AUSTRIAN ACADEMY OF SCIENCES

MARIETTA BLAU INSTITUTE FOR PARTICLE PHYSICS

Higgs Signal Strength Estimation with Machine Learning under Systematic Uncertainties

Daohan Wang

Marietta Blau Institute for Particle Physics (MBI), Austrian Academy of Sciences (OeAW)

December 17, 2025

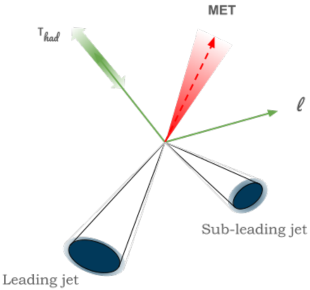
Based on 2505.05544 and 2509.00672

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
- Fair Universe HiggsML Uncertainty Challenge, a NeurIPS 2024 competition

Higgs Uncertainty Challenge



Goal: given a pseudo-experiment with given signal strength, provide a Confidence Interval on signal strength taking into account systematics uncertainties.

Process	Number Generated	LHC Events	Label
Higgs	52101127	1015	**signal**
Z Boson	221724480	1002395	**background**
Di-Boson	2105415	3783	**background**
$t\bar{t}$	12073068	44190	**background**

Data: 28 input features.

The six nuisance parameters lead to unknown nonlinear variations in the 28 input features via simulation and reconstruction effects.

2410.02867

28 event features:

Symbol	Description	Symbol	Description
Primary Features (PRI)			
p_T^ℓ	Transverse momentum of the lepton	η^ℓ	Pseudorapidity of the lepton
ϕ^ℓ	Azimuthal angle of the lepton	$p_T^{\tau_{had}}$	Transverse momentum of the hadronic τ
$\eta^{\tau_{had}}$	Pseudorapidity of the hadronic τ	$\phi^{\tau_{had}}$	Azimuthal angle of the hadronic τ
$p_T^{j_1}$	Transverse momentum of the leading jet	η^{j_1}	Pseudorapidity of the leading jet
ϕ^{j_1}	Azimuthal angle of the leading jet	$p_T^{j_2}$	Transverse momentum of the subleading jet
η^{j_2}	Pseudorapidity of the subleading jet	ϕ^{j_2}	Azimuthal angle of the subleading jet
N_j	Number of reconstructed jets	$\sum_{\text{jets}} p_T$	Scalar sum of transverse momenta of all jets
$\vec{p}_T^{\,\text{miss}}$	Missing transverse momentum	ϕ^{miss}	Azimuthal angle of missing transverse momentum
Derived Features (DER)			
Symbol	**Description**	**Symbol**	**Description**
$m_T(\ell, \vec{p}_T^{\,\text{miss}})$	Transverse mass of lepton and $\vec{p}_T^{\,\text{miss}}$	m_{vis}	Visible invariant mass of τ_{had} and ℓ
p_T^H	Vector sum of $p_T^{\tau_{had}}, p_T^\ell, \vec{p}_T^{\,\text{miss}}$	$m^{j_1 j_2}$	Invariant mass of the two leading jets
$\Delta\eta^{j_1 j_2}$	Pseudorapidity separation of leading jets	$\Delta R(\tau_{\text{had}}, \ell)$	Angular distance between τ_{had} and ℓ
p_T^{tot}	Vector sum of all visible momenta and $\vec{p}_T^{\,\text{miss}}$	$\sum p_T$	Scalar sum of all visible momenta and $\vec{p}_T^{\,\text{miss}}$
C_ϕ^{miss}	Azimuthal centrality of $\vec{p}_T^{\,\text{miss}}$ w.r.t. ℓ, τ_{had}	C_η^ℓ	Pseudorapidity centrality of the lepton w.r.t. the jets
$p_T^\ell / p_T^{\tau_{had}}$	Transverse–momentum ratio of lepton to τ_{had}	$\eta^{j_1} \cdot \eta^{j_2}$	Product of leading jet pseudorapidities

Higgs Uncertainty Challenge

Feature distortions:

Tau Energy Scale (and correlated MET)

$$P_{\text{had}}^{\text{biased}} = \alpha_{\text{tes}} P_{\text{had}}$$

Jet Energy Scale (and correlated MET impact)

$$P_{\text{j}_{1/2}}^{\text{biased}} = \alpha_{\text{jes}} P_{\text{j}_{1/2}}$$

Additional randomized Soft MET

$$P_{\text{met}}^{\text{biased}} = P_{\text{met}} + \begin{pmatrix} \text{Gauss}(0, \text{ET}_{\text{soft}}) \\ \text{Gauss}(0, \text{ET}_{\text{soft}}) \end{pmatrix}$$

Event category normalization:

Background overall normalization

$$w_{Z\to\tau\tau}^{\text{biased}} = \alpha_{\text{bkg}} \times w_{Z\to\tau\tau}$$

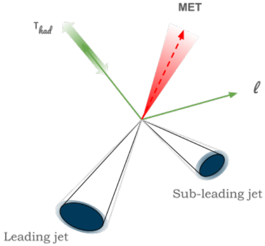
Diboson background normalization

$$w_{\text{VV}}^{\text{biased}} = \alpha_{\text{bkg}} \times \alpha_{\text{VV}} \times w_{\text{VV}}$$

ttbar background normalization

$$w_{t\bar{t}}^{\text{biased}} = \alpha_{\text{bkg}} \times \alpha_{t\bar{t}} \times w_{t\bar{t}}$$

Variable	Mean	Sigma	Range
α_{tes}	1.	0.01	[0.9, 1.1]
α_{jes}	1.	0.01	[0.9, 1.1]
α_{met}	0.	1.	[0., 5.]
$\alpha_{t\bar{t}}$	1.	0.02	[0.8, 1.2]
α_{VV}	1.	0.25	[0., 2.]
α_{bkg}	1.	0.001	[0.99, 1.01]



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

(GOLLUM goes Higgs Uncertainty Challenge)

GOLLUM

2505.05544

From an idea to a joint effort

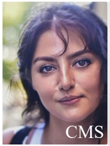
We should use your unbinned method [2406.19076] for that challenge...

Claudius Krause

Tell me more!

Robert Schöfbeck

Daohan Wang



Cristina Giordano

Lisa Benato

Dennis Schwarz

Ang Li

Maryam Shooshtari

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

Analysis 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`	$p_T^{j1} > 50$ GeV $p_T^{j2} > 30$ GeV $m_T \leq 70$ GeV	UB, SR	225.8	41 280.4	15 425.9	313.4	$3.96 \cdot 10^{-3}$
`highMT-VBFJet`	$p_T^{j1} > 50$ GeV $p_T^{j2} > 30$ 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	$3.14 \cdot 10^{-3}$
`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	$7.62 \cdot 10^{-4}$
`highMT-noVBFJet`	$m_T > 70$ GeV veto on `VBFJet`	CR	26.0	3 826.9	5 054.3	1 409.4	$2.53 \cdot 10^{-3}$

Main SR for VBF

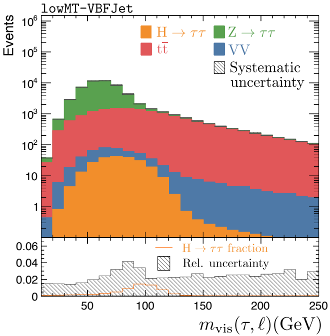
Unbinned CR

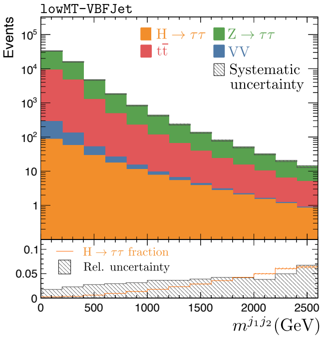
SR for ggH

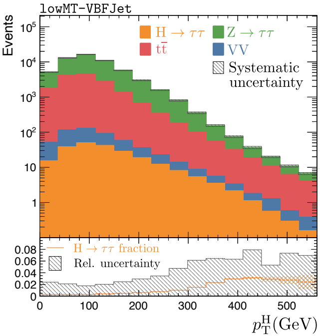
1 bin for bkg norm

Split off tt and VV CR (2 bins)

The Signal is buried in the systematics!







The total systematic uncertainty is obtained by combining all the $\pm 1\sigma$ variations of 6 nuisances.

Test Statistic

Profiled log-likelihood ratio test statistic

$$q_{\mu}(\mathcal{D}) = -2\log\frac{\max_{\nu} \quad L(\mathcal{D}|\mu,\nu)}{\max_{\mu,\nu} L(\mathcal{D}|\mu,\nu)} = -2(\max_{\nu}\log\frac{L(\mathcal{D}|\mu,\nu)}{L(\mathcal{D}|1,0)} - \max_{\mu,\nu}\log\frac{L(\mathcal{D}|\mu,\nu)}{L(\mathcal{D}|1,0)}) = \min_{\nu} u(\mathcal{D}|\mu,\nu) - \min_{\mu,\nu} u(\mathcal{D}|\mu,\nu) \tag{1}$$

Extended likelihood (for each selection)

$$L(\mathcal{D}|\mu,\nu) = \mathrm{P}(N_{\mathrm{obs}}|\mathcal{L}\sigma(\mu,\nu))\prod_{i=1}^{N_{\mathrm{obs}}} p(\boldsymbol{x}_i|\mu,\nu), \quad \mathrm{d}\sigma(\boldsymbol{x}|\mu,\nu) = \sigma(\mu,\nu)p(\boldsymbol{x}|\mu,\nu)\mathrm{d}\boldsymbol{x} \tag{2}$$

Log-likelihood ratio to the reference

$$log\frac{L(\mathcal{D}|\mu,\nu)}{L(\mathcal{D}|1,0)} = -\mathcal{L}(\sigma(\mu,\nu) - \sigma(1,\mathbf{0})) + \sum_{i=1}^{N_{\mathrm{obs}}}\log\left(\frac{\mathrm{d}\sigma(\boldsymbol{x}_i|\mu,\nu)}{\mathrm{d}\sigma(\boldsymbol{x}_i|1,\mathbf{0})}\right) \tag{3}$$

inclusive cross section of $\mathcal{D}$ differential cross section ratio (DCR)

Rewrite DCR in terms of known quantities

The DCR splits into signal and background components

$$\frac{d\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu})}{d\sigma(\boldsymbol{x}|1,\boldsymbol{0})} = \mu\frac{d\sigma_H(\boldsymbol{x}|\boldsymbol{\nu}_{\text{calib}})}{d\sigma(\boldsymbol{x}|1,\boldsymbol{0})} + (1+\alpha_{\text{bkg}})^{\nu_{\text{bkg}}}\frac{d\sigma_Z(\boldsymbol{x}|\boldsymbol{\nu}_{\text{calib}})}{d\sigma(\boldsymbol{x}|1,\boldsymbol{0})} + (1+\alpha_{t\bar{t}})^{\nu_{t\bar{t}}}\frac{d\sigma_{t\bar{t}}(\boldsymbol{x}|\boldsymbol{\nu}_{\text{calib}})}{d\sigma(\boldsymbol{x}|1,\boldsymbol{0})} + (1+\alpha_{VV})^{\nu_{VV}}\frac{d\sigma_{VV}(\boldsymbol{x}|\boldsymbol{\nu}_{\text{calib}})}{d\sigma(\boldsymbol{x}|1,\boldsymbol{0})}. \tag{4}$$

Relate to reference model factorize the calibration-type nuisances:

$$\frac{d\sigma_p(\boldsymbol{x}|\boldsymbol{\nu}_{\text{calib}})}{d\sigma(\boldsymbol{x}|1,\boldsymbol{0})} = \underbrace{\frac{d\sigma_p(\boldsymbol{x}|\boldsymbol{\nu}_{\text{calib}})}{d\sigma_p(\boldsymbol{x}|1,\boldsymbol{0})}}_{\substack{\text{shape response}\\ \text{to calibration nuisances}}} \underbrace{\frac{d\sigma_p(\boldsymbol{x}|1,\boldsymbol{0})}{d\sigma(\boldsymbol{x}|1,\boldsymbol{0})}}_{\substack{\text{relative rate}\\ \text{of process } p}} \simeq \underbrace{\hat{S}_p(\boldsymbol{x}|\boldsymbol{\nu}_{\text{calib}})}_{\substack{\text{Parametric}\\ \text{ML surrogate}}} \underbrace{\hat{g}_p(\boldsymbol{x})}_{\text{Multi-classifier}}$$

The first ingredient to the DCR: the multi classifier: $\hat{g}_p(\mathbf{x})$

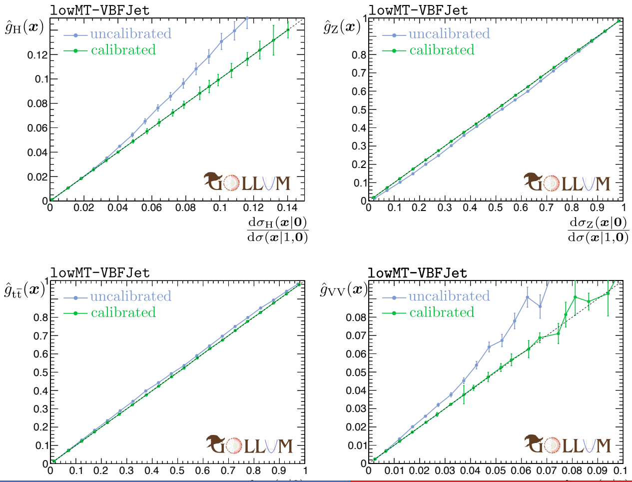
$$\hat{g}_p(\boldsymbol{x}) = \frac{\mathrm{d}\sigma_p(\boldsymbol{x}|\mathbf{0})}{\mathrm{d}\sigma(\boldsymbol{x}|1,\mathbf{0})} = \frac{\mathrm{d}\sigma_p(\boldsymbol{x}|\mathbf{0})}{\sum_q \mathrm{d}\sigma_q(\boldsymbol{x}|\mathbf{0})} \tag{5}$$

Predicts the $\mathbf{x}$-dependent DCR of process p to the total differential cross section.
Approximate $\hat{g}_p(\boldsymbol{x})$ with a four-class classifier based on nominal dataset.

$$L_{\mathrm{CE}}[\hat{f}_p] = -\sum_{p=1}^{N_p} \int \frac{\mathrm{d}\sigma_p(\boldsymbol{x})}{\sigma_p} \log\left(\hat{f}_p(\boldsymbol{x})\right) \approx -\sum_{p=1}^{N_p} \sum_{\{\boldsymbol{x}_i, w_i\} \in \mathcal{D}_p} \frac{w_i}{\mathcal{L}\sigma_p} \log\left(\hat{f}_p(\boldsymbol{x}_i)\right) \tag{6}$$

Train one multiclassifier for each of the three regions marked with UB.
Tensorflow NN with 28/128/128/4, ADAM, moderate L1/L2 regularization, CE loss

The first ingredient to the DCR: the multi classifier: $\hat{g}_p(\mathrm{x})$



Multiclassifer is not well calibrated!

Events in bin $0.19 < g_p(x) < 0.21$ need to have $\sim 20\%$ fraction of p

But g_H is off by 7%, g_Z is off by -2%

Isotonic Regression saves us

$$\hat{g}_p(\boldsymbol{x}) = \begin{cases} \mathrm{IREG}(g_H^*(\boldsymbol{x})), & \text{if } p = H \\ \frac{\mathrm{IREG}(g_p^*(\boldsymbol{x}))(1-\mathrm{IREG}(g_H^*(\boldsymbol{x})))}{\sum_{q=Z,t\bar{t},VV} \mathrm{IREG}(g_q^*(\boldsymbol{x}))}, & \text{otherwise.} \end{cases} \quad (7)$$

The second ingredient to the DCR: the parametric NN $\hat{S}_p(\mathbf{x}|\boldsymbol{\nu})$

$$\hat{S}_p(\boldsymbol{x}|\boldsymbol{\nu}_{\text{calib}}) \simeq \frac{\mathrm{d}\sigma_p(\boldsymbol{x}|\boldsymbol{\nu}_{\text{calib}})}{\mathrm{d}\sigma_p(\boldsymbol{x}|\mathbf{0})} \tag{8}$$

Predicts the relative dependence of the differential cross-section of a process p on $\boldsymbol{\nu}_{\text{calib}}$ and $\boldsymbol{x}$. Train a binary classifier between a specific $\mathcal{D}_{\boldsymbol{\nu}_{\text{calib}}}$ and the nominal with

$$L_{\text{CE}}[\hat{f}] = -\int \mathrm{d}\sigma(\boldsymbol{x}|\mathbf{0}) \log \hat{f}(\boldsymbol{x}) - \int \mathrm{d}\sigma(\boldsymbol{x}|\boldsymbol{\nu}) \log(1 - \hat{f}(\boldsymbol{x})),$$

the network $\hat{f}(\boldsymbol{x})$ attains its minimum at $\left(1 + \frac{\mathrm{d}\sigma(\boldsymbol{x}|\boldsymbol{\nu})}{\mathrm{d}\sigma(\boldsymbol{x}|\mathbf{0})}\right)^{-1}$

The second ingredient to the DCR: the parametric NN $\hat{S}_p(\mathbf{x}|\boldsymbol{\nu})$

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

$$f_{\boldsymbol{\nu}}(\boldsymbol{x}) = \frac{1}{1 + S_p(\boldsymbol{x}|\boldsymbol{\nu})} \tag{9}$$

$$\begin{aligned} \log S_p(\boldsymbol{x}|\boldsymbol{\nu}) = {} & \nu_{\text{tes}}\,\hat{\Delta}_{\text{tes}}(\boldsymbol{x}) + \nu_{\text{jes}}\,\hat{\Delta}_{\text{jes}}(\boldsymbol{x}) + \nu_{\text{met}}\,\hat{\Delta}_{\text{met}}(\boldsymbol{x}) \\ & + \nu_{\text{tes}}^2\,\hat{\Delta}_{\text{tes,tes}}(\boldsymbol{x}) + \nu_{\text{jes}}^2\,\hat{\Delta}_{\text{jes,jes}}(\boldsymbol{x}) + \nu_{\text{met}}^2\,\hat{\Delta}_{\text{met,met}}(\boldsymbol{x}) + \nu_{\text{tes}}\nu_{\text{jes}}\,\hat{\Delta}_{\text{tes,jes}}(\boldsymbol{x}) \\ & + \nu_{\text{jes}}\nu_{\text{met}}\,\hat{\Delta}_{\text{jes,met}}(\boldsymbol{x}) + \nu_{\text{tes}}\nu_{\text{met}}\,\hat{\Delta}_{\text{tes,met}}(\boldsymbol{x}). \end{aligned} \tag{10}$$

Tensorflow NN with 28 / 256 / 256 / 9, ADAM, CE loss
Training data: All mixed samples with $|\boldsymbol{\nu}_{jes}| + |\boldsymbol{\nu}_{tes}| + |\boldsymbol{\nu}_{met}| \leq 2$, and single $|\boldsymbol{\nu}_i| \leq 3$.

The second ingredient to the DCR: the parametric NN $\hat{S}_p(\mathbf{x}|\boldsymbol{\nu})$

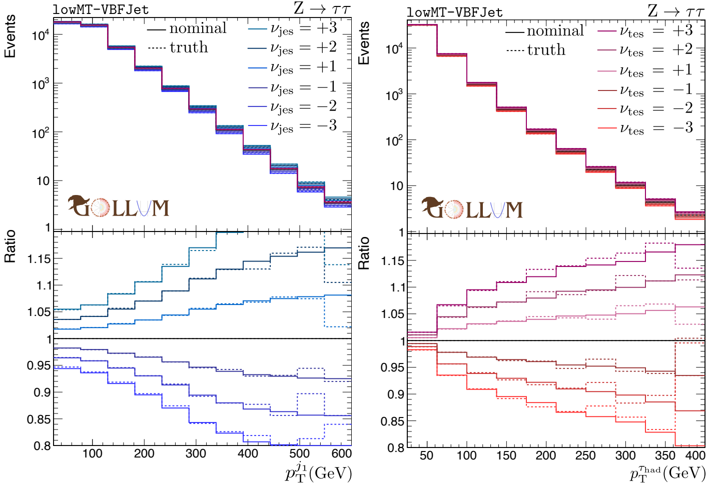
To train all coefficient functions simultaneously, we sum the CE loss over $\boldsymbol{\nu} \in \mathcal{V}$.

$$L[\hat{\Delta}_A] = \sum_{\boldsymbol{\nu}\in\mathcal{V}} \left[\int \mathrm{d}\sigma(\boldsymbol{x}|\mathbf{0}) \log(1 + S_p(\boldsymbol{x}|\boldsymbol{\nu})) + \int \mathrm{d}\sigma(\boldsymbol{x}|\boldsymbol{\nu}) \log(1 + \frac{1}{S_p(\boldsymbol{x}|\boldsymbol{\nu})}) \right]$$
$$\approx \sum_{\boldsymbol{\nu}\in\mathcal{V}} \left[\sum_{\{\boldsymbol{x}_i, w_i\}\in\mathcal{D}_{\mathbf{0}}} w_i \log(1 + S_p(\boldsymbol{x}_i|\boldsymbol{\nu})) + \sum_{\{\boldsymbol{x}_i, w_i\}\in\mathcal{D}_{\boldsymbol{\nu}}} w_i \frac{1}{S_p(\boldsymbol{x}_i|\boldsymbol{\nu}))} \right]. \tag{11}$$

44 systematically varied datasets VS nominal dataset

12 Parametrized Neural Network separately for each region and process.

The second ingredient to the DCR: the parametric NN $\hat{S}_p(\mathbf{x}|\boldsymbol{\nu})$



ML Surrogate (no need to retrain everything!)

$$\frac{d\sigma(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\nu})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})} \simeq \hat{R}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\nu}) = \mu\,\hat{g}_H(\boldsymbol{x})\hat{S}_H(\boldsymbol{x}|\boldsymbol{\nu}_{\text{calib}}) + (1+\alpha_{\text{bkg}})^{\nu_{\text{bkg}}}\Big(\hat{g}_Z(\boldsymbol{x})\hat{S}_Z(\boldsymbol{x}|\boldsymbol{\nu}_{\text{calib}}) + (1+\alpha_{t\bar{t}})^{\nu_{t\bar{t}}}\hat{g}_{t\bar{t}}(\boldsymbol{x})\hat{S}_{t\bar{t}}(\boldsymbol{x}|\boldsymbol{\nu}_{\text{calib}}) + (1+\alpha_{VV})^{\nu_{VV}}\hat{g}_{VV}(\boldsymbol{x})\hat{S}_{VV}(\boldsymbol{x}|\boldsymbol{\nu}_{\text{calib}})\Big). \tag{12}$$

Add a new $\boldsymbol{\nu}'_{\text{calib}}$: $\hat{S}_p^{\text{updated}} = \hat{S}_p(\boldsymbol{x}|\boldsymbol{\nu}_{\text{calib}})\hat{S}'_p(\boldsymbol{x}|\boldsymbol{\nu}'_{\text{calib}})$

Add a new background $d\sigma'(\boldsymbol{x}|\mu,\boldsymbol{\nu}) = d\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu}) + d\sigma_{\text{new}}(\boldsymbol{x}|\boldsymbol{\nu})$:

$$\frac{d\sigma'(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\nu})}{d\sigma'(\boldsymbol{x}|1,\mathbf{0})} = \frac{d\sigma(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\nu}) + d\sigma_{\text{new}}(\boldsymbol{x}|\boldsymbol{\nu})}{d\sigma(\boldsymbol{x}|1,\mathbf{0}) + d\sigma_{\text{new}}(\boldsymbol{x}|\mathbf{0})} = \frac{\frac{d\sigma(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\nu})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})} + \frac{d\sigma_{\text{new}}(\boldsymbol{x}|\boldsymbol{\nu})}{d\sigma_{\text{new}}(\boldsymbol{x}|\mathbf{0})}\frac{d\sigma_{\text{new}}(\boldsymbol{x}|\mathbf{0})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})}}{1 + \frac{d\sigma_{\text{new}}(\boldsymbol{x}|\mathbf{0})}{d\sigma(\boldsymbol{x}|1,\mathbf{0})}}$$

$$\simeq \frac{\hat{R}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\nu}) + \hat{S}_{\text{new}}(\boldsymbol{x}|\boldsymbol{\nu})\hat{g}_{\text{new}}(\boldsymbol{x})}{1 + \hat{g}_{\text{new}}(\boldsymbol{x})}. \tag{13}$$

Inclusive cross-section parametrization

$$\mathcal{L}(\sigma(\boldsymbol{\mu},\boldsymbol{\nu})-\sigma(1,\mathbf{0}))=\mathcal{L}\int\left(\frac{\mathrm{d}\sigma(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\nu})}{\mathrm{d}\sigma(\boldsymbol{x}|1,\mathbf{0})}-1\right)\frac{\mathrm{d}\sigma(\boldsymbol{x}|1,\mathbf{0})}{\mathrm{d}\boldsymbol{x}}\mathrm{d}\boldsymbol{x}\approx\sum_{\{w_i,\boldsymbol{x}_i\}\in\mathcal{D}_{1,\mathbf{0}}}w_i\left(\hat{R}(\boldsymbol{x}_i|\boldsymbol{\mu},\boldsymbol{\nu})-1\right). \tag{14}$$

The event weights ω_i implement the cross section normalization $\sum_{D_{1,\mathbf{0}}^{(p)}} w_i = \mathcal{L}\sigma_p(\mathbf{0})$

$$\begin{aligned}\sum_{\{w_i,\boldsymbol{x}_i\}\in\mathcal{D}_{1,\boldsymbol{\nu}}}\left(\hat{R}(\boldsymbol{x}_i|\boldsymbol{\mu},\boldsymbol{\nu})-1\right)w_i =\; & \boldsymbol{\mu}\,\mathrm{CSI}_H(\boldsymbol{\nu}_{\mathrm{calib}})+(1+\alpha_{\mathrm{bkg}})^{\boldsymbol{\nu}_{\mathrm{bkg}}}\left[\mathrm{CSI}_Z(\boldsymbol{\nu}_{\mathrm{calib}})+(1+\alpha_{t\bar{t}})^{\boldsymbol{\nu}_{t\bar{t}}}\mathrm{CSI}_{t\bar{t}}(\boldsymbol{\nu}_{\mathrm{calib}})+(1+\alpha_{VV})^{\boldsymbol{\nu}_{VV}}\mathrm{CSI}_{VV}(\boldsymbol{\nu}_{\mathrm{calib}})\right]\\ & +(\boldsymbol{\mu}-1)\,\mathrm{CSIC}_H+\left((1+\alpha_{\mathrm{bkg}})^{\boldsymbol{\nu}_{\mathrm{bkg}}}-1\right)\mathrm{CSIC}_Z\\ & +\left((1+\alpha_{\mathrm{bkg}})^{\boldsymbol{\nu}_{\mathrm{bkg}}}(1+\alpha_{t\bar{t}})^{\boldsymbol{\nu}_{t\bar{t}}}-1\right)\mathrm{CSIC}_{t\bar{t}}+\left((1+\alpha_{\mathrm{bkg}})^{\boldsymbol{\nu}_{\mathrm{bkg}}}(1+\alpha_{VV})^{\boldsymbol{\nu}_{VV}}-1\right)\mathrm{CSIC}_{VV}.\end{aligned} \tag{15}$$

Inclusive cross-section parametrization

$$\mathrm{CSI}_p(\boldsymbol{\nu}_{\mathrm{calib}}) = \sum_{\{w_i,\boldsymbol{x}_i\}\in\mathcal{D}_{1,0}} w_i\,\hat{g}_p(\boldsymbol{x})(\hat{S}_p(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})-1), \mathrm{CSIC}_p = \sum_{\{w_i,\boldsymbol{x}_i\}\in\mathcal{D}_{1,0}} w_i\,\hat{g}_p(\boldsymbol{x}).$$

The dependence on $\boldsymbol{\nu}_{\mathrm{calib}}$ is modeled using a 3D quintic spline interpolation over 6561 grid points based on trained $\hat{S}_p$, enabling fast and smooth interpolation for each process.

$$\begin{aligned} log\frac{L(\mathcal{D}|\boldsymbol{\mu},\boldsymbol{\nu})}{L(\mathcal{D}|1,0)} &= \sum_{r=1}^{N_r}\left[-\mathcal{L}(\sigma_r(\boldsymbol{\mu},\boldsymbol{\nu})-\sigma_r(1,0)) + \sum_{\boldsymbol{x}_i\in\mathcal{D}}\log\left(\frac{\mathrm{d}\sigma_r(\boldsymbol{x}_i|\boldsymbol{\mu},\boldsymbol{\nu})}{\mathrm{d}\sigma_r(\boldsymbol{x}_i|1,0)}\right)\right] \\ &\approx \sum_{r=1}^{N_r}\left[-\sum_{\{w_i,\boldsymbol{x}_i\}\in\mathcal{D}_{1,0}\cap r} w_i\left(\hat{R}_r(\boldsymbol{x}_i|\boldsymbol{\mu},\boldsymbol{\nu})-1\right) + \sum_{\boldsymbol{x}_i\in\mathcal{D}}\log\hat{R}_r(\boldsymbol{x}_i|\boldsymbol{\mu},\boldsymbol{\nu})\right]. \end{aligned} \tag{16}$$

The first DCR is based on nominal training dataset and the second DCR is based on observed pseudo dataset!

Binned control regions

Region	Requirements	Type	Poisson yield $\mathcal{L}\sigma$ $H\to\tau\tau$	$H\to\tau\tau$	$t\bar{t}$	VV	S/B
`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	$7.62 \cdot 10^{-4}$
`highMT-noVBFJet-tt`*	$m_T > 70$ GeV veto on `VBFJet` $\hat{f}_{t\bar{t}} > 0.4$	CR	3.2	203.7	3 821.8	258.5	$7.36 \cdot 10^{-4}$
`highMT-noVBFJet-VV`*	$m_T > 70$ GeV veto on `VBFJet` $\hat{f}_{t\bar{t}} \leq 0.4$ $\hat{f}_{VV} > 0.5$	CR	2.8	165.7	292.2	724.5	$2.3 \cdot 10^{-3}$

$$-\frac{1}{2}u_{\mathrm{CR}}(\mathcal{D}|\boldsymbol{\mu},\boldsymbol{\nu}) = \sum_{i=1}^{N_{\mathrm{CR}}}\left[-\sum_{p}\mathcal{L}\left(\sigma_{i,p}(\boldsymbol{\mu},\boldsymbol{\nu}) - \sigma_{i,p}(1,\mathbf{0})\right) + N_{i,\mathrm{obs}}\log\left(\frac{\sum_p \sigma_{i,p}(\boldsymbol{\mu},\boldsymbol{\nu})}{\sum_p \sigma_{i,p}(1,\mathbf{0})}\right)\right]. \tag{17}$$

$\sigma_i(\boldsymbol{\mu},\boldsymbol{\nu}) = \mu\,\sigma_{i,H}(\boldsymbol{\nu}_{\mathrm{calib}}) + \sum_{p=Z,t\bar{t},VV}\sigma_{i,p}(\boldsymbol{\nu}_{\mathrm{calib}})$ is parametrized with COMBINE package based on standard binned analyses.

Putting all pieces together

$$
\begin{aligned}
-\frac{1}{2}u_{\mathrm{UB}}(\mathcal{D}|\boldsymbol{\mu},\boldsymbol{\nu}) = \log\frac{L(\mathcal{D}|\boldsymbol{\mu},\boldsymbol{\nu})}{L(\mathcal{D}|1,\mathbf{0})} &= \sum_{r=1}^{N_r}\left[-\mathcal{L}\left(\sigma_r(\boldsymbol{\mu},\boldsymbol{\nu})-\sigma_r(1,\mathbf{0})\right)+\sum_{\boldsymbol{x}_i\in\mathcal{D}}\log\left(\frac{\mathrm{d}\sigma_r(\boldsymbol{x}_i|\boldsymbol{\mu},\boldsymbol{\nu})}{\mathrm{d}\sigma_r(\boldsymbol{x}_i|1,\mathbf{0})}\right)\right] \\
&\approx \sum_{r=1}^{N_r}\left[-\sum_{\{w_i,\boldsymbol{x}_i\}\in\mathcal{D}_{1,\mathbf{0}}\cap r} w_i\left(\hat{R}_r(\boldsymbol{x}_i|\boldsymbol{\mu},\boldsymbol{\nu})-1\right)+\sum_{\boldsymbol{x}_i\in\mathcal{D}}\log\hat{R}_r(\boldsymbol{x}_i|\boldsymbol{\mu},\boldsymbol{\nu})\right]. \qquad (18)
\end{aligned}
$$

$$
-\tfrac{1}{2}u_{\mathrm{CR}}(\mathcal{D}|\boldsymbol{\mu},\boldsymbol{\nu}) = \sum_{i=1}^{N_{\mathrm{CR}}}\left[-\sum_{p}\mathcal{L}\left(\sigma_{i,p}(\boldsymbol{\mu},\boldsymbol{\nu})-\sigma_{i,p}(1,\mathbf{0})\right)+N_{i,\mathrm{obs}}\log\left(\frac{\sum_p\sigma_{i,p}(\boldsymbol{\mu},\boldsymbol{\nu})}{\sum_p\sigma_{i,p}(1,\mathbf{0})}\right)\right]. \qquad (19)
$$

Penalty term $u_{\mathrm{P}} = \nu_{\mathrm{bkg}}^2+\nu_{t\bar{t}}^2+\nu_{VV}^2+\nu_{\mathrm{tes}}^2+\nu_{\mathrm{jes}}^2+\nu_{\mathrm{met}}^2$

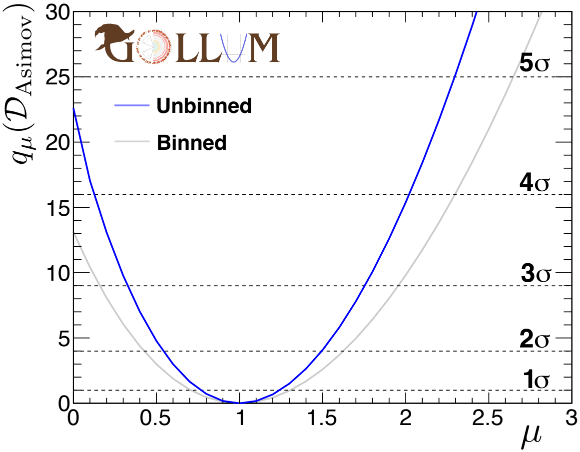
Putting all pieces together

$$u_{\text{unbinned}} = u_{\text{UB}}(\mathcal{D}|\mu,\nu) + u_{\text{CR}}(\mathcal{D}|\mu,\nu) + u_{\text{P}}(\mathcal{D}|\nu) \tag{20}$$

$$q_{\mu}(\mathcal{D}) = \min_{\nu} u_{\text{unbinned}}(\mathcal{D}|\mu,\nu) - \min_{\mu,\nu} u_{\text{unbinned}}(\mathcal{D}|\mu,\nu) \tag{21}$$

The profiling is performed with the MINUIT package.

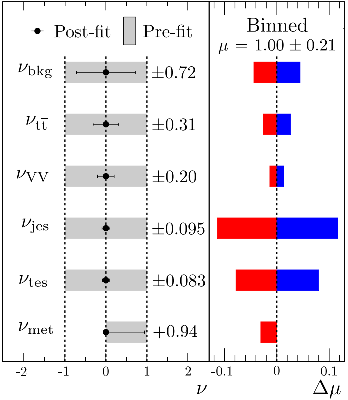
Results I: Asimov Dataset

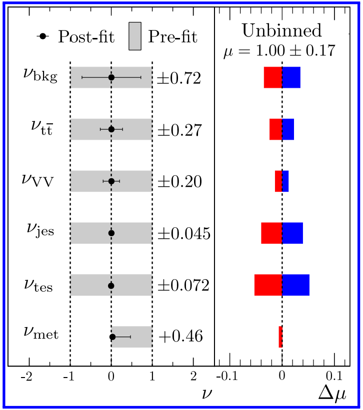


Binned: $\mu = 1.00 \pm 0.21$

Unbinned: $\mu = 1.00 \pm 0.17$

Results II: Impacts

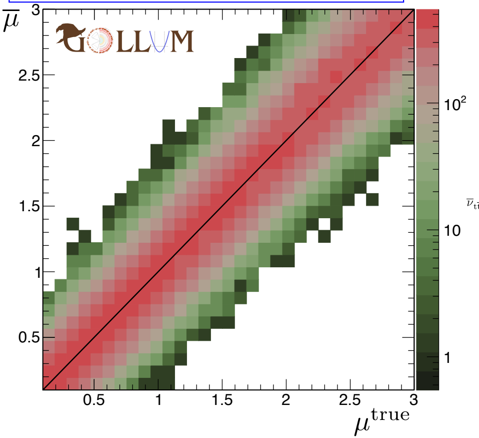


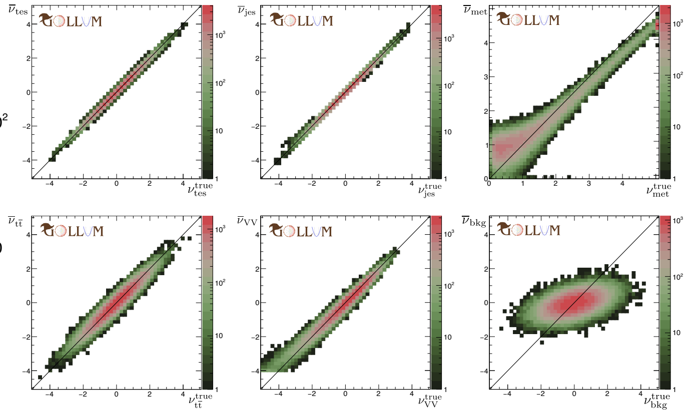


Post-fit results are statistically dominated, with all impacts below $\sim$5%.

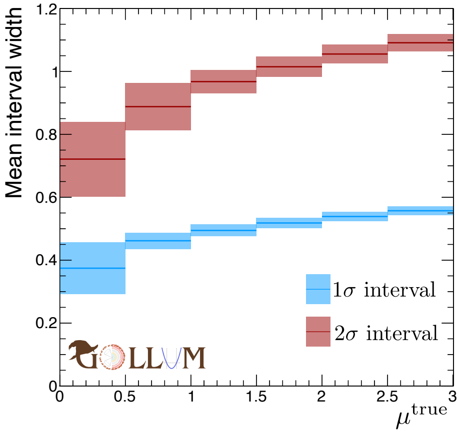
Results III: Toy Studies

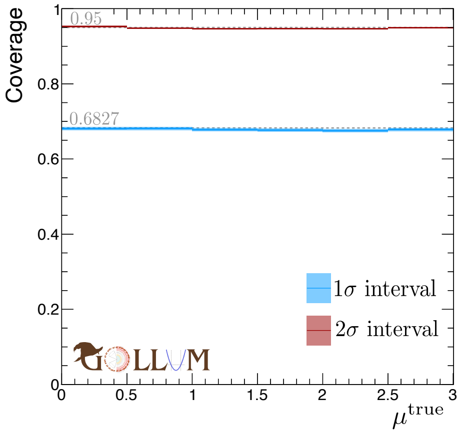
Study of 50k toys, sampling μ uniformly and ν from their priors.



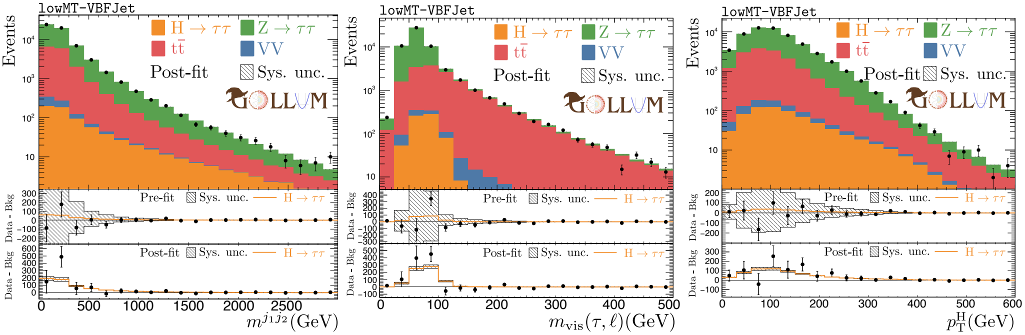


Results IV: Interval and Coverage





Results V: Post-fit Distributions



$$\mu^{\text{true}} = 3, \quad \mu^{\text{MLE}} = 3.24 \tag{22}$$

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.

Papers documenting the winning solutions are being prepared and will be linked here. The dataset will be released permanently on Zenodo to serve as a permanent benchmark. All winners will present at Fair Universe HiggsML Uncertainty CERN workshop.

Higgs Signal Strength Estimation with Machine Learning under Systematic Uncertainties

(SAGE goes Higgs Uncertainty Challenge)

2509.00672

Overall Model Architecture.

12 epochs, ~2 days on a single NVIDIA Tesla~V100 GPU.

We use trilinear interpolation to model the remaining ν −dependence.

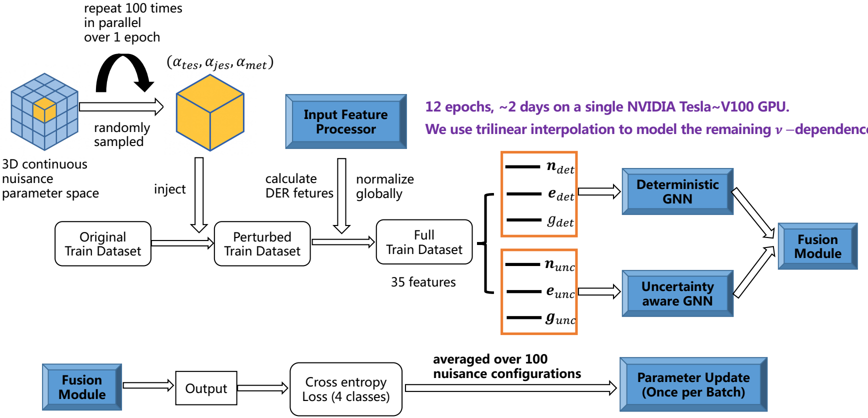


Table: Node feature assignment for final-state objects. Each node is represented by a one-hot encoding indicating its type. Deterministic and uncertainty-aware features are assigned separately according to their sensitivity to systematic variations. These two types of features are represented as $\vec{n}_{\text{det}}$ and $\vec{n}_{\text{unc}}$.

Node Type	Feature Type	Variables Used	Feature Role	Node Encoding
Lepton	Deterministic	p_T, η, ϕ	Fully specified	[1, 0, 0, 0, 0]
Tau-jet	Deterministic	η, ϕ	Spatial info	[0, 1, 0, 0, 0]
	Uncertainty-aware	p_T	Affected by α_{tes}	[0, 1, 0, 0, 0]
Leading Jet	Deterministic	η, ϕ	Spatial info	[0, 0, 1, 0, 0]
	Uncertainty-aware	p_T	Affected by α_{jes}	[0, 0, 1, 0, 0]
Subleading Jet	Deterministic	η, ϕ	Spatial info	[0, 0, 0, 1, 0]
	Uncertainty-aware	p_T	Affected by α_{jes}	[0, 0, 0, 1, 0]
MET	Uncertainty-aware	p_T, ϕ	Affected by α_{met}	[0, 0, 0, 0, 1]

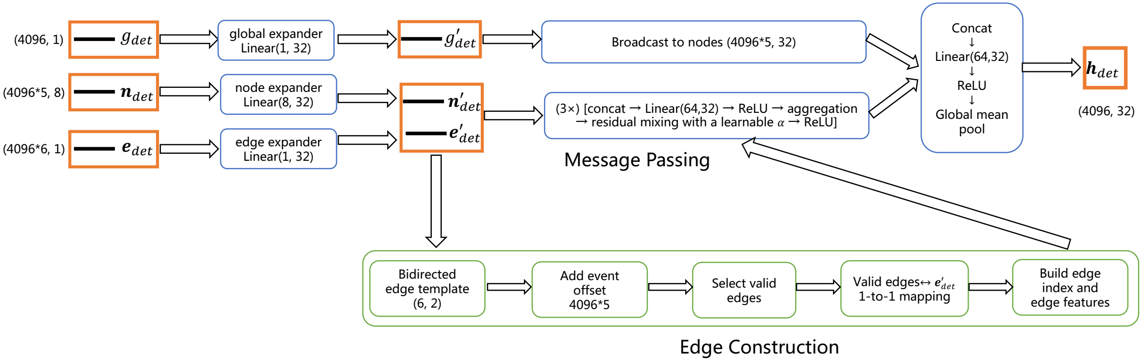
Table: Edge features constructed between pairs of final-state objects, which are categorized as deterministic or uncertainty-aware based on their sensitivity to systematic variations. These two types of features are represented as $\vec{e}_{\mathrm{det}}$ and $\vec{e}_{\mathrm{unc}}$.

Edge Type	**Feature Name**	**Feature Type**
Tau-jet – MET	$\Delta R(\tau_{\mathrm{had}}, \mathrm{miss})$	Uncertainty-aware
Lepton – MET	$\Delta R(\ell, \mathrm{miss})$	Uncertainty-aware
	$m_{\mathrm{T}}(\ell, p_{\mathrm{T}}^{\mathrm{miss}})$	Uncertainty-aware
Jet1 – MET	$\Delta R(j_1, \mathrm{miss})$	Uncertainty-aware
Jet2 – MET	$\Delta R(j_2, \mathrm{miss})$	Uncertainty-aware
Tau-jet – Jet1	$\Delta R(\tau_{\mathrm{had}}, j_1)$	Deterministic
Tau-jet – Jet2	$\Delta R(\tau_{\mathrm{had}}, j_2)$	Deterministic
Lepton – Jet1	$\Delta R(\ell, j_1)$	Deterministic
Lepton – Jet2	$\Delta R(\ell, j_2)$	Deterministic
Jet1 – Jet2	$\Delta R(j_1, j_2)$	Deterministic
	$m^{j_1 j_2}$	Uncertainty-aware
Tau-jet – Lepton	$\Delta R(\tau_{\mathrm{had}}, \ell)$	Deterministic
	$p_{\mathrm{T}}^{\ell} / p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$	Uncertainty-aware
	$m_{\mathrm{vis}}(\tau_{\mathrm{had}}, \ell)$	Uncertainty-aware

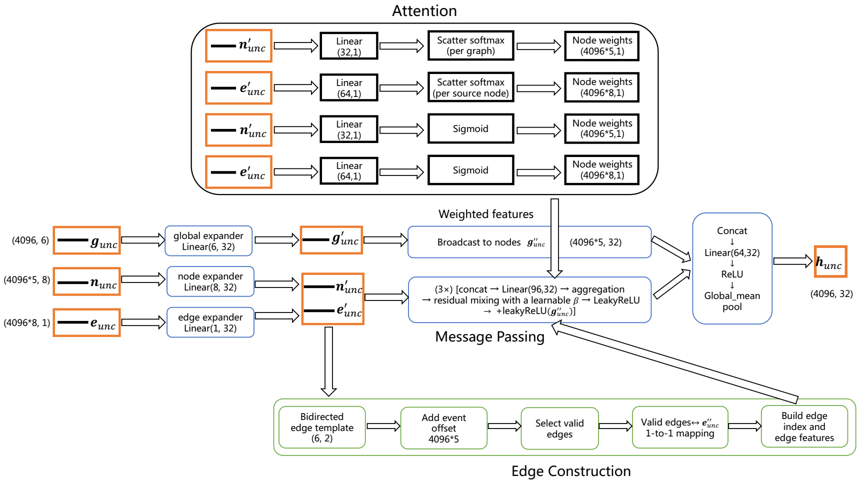
Table: Global features constructed from three or more final-state objects. Each feature is categorized by its defining particles and its sensitivity to systematic uncertainties. These two types of features are represented as $\vec{g}_{\text{det}}$ and $\vec{g}_{\text{unc}}$.

Particles Involved	**Feature Name**	**Feature Type**
All QCD jets	N_j	Uncertainty-aware
Tau-jet, Lepton, Jet1, Jet2, MET	$p_{\mathrm{T}}^{\text{tot}}$	Uncertainty-aware
Tau-jet, Lepton, MET	p_{T}^{H}	Uncertainty-aware
Tau-jet, Lepton, Jet1, Jet2, MET	$\sum p_{\mathrm{T}}$	Uncertainty-aware
Tau-jet, Lepton, MET	C_{ϕ}^{miss}	Uncertainty-aware
All QCD jets	$\sum_{\text{jets}} p_{\mathrm{T}}$	Uncertainty-aware
Lepton, Jet1, Jet2	C_{η}^{ℓ}	Deterministic

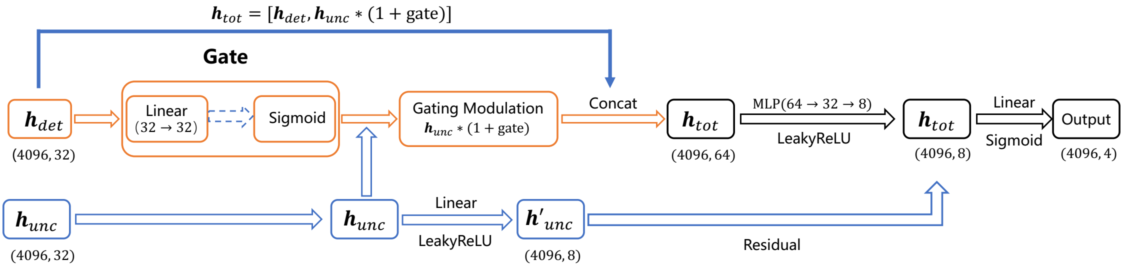
Deterministic GNN Branch



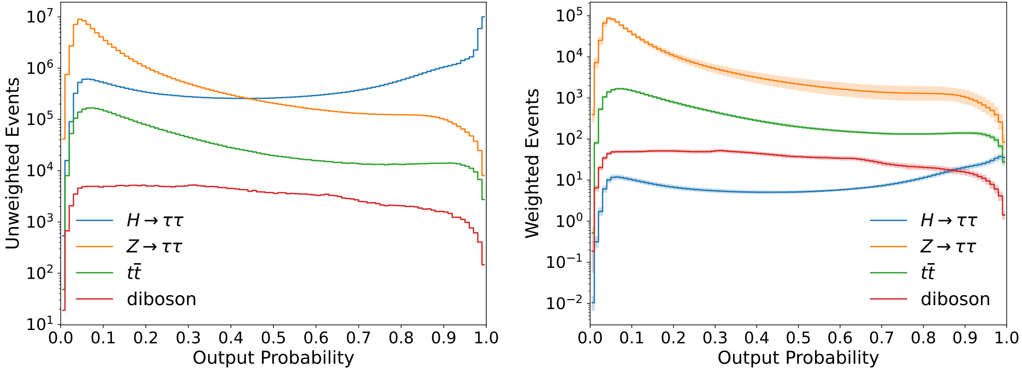
Uncertainty-aware GNN Branch



Fusion Module



Histogram of Output Probabilities



The solid lines indicate the average count of events per bin evaluated across a grid of 11849 fixed nuisance parameter settings, uniformly spanning $\alpha_{\text{tes}} \in [0.96, 1.04]$, $\alpha_{\text{jes}} \in [0.96, 1.04]$, $\alpha_{\text{soft\_met}} \in [0, 5]$.

Analysis Regions

Region	Requirements	Type	Poisson yield $\mathcal{L}\sigma$				S/B
			$H \to \tau\tau$	$Z \to \tau\tau$	$t\bar{t}$	VV	
`inclusive`	–	–	966.0	901 137.5	41 283.4	3 433.5	1.02×10^{-3}
`highMT-VBFJet`	$p_T^{j_1} > 50$ GeV $p_T^{j_2} > 30$ GeV $m_T > 70$ GeV	CR1	14.7	721.7	16 768.6	193.2	8.30×10^{-4}
`highMT-noVBFJet-tt`	$m_T > 70$ GeV veto on `VBFJet` $\hat{f}_{t\bar{t}} > 0.4$	CR2	2.7	202.7	3 607.1	268.8	6.62×10^{-4}
`highMT-noVBFJet-VV`	$m_T > 70$ GeV veto on `VBFJet` $\hat{f}_{VV} > 0.2$	CR3	1.8	189.5	207.6	597.4	1.8×10^{-3}

Signal Strength Profiling via Surrogate Likelihood

Goal: Extract the signal strength μ while profiling six nuisance parameters $\nu = (\nu_{\text{tes}}, \nu_{\text{jes}}, \nu_{\text{met}}, \nu_{\text{bkg}}, \nu_{\text{tt}}, \nu_{\text{VV}})$.

Inputs: Trained classifier $\rightarrow$ binned event counts of the signal-class probability distributions.

Process-wise Vectorized Trilinear Interpolation

Interpolate to obtain process-wise binned yields as functions of the nuisance parameters.

TES, JES $\in [0.96, 1.04]$, MET $\in [0, 5]$, $17 \times 17 \times 41 = 11\,849$ systematic nodes per region (up to 4 σ)

Adaptive binning ensures every bin contains at least 10 weighted counts.

$$(\text{tes,jes,met}) \xrightarrow{\texttt{\_trilinear\_multi()}} \{S_{\text{raw}}, tt_{\text{raw}}, VV_{\text{raw}}, Z_{\text{raw}}\}$$

$$\lambda_{\text{Full}} = \mu\, S_{\text{raw}} + \alpha_{\text{tt}}\alpha_{\text{bkg}}\, tt_{\text{raw}} + \alpha_{\text{VV}}\alpha_{\text{bkg}}\, VV_{\text{raw}} + \alpha_{\text{bkg}}\, Z_{\text{raw}}$$

$$\lambda_{\text{CR}} = \alpha_{\text{tt}}\alpha_{\text{bkg}}\, tt_{\text{raw}} + \alpha_{\text{VV}}\alpha_{\text{bkg}}\, VV_{\text{raw}} + \alpha_{\text{bkg}}\, Z_{\text{raw}}$$

Signal Strength Profiling via Surrogate Likelihood

Negative Log Lokelihood: $\mathrm{NLL}(\mu, \boldsymbol{\nu}) = \sum_{b=1}^{B} \left[\lambda_b(\mu, \boldsymbol{\nu}) - n_b(\boldsymbol{\nu}) \log \lambda_b(\mu, \boldsymbol{\nu})\right] + \frac{1}{2}\|\boldsymbol{\nu}\|^2 + \mathrm{const.}$

Scan grid: $\mu \in [0, 3]$ (100 points). Compute profile NLL. The first point uses a fresh global optimisation based on `dual_annealing`; others warm-start from the previous $\hat{\nu}$.

Local refinement: each profile step finishes with `LBFGS` via PyTorch autograd and up to five random restarts.

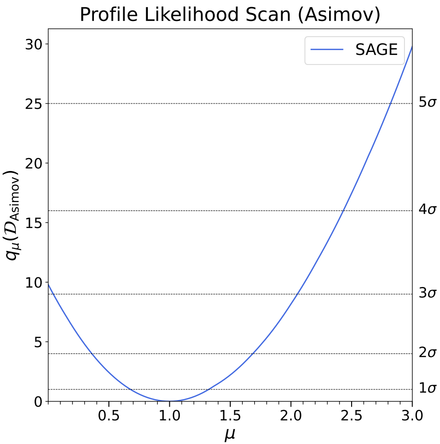
68% CL limits: $\Delta\mathrm{NLL}(\mu) = \mathrm{NLL}(\mu) - \mathrm{NLL}_{\mathrm{min}} = 0.5$

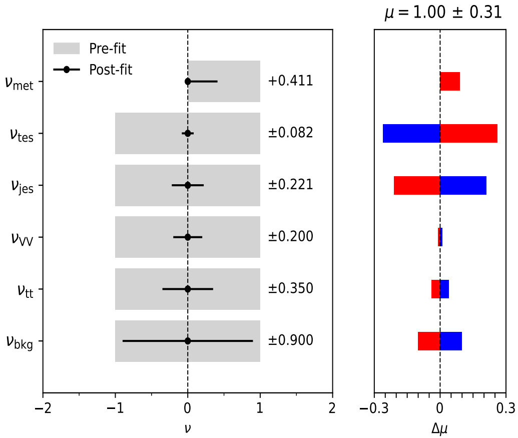
Output: Signal strength estimate $\hat{\mu}$ with 68% CI $[p_{16},\ p_{84}]$, along with fitted nuisance parameters $\hat{\alpha}_i$.

Runtime: one pseudo-experiment $\sim$5 minutes on CPU, $\sim$2 minutes on a Tesla V100 GPU.

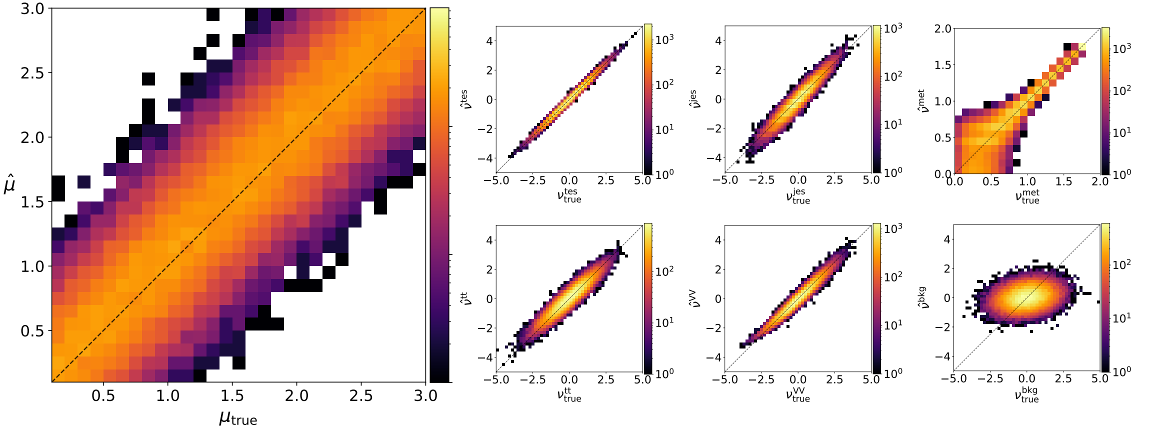
Extension: More efficient approaches (e.g. Normalization Flows) may be needed for high-dim interpolation.

Asimov Fitting

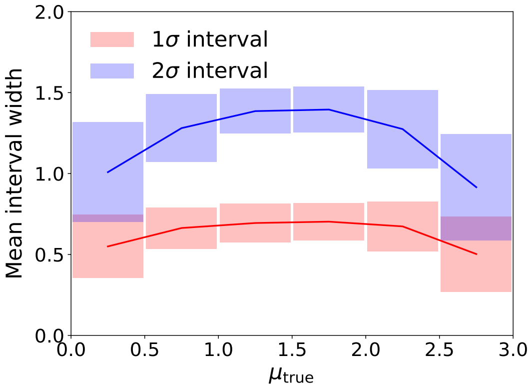


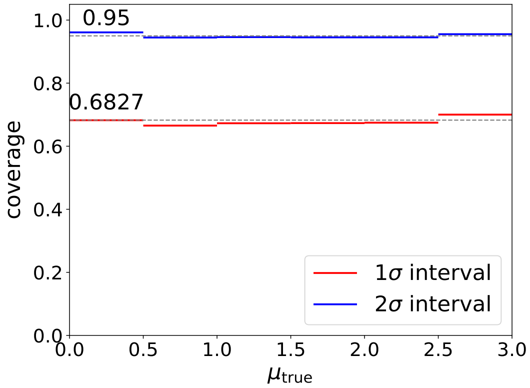


Toy Studies


$\hat{\mu}$
μ_{true}
$\hat{\nu}^{\text{tes}}$
$\nu^{\text{tes}}_{\text{true}}$
$\hat{\nu}^{\text{jes}}$
$\nu^{\text{jes}}_{\text{true}}$
$\hat{\nu}^{\text{met}}$
$\nu^{\text{met}}_{\text{true}}$
$\hat{\nu}^{\text{tt}}$
$\nu^{\text{tt}}_{\text{true}}$
$\hat{\nu}^{\text{VV}}$
$\nu^{\text{VV}}_{\text{true}}$
$\hat{\nu}^{\text{bkg}}$
$\nu^{\text{bkg}}_{\text{true}}$

Interval and Coverage





Summary

GOLLUM

2505.05544

- We developed an unbinned likelihood surrogate in μ and ν using machine-learned approximations of systematics.
- we reconstruct a high-dimensional likelihood ratio test statistic that captures both known analytic and unknown data-driven dependencies on nuisance parameters.
- Method is refinable and workflow-friendly, making it well-suited for experimental collaborations.
- Our performance is state-of-the-art, winning the Higgs Uncertainty Challenge ex aequo.

SAGE

2509.00672

- We developed a binned likelihood surrogate based on dual-branch GNN and trilinear interpolation.
- Training: 12 epochs, $\sim$2 days on a single NVIDIA Tesla V100 GPU.
- Inference: one pseudo-experiment $\sim$5 minutes on CPU, $\sim$2 minutes on a Tesla V100 GPU.
- **Performance** – Based on public leaderboard trends, our method likely ranks around **3rd place**.
- **Outlook** – Well suited for complex topologies + Largely insensitive to the number of nuisance parameters.