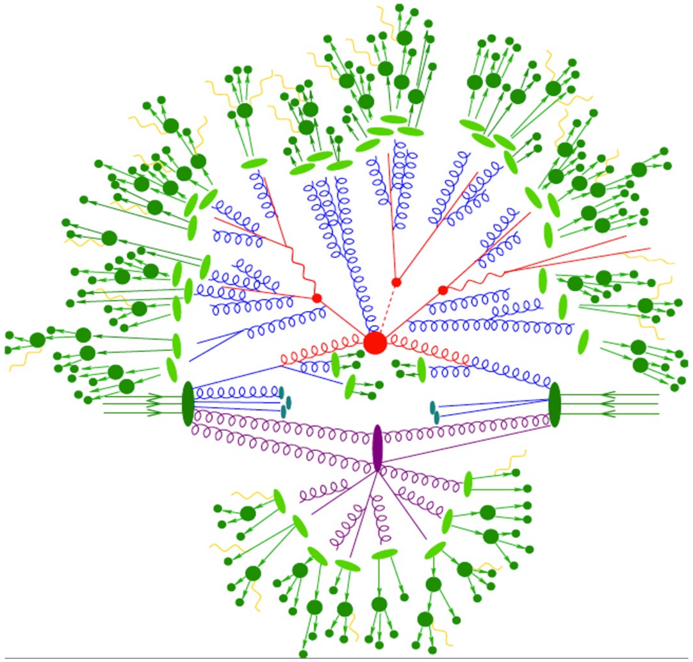
Hypothesis $\mu_{truth} = 1$

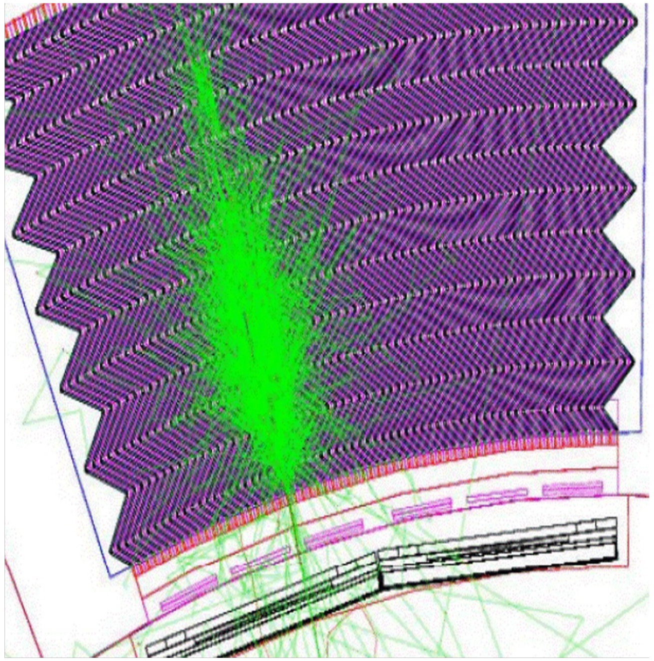
Comparison of NLL curves

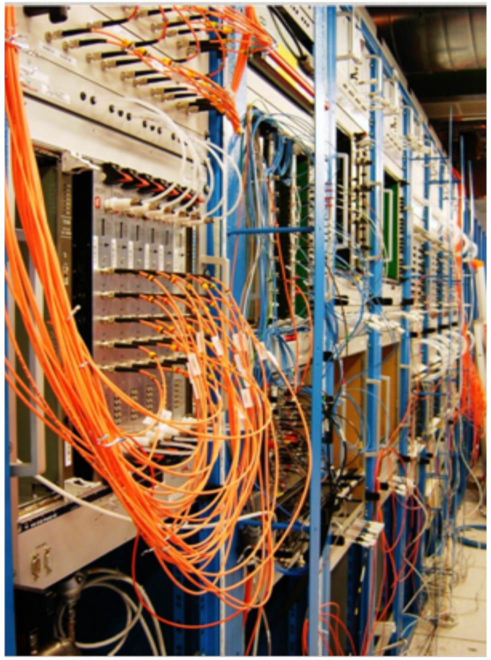
Systematic Uncertainties

Systematic Uncertainties

(also called epistemic uncertainties in Computer Science)







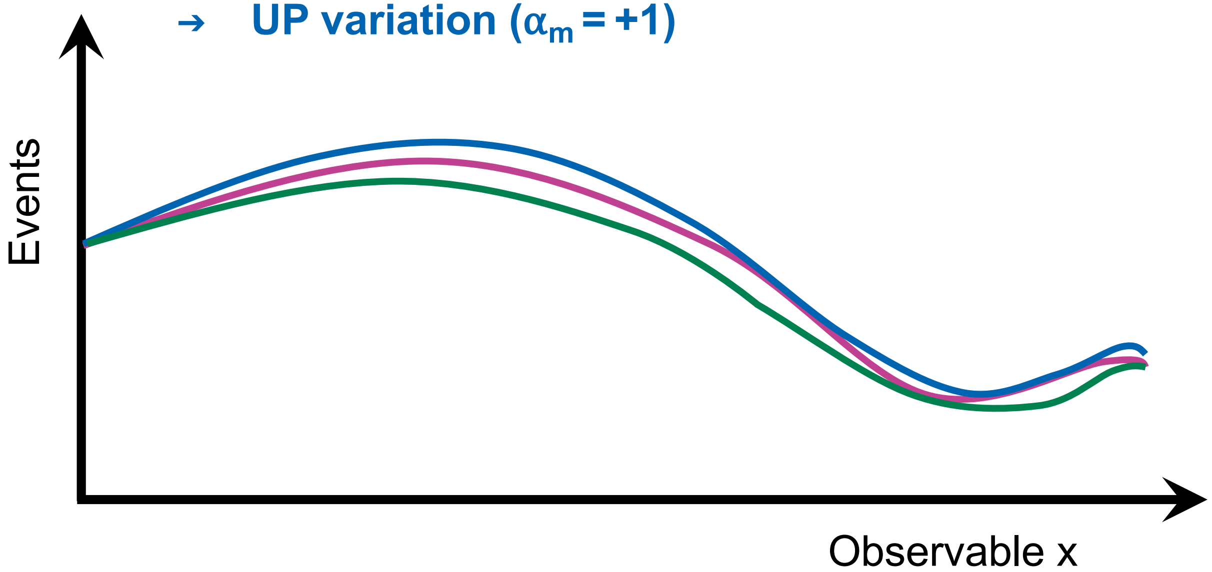
Systematic Uncertainties (Epistemic)

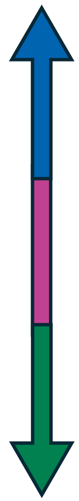
- There are experimental and theoretical uncertainties.
- In our probability model, systematic uncertainties are modelled by Nuisance Parameters α

  → **DOWN variation ($\alpha_m = -1$)**

  → **Nominal ($\alpha_m = 0$)**

  → **UP variation ($\alpha_m = +1$)**





UP variation

$$W_X(\alpha_m^+) = \frac{\nu_X(\alpha_m = +1)}{\nu_X}$$

Nominal

DOWN variation

$$W_X(\alpha_m^-) = \frac{\nu_X(\alpha_m = -1)}{\nu_X}$$

Systematic Uncertainties (Epistemic)

Yield term: gives the factor of change in yield due to the NP

The density ration can be modified with Nuisance parameter NP,

Which then can be profiled

$$\frac{p(x_i|\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu,\alpha)} \sum_X^C f_X(\mu)\cdot \nu_X \cdot \frac{p_X(x_i)}{p_{ref}(x_i)} \prod_k^{N_{syst}} W_X(\alpha_m)\, g_X(x_i,\alpha_m)$$

Gives the shape change due to the NP

Systematics - Interpolation

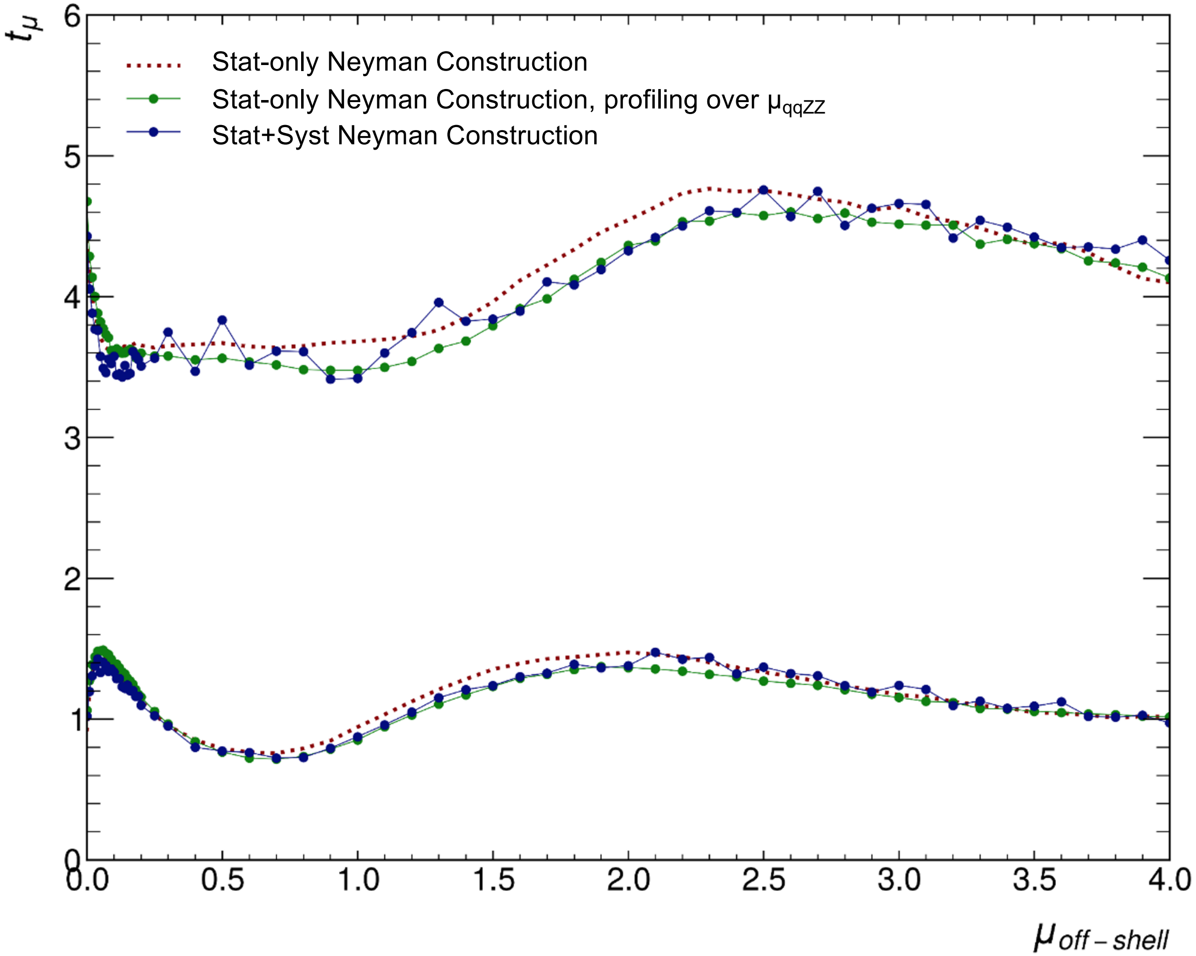
- The G(α) is the then fitted from ratio of the yields
- gx(α +) and gx(α -) is density ratio from NN Classifiers (+/- vs nominal)
- g(α) is a polynomial fit between them

$$W_X(\alpha_m) = \begin{cases} \left(\frac{\nu_X(\alpha_m{}^{(+)})}{\nu_X(\alpha_m{}^{0})}\right)^{\alpha_m}, & \text{if } \alpha_m > 1 \\ 1 + \sum_{n=1}^{3} c_n {\alpha^n}_k, & \text{if } -1 \leq \alpha_m \leq 1 \\ \left(\frac{\nu_X(\alpha_m{}^{(-)})}{\nu_X(\alpha_m{}^{0})}\right)^{-\alpha_m}, & \text{if } \alpha_m < -1 \end{cases}$$

Estimated using a NN Classifier

$$g_X(x_i, \alpha_m) = \begin{cases} g_X(x_i, \alpha_m{}^{(+)})^{\alpha_m}, & \text{if } \alpha_m > 1 \\ 1 + \sum_{n=1}^{3} c_n {\alpha^n}_k, & \text{if } -1 \leq \alpha_m \leq 1 \\ g_X(x_i, \alpha_m{}^{(-)})^{-\alpha_m}, & \text{if } \alpha_m < -1 \end{cases}$$

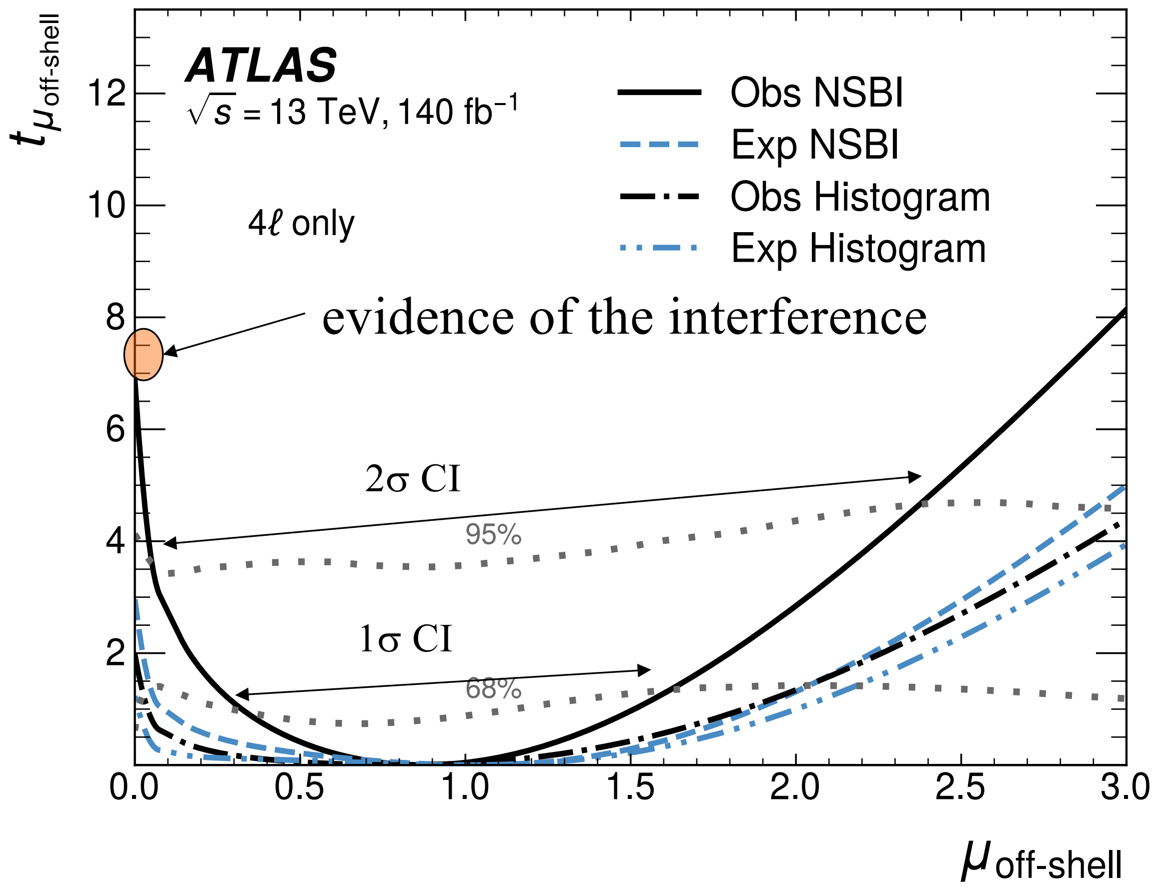
Neyman Construction for Stat+Syst



- Similar Confidence Intervals
- Stat-only CIs are sufficient
- Stat+Syst verification useful nevertheless

Results

$\mu_{\text{off-shell}} = 0.87^{+0.75}_{-0.54}$ With Systematic/epistemic effects



« If you need sophisticated statistics techniques, what you really need is more data »

Famous quote, but…

Conclusion

- The first implementation of Neural Simulation Based Inference (NSBI) in Particle Physics (not just a proof-of-concept) arXiv:2412.01548 arXiv:2412.01600
- Work started in 2020….first NSBI PoC papers 2016 (Louppe 2016)….
- NSBI: well-suited for similar problems with non linear impact of the parameter of interest (λ_{HHH} in diHiggs studies, SMEFT…)
- Challenges:
  - Proper likelihood parameterisation as a function of samples
  - Choice of reference sample
  - Handling low MC stat
  - NN model with calibration/sensitivity compromise
- Reading material :Kyle Cranmer's arXiv:2010.06439 Jay Sandesara's 4 courses at PhyStat 2025 and NSBI framework using Fair Universe HiggsML Htautau with systematics open dataset