AUSTRIAN ACADEMY OF SCIENCES

HEPHY INSTITUTE OF HIGH ENERGY PHYSICS

Unbinned inclusive cross-section measurements with machine-learned systematic uncertainties

(a.k.a. Gollum goes Higgs Uncertainty Challenge)

Claudius Krause

GOLLUM

2505.05544

ÖAW AUSTRIAN ACADEMY OF SCIENCES

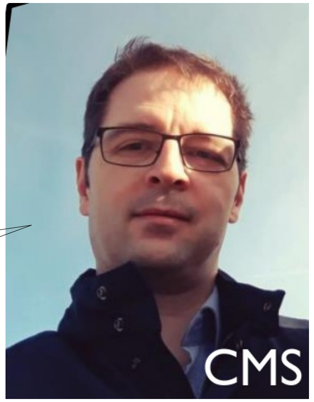
HEPHY INSTITUTE OF HIGH ENERGY PHYSICS

It all started in November 2024 ...

Claudius Krause

We should use your unbinned method [2406.19076] for that challenge …

Tell me more!



Robert Schöfbeck

Daohan Wang

Dennis Schwarz

Lisa Benato

Ang Li

Cristina Giordano

Maryam Shooshtari

Joint venture of the CMS Data Analysis group & Machine learning group @ HEPHY (Vienna)

Gollum goes Higgs Uncertainty Challenge

1) FAIR Universe Higgs Uncertainty Challenge

2) "Guaranteed Optimal
Log-Likelihood-based
Unbinned Method"
(GOLLUM)

3) Results

The FAIR Universe Project https://fair-universe.lbl.gov/

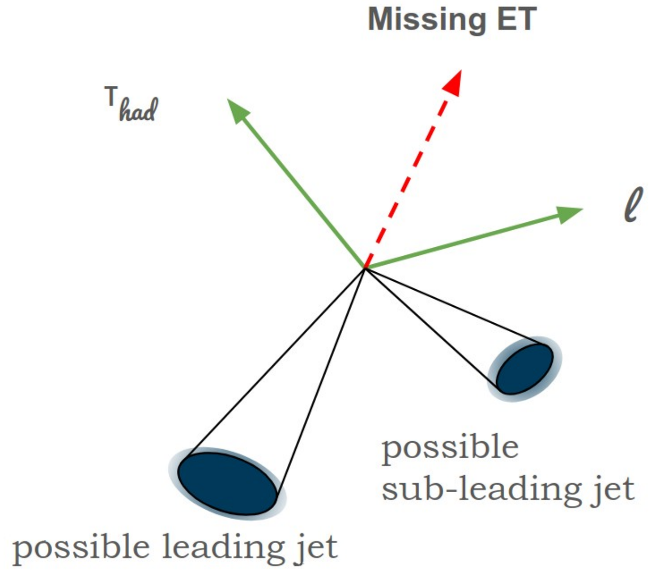
- 3 year US Dept. of Energy, AI for HEP project. (U Berkeley, U Washington and Chalearn)
  - Provide an open, large-compute-scale AI ecosystem for sharing datasets, training large models, fine-tuning those models, and hosting challenges and benchmarks.
  - Organize a challenge series, progressively rolling in tasks of increasing difficulty, based on novel datasets.
  - Tasks will focus on measuring and minimizing the effects of systematic uncertainties in HEP (particle physics and cosmology).

- Broad team in HEP, ML and computing involved in several previous challenges and benchmarks for HEP (e.g. HiggsML and TrackML, LHC Olympics, Fast Calorimeter Simulation Challenge) and wider (e.g NeurIPS competition track, MLPerf HPC); as well as Uncertainty aware learning in HEP

- Now: Fair Universe HiggsML Uncertainty Challenge, a NeurIPS 2024 competition

Fair Universe: HiggsML Uncertainty Challenge

Extension of previous HiggsML challenge from 2014 (a classification problem for Higgs decaying to Tau leptons based on final state 3-momenta and derived quantities)



⇒ new Fair Universe dataset, with following improvements

- Use (much) faster simulation, less accurate (but does not matter)
- Numbers of events 800.000 ⇒ ~300 millions
- Parametrised systematics (Nuisance Parameters)

Composed of Signal (H→ττ) and three backgrounds (Z→ττ, tt, VV)

Task : given a pseudo-experiment with given amount of signal, provide a Confidence Interval on signal strength.

2410.02867

Run as an official NeurIPS 2024 competition till 14th March 2025

Fair Universe: HiggsML Uncertainty Challenge

28 features:

Symbol	Description	Symbol	Description
$p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$	Transverse momentum of the τ_{had}	p_{T}^{ℓ}	Transverse momentum of the lepton
$\eta^{\tau_{\mathrm{had}}}$	Pseudorapidity of the τ_{had}	η^{ℓ}	Pseudorapidity of the lepton
$\phi^{\tau_{\mathrm{had}}}$	Azimuthal angle of the τ_{had}	ϕ^{ℓ}	Azimuthal angle of the lepton
$\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse momentum	ϕ^{miss}	Azimuthal angle of missing transverse momentum
$p_{\mathrm{T}}^{j_1}$	Transverse momentum of the leading jet	$p_{\mathrm{T}}^{j_2}$	Transverse momentum of the subleading jet
η^{j_1}	Pseudorapidity of the leading jet	η^{j_2}	Pseudorapidity of the subleading jet
ϕ^{j_1}	Azimuthal angle of the leading jet	ϕ^{j_2}	Azimuthal angle of the subleading jet
N_j	Number of reconstructed jets	$\sum_{\mathrm{jets}} p_{\mathrm{T}}$	Sum of transverse momenta of all jets
$m_{\mathrm{T}}(\ell, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$	Transverse mass of lepton and missing p_{T}	m_{vis}	Visible invariant mass of τ_{had} and ℓ
$p_{\mathrm{T}}^{\mathrm{H}}$	Modulus of vector sum of $\tau_{\mathrm{had}}, \ell, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}$	$m^{j_1 j_2}$	Invariant mass of the two leading jets
$\Delta\eta^{j_1 j_2}$	Pseudorapidity separation of leading jets	$\eta^{j_1} \cdot \eta^{j_2}$	Product of leading jet pseudorapidities
$p_{\mathrm{T}}^{\mathrm{tot}}$	Modulus of vector sum of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$, $p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$, p_{T}^{ℓ}, $p_{\mathrm{T}}^{j_1}$, $p_{\mathrm{T}}^{j_2}$	$\sum p_{\mathrm{T}}$	Scalar sum of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$, $p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$, p_{T}^{ℓ}, and all jets
C_{ϕ}^{miss}	Azimuthal centrality of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ w.r.t. $\tau_{\mathrm{had}}, \ell$	C_{η}^{ℓ}	Pseudorapidity centrality of lepton w.r.t. jets
$\Delta R(\tau_{\mathrm{had}}, \ell)$	Angular separation between τ_{had} and ℓ	$p_{\mathrm{T}}^{\ell}/p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$	Ratio of transverse momenta of lepton and τ_{had}

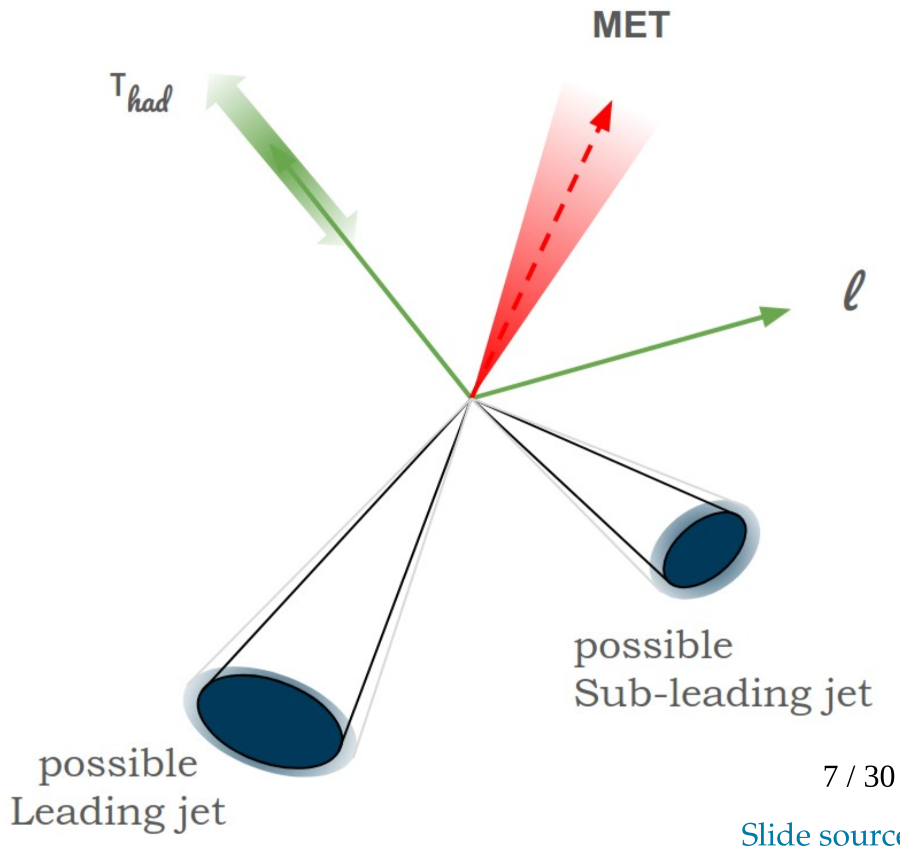
Systematic Uncertainties: parameterized by nuisance parameters ν

Three primary Feature distortions (and correlated impact on derived features):

- Tau Energy Scale (and correlated MET)
- Jet Energy Scale (and correlated MET)
- Additional randomized Soft MET

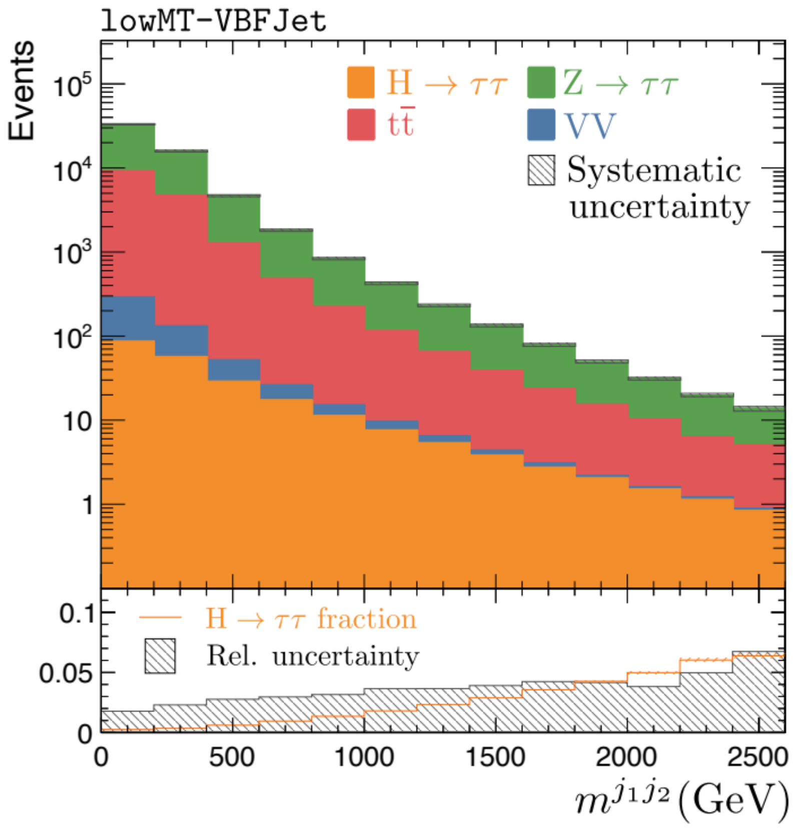
Three Event category normalizations

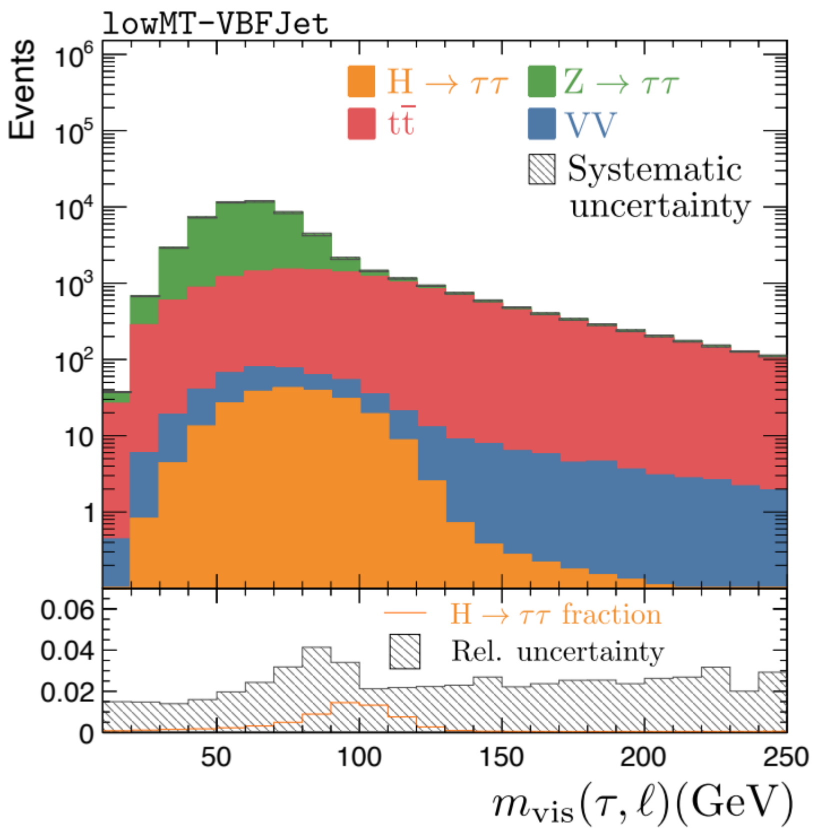
- Overall background normalization
- Di-boson background normalization
- ttbar background normalization

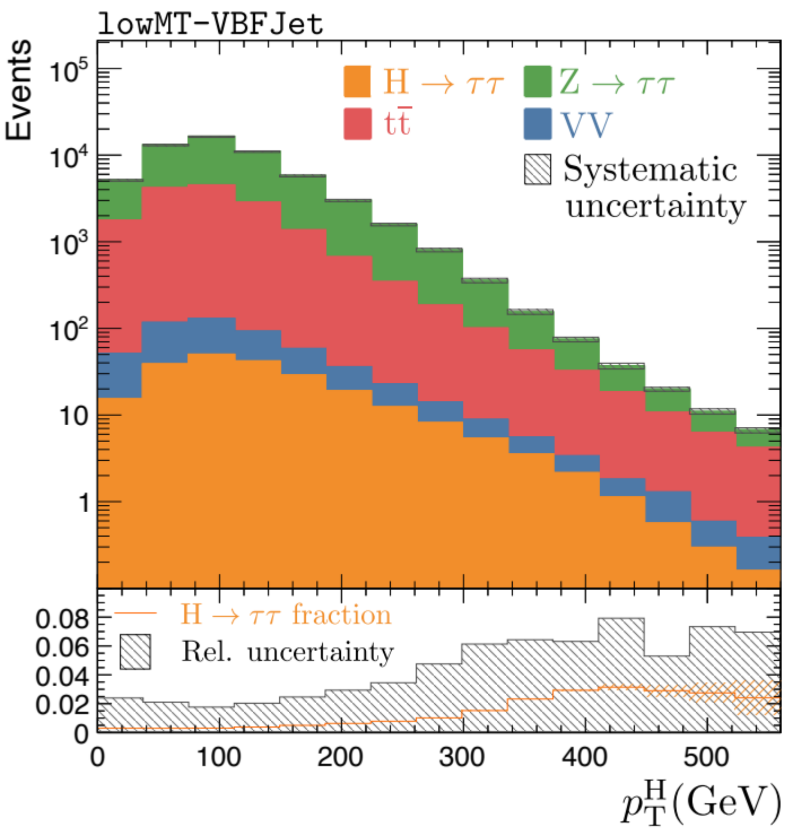


Slide source

The Signal is buried in the systematics!







Gollum goes Higgs Uncertainty Challenge

1) FAIR Universe Higgs Uncertainty Challenge

2) "Guaranteed Optimal Log-Likelihood-based Unbinned Method" (GOLLUM)

3) Results

How can we extract the signal from this dataset?

We divide and conquer!

How can we extract the signal from this dataset?

We divide and conquer!

- Apply physics domain knowledge
- Break it down to smaller pieces
- Use analytic formulas when possible

How can we extract the signal from this dataset?

We divide and conquer!

- Apply physics domain knowledge
- Break it down to smaller pieces
- Use analytic formulas when possible

- Base everything on likelihood test statistic
- Use Machine Learning in remaining parts
- Use many small instead of one big network

First, we divide phase space in signal and control regions

Region	Requirements	Type	Poisson yield $\mathcal{L}\sigma$ $H \to \tau\tau$	$Z \to \tau\tau$	$t\bar{t}$	VV	S/B	
`lowMT-VBFJet`	$p_T^{j_1} > 50$ GeV $p_T^{j_2} > 30$ GeV $m_T \leq 70$ GeV	UB, SR	225.8	41 280.4	15 425.9	313.4	$3.96 \cdot 10^{-3}$	→ VBF SR
`highMT-VBFJet`	$p_T^{j_1} > 50$ GeV $p_T^{j_2} > 30$ GeV $m_T > 70$ GeV	UB, CR	14.7	721.7	16 768.6	193.2	$8.30 \cdot 10^{-4}$	
`lowMT-noVBFJet-ptH100`	$p_T^{H} > 100$ GeV $m_T \leq 70$ GeV veto on `VBFJet`	UB, SR	57.0	17 379.8	674.6	79.0	$3.14 \cdot 10^{-3}$	→ ggH SR
`lowMT-noVBFJet-ptH0to100`	$p_T^{H} \leq 100$ GeV $m_T \leq 70$ GeV veto on `VBFJet`	CR	642.5	837 928.7	3 360.1	1 438.5	$7.62 \cdot 10^{-4}$	→ 90% of data, bkg CR
`highMT-noVBFJet`	$m_T > 70$ GeV veto on `VBFJet`	CR	26.0	3 826.9	5 054.3	1 409.4	$2.53 \cdot 10^{-3}$	

Then, we define our test statistic

Profile log-likelihood test statistic

$$q_{\mu}(\mathcal{D}) = -2\log\frac{\max_{\nu} L(\mathcal{D}|\mu,\nu)}{\max_{\mu,\nu} L(\mathcal{D}|\mu,\nu)}$$

Then, we define our test statistic

Profile log-likelihood test statistic

$$q_{\mu}(\mathcal{D}) = -2\log\frac{\max_{\boldsymbol{\nu}} \quad L(\mathcal{D}|\mu,\boldsymbol{\nu})}{\max_{\mu,\boldsymbol{\nu}} L(\mathcal{D}|\mu,\boldsymbol{\nu})}$$

Extended likelihood (for each selection)

$$L(\mathcal{D}|\mu,\boldsymbol{\nu}) = \mathrm{P}(N_{\mathrm{obs}}|\mathcal{L}\sigma(\mu,\boldsymbol{\nu}))\prod_{i=1}^{N_{\mathrm{obs}}} p(\boldsymbol{x}_i|\mu,\boldsymbol{\nu})$$

Then, we define our test statistic

Profile log-likelihood test statistic

$$q_{\mu}(\mathcal{D}) = -2\log\frac{\max_{\boldsymbol{\nu}} \quad L(\mathcal{D}|\mu,\boldsymbol{\nu})}{\max_{\mu,\boldsymbol{\nu}} L(\mathcal{D}|\mu,\boldsymbol{\nu})}$$

$$p(\boldsymbol{x}|\mu,\boldsymbol{\nu}) = \frac{1}{\sigma(\mu,\boldsymbol{\nu})}\frac{\mathrm{d}\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu})}{\mathrm{d}\boldsymbol{x}}$$

Extended likelihood (for each selection)

$$L(\mathcal{D}|\mu,\boldsymbol{\nu}) = \mathrm{P}(N_{\mathrm{obs}}|\mathcal{L}\sigma(\mu,\boldsymbol{\nu}))\prod_{i=1}^{N_{\mathrm{obs}}} p(\boldsymbol{x}_i|\mu,\boldsymbol{\nu})$$

Then, we define our test statistic

Profile log-likelihood test statistic

$$q_\mu(\mathcal{D}) = -2\log\frac{\max_{\boldsymbol{\nu}} \; L(\mathcal{D}|\mu,\boldsymbol{\nu})}{\max_{\mu,\boldsymbol{\nu}} L(\mathcal{D}|\mu,\boldsymbol{\nu})}$$

$$p(\boldsymbol{x}|\mu,\boldsymbol{\nu}) = \frac{1}{\sigma(\mu,\boldsymbol{\nu})}\frac{\mathrm{d}\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu})}{\mathrm{d}\boldsymbol{x}}$$

Extended likelihood (for each selection)

$$L(\mathcal{D}|\mu,\boldsymbol{\nu}) = \mathrm{P}(N_{\mathrm{obs}}|\mathcal{L}\sigma(\mu,\boldsymbol{\nu}))\prod_{i=1}^{N_{\mathrm{obs}}} p(\boldsymbol{x}_i|\mu,\boldsymbol{\nu})$$

… and its (log-) ratio

$$\log\frac{L(\mathcal{D}|\mu,\boldsymbol{\nu})}{L(\mathcal{D}|1,\mathbf{0})} = -\mathcal{L}(\sigma(\mu,\boldsymbol{\nu})-\sigma(1,\mathbf{0})) + \sum_{i=1}^{N_{\mathrm{obs}}}\log\left(\frac{\mathrm{d}\sigma(\boldsymbol{x}_i|\mu,\boldsymbol{\nu})}{\mathrm{d}\sigma(\boldsymbol{x}_i|1,\mathbf{0})}\right)$$

Then, we define our test statistic

Profile log-likelihood test statistic

$$q_{\mu}(\mathcal{D}) = -2\log\frac{\max_{\boldsymbol{\nu}} \quad L(\mathcal{D}|\mu,\boldsymbol{\nu})}{\max_{\mu,\boldsymbol{\nu}} L(\mathcal{D}|\mu,\boldsymbol{\nu})}$$

$$p(\boldsymbol{x}|\mu,\boldsymbol{\nu}) = \frac{1}{\sigma(\mu,\boldsymbol{\nu})}\frac{\mathrm{d}\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu})}{\mathrm{d}\boldsymbol{x}}$$

Extended likelihood (for each selection)

$$L(\mathcal{D}|\mu,\boldsymbol{\nu}) = \mathrm{P}(N_{\mathrm{obs}}|\mathcal{L}\sigma(\mu,\boldsymbol{\nu}))\prod_{i=1}^{N_{\mathrm{obs}}} p(\boldsymbol{x}_i|\mu,\boldsymbol{\nu})$$

… and its (log-) ratio

$$\log\frac{L(\mathcal{D}|\mu,\boldsymbol{\nu})}{L(\mathcal{D}|1,\mathbf{0})} = -\mathcal{L}(\sigma(\mu,\boldsymbol{\nu}) - \sigma(1,\mathbf{0})) + \sum_{i=1}^{N_{\mathrm{obs}}}\log\left(\boxed{\frac{\mathrm{d}\sigma(\boldsymbol{x}_i|\mu,\boldsymbol{\nu})}{\mathrm{d}\sigma(\boldsymbol{x}_i|1,\mathbf{0})}}\right)$$

Differential Cross section Ratio: DCR

And we re-write it in terms of known quantities

The DCR splits into signal and background components

$$\frac{d\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})} = \mu\frac{d\sigma_{\mathrm{H}}(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})} + (1+\alpha_{\mathrm{bkg}})^{\nu_{\mathrm{bkg}}}\left(\frac{d\sigma_{\mathrm{Z}}(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})}\right.$$
$$\left. + (1+\alpha_{\mathrm{t\bar{t}}})^{\nu_{\mathrm{t\bar{t}}}}\frac{d\sigma_{\mathrm{t\bar{t}}}(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})} + (1+\alpha_{\mathrm{VV}})^{\nu_{\mathrm{VV}}}\frac{d\sigma_{\mathrm{VV}}(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})}\right)$$

with analytic dependence on the normalization-type nuisances and signal strength μ.

And we re-write it in terms of known quantities

The DCR splits into signal and background components

$$\frac{d\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})} = \mu\frac{d\sigma_{\mathrm{H}}(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})} + (1+\alpha_{\mathrm{bkg}})^{\nu_{\mathrm{bkg}}}\left(\frac{d\sigma_{\mathrm{Z}}(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})}\right.$$
$$\left. + (1+\alpha_{\mathrm{t\bar{t}}})^{\nu_{\mathrm{t\bar{t}}}}\frac{d\sigma_{\mathrm{t\bar{t}}}(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})} + (1+\alpha_{\mathrm{VV}})^{\nu_{\mathrm{VV}}}\frac{d\sigma_{\mathrm{VV}}(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})}\right)$$

with analytic dependence on the normalization-type nuisances and signal strength μ.

The calibration-type nuisances are tackled by inserting a “factor 1”:

$$\frac{d\sigma_p(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})} = \frac{d\sigma_p(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})}{d\sigma_p(\boldsymbol{x}|\mathbf{0})}\frac{d\sigma_p(\boldsymbol{x}|\mathbf{0})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})}$$
$$\simeq \hat{S}_p(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})\hat{g}_p(\boldsymbol{x})$$

And we re-write it in terms of known quantities

The DCR splits into signal and background components

$$\frac{d\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})} = \mu\frac{d\sigma_{\mathrm{H}}(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})} + (1+\alpha_{\mathrm{bkg}})^{\nu_{\mathrm{bkg}}}\Bigg(\frac{d\sigma_{\mathrm{Z}}(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})} + (1+\alpha_{\mathrm{t\bar{t}}})^{\nu_{\mathrm{t\bar{t}}}}\frac{d\sigma_{\mathrm{t\bar{t}}}(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})} + (1+\alpha_{\mathrm{VV}})^{\nu_{\mathrm{VV}}}\frac{d\sigma_{\mathrm{VV}}(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})}\Bigg)$$

with analytic dependence on the normalization-type nuisances and signal strength μ.

The calibration-type nuisances are tackled by inserting a "factor 1":

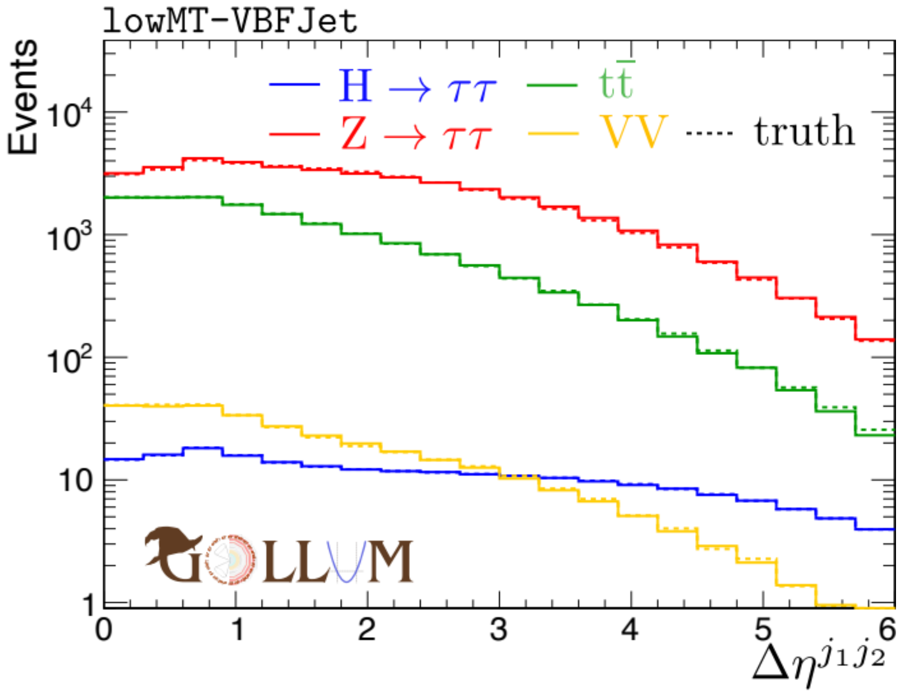
$$\frac{d\sigma_p(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})} = \frac{d\sigma_p(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})}{d\sigma_p(\boldsymbol{x}|\mathbf{0})}\frac{d\sigma_p(\boldsymbol{x}|\mathbf{0})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})} \simeq \hat{S}_p(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})\hat{g}_p(\boldsymbol{x})$$

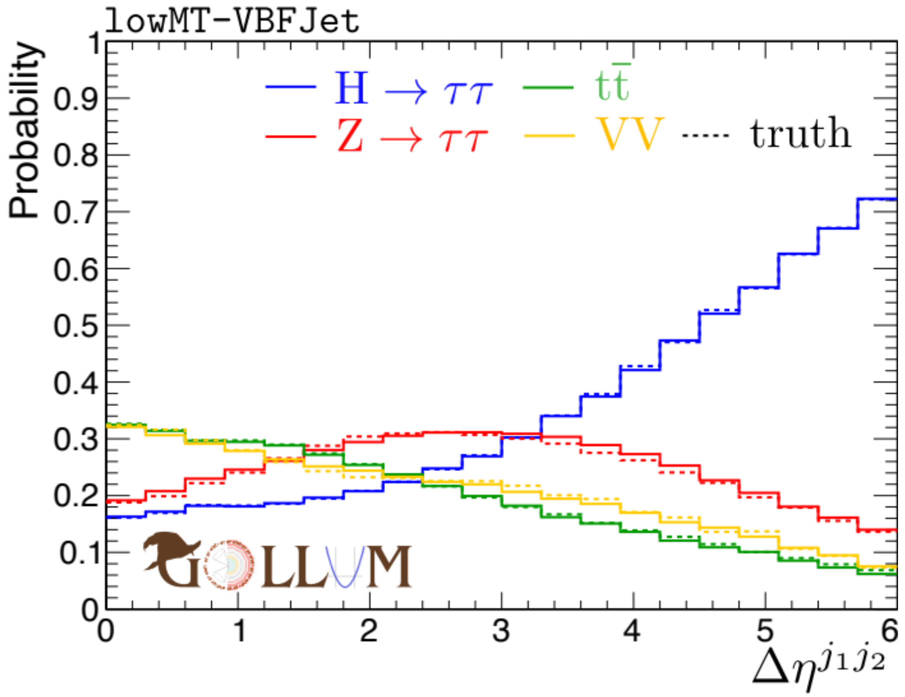
Effect of systematics per process

Share of process p on total xsec

The first ingredient to the DCR: the multi classifier $g_p(x)$

- Tensorflow NN with 28/128/128/4, ADAM, moderate L1/L2 regularization, CE loss
- Data highly imbalanced – needed class weights (based on σ)

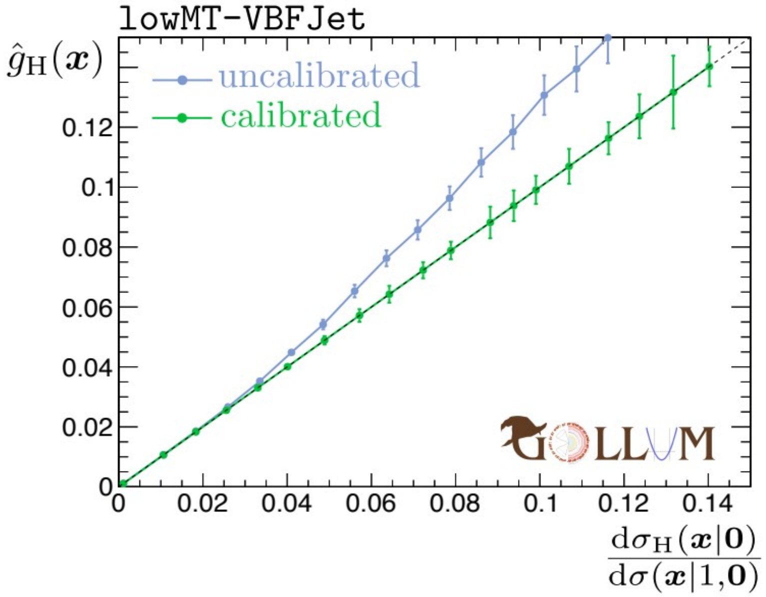


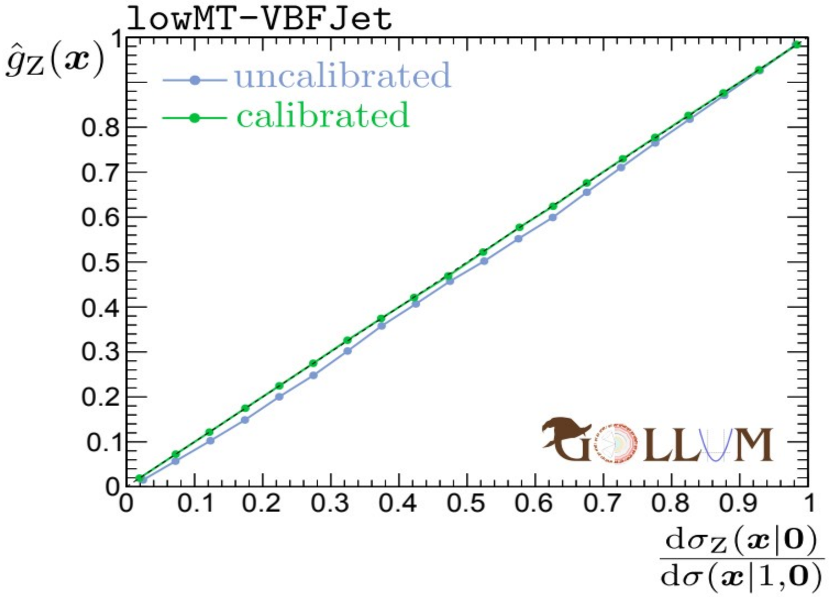


The first ingredient to the DCR: the multi classifier $g_p(x)$

Calibration is important!

- Events in the bin $0.19 < g_p(x) < 0.21$ need to have ~20% fraction of p
- But: g_H is off by ~7%, g_Z is off by -2% (gets amplified x10-x100!)

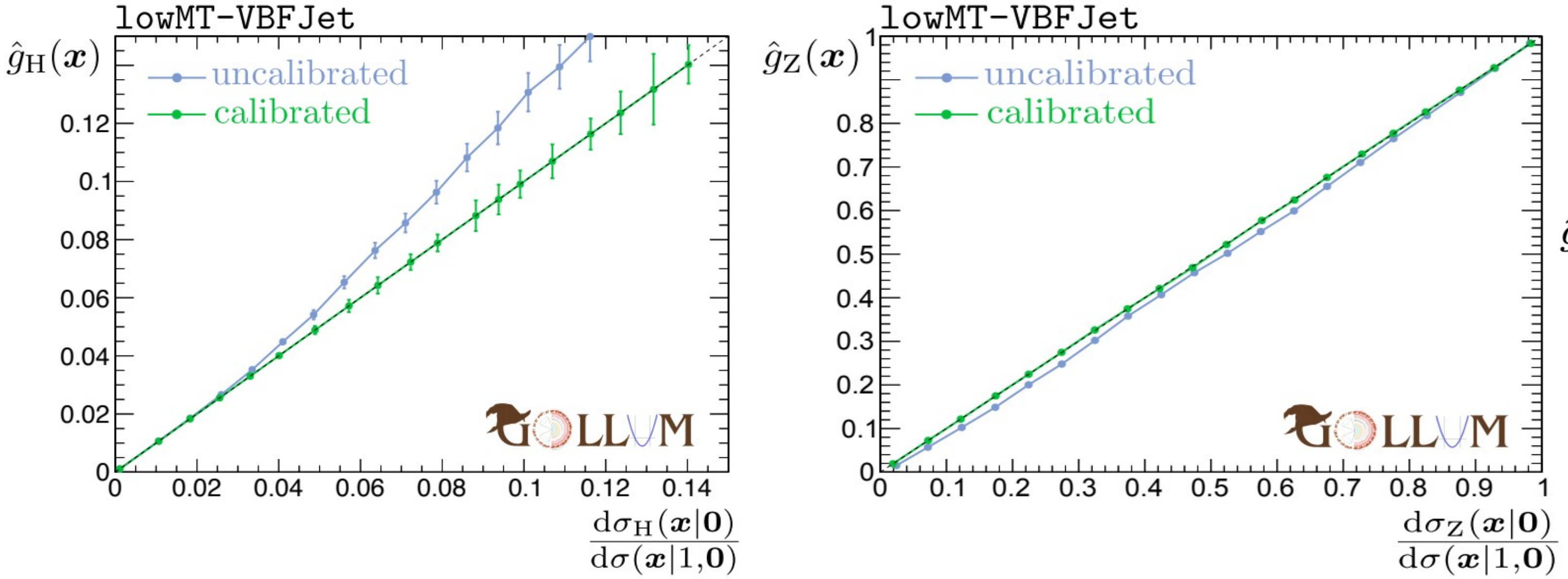




The first ingredient to the DCR: the multi classifier $g_p(x)$

Calibration is important!

- Events in the bin $0.19 < g_p(x) < 0.21$ need to have ~20% fraction of p
- But: g_H is off by ~7%, g_Z is off by -2% (gets amplified x10-x100!)



Isotonic Regression saves us

$$\hat{g}_p(\boldsymbol{x}) = \begin{cases} \mathrm{IREG}(g^*_{\mathrm{H}}(\boldsymbol{x})), & \text{if } p = \mathrm{H} \\ \frac{\mathrm{IREG}(g^*_p(\boldsymbol{x}))(1-\mathrm{IREG}(g^*_{\mathrm{H}}(\boldsymbol{x})))}{\sum_{q=\mathrm{Z},\mathrm{t\bar{t}},\mathrm{VV}} \mathrm{IREG}(g^*_q(\boldsymbol{x}))}, & \text{otherwise.} \end{cases}$$

The second ingredient to the DCR: the parametric NN $S_p(x|\nu)$

- A network trained with cross entropy loss learns the density ratio: [2406.19076]

$$f_{\mathrm{CE}}(\mathbf{x}) = \frac{1}{1 + \frac{\mathrm{d}\sigma(\mathbf{x}|\nu)}{\mathrm{d}\sigma(\mathbf{x}|0)}}$$

The second ingredient to the DCR: the parametric NN $S_p(x|\nu)$

- A network trained with cross entropy loss learns the density ratio: [2406.19076]

$$f_{\mathrm{CE}}(\mathbf{x}) = \frac{1}{1 + \frac{d\sigma(\mathbf{x}|\nu)}{d\sigma(\mathbf{x}|0)}}$$

- We make a specific Ansatz, up to log-quadratic order:

$$f_\nu(\mathbf{x}) = \frac{1}{1 + S_p(\mathbf{x}|\nu)}$$

$$\begin{aligned}\log S_p(\mathbf{x}|\nu) = {} & \nu_{\mathrm{tes}}\, \hat{\Delta}_{\mathrm{tes}}(\mathbf{x}) + \nu_{\mathrm{jes}}\, \hat{\Delta}_{\mathrm{jes}}(\mathbf{x}) + \nu_{\mathrm{met}}\, \hat{\Delta}_{\mathrm{met}}(\mathbf{x}) \\ & + \nu_{\mathrm{tes}}^2\, \hat{\Delta}_{\mathrm{tes,tes}}(\mathbf{x}) + \nu_{\mathrm{jes}}^2\, \hat{\Delta}_{\mathrm{jes,jes}}(\mathbf{x}) + \nu_{\mathrm{met}}^2\, \hat{\Delta}_{\mathrm{met,met}}(\mathbf{x}) \\ & + \nu_{\mathrm{tes}}\nu_{\mathrm{jes}}\, \hat{\Delta}_{\mathrm{tes,jes}}(\mathbf{x}) + \nu_{\mathrm{jes}}\nu_{\mathrm{met}}\, \hat{\Delta}_{\mathrm{jes,met}}(\mathbf{x}) + \nu_{\mathrm{tes}}\nu_{\mathrm{met}}\, \hat{\Delta}_{\mathrm{tes,met}}(\mathbf{x})\end{aligned}$$

ÖAW AUSTRIAN ACADEMY OF SCIENCES

HEPHY INSTITUTE OF HIGH ENERGY PHYSICS

The second ingredient to the DCR: the parametric NN $S_p(x|\nu)$

[2406.19076]

- A network trained with cross entropy loss learns the density ratio:

$$f_{\mathrm{CE}}(\mathbf{x}) = \frac{1}{1 + \frac{\mathrm{d}\sigma(\mathbf{x}|\nu)}{\mathrm{d}\sigma(\mathbf{x}|0)}}$$

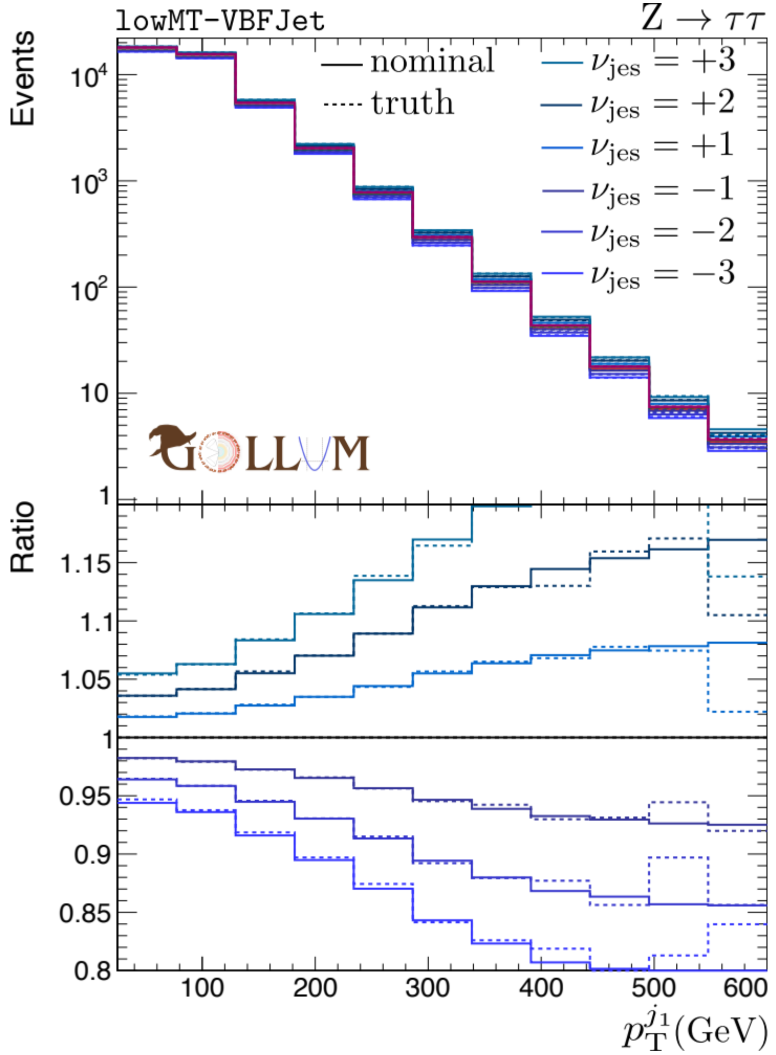
- We make a specific Ansatz, up to log-quadratic order:

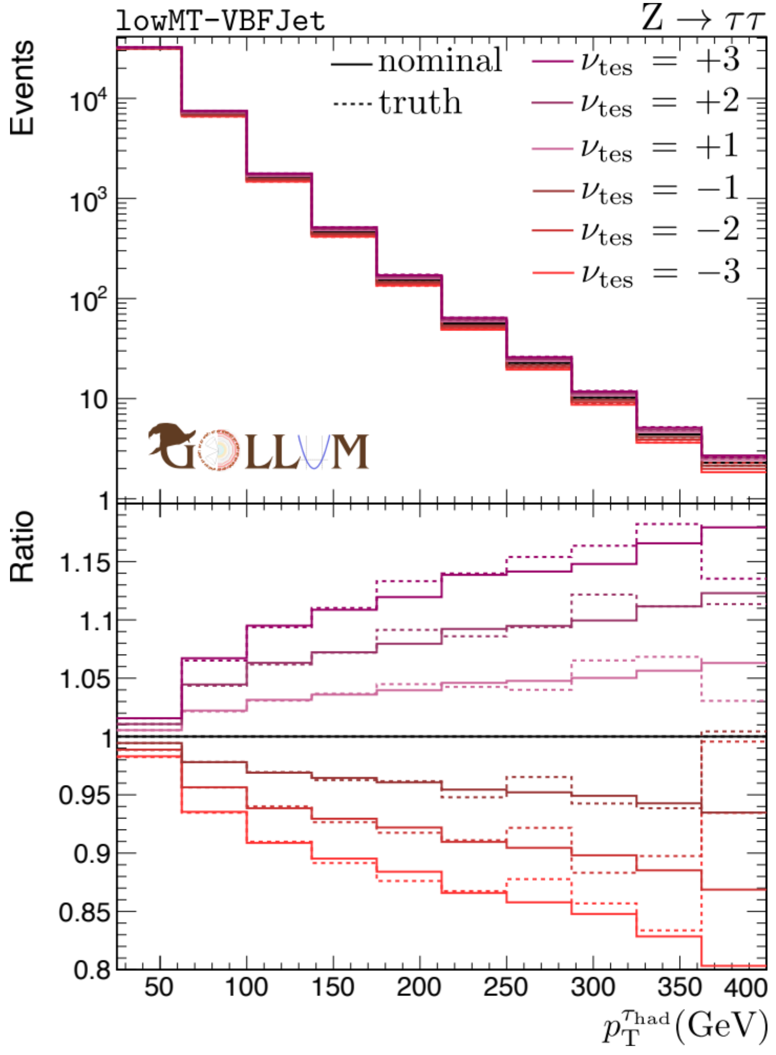
$$f_\nu(\mathbf{x}) = \frac{1}{1 + S_p(\mathbf{x}|\nu)}$$

$$\begin{aligned}\log S_p(\mathbf{x}|\nu) = {} & \nu_{\mathrm{tes}}\,\hat{\Delta}_{\mathrm{tes}}(\mathbf{x}) + \nu_{\mathrm{jes}}\,\hat{\Delta}_{\mathrm{jes}}(\mathbf{x}) + \nu_{\mathrm{met}}\,\hat{\Delta}_{\mathrm{met}}(\mathbf{x}) \\ & + \nu_{\mathrm{tes}}^2\,\hat{\Delta}_{\mathrm{tes,tes}}(\mathbf{x}) + \nu_{\mathrm{jes}}^2\,\hat{\Delta}_{\mathrm{jes,jes}}(\mathbf{x}) + \nu_{\mathrm{met}}^2\,\hat{\Delta}_{\mathrm{met,met}}(\mathbf{x}) \\ & + \nu_{\mathrm{tes}}\nu_{\mathrm{jes}}\,\hat{\Delta}_{\mathrm{tes,jes}}(\mathbf{x}) + \nu_{\mathrm{jes}}\nu_{\mathrm{met}}\,\hat{\Delta}_{\mathrm{jes,met}}(\mathbf{x}) + \nu_{\mathrm{tes}}\nu_{\mathrm{met}}\,\hat{\Delta}_{\mathrm{tes,met}}(\mathbf{x})\end{aligned}$$

- Tensorflow NN with 28 / 256 / 256 / 9, ADAM, CE loss
- Training data:
  All mixed samples with $|\nu_{\mathrm{jes}}| + |\nu_{\mathrm{tes}}| + |\nu_{\mathrm{met}}| \leq 2$, and single $|\nu_{\mathrm{i}}| \leq 3$

The second ingredient to the DCR: the parametric NN $S_p(x|\nu)$





Added bonus: this setup is refinable!

If we need to add a new background component, the existing terms remain valid.

$$d\sigma'(\boldsymbol{x}|\mu,\boldsymbol{\nu}) = d\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu}) + d\sigma_{\mathrm{new}}(\boldsymbol{x}|\boldsymbol{\nu})$$

Added bonus: this setup is refinable!

If we need to add a new background component, the existing terms remain valid.

$$\mathrm{d}\sigma'(\boldsymbol{x}|\mu,\boldsymbol{\nu}) = \mathrm{d}\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu}) + \mathrm{d}\sigma_{\mathrm{new}}(\boldsymbol{x}|\boldsymbol{\nu})$$

$$\boxed{\begin{aligned}\frac{\mathrm{d}\sigma'(\boldsymbol{x}|\mu,\boldsymbol{\nu})}{\mathrm{d}\sigma'(\boldsymbol{x}|1,\mathbf{0})} &= \frac{\mathrm{d}\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu}) + \mathrm{d}\sigma_{\mathrm{new}}(\boldsymbol{x}|\boldsymbol{\nu})}{\mathrm{d}\sigma(\boldsymbol{x}|1,\mathbf{0}) + \mathrm{d}\sigma_{\mathrm{new}}(\boldsymbol{x}|\mathbf{0})} \\ &= \frac{\frac{\mathrm{d}\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu})}{\mathrm{d}\sigma(\boldsymbol{x}|1,\mathbf{0})} + \frac{\mathrm{d}\sigma_{\mathrm{new}}(\boldsymbol{x}|\boldsymbol{\nu})}{\mathrm{d}\sigma_{\mathrm{new}}(\boldsymbol{x}|\mathbf{0})}\frac{\mathrm{d}\sigma_{\mathrm{new}}(\boldsymbol{x}|\mathbf{0})}{\mathrm{d}\sigma(\boldsymbol{x}|1,\mathbf{0})}}{1 + \frac{\mathrm{d}\sigma_{\mathrm{new}}(\boldsymbol{x}|\mathbf{0})}{\mathrm{d}\sigma(\boldsymbol{x}|1,\mathbf{0})}} \\ &\simeq \frac{\hat{R}(\boldsymbol{x}|\mu,\boldsymbol{\nu}) + \hat{S}_{\mathrm{new}}(\boldsymbol{x}|\boldsymbol{\nu})\hat{g}_{\mathrm{new}}(\boldsymbol{x})}{1 + \hat{g}_{\mathrm{new}}(\boldsymbol{x})}\end{aligned}}$$

Added bonus: this setup is refinable!

If we need to add a new background component, the existing terms remain valid.

$$\mathrm{d}\sigma'(\boldsymbol{x}|\mu,\boldsymbol{\nu}) = \mathrm{d}\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu}) + \mathrm{d}\sigma_{\mathrm{new}}(\boldsymbol{x}|\boldsymbol{\nu})$$

$$\boxed{\begin{aligned}\frac{\mathrm{d}\sigma'(\boldsymbol{x}|\mu,\boldsymbol{\nu})}{\mathrm{d}\sigma'(\boldsymbol{x}|1,\mathbf{0})} &= \frac{\mathrm{d}\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu}) + \mathrm{d}\sigma_{\mathrm{new}}(\boldsymbol{x}|\boldsymbol{\nu})}{\mathrm{d}\sigma(\boldsymbol{x}|1,\mathbf{0}) + \mathrm{d}\sigma_{\mathrm{new}}(\boldsymbol{x}|\mathbf{0})} \\ &= \frac{\frac{\mathrm{d}\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu})}{\mathrm{d}\sigma(\boldsymbol{x}|1,\mathbf{0})} + \frac{\mathrm{d}\sigma_{\mathrm{new}}(\boldsymbol{x}|\boldsymbol{\nu})}{\mathrm{d}\sigma_{\mathrm{new}}(\boldsymbol{x}|\mathbf{0})}\frac{\mathrm{d}\sigma_{\mathrm{new}}(\boldsymbol{x}|\mathbf{0})}{\mathrm{d}\sigma(\boldsymbol{x}|1,\mathbf{0})}}{1 + \frac{\mathrm{d}\sigma_{\mathrm{new}}(\boldsymbol{x}|\mathbf{0})}{\mathrm{d}\sigma(\boldsymbol{x}|1,\mathbf{0})}} \\ &\simeq \frac{\hat{R}(\boldsymbol{x}|\mu,\boldsymbol{\nu}) + \hat{S}_{\mathrm{new}}(\boldsymbol{x}|\boldsymbol{\nu})\hat{g}_{\mathrm{new}}(\boldsymbol{x})}{1 + \hat{g}_{\mathrm{new}}(\boldsymbol{x})}\end{aligned}}$$

→ no need for retraining everything, only 2 new pieces are needed!

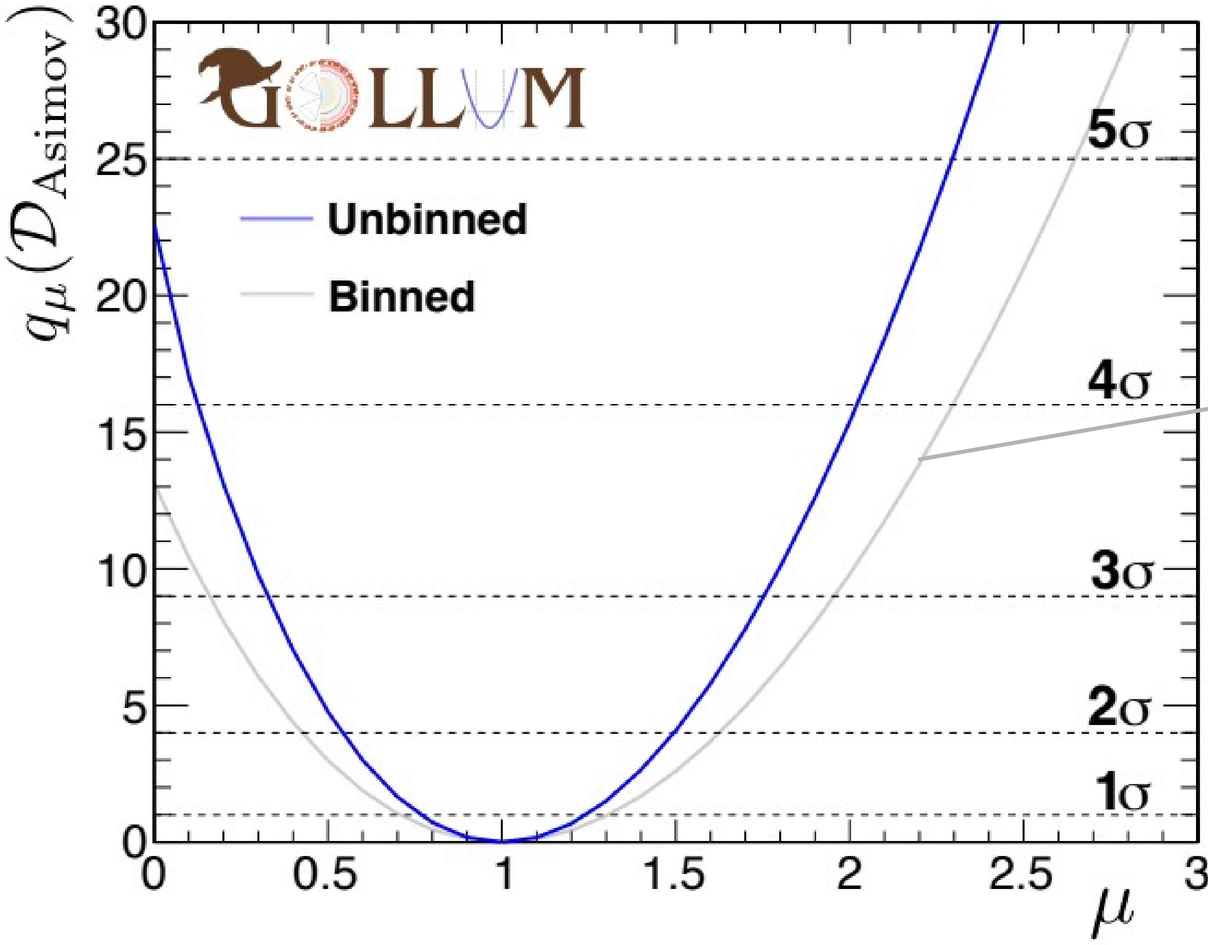
Gollum goes Higgs Uncertainty Challenge

1) FAIR Universe Higgs Uncertainty Challenge

2) "Guaranteed Optimal
Log-Likelihood-based
Unbinned Method"
(GOLLUM)

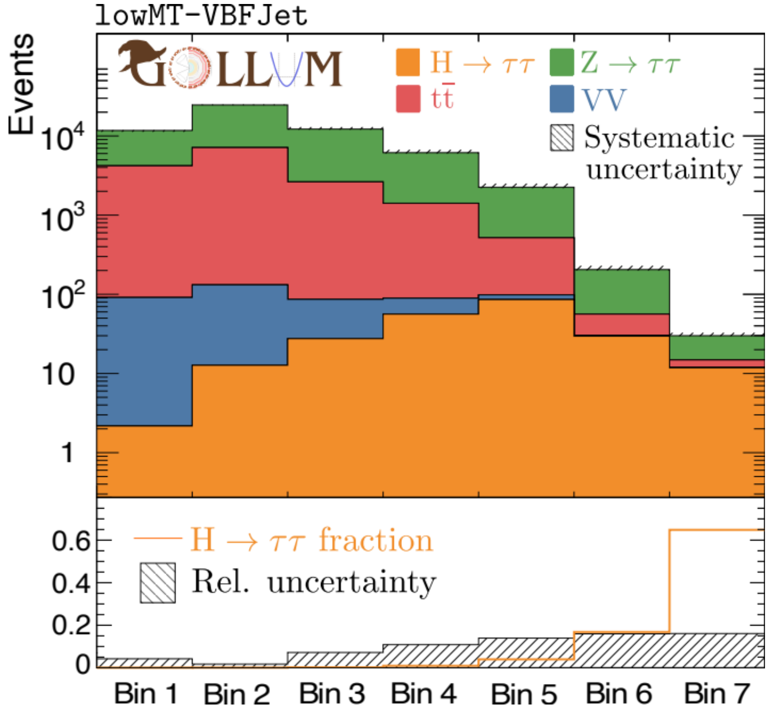
3) Results

Results I: Asimov Dataset

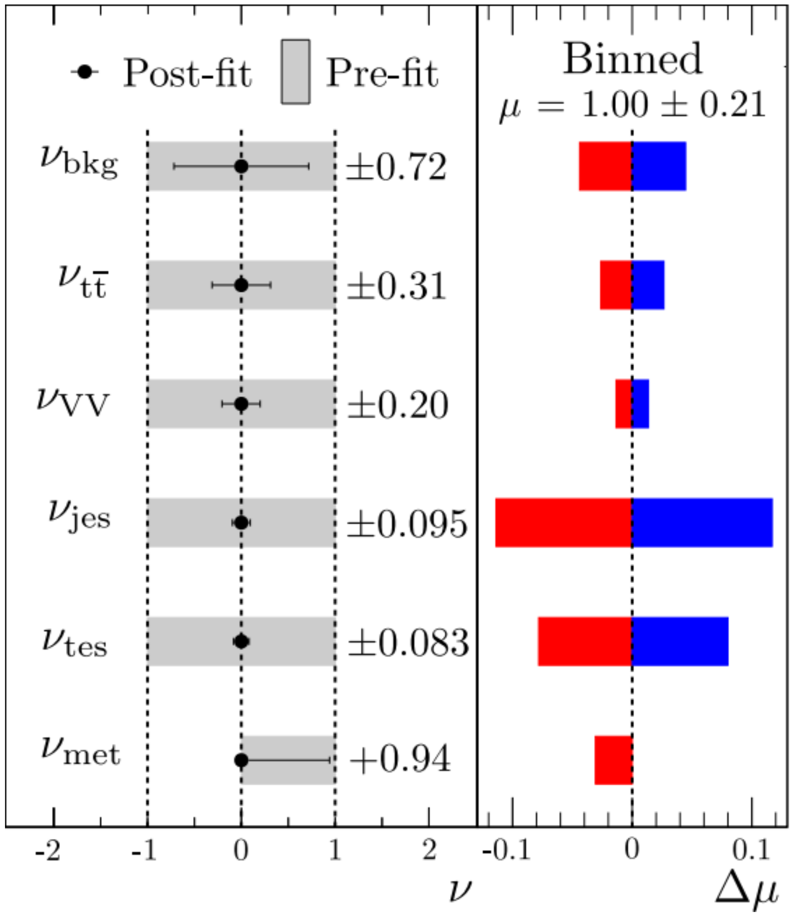


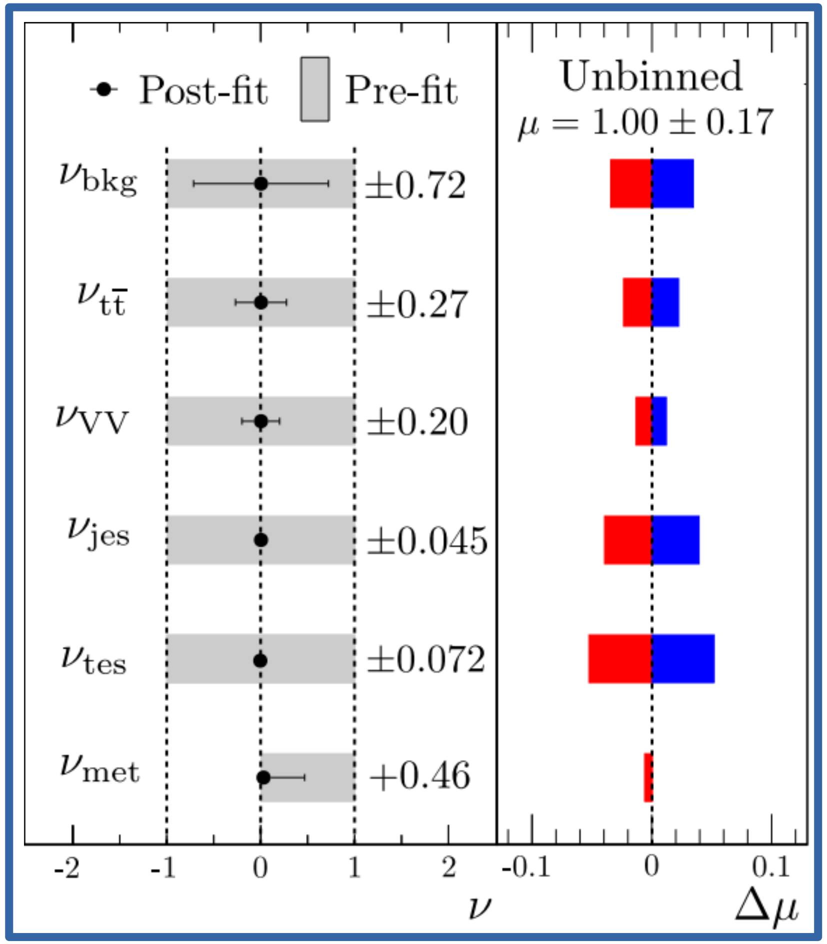
Binned reference:

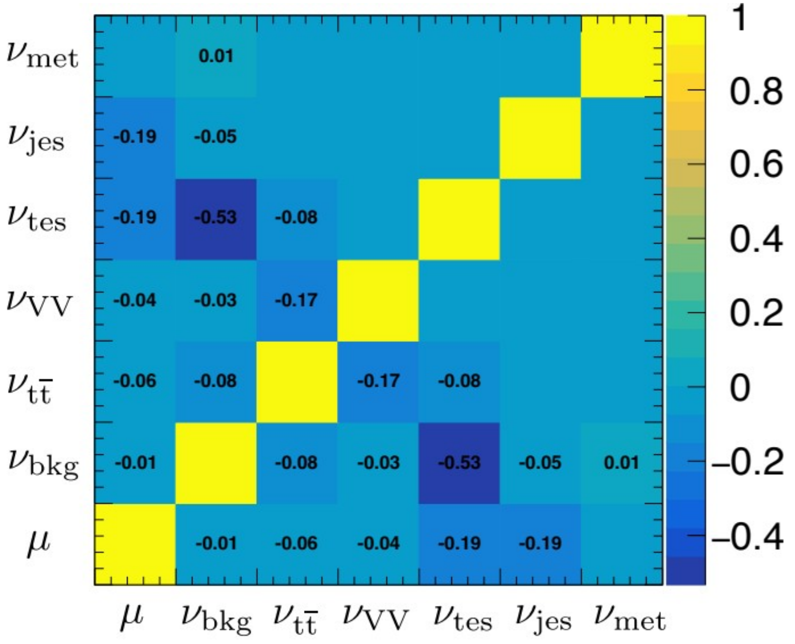
- In output node of multiclassifier
- Stable w.r.t. bin choices



Results II: Impacts

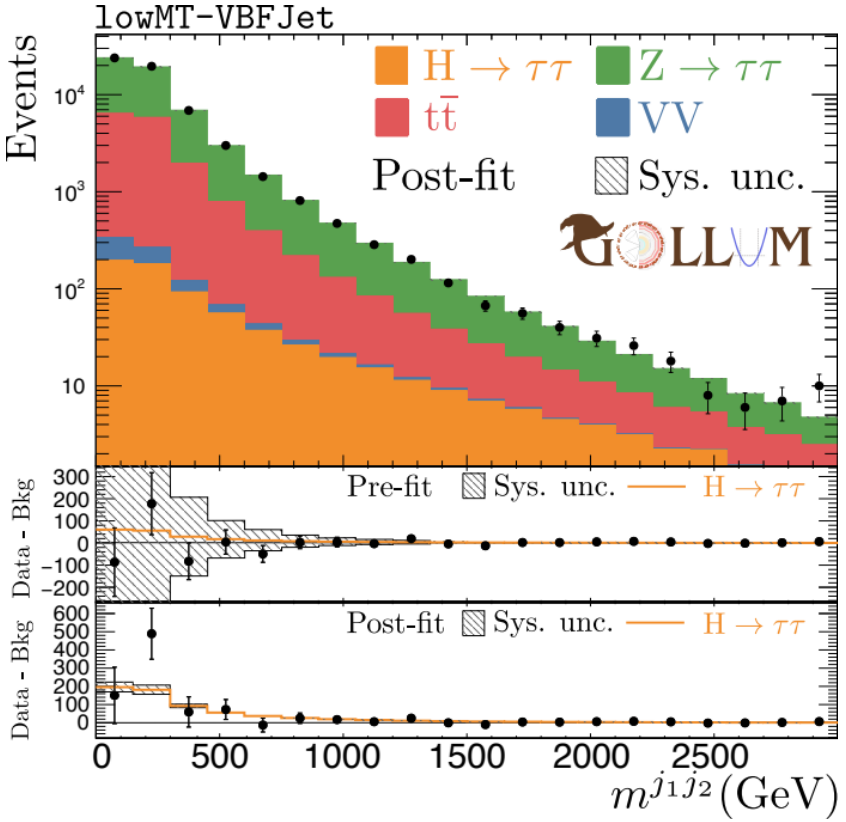


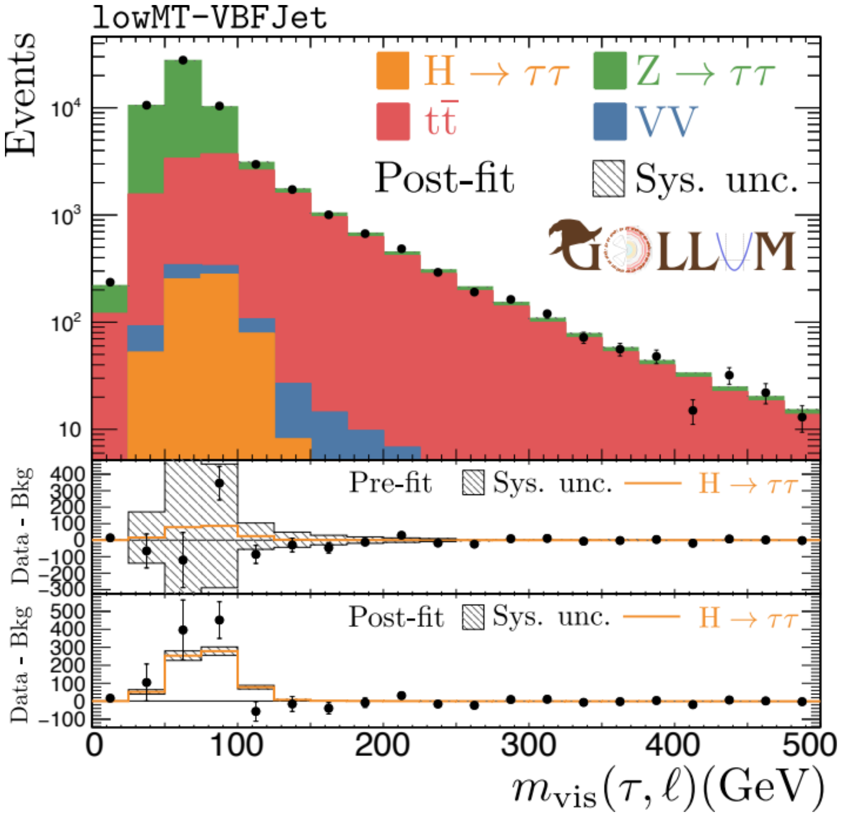


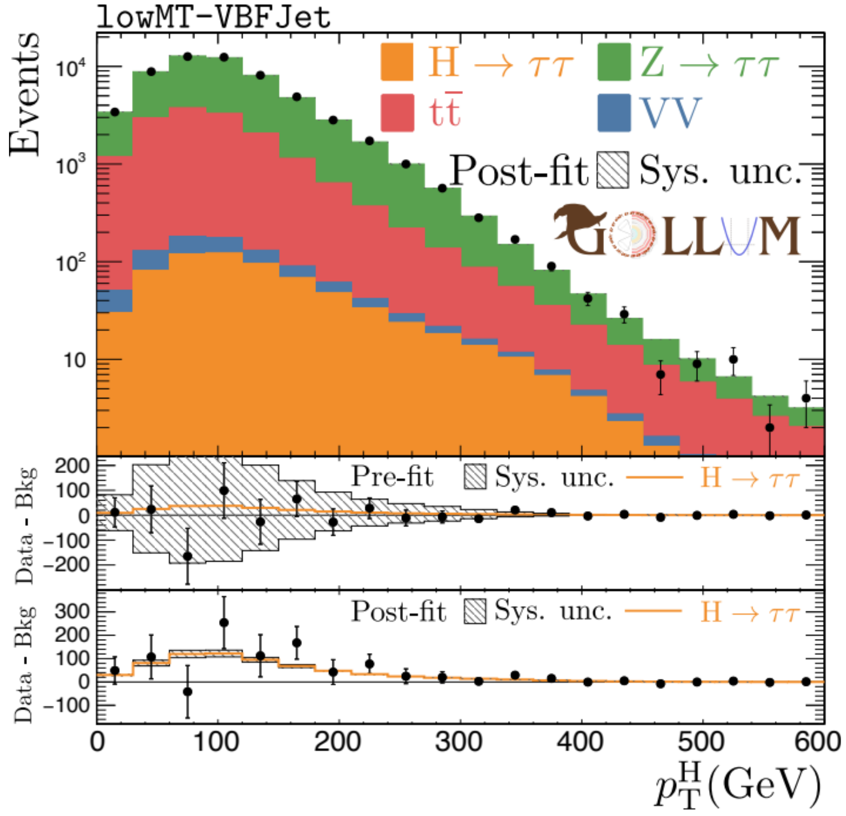


- Dominated by statistics
- Impacts below 5%

Results III: Postfit Distributions



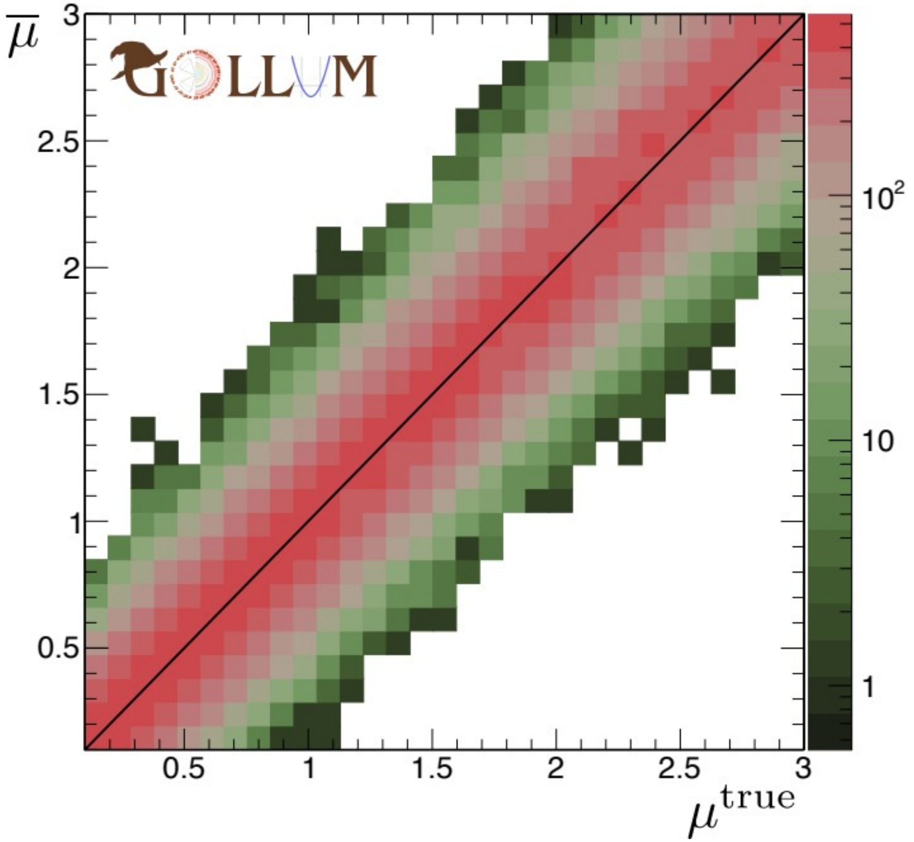


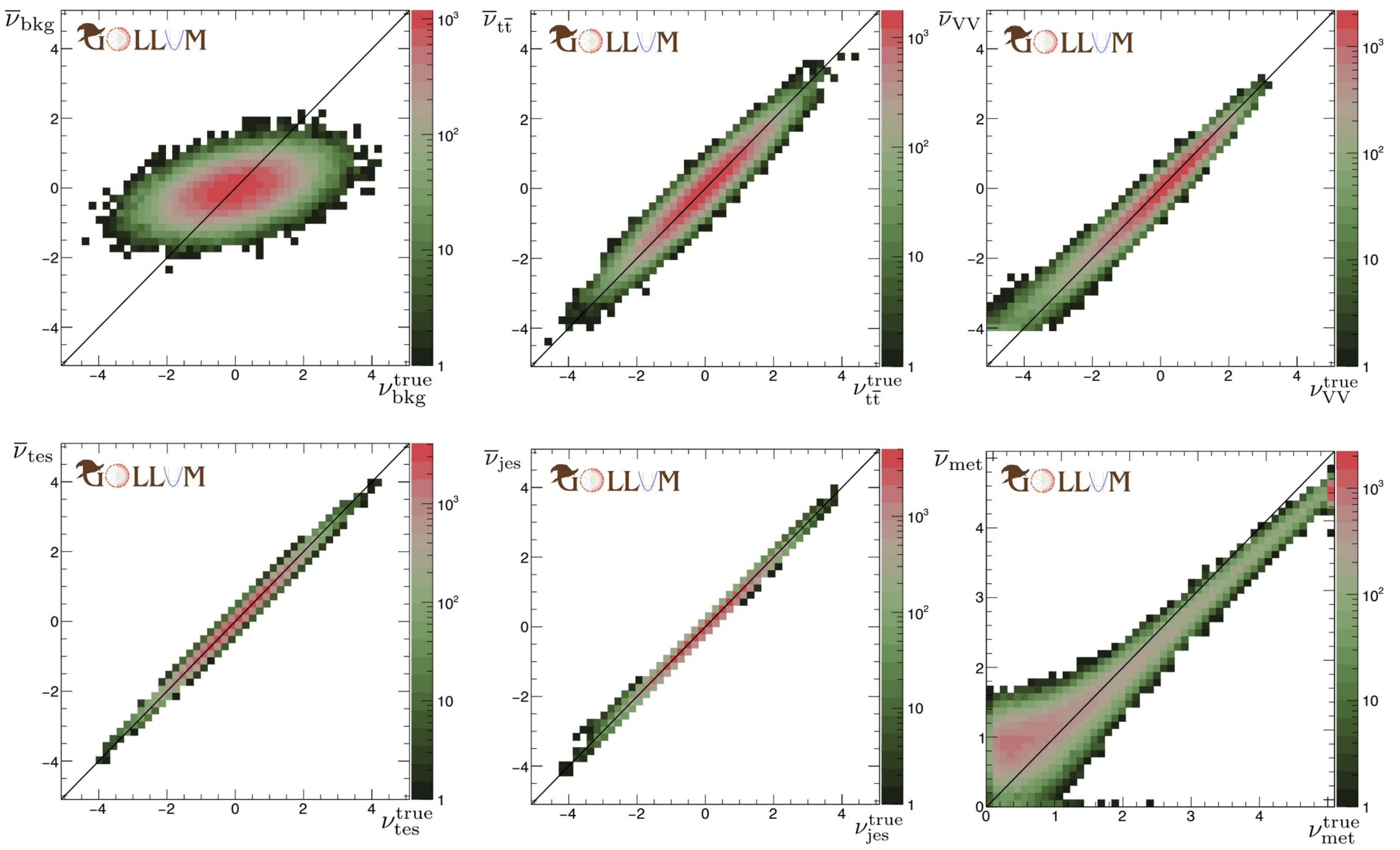


$\mu_{true}=3 \qquad \mu_{MLE}=3.24$

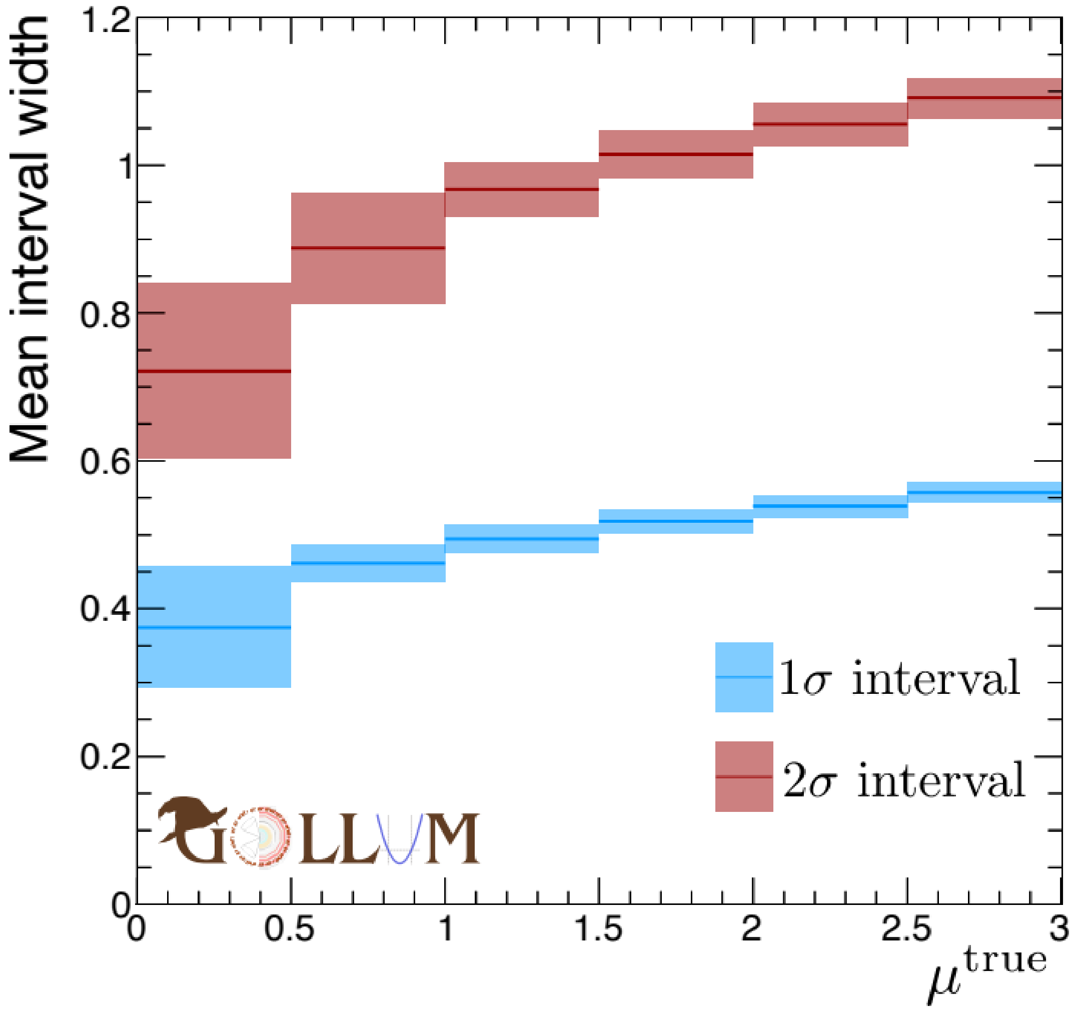
Results IV: Toy Studies

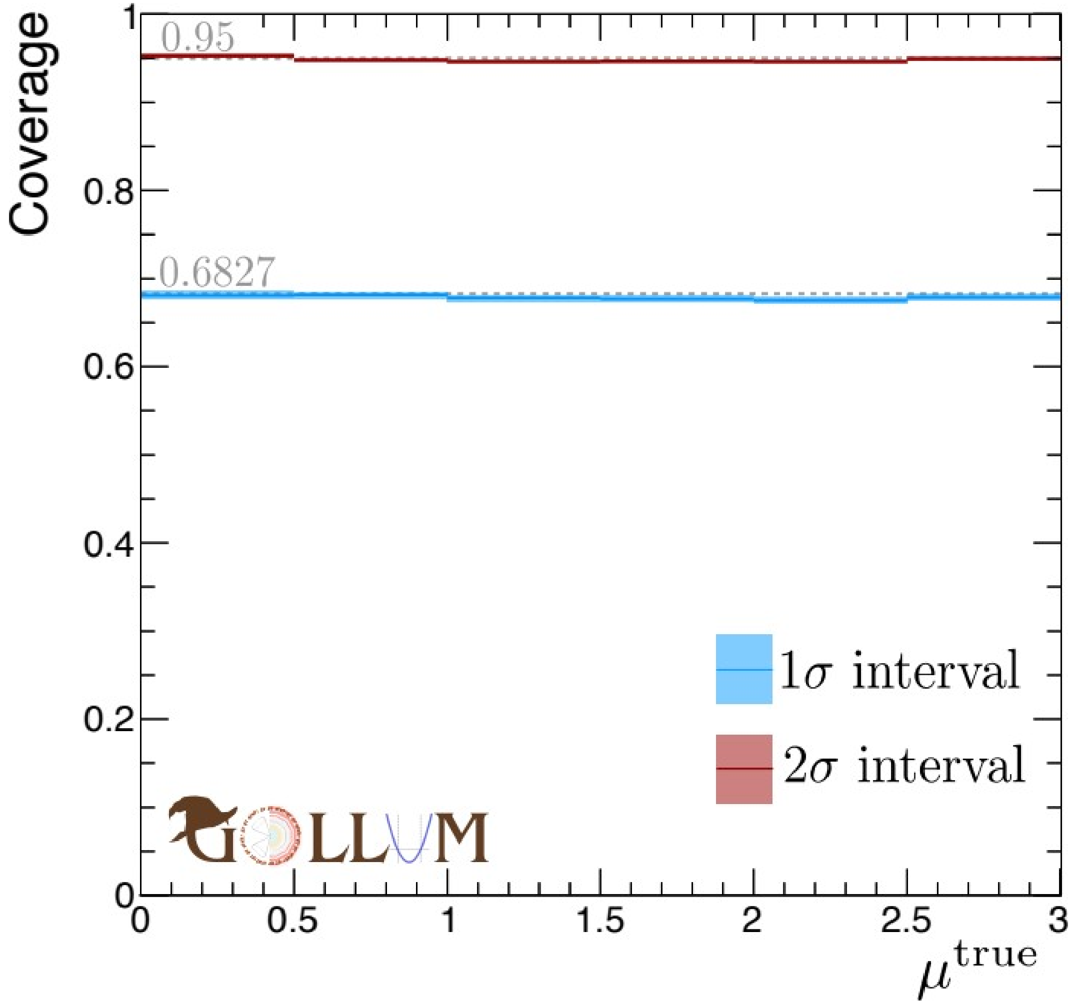
Study of 50k toys,
Sampling μ and ν
according to Challenge





Results V: Interval Size & Coverage





ÖAW AUSTRIAN ACADEMY OF SCIENCES

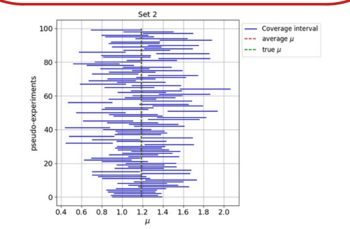
HEPHY INSTITUTE OF HIGH ENERGY PHYSICS

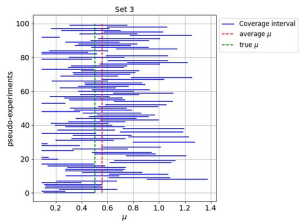
Official Results: pseudo experiments (p-e)

The official evaluation is done using pseudo-experiments: p-e

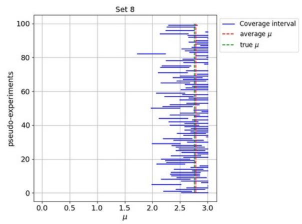
"A dataset representative of what would be measured from $10 fb^{-1}$ ~800 billion LHC pp collisions for a given value of μ and ν."

Case 1, $\mu=\mu_1$ Randomized ν	Case 2, $\mu=\mu_2$ Randomized ν	…	Case N, $\mu=\mu_N$ Randomized ν
p-e p-e p-e p-e	p-e p-e p-e p-e		p-e p-e p-e p-e





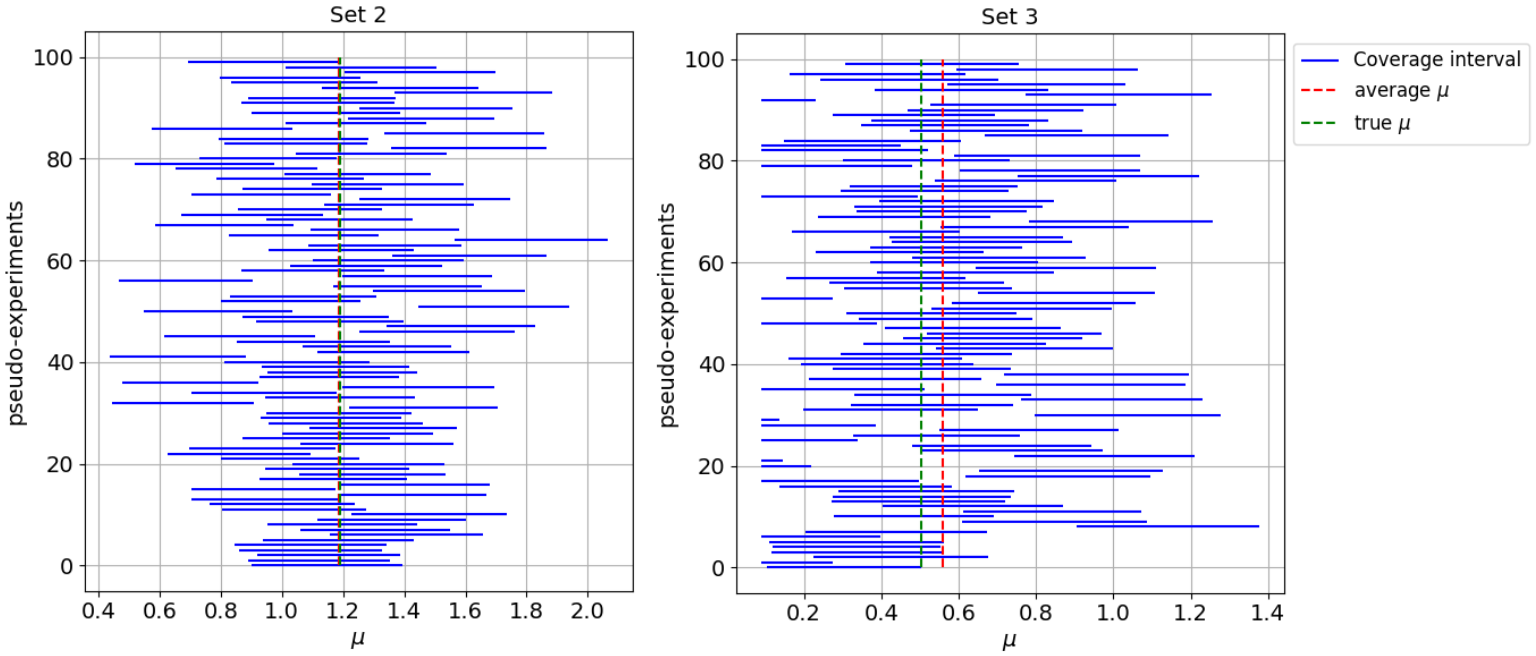
2410.02867



Slide source

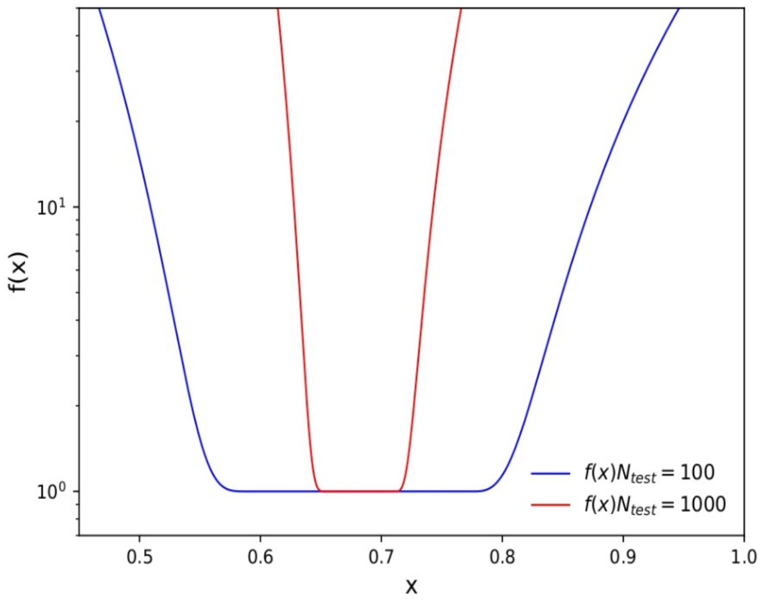
Official Results: Scoring

2410.02867



Online leader board:
10 trials x 100 p-e

Final leader board:
1000 trails x 100 p-e



Score = - ln ([interval size] x [coverage penalty])

Official Results: Public Leader Board

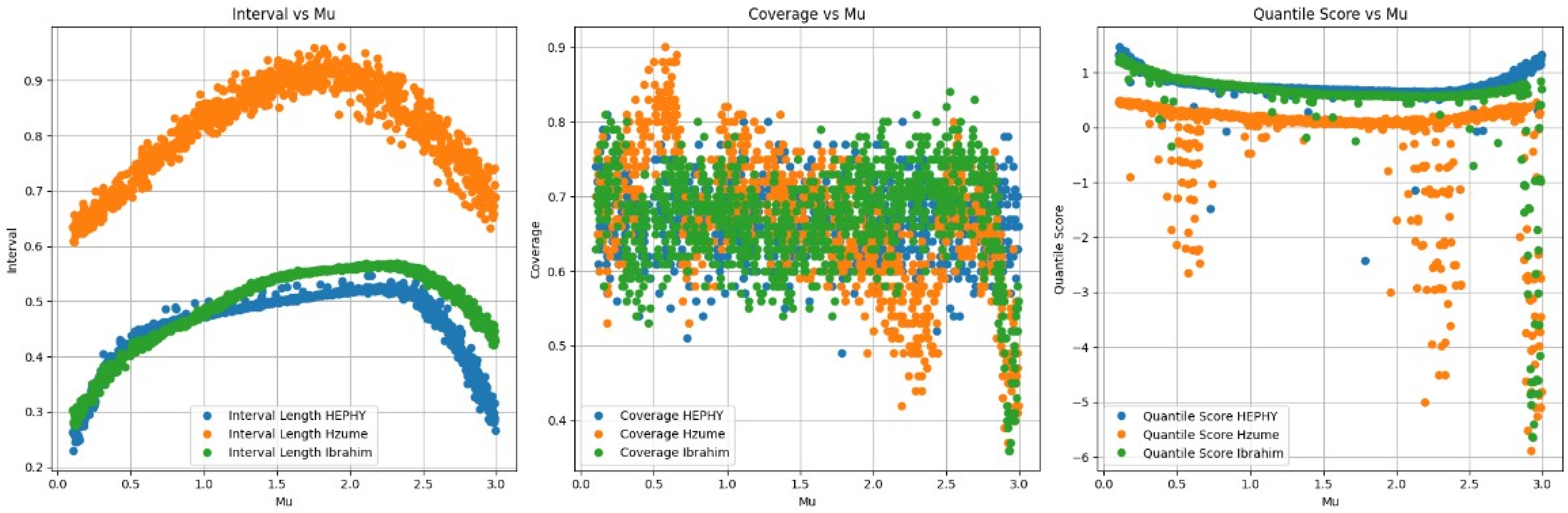
Task:					Results Fact Sheet Answers			Higgs NeurIPS T	
#	Participant	Entries	Date	ID	Method Name	Quantile Score	Interval	Coverage	RMSE
1	HEPHY	1	2025-03-12 18:26	244525	GOLLUM_calib-v5_v2-8	**0.878**	0.415	0.693	0.2
2	HEPHY	1	2025-03-10 20:20	243324	GOLLUM_v10-3	**0.833**	0.434	0.673	0.22
3	ibrahime	1	2025-03-12 14:47	244411	AdvnF-2j-4nf-edge	**0.823**	0.438	0.672	0.191
4	HEPHY	1	2025-03-10 20:19	243323	GOLLUM_v10-2	**0.803**	0.447	0.651	0.213
5	HEPHY	1	2025-03-11 15:23	243664	GOLLUM_calib-v5_v2-6	**0.797**	0.449	0.65	0.238
6	HEPHY	1	2025-03-06 10:11	241784	GOLLUM_v7	**0.776**	0.459	0.681	0.207

Official Results: Public Leader Board

2505.08709

Task:				Results	Fact Sheet Answers			Higgs NeurIPS T	
#	Participant	Entries	Date	ID	Method Name	Quantile Score	Interval	Coverage	RMSE
	…HY	1	2025-03-12 18:26	244525	GOLLUM_calib-v5_v2-8	**0.878**	0.415	0.693	0.2
2	HEPHY	1	2025-03-10 20:20	243324	GOLLUM_v10-3	**0.833**	0.434	0.673	0.22
3	ibrahime	1	2025-03-12 14:47	244411	AdvnF-2j-4nf-edge	**0.823**	0.438	0.672	0.191
4	HEPHY	1	2025-03-10 20:19	243323	GOLLUM_v10-2	**0.803**	0.447	0.651	0.213
5	HEPHY	1	2025-03-11 15:23	243664	GOLLUM_calib-v5_v2-6	**0.797**	0.449	0.65	0.238
6	HEPHY	1	2025-03-06 10:11	241784	GOLLUM_v7	**0.776**	0.459	0.681	0.207

Official Results: Final Evaluation



Official Results: Final Leader Board

WINNERS

NeurIPS 2024 Higgs Uncertainty Challenge has come to an end. We have announced the winners of the competition. After extensive studies on a new hold-out data set, HEPHY and IBRAHIME cannot be separated in a significant way and are declared joint first. HZUME has secured third position.

Medal	Rank	Team	Avg Coverage	Avg Interval	Avg Quantile Score
	1 (Tie)	HEPHY	0.6683	0.4599	-0.5823
	1 (Tie)	IBRAHIME	0.6698	0.4974	-0.5761
	3	HZUME	0.6659	0.8134	-2.1650

- HEPHY (Lisa Benato, Cristina Giordano, Claudius Krause, Ang Li, Robert Schöfbeck, Dennis Schwarz, Maryam Shooshtari, Daohan Wang) from Vienna's Institute of High Energy Physics (HEPHY) in Austria will win $2000.
- IBRAHIME (Ibrahim Elsharkawy) from University of Illinois at Urbana-Champaign will win $2000.
- HZUME (Hashizume Yota) from Kyoto University Japan will win $500.

ÖAW AUSTRIAN ACADEMY OF SCIENCES

HEPHY INSTITUTE OF HIGH ENERGY PHYSICS

Gollum goes Higgs Uncertainty Challenge

2505.05544

- We developed an unbinned likelihood surrogate in μ and ν.
- We maximized physics input and used analytic functions wherever possible.
- Method is refinable, making it well-suited for experimental collaborations.

- The challenge was well designed and fairly realistic (missing fakes, non-prompts).
  → challenging, but doable

- Our performance is state-of-the-art, winning the challenge ex aequo.