Higgs Signal Strength Estimation with a Dual-Branch GNN under Systematic Uncertainties

Daohan Wang

Institute of High Energy Physics (HEPHY), Austrian Academy of Sciences (OeAW)

July 9, 2025

Collaborated with Minxuan He and Claudius Krause

Higgs Uncertainty Challenge

FAIR Universe
HiggsML Uncertainty Challenge Competition

Wahid Bhimji[1], Paolo Calafiura[1], Ragansu Chakkappai[2,6], Po-Wen Chang[1], Yuan-Tang Chou[3], Sascha Diefenbacher[1], Jordan Dudley[4,1], Steven Farrell[1], Aishik Ghosh[5,1], Isabelle Guyon[6], Chris Harris[1], Shih-Chieh Hsu[3], Elham E Khoda[7,3,1], Rémy Lyscar[2], Alexandre Michon[2], Benjamin Nachman[1], Peter Nugent[1], Mathis Reymond[8], David Rousseau[2,6], Benjamin Sluijter[9,1], Benjamin Thorne[1], Ihsan Ullah[6], and Yulei Zhang[3]

[1]Lawrence Berkeley National Laboratory
[2]Université Paris-Saclay, CNRS/IN2P3, IJCLab
[3]University of Washington, Seattle
[4]University of California, Berkeley
[5]University of California, Irvine
[6]ChaLearn
[7]University of California, San Diego
[8]Université Paris-Saclay
[9]Universiteit Leiden

fair-universe@lbl.gov
https://fair-universe.lbl.gov

Compiled: December 19, 2024

Abstract

The FAIR Universe – HiggsML Uncertainty Challenge focuses on measuring the physics properties of elementary particles with imperfect simulators due to differences in modelling systematic errors. Additionally, the challenge is leveraging a large-compute-scale AI platform for sharing datasets, training models, and hosting machine learning competitions. Our challenge brings together the physics and machine learning communities to advance our understanding and methodologies in handling systematic (epistemic) uncertainties within AI techniques.

2410.02867

- Successor to the **2014 Higgs ML Challenge**, now targeting parameter estimation rather than pure classification.
- Objective: deliver a confidence interval on the signal strength μ while being **robust to systematics & able to quantify them**.
- Improvements: Larger dataset (from 800k to $\sim$300 M events).
- Parameterised systematics (multiple nuisance parameters).
- Challenge: scale modern ML techniques to this high-dimensional systematic landscape.

ÖAW AUSTRIAN ACADEMY OF SCIENCES

HEPHY INSTITUTE OF HIGH ENERGY PHYSICS

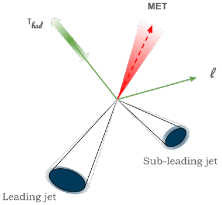
Higgs Uncertainty Challenge

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

Variable	Mean	Sigma	Range
α_{tes}	1.	0.01	[0.9, 1.1]
α_{jes}	1.	0.01	[0.9, 1.1]
$\alpha_{\text{soft\_met}}$	0.	1.	[0., 5.]
$\alpha_{\text{ttbar\_scale}}$	1.	0.02	[0.8, 1.2]
$\alpha_{\text{diboson\_scale}}$	1.	0.25	[0., 2.]
$\alpha_{\text{bkg\_scale}}$	1.	0.001	[0.99, 1.01]

Data: 28 Input Features

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



Higgs Uncertainty Challenge

- Pseudo-experiments
  - dataset representative of what would be measured from 10 fb^{-1} for a given value of μ and of the Nuisance Parameters
- Objective
  - Measure signal strength parameter μ
  - Give correct and small 68% CI on the measurement
- Evaluation Metrics
  - **Interval width (ω)** averaged over N test sets $\omega = \frac{1}{N_{test}} \sum_{i=1}^{N} |\mu_{84,i} - \mu_{16,i}|$
  - **Coverage (c):** fraction of time μ is contained $c = \frac{1}{N_{test}} \sum_{i=1}^{N} 1 \text{ if } \mu_{\text{true},i} \in [\mu_{16,i} - \mu_{84,i}]$

Primary Features (PRI)			
Symbol	**Description**	**Symbol**	**Description**
p_{T}^{ℓ}	Transverse momentum of the lepton	η^{ℓ}	Pseudorapidity of the lepton
ϕ^{ℓ}	Azimuthal angle of the lepton	$p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$	Transverse momentum of the hadronic τ
$\eta^{\tau_{\mathrm{had}}}$	Pseudorapidity of the hadronic τ	$\phi^{\tau_{\mathrm{had}}}$	Azimuthal angle of the hadronic τ
$p_{\mathrm{T}}^{j_1}$	Transverse momentum of the leading jet	η^{j_1}	Pseudorapidity of the leading jet
ϕ^{j_1}	Azimuthal angle of the leading jet	$p_{\mathrm{T}}^{j_2}$	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	$\sum_{\mathrm{jets}} p_{\mathrm{T}}$	Scalar sum of transverse momenta of all jets
$\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse momentum	ϕ^{miss}	Azimuthal angle of missing transverse momentum
Derived Features (DER)			
Symbol	**Description**	**Symbol**	**Description**
$m_{\mathrm{T}}(\ell, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$	Transverse mass of lepton and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$	m_{vis}	Visible invariant mass of τ_{had} and ℓ
p_{T}^{H}	Vector sum of $p_{\mathrm{T}}^{\tau_{\mathrm{had}}}, p_{\mathrm{T}}^{\ell}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}$	$m^{j_1 j_2}$	Invariant mass of the two leading jets
$\Delta R(\ell, miss)$	Angular distance between ℓ and MET	$\Delta R(\tau_{\mathrm{had}}, miss)$	Angular distance between τ_{had} and MET
$\Delta R(j_1, miss)$	Angular distance between j_1 and MET	$\Delta R(j_2, miss)$	Angular distance between j_2 and MET
$\Delta R(j_1, \tau_{\mathrm{had}})$	Angular distance between j_1 and τ_{had}	$\Delta R(j_2, \tau_{\mathrm{had}})$	Angular distance between j_2 and τ_{had}
$\Delta R(j_1, \ell)$	Angular distance between j_1 and ℓ	$\Delta R(j_2, \ell)$	Angular distance between j_2 and ℓ
$\Delta R(j_1, j_2)$	Angular distance between j_1 and j_2	$\Delta R(\tau_{\mathrm{had}}, \ell)$	Angular distance between τ_{had} and ℓ
$p_{\mathrm{T}}^{\mathrm{tot}}$	Vector sum of all visible momenta and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$	$\sum p_{\mathrm{T}}$	Scalar sum of all visible momenta and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$
C_{ϕ}^{miss}	Azimuthal centrality of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ w.r.t. $\ell, \tau_{\mathrm{had}}$	C_{η}^{ℓ}	Pseudorapidity centrality of the lepton w.r.t. the jets
$p_{\mathrm{T}}^{\ell} / p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$	Transverse–momentum ratio of lepton to τ_{had}		

Overall Model Architecture.

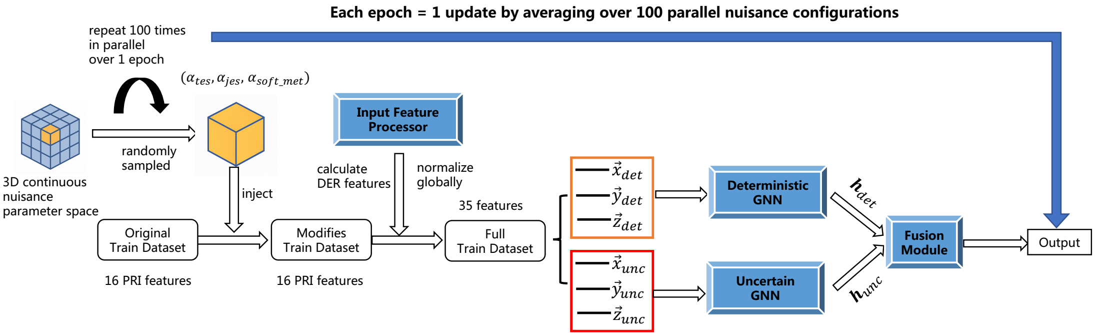


Table: Node feature assignment for final-state objects. Each node is represented by a one-hot encoding indicating its type. Deterministic and uncertainty-aware features are assigned separately according to their sensitivity to systematic variations. These two types of features are represented as $\vec{x}_{\text{det}}$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

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]

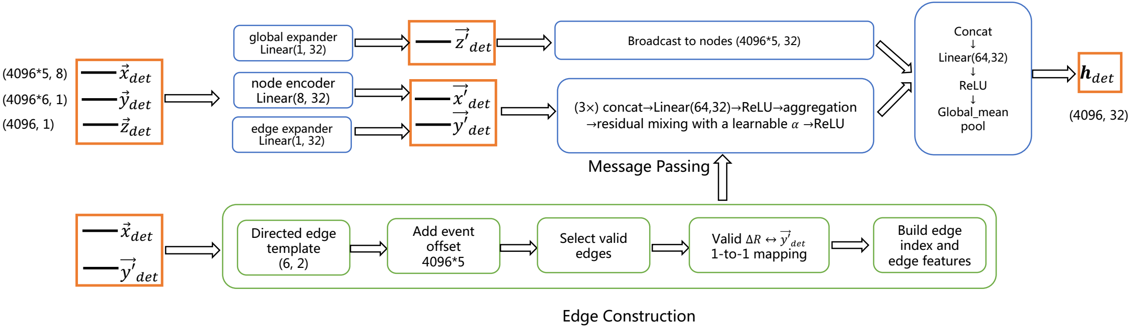
Table: Edge features constructed between pairs of final-state objects, which are categorized as deterministic or uncertainty-aware based on their sensitivity to systematic variations. These two types of features are represented as $\vec{y}_{\text{det}}$ and $\vec{y}_{\text{unc}}$.

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

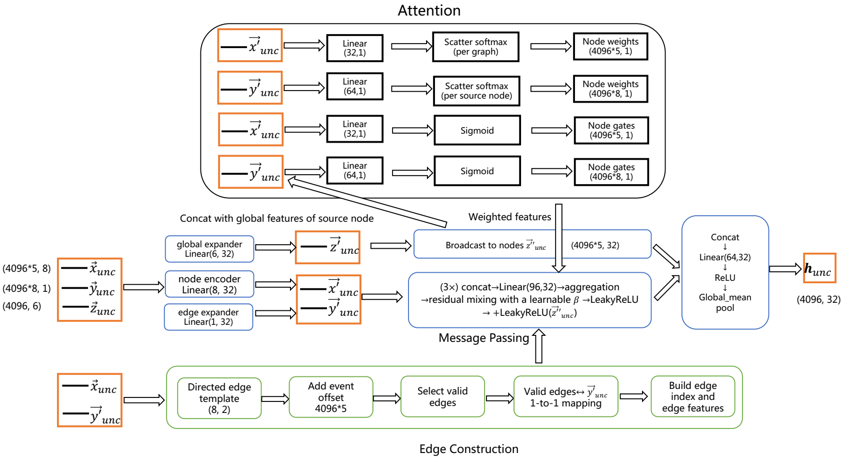
Table: Global features constructed from three or more final-state objects. Each feature is categorized by its defining particles and its sensitivity to systematic uncertainties. These two types of features are represented as $\vec{z}_{\text{det}}$ and $\vec{z}_{\text{unc}}$.

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

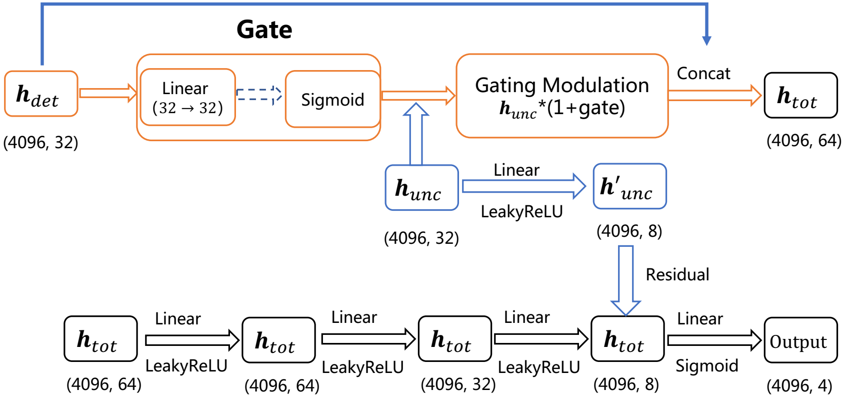
Deterministic GNN Branch



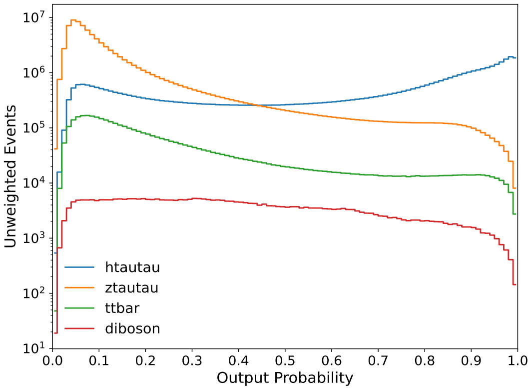
Uncertain-aware GNN Branch

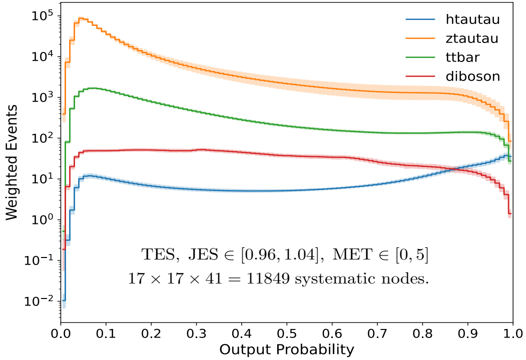


Fusion Module



Histogram of Output Probabilities





Relative Change in Event Yields

Process	Fraction [%]	TES [%]	ν [%]
$Z \to \tau\tau$	95.3	6.5	0.5
$t\bar{t}$	4.4	3.8	10
VV	0.3	5.3	100

Table: Relative change in event yields (up to 5σ variations) due to TES and normalization parameters for $Z \to \tau\tau$, $t\bar{t}$, and VV process in inclusive region.

Variable	Mean	Sigma	Range
α_{tes}	1.	0.01	[0.9, 1.1]
α_{jes}	1.	0.01	[0.9, 1.1]
$\alpha_{\text{soft\_met}}$	0.	1.	[0., 5.]
$\alpha_{\text{ttbar\_scale}}$	1.	0.02	[0.8, 1.2]
$\alpha_{\text{diboson\_scale}}$	1.	0.25	[0., 2.]
$\alpha_{\text{bkg\_scale}}$	1.	0.001	[0.99, 1.01]

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`	$p_{\mathrm{T}}^{j_1} > 50$ GeV $p_{\mathrm{T}}^{j_2} > 30$ GeV $m_{\mathrm{T}} > 70$ GeV	CR1	14.7	721.7	16 768.6	193.2	8.30×10^{-4}
`highMT-noVBFJet-tt`	$m_{\mathrm{T}} > 70$ 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`	$m_{\mathrm{T}} > 70$ GeV veto on `VBFJet` $\hat{f}_{VV} > 0.2$	CR3	1.8	189.5	207.6	597.4	1.8×10^{-3}

Region-wise Interpolation Tables

- **Regions:** 1 inclusive Full Region + 3 Control Regions (CR1, CR2, CR3).
- Objective: extract the signal strength μ while profiling six nuisance parameters $\nu = (\nu_{\text{tes}}, \nu_{\text{jes}}, \nu_{\text{met}}, \nu_{\text{bkg}}, \nu_{\text{tt}}, \nu_{\text{VV}})$.
- TES, JES $\in [0.96, 1.04]$, MET $\in [0, 5]$, $17 \times 17 \times 41 = 11\,849$ systematic nodes per region
- **Adaptive observable binning**
  1. Start from 100 uniform bins in the signal-class output probability.
  2. Greedy merge neighbouring bins until every remaining bin contains at least 10 weighted counts simultaneously in `tt`, `vv`, `z`.
  3. The procedure is run independently for each region, producing region-specific bin edges: 52 bins for FR, 14 for CR1, 15 for CR2, and 13 for CR3 after merging.

Parameterisation & Trilinear Interpolation

Nuisance re-mapping $\alpha \to \nu$

$$\alpha_{tes} = 1 + 0.01\,\nu_{tes}, \qquad \alpha_{jes} = 1 + 0.01\,\nu_{jes},$$
$$\alpha_{met} = e^{\nu_{met}} - 1, \qquad \alpha_{bkg} = e^{\sigma_{bkg} * \nu_{bkg}},$$
$$\alpha_{tt} = e^{\sigma_{tt} * \nu_{tt}}, \qquad \alpha_{VV} = e^{\sigma_{VV} * \nu_{VV}}$$

Vectorised $17 \times 17 \times 41$ trilinear interpolation (up to 4 σ (99.9937%))

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

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

Prior widths:$\sigma_{tes,jes,met} = 1$; $\sigma_{tt} = 0.02$; $\sigma_{vv} = 0.25$; $\sigma_{bkg} = 0.001$.
The yields of all four regions are concatenated into one vector $\boldsymbol{\lambda}$.

Negative Log-Likelihood

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

- **Profiling ν at fixed μ**
  1. Global search: `dual_annealing` on $\nu_k \in [-5,5]$ (met bound $[\ln 1.001, \ln 6]$).
  2. All nuisance parameters mapped to physical domain.
  3. Local refinement: `L-BFGS-B` + 5 random restarts.

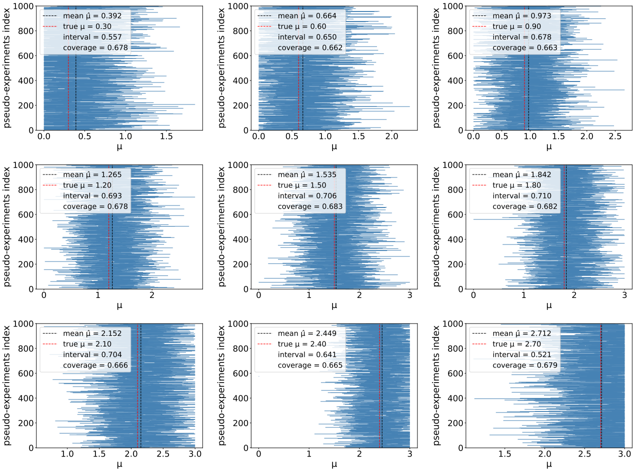
μ Scan and Confidence Interval

- **Scan grid:** $\mu \in [0, 3]$ (150 points). Compute profile NLL. Every 10th 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 `L-BFGS-B` (`maxiter`=150, `ftol`=1e-12) and up to five random restarts.
- Uses analytical gradients (via PyTorch autograd)
- **68% CL limits:**

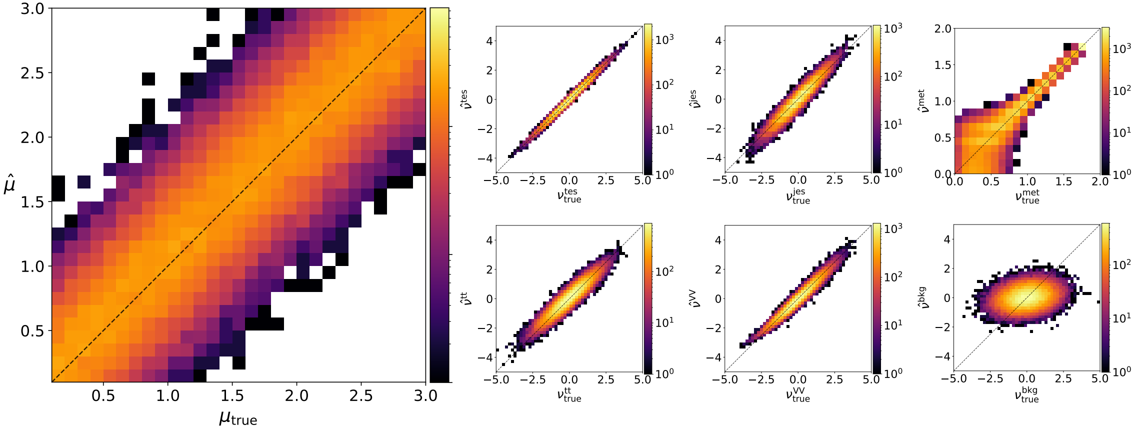
$$\Delta\text{NLL}(\mu) = \text{NLL}(\mu) - \text{NLL}_{\text{min}} = 0.5$$

  Roots (p_{16}, p_{84}) are located with Brent's method; spline and quadratic fall-backs guarantee a solution.
- **Output:** Signal strength estimate $\hat{\mu}$ with 68% confidence interval $[p_{16},\ p_{84}]$, along with fitted nuisance parameters $\hat{\alpha}_i$.

Toy Studies

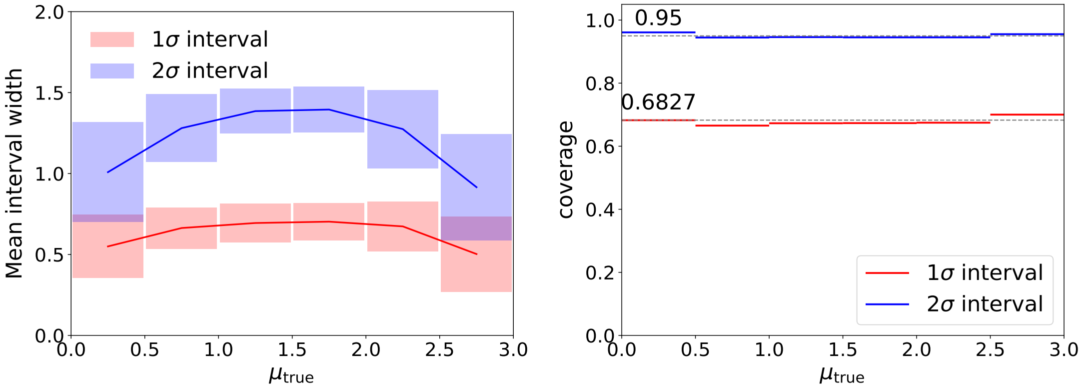

mean μ̂ = 0.392
true μ = 0.30
interval = 0.557
coverage = 0.678
mean μ̂ = 0.664
true μ = 0.60
interval = 0.650
coverage = 0.662
mean μ̂ = 0.973
true μ = 0.90
interval = 0.678
coverage = 0.663
mean μ̂ = 1.265
true μ = 1.20
interval = 0.693
coverage = 0.678
mean μ̂ = 1.535
true μ = 1.50
interval = 0.706
coverage = 0.683
mean μ̂ = 1.842
true μ = 1.80
interval = 0.710
coverage = 0.682
mean μ̂ = 2.152
true μ = 2.10
interval = 0.704
coverage = 0.666
mean μ̂ = 2.449
true μ = 2.40
interval = 0.641
coverage = 0.665
mean μ̂ = 2.712
true μ = 2.70
interval = 0.521
coverage = 0.679
pseudo-experiments index
μ

Toy Studies



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

Interval and Coverage



Summary

- **Architecture** – Dual-branch GNN:
  Deterministic GNN branch for features unaffected by nuisance variations
  Uncertainty-aware GNN branch for systematics-perturbed features
- **Training** – The uncertainty-aware GNN branch averages over 100 different nuisance configurations for each update.
- **Interpolation** – We interpolate event yields across a dense nuisance grid using region-specific templates, building a smooth surrogate likelihood that combines all four analysis regions.
- **Measurement** – The signal strength μ is extracted via profile likelihood scanning over nuisance parameters. The 68% confidence interval is defined by $\Delta\mathrm{NLL} = 0.5$.
- **Performance** – Large-scale pseudo-experiments demonstrate accurate coverage and consistently narrow intervals, validating the method's reliability under systematic uncertainties.
  Based on public leaderboard trends, our method likely ranks around **3rd place**.