Unbinned measurements with machine-learned systematic uncertainties

Ricardo Barrué, Lisa Benato, Cristina Giordano, Ang Li, Robert Schöfbeck, Dennis Schwarz, Maryam Shooshtari, Daohan Wang

CMS

MBI Marietta-Blau-Institut für Teilchenphysik

ÖAW Österreichische Akademie der Wissenschaften

Introduction

In stat-dominated analyses, SBI can improve sensitivity without solid treatment of systematics

In systematics-dominated analyses, need SBI methods to account for:

- O(100-1000) nuisance parameters
- highly non-trivial impact on event kinematics
- controlled computational costs

This presentation: efficient, refinable methodology for modeling systematic uncertainties in SBI

- Introduced in MLST 6 015007 (2025), tested in Phys.Rev.D 112 (2025) 5, 052006

Extended likelihood and unbinned likelihood ratio

Extended likelihood: $$L(\mathcal{D}|\boldsymbol{\theta},\boldsymbol{\nu}) = \mathrm{P}_{\mathcal{L}(\boldsymbol{\nu})\sigma(\boldsymbol{\theta},\boldsymbol{\nu})}(N)\prod_{i=1}^{N} p(\boldsymbol{x}_i|\boldsymbol{\theta},\boldsymbol{\nu}) = \mathrm{P}_{\mathcal{L}(\boldsymbol{\nu})\sigma(\boldsymbol{\theta},\boldsymbol{\nu})}(N)\prod_{i=1}^{N}\frac{1}{\sigma(\boldsymbol{\theta},\boldsymbol{\nu})}\frac{\mathrm{d}\Sigma(\boldsymbol{x}_i|\boldsymbol{\theta},\boldsymbol{\nu})}{\mathrm{d}\boldsymbol{x}},$$

Profile likelihood ratio ($q_{\boldsymbol{\theta}} \equiv t_\mu$ in ATLAS SBI paper):

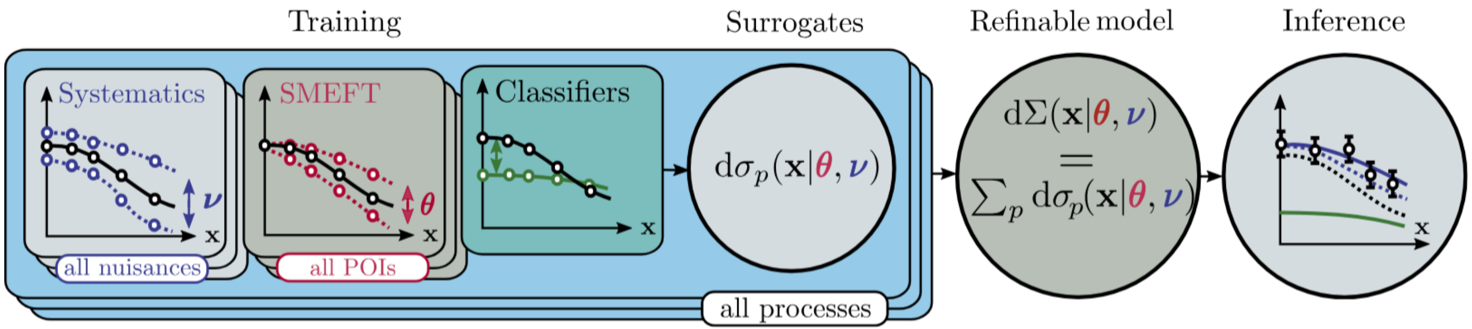
$$q_{\boldsymbol{\theta}}(\mathcal{D}) = \min_{\boldsymbol{\nu}} u(\mathcal{D},\mathcal{A}|\boldsymbol{\nu},\boldsymbol{\theta}) - \min_{\boldsymbol{\nu},\boldsymbol{\theta}} u(\mathcal{D},\mathcal{A}|\boldsymbol{\nu},\boldsymbol{\theta}),$$

where u ($\equiv \lambda$ in ATLAS SBI paper) is given by:

$$-\frac{1}{2}u(\mathcal{D},\mathcal{A}|\boldsymbol{\nu},\boldsymbol{\theta}) = -\mathcal{L}(\boldsymbol{\nu})\,\sigma(\boldsymbol{\theta},\boldsymbol{\nu}) + \mathcal{L}_0\,\sigma(\mathrm{SM}) + \sum_{i=1}^{N(\mathcal{D})}\log\left(\frac{\mathcal{L}(\boldsymbol{\nu})}{\mathcal{L}_0}\frac{\mathrm{d}\Sigma(\boldsymbol{x}_i|\boldsymbol{\theta},\boldsymbol{\nu})}{\mathrm{d}\Sigma(\boldsymbol{x}_i|\mathrm{SM})}\right) - \frac{1}{2}\sum_{k=1}^{K}\nu_k^2,$$

Requires differential cross-section ratio in observed events - **intractable**

A surrogate for the unbinned likelihood ratio



$$\frac{d\Sigma(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{\nu})}{d\Sigma(\boldsymbol{x}|\mathrm{SM})} = \sum_{p=1}^{N_p} \hat{R}_p(\boldsymbol{x}|\boldsymbol{\theta})\, \alpha_{\mathrm{norm},\,p}^{\nu_{\mathrm{norm},\,p}}\, \hat{S}_p(\boldsymbol{x}|\boldsymbol{\nu})\, \hat{g}_p(\boldsymbol{x})$$

$$\hat{R}_p(\boldsymbol{x}|\boldsymbol{\theta}) \simeq \frac{d\sigma_p(\boldsymbol{x}|\boldsymbol{\theta},\mathbf{0})}{d\sigma_p(\boldsymbol{x}|\mathrm{SM})} \qquad \hat{S}_p(\boldsymbol{x}|\boldsymbol{\nu}) \simeq \frac{d\sigma_p(\boldsymbol{x}|\mathbf{0},\boldsymbol{\nu})}{d\sigma_p(\boldsymbol{x}|\mathrm{SM})}. \qquad \hat{g}_p(\boldsymbol{x}) \simeq \frac{d\sigma_p(\boldsymbol{x}|\mathrm{SM})}{\sum_q d\sigma_q(\boldsymbol{x}|\mathrm{SM})}$$

Will **not** discuss unbinned POI parametrizations, focusing on systematics surrogate *S*

Unbinned surrogate for systematics

First implementation in full experimental analysis done by ATLAS (see Tae's presentation)

Has some limitations:

- fixed order in ν ($\equiv \alpha$ in ATLAS SBI paper)
- two classifiers per nuisance parameter
- does not handle non-factorizable uncertainties
- kinematics not taken into account for $|\alpha| < 1$
- does not use joint likelihood ratio information

How to extend this approach?

Systematic surrogates in binned scenario

In the binned scenario, sum over observed events written as $\sum_{i=1}^{N_{\text{bins}}} N_{\text{obs},i} \log\left(\frac{\mathcal{L}(\nu)}{\mathcal{L}_0}\frac{\sigma_i(\theta,\nu)}{\sigma_i(\text{SM})}\right)$

Following factorization in slide 4 leads to terms $\log\frac{\sigma_{i,p}(\nu_{\text{calib}})}{\sigma_{i,p}(0)}$

For nuisance parameters != overall normalization NPs, explore parametric ansatz:

$$\log\frac{\sigma_{i,p}(\nu_{\text{calib}})}{\sigma_{i,p}(0)} = \nu_A \hat{\Delta}_{i,p,A} \qquad \text{e.g. } A = [(\mu_R),(\mu_F),(\mu_R,\mu_R),(\mu_F,\mu_F),(\mu_R,\mu_F),\ldots]$$

MBI

Systematic parametrizations in binned scenario - II

How to determine binned coefficients Δ ?

- Take events simulated at basis points **\nu** = (**0, \nu_1, \nu_2**, ...) and minimize

$$\chi^2 = \sum_{\boldsymbol{\nu}\in\mathcal{V}} \left(\log \frac{\sigma_{i,p}(\boldsymbol{\nu})}{\sigma_{i,p}(\mathbf{0})} - \nu_A \hat{\Delta}_{i,p,A} \right)^2 \qquad \hat{\Delta}_{i,p,A} = \left[\sum_{\boldsymbol{\nu}\in\mathcal{V}} \boldsymbol{\nu}\boldsymbol{\nu}^{\mathrm{T}} \right]^{-1}_{AB} \left[\sum_{\boldsymbol{\nu}\in\mathcal{V}} \boldsymbol{\nu} \log \frac{\sigma_{i,p}(\boldsymbol{\nu})}{\sigma_{i,p}(\mathbf{0})} \right]_B .$$

Main features:

- Can derive separately the coefficients for factorizable NPs
- Can take into account non-factorizable NPs
- Can be extended to arbitrary orders

Refinable modeling for unbinned analyses

Goal: promote coefficients $\Delta_{i,p,A}$ to functions $\Delta_{p,A}(\boldsymbol{x})$, learned using ML estimators.

Start from cross-entropy loss, but taking as ansatz $$\hat{f}_{\boldsymbol{\nu}}(\boldsymbol{x}) = \frac{1}{1+\exp(\nu_A \hat{\Delta}_A(\boldsymbol{x}))},$$

This leads to $$L[\hat{\Delta}_A] \approx \sum_{\boldsymbol{\nu}\in\mathcal{V}} \Bigg[\sum_{\{\boldsymbol{x}_i, w_i\}\in\mathcal{D}_{\mathbf{0}}} w_i \,\mathrm{Soft}^+(\nu_A \hat{\Delta}_A(\boldsymbol{x}_i) \;+ \sum_{\{\boldsymbol{x}_i, w_i\}\in\mathcal{D}_{\boldsymbol{\nu}}} w_i \,\mathrm{Soft}^+(-\nu_A \hat{\Delta}_A(\boldsymbol{x}_i)) \Bigg], \quad \mathrm{Soft}^+(x) = \log(1+\exp(x)).$$

Bonus: joint likelihood ratio can be incorporated (improve training efficiency)

Refinable modeling for unbinned analyses - II

This procedure inherits all the characteristics of the binned approach:

- Can derive separately coefficient functions for factorizable NPs
- Can take into account non-factorizable NPs
- Can be extended to arbitrary orders

Additional advantages w.r.t. ATLAS approach:

- Reduces the number of estimators to learn (if using log-linear parametrization)
- Also uses kinematic features to interpolate effects in $|\alpha| < 1$
- Makes use of joint likelihood ratios where available

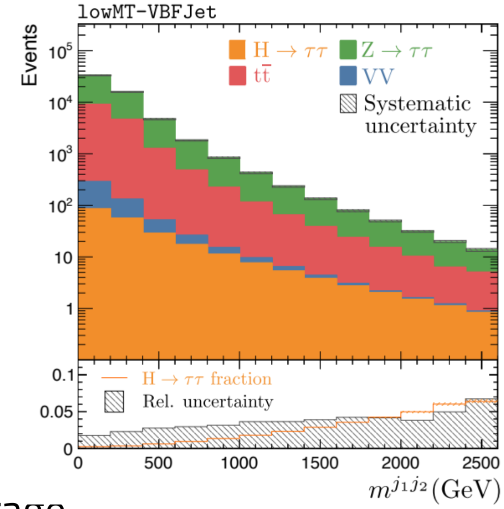
Quality of estimators can be assessed using standard methods.

Case study

Introduction

Based on FAIR Universe Higgs Uncertainty Challenge (arXiv:2410.02867):

- Signal: H->tau_lep tau_had
- Backgrounds: Z->tautau, ttbar, VV
- Systematics:
  - background normalizations - per-process + total
  - “calibration” uncertainties: MET, JES and TES
- Features, event selection and categorization in backup
- Target: Inclusive signal strength, central value and 68%CL



Goal: minimize average width of 68%CL interval, maintaining coverage

Differential model

Writing the differential cross-sections as

$$d\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu}) = \mu\, d\sigma_{\mathrm{H}}(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}}) + \sum_{p=\mathrm{Z},\mathrm{t\bar{t}},\mathrm{VV}} d\sigma_p(\boldsymbol{x}|\boldsymbol{\nu}).$$

$$d\sigma_{\mathrm{Z}}(\boldsymbol{x}|\boldsymbol{\nu}) = (1+\alpha_{\mathrm{bkg}})^{\nu_{\mathrm{bkg}}} d\sigma_{\mathrm{Z}}(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}}) \qquad d\sigma_{\mathrm{t\bar{t}}}(\boldsymbol{x}|\boldsymbol{\nu}) = (1+\alpha_{\mathrm{bkg}})^{\nu_{\mathrm{bkg}}}(1+\alpha_{\mathrm{t\bar{t}}})^{\nu_{\mathrm{t\bar{t}}}} d\sigma_{\mathrm{t\bar{t}}}(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})$$

$$d\sigma_{\mathrm{VV}}(\boldsymbol{x}|\boldsymbol{\nu}) = (1+\alpha_{\mathrm{bkg}})^{\nu_{\mathrm{bkg}}}(1+\alpha_{\mathrm{VV}})^{\nu_{\mathrm{VV}}} d\sigma_{\mathrm{VV}}(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}})$$

The differential cross-section ratio is written as

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

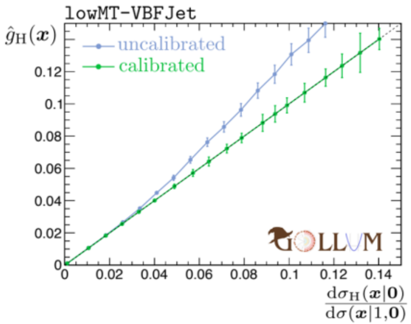
Process fraction estimators

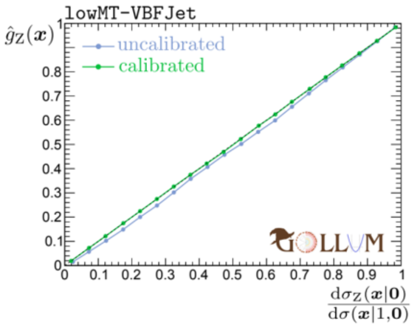
g implemented with multi-classifier

$$L_{\mathrm{CE}}[\hat{f}_p] \approx -\sum_{p=1}^{N_p} \sum_{\{\boldsymbol{x}_i, w_i\} \in \mathcal{D}_p} \frac{w_i}{\mathcal{L}\sigma_p} \log\Big(\hat{f}_p(\boldsymbol{x}_i)\Big). \qquad \hat{g}^*_p(\boldsymbol{x}) = \frac{\hat{f}^*_p(\boldsymbol{x})\sigma_p}{\sum_q \hat{f}^*_q(\boldsymbol{x})\sigma_q} \simeq \frac{\mathrm{d}\sigma_p(\boldsymbol{x}|\mathbf{0})}{\sum_q \mathrm{d}\sigma_q(\boldsymbol{x}|\mathbf{0})}$$

Calibrated using multi class version of isotonic regression

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





Large miscalibrations in signal, smaller in main backgrounds

- lead to large impact on final result (discussed later)

Systematics model and surrogates

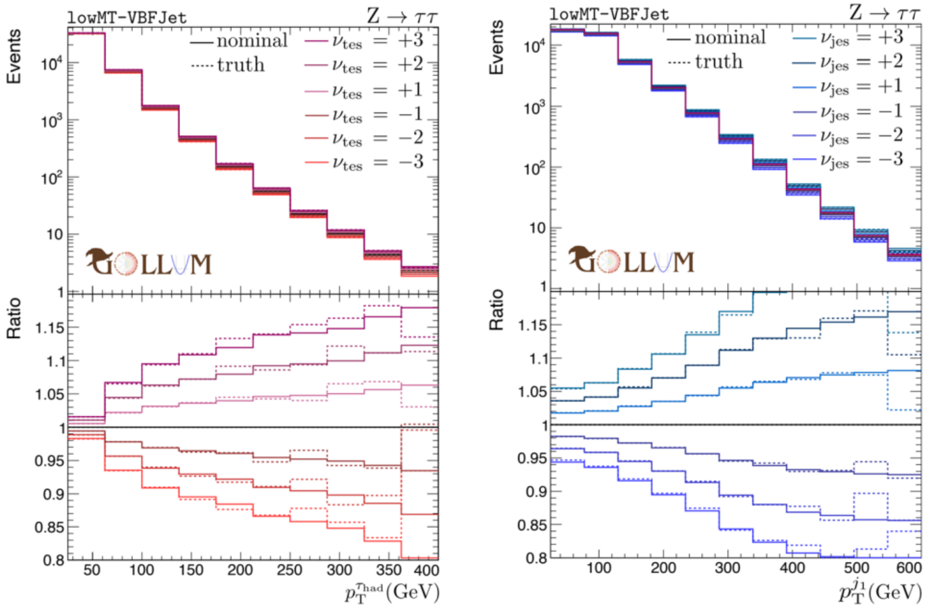
Quadratic parametrization inc. cross-terms

- 9 coefficient functions derived with NNs
- one NNs per process and region

Closure plots (left): excellent closure in bulk, interpolates smoothly in tails

MLST 6 015007 (2025) also includes:

- additional systematics
- validation with Classifier Two Sample Tests

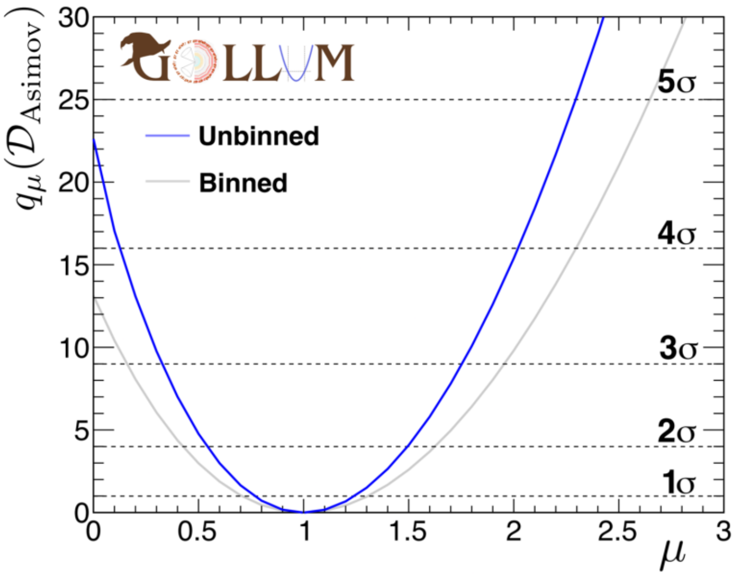


Results

No bias in minimum within 10^{-3} when scanning μ from 0.1-3

Overall quality of surrogates critical !

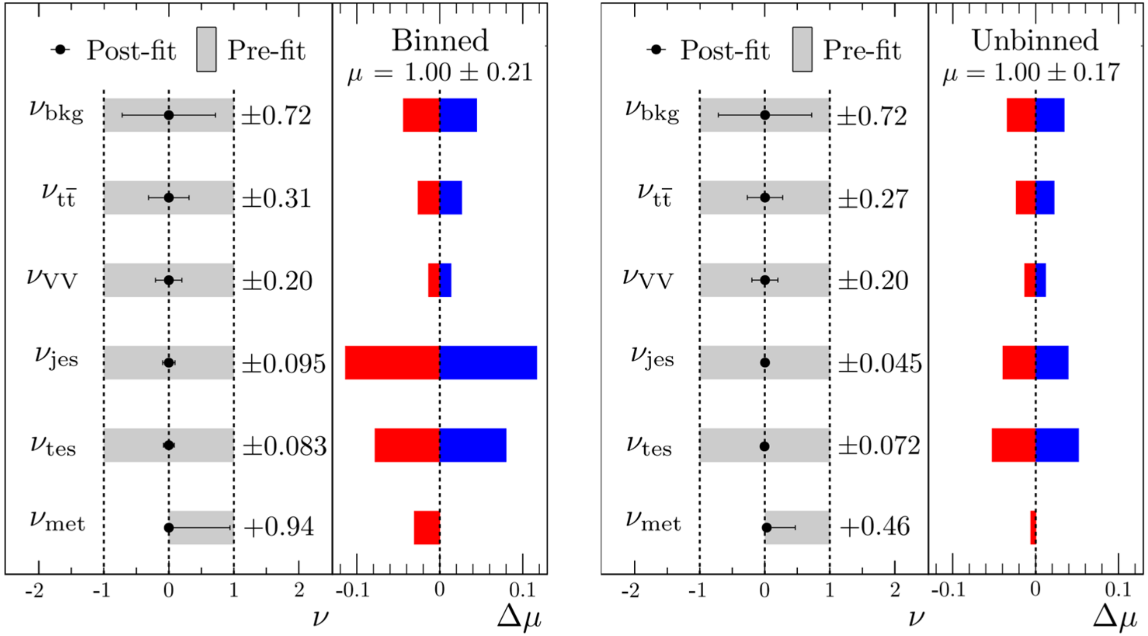
- removing calibration from multiclassifier leads to bias on μ within 0.3-0.4, with **strong dependence on** ν_{true}



Improvement of 20% on 1σ bounds with unbinned fit w.r.t. binned

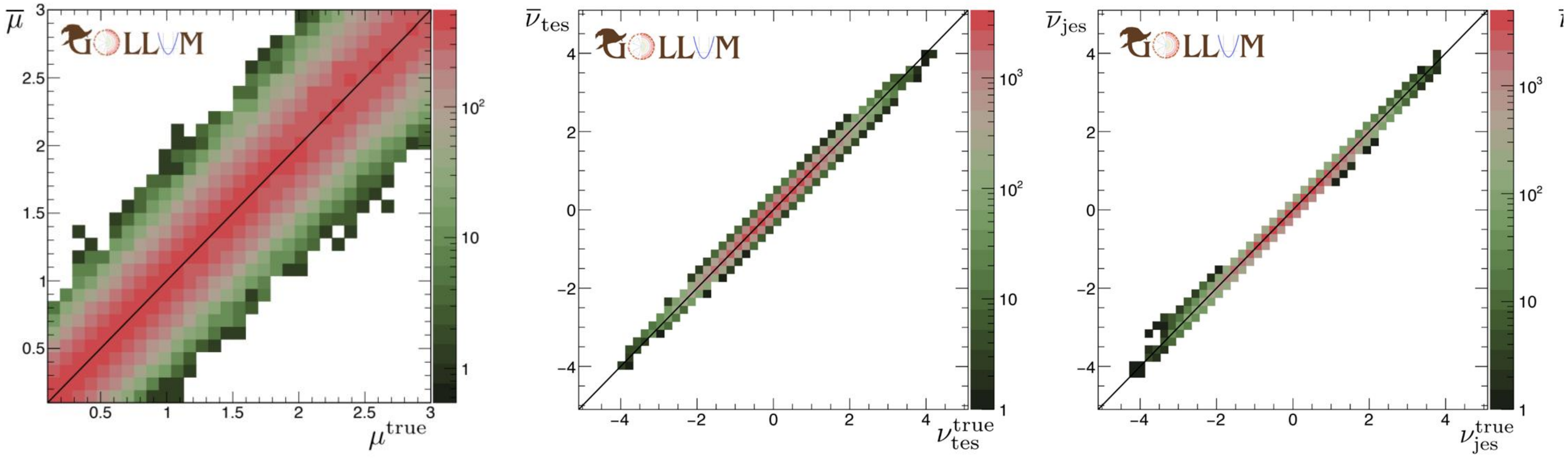
- higher expected improvement for higher dimensionality of parameter space

Impacts



Unbinned fit shows tighter constraints in calibration systematics.

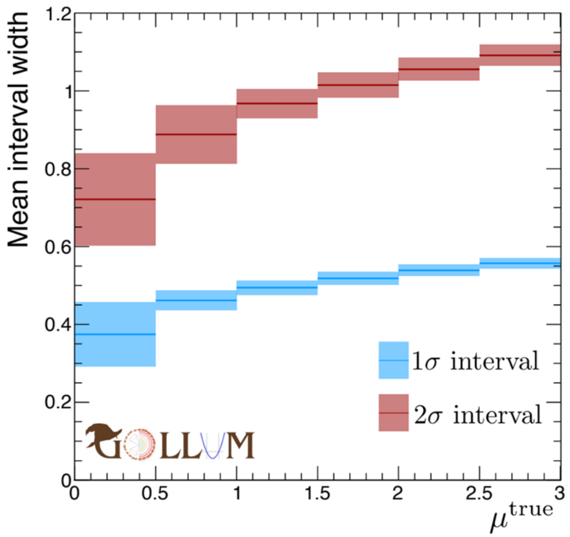
Toy studies

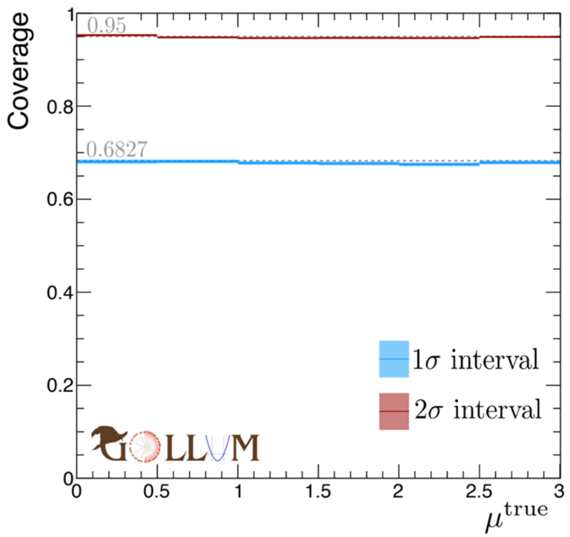


μ plot shows **excellent calibration,** same for ν_{tes} and ν_{jes} (highest impact)

- others show imperfect calibration (see backup), but less impactful/constrained

Mean interval width and coverage





Model provides correct coverage across the probed values of μ_{true}

- Won first place in challenge !

Conclusions

Showed efficient, refinable methodology for modeling systematics in SBI

- Can derive separately coefficient functions for factorizable NPs
- Can take into account non-factorizable NPs
- Can be extended to arbitrary orders

Showed application of methodology to FAIR Higgs Universe Challenge

- Improvement of 20% on 1σ bounds w.r.t. binned fit
- Provides correct coverage across probed values of POI

Open questions

- To achieve closure/calibration, should one first test a higher order parametrization or immediately explore more complex architectures (inc. ensembling) ?
- What to do with nuisances with very small effects?
  - N-dimensional smoothing? Accept smooth interpolation from estimators?
  - Can one say “I don’t need excellent closure as this NP has a negligible impact”?
- Cutoff between constraining systematics because using more information vs. needing to adjust nuisance model ?
- What to do in sparsely populated phase-space regions?
  - Idea: binned approach (overflow bin) explored in MLST 6 015007 (2025)

Thank you!

CMS

MBI Marietta-Blau-Institut für Teilchenphysik

ÖAW Österreichische Akademie der Wissenschaften

Backup

Event features

TABLE I. List of event features used in the analysis. Primary features correspond to direct kinematic observables, while derived features incorporate correlations and high-level event properties.

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

Regions + impact of systematics on total yields

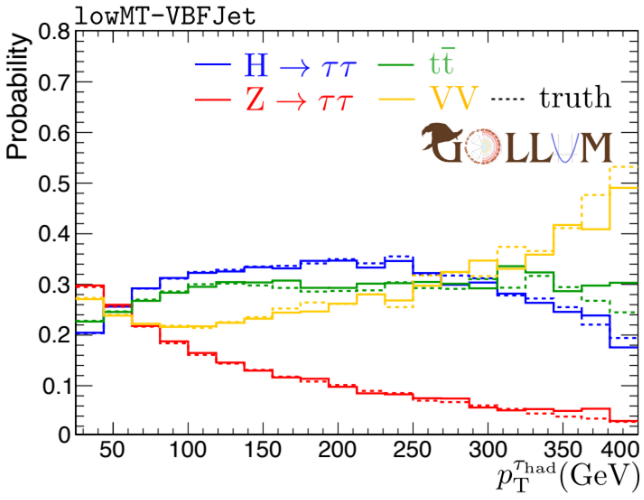
TABLE II. Summary of the event selections. The "UB" label indicates whether the selection is included in the unbinned analysis. Signal regions (SR) and control regions (CR) are treated the same. The predicted Poisson yields correspond to $\mathcal{L}\sigma$ where $\mathcal{L}$ denotes the luminosity and are reported for each sub-process and region. The last column shows the signal-to-background ratio. The total yields in the "inclusive" row are slightly lower than the total yields in Ref. [22] because of selections on p_T^H and p_T^ℓ mandated by FAIR-HUC. The regions marked with $*$ are subsets of the `highMT-noVBFJet` region and target the $t\bar{t}$ and VV normalization via requirements on multivariate discriminants explained in Sec. V A.

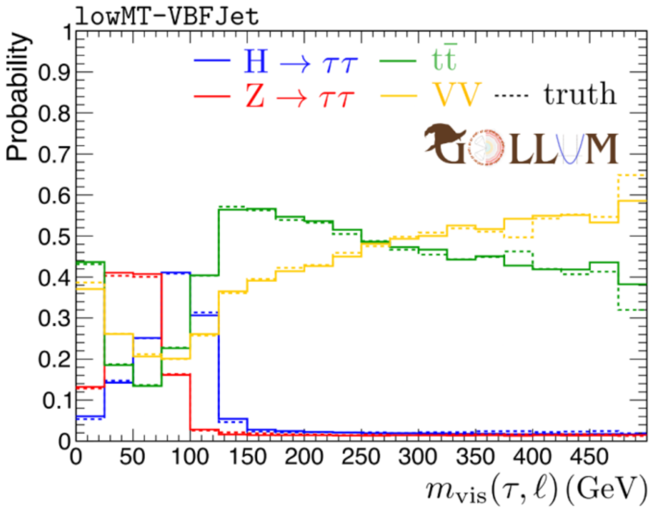
Region	Requirements	Type	Poisson yield $\mathcal{L}\sigma$				S/B
			$H \to \tau\tau$	$Z \to \tau\tau$	$t\bar{t}$	VV	
`lowMT-VBFJet`	$p_T^{j1} > 50$ GeV $p_T^{j2} > 30$ GeV $m_T \leq 70$ GeV	UB, SR	225.8	41 280.4	15 425.9	313.4	$3.96 \cdot 10^{-3}$
`highMT-VBFJet`	$p_T^{j1} > 50$ GeV $p_T^{j2} > 30$ 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	$2.53 \cdot 10^{-3}$
`inclusive`			966.0	901 137.5	41 283.4	3 433.5	$1.02 \cdot 10^{-3}$
`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	$2.3 \cdot 10^{-3}$

TABLE III. Systematic dependence of the four processes ($H \to \tau\tau$, $Z \to \tau\tau$, $t\bar{t}$, and VV) in the five regions. Each row represents the coefficients of a quadratic polynomial describing the impact of the nuisance parameters ν_{tes}, ν_{jes}, and ν_{met}.

	$\nu_{tes} \cdot 10^{-3}$	$\nu_{jes} \cdot 10^{-3}$	$\nu_{met} \cdot 10^{-5}$	$\nu_{tes}^2 \cdot 10^{-5}$	$\nu_{jes}^2 \cdot 10^{-5}$	$\nu_{met}^2 \cdot 10^{-5}$	$\nu_{tes}\nu_{jes} \cdot 10^{-5}$	$\nu_{jes}\nu_{met} \cdot 10^{-5}$	$\nu_{tes}\nu_{met} \cdot 10^{-5}$
`lowMT-VBFJet`									
$H \to \tau\tau$	6.03	13.0	-3.40	-9.91	-16.5	-9.15	4.63	5.28	-0.451
$Z \to \tau\tau$	9.89	22.0	2.57	-12.8	-16.0	-6.87	7.25	2.03	-1.78
$t\bar{t}$	8.78	7.15	-6.36	-8.89	-18.1	-13.7	2.38	0.466	-3.84
VV	11.0	20.4	11.3	-6.43	-4.85	-6.97	2.60	-59.9	14.5
`highMT-VBFJet`									
$H \to \tau\tau$	-1.95	8.85	29.3	23.3	67.7	148	-59.3	-10.3	2.60
$Z \to \tau\tau$	8.32	3.60	22.5	34.8	149	333	-119	-11.7	29.5
$t\bar{t}$	6.25	6.50	-0.402	-6.40	-8.06	14.6	-2.05	-1.05	-0.172
VV	6.57	16.5	-24.6	-4.91	-18.7	13.4	4.71	3.45	-10.9
`highMT-noVBFJet`									
$H \to \tau\tau$	-10.2	-3.83	-18.4	8.18	27.9	351	-18.6	2.41	-6.94
$Z \to \tau\tau$	-7.65	-2.56	46.7	2.35	46.4	930	-37.5	5.82	20.2
$t\bar{t}$	5.32	-24.0	-23.9	-5.14	9.06	7.79	-3.07	-9.88	3.33
VV	6.99	-3.16	7.56	-11.7	-2.86	5.61	-6.29	8.78	5.92
`lowMT-noVBFJet-ptH100`									
$H \to \tau\tau$	5.38	2.49	5.27	-7.84	-11.9	24.4	9.03	-4.20	1.61
$Z \to \tau\tau$	7.06	12.9	-9.71	-8.88	-19.9	51.6	9.48	-5.45	-1.04
$t\bar{t}$	7.50	-2.96	82.6	-4.46	-1.11	4.63	12.1	26.5	5.17
VV	10.2	9.49	341	27.6	-19.0	-72.5	9.88	66.8	33.3
`lowMT-noVBFJet-ptH0to100`									
$H \to \tau\tau$	6.42	-4.83	-0.726	-11.1	-0.15	-16.2	-0.20	0.205	-0.44
$Z \to \tau\tau$	12.5	-1.34	-2.28	-22.4	-0.644	-4.67	0.137	-0.055	-1.57
$t\bar{t}$	8.40	-28.4	15.0	-10.9	2.93	-11.3	-3.67	6.35	-8.47
VV	14.2	-4.07	-4.42	-10.8	-0.172	-8.79	5.44	0.067	1.59

Process fraction estimators - II





Binned 1-dimensional projections showing calibrated classifiers across phase-space.

Cross-section interpolation

$$
\begin{aligned}
&\mathcal{L}(\sigma(\mu,\boldsymbol{\nu}) - \sigma(1,\mathbf{0})) \\
&\quad = \mathcal{L}\int\left(\frac{\mathrm{d}\sigma(\boldsymbol{x}|\mu,\boldsymbol{\nu})}{\mathrm{d}\sigma(\boldsymbol{x}|1,\mathbf{0})} - 1\right)\frac{\mathrm{d}\sigma(\boldsymbol{x}|1,\mathbf{0})}{\mathrm{d}\boldsymbol{x}}\mathrm{d}\boldsymbol{x} \\
&\quad \approx \sum_{\{w_i,\boldsymbol{x}_i\}\in\mathcal{D}_{1,\mathbf{0}}} w_i\left(\hat{R}(\boldsymbol{x}_i|\mu,\boldsymbol{\nu}) - 1\right). \qquad (26)
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{\{w_i,\boldsymbol{x}_i\}\in\mathcal{D}_{1,\boldsymbol{\nu}}}\left(\hat{R}(\boldsymbol{x}_i|\mu,\boldsymbol{\nu}) - 1\right) w_i \\
&= \mu\,\mathrm{CSI}_{\mathrm{H}}(\boldsymbol{\nu}_{\mathrm{calib}}) \\
&\quad + (1+\alpha_{\mathrm{bkg}})^{\nu_{\mathrm{bkg}}}\,\mathrm{CSI}_{\mathrm{Z}}(\boldsymbol{\nu}_{\mathrm{calib}}) \\
&\quad + (1+\alpha_{\mathrm{bkg}})^{\nu_{\mathrm{bkg}}}(1+\alpha_{\mathrm{t\bar{t}}})^{\nu_{\mathrm{t\bar{t}}}}\,\mathrm{CSI}_{\mathrm{t\bar{t}}}(\boldsymbol{\nu}_{\mathrm{calib}}) \\
&\quad + (1+\alpha_{\mathrm{bkg}})^{\nu_{\mathrm{bkg}}}(1+\alpha_{\mathrm{VV}})^{\nu_{\mathrm{VV}}}\,\mathrm{CSI}_{\mathrm{VV}}(\boldsymbol{\nu}_{\mathrm{calib}}) \\
&\quad + (\mu - 1)\,\mathrm{CSIC}_{\mathrm{H}} \\
&\quad + ((1+\alpha_{\mathrm{bkg}})^{\nu_{\mathrm{bkg}}} - 1)\,\mathrm{CSIC}_{\mathrm{Z}} \\
&\quad + ((1+\alpha_{\mathrm{bkg}})^{\nu_{\mathrm{bkg}}}\,(1+\alpha_{\mathrm{t\bar{t}}})^{\nu_{\mathrm{t\bar{t}}}} - 1)\mathrm{CSIC}_{\mathrm{t\bar{t}}} \\
&\quad + ((1+\alpha_{\mathrm{bkg}})^{\nu_{\mathrm{bkg}}}\,(1+\alpha_{\mathrm{VV}})^{\nu_{\mathrm{VV}}} - 1)\mathrm{CSIC}_{\mathrm{VV}} \qquad (28)
\end{aligned}
$$

CSIC: evaluate surrogates ahead of time, sum using numerically stable methods (e.g. expm1).

$$
\mathrm{CSI}_p(\boldsymbol{\nu}_{\mathrm{calib}}) = \sum_{\{w_i,\boldsymbol{x}_i\}\in\mathcal{D}_{1,\mathbf{0}}} w_i\,\hat{g}_p(\boldsymbol{x})(\hat{S}_p(\boldsymbol{x}|\boldsymbol{\nu}_{\mathrm{calib}}) - 1), \qquad (29a)
$$

$$
\mathrm{CSIC}_p = \sum_{\{w_i,\boldsymbol{x}_i\}\in\mathcal{D}_{1,\mathbf{0}}} w_i\,\hat{g}_p(\boldsymbol{x}). \qquad (29b)
$$

CSI: use pre-computed g surrogates, evaluate S in grid (fine for $|\nu| < 3$, coarse outside), interpolate with quintic spline.

Alternative: separate by process, and use COMBINE ansatz for calibration systematics.

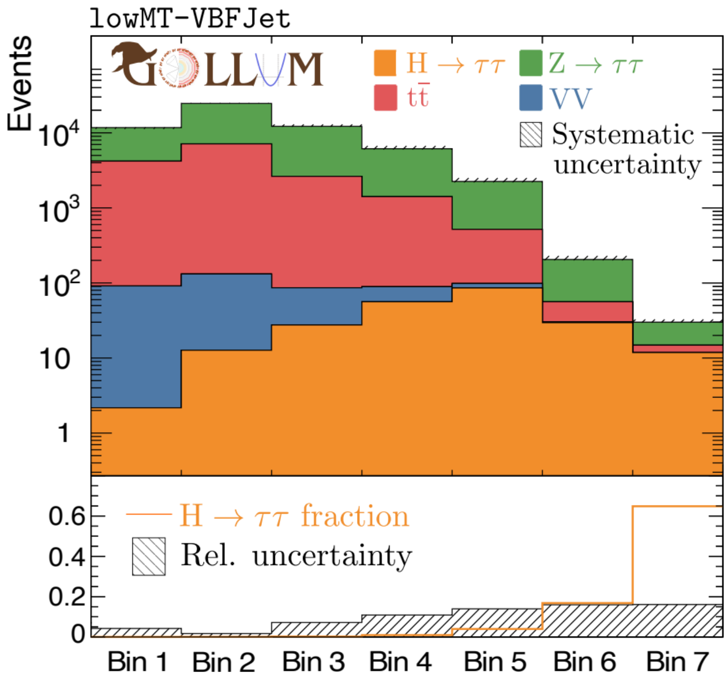
Binning unbinned regions

The distribution of the multi-classifier signal node, f_H, is binned with iterative procedure:

- starting from f_H = 1 find bins that maximize s = S/sqrt(S/B), with s <= 2

Likelihood constructed using standard binned procedures.

Finer binnings led to similar results.



Classifier Two Sample Tests

“To evaluate this result, we merge D_reweighted and D_SM, randomize labels, and train 1000 classifiers on pairs of identical subsets”

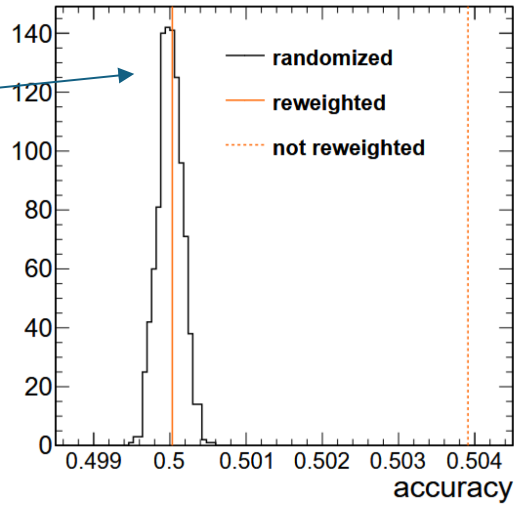
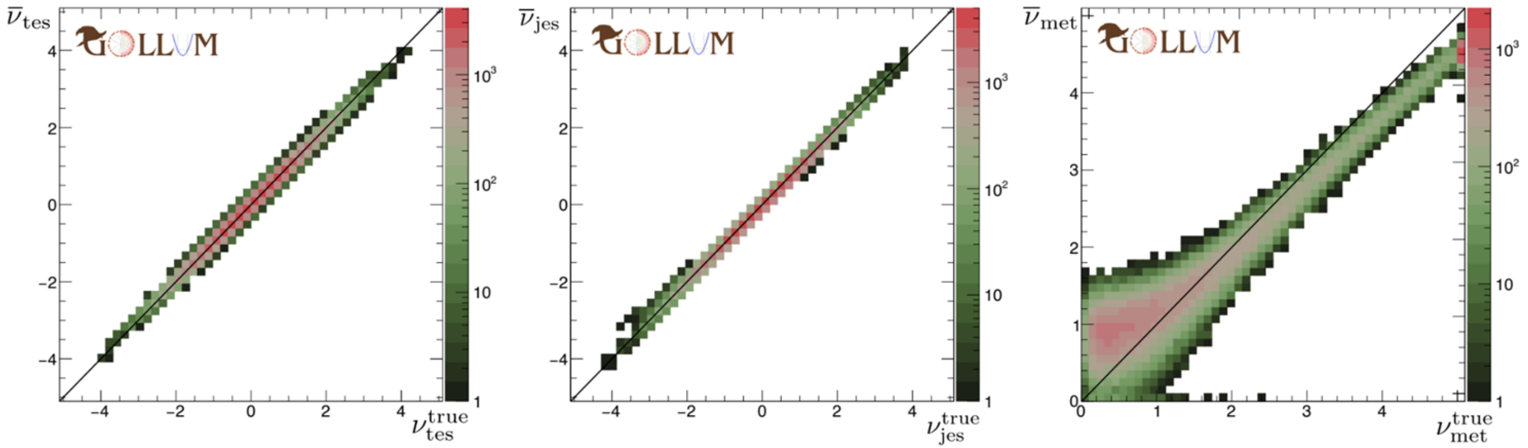


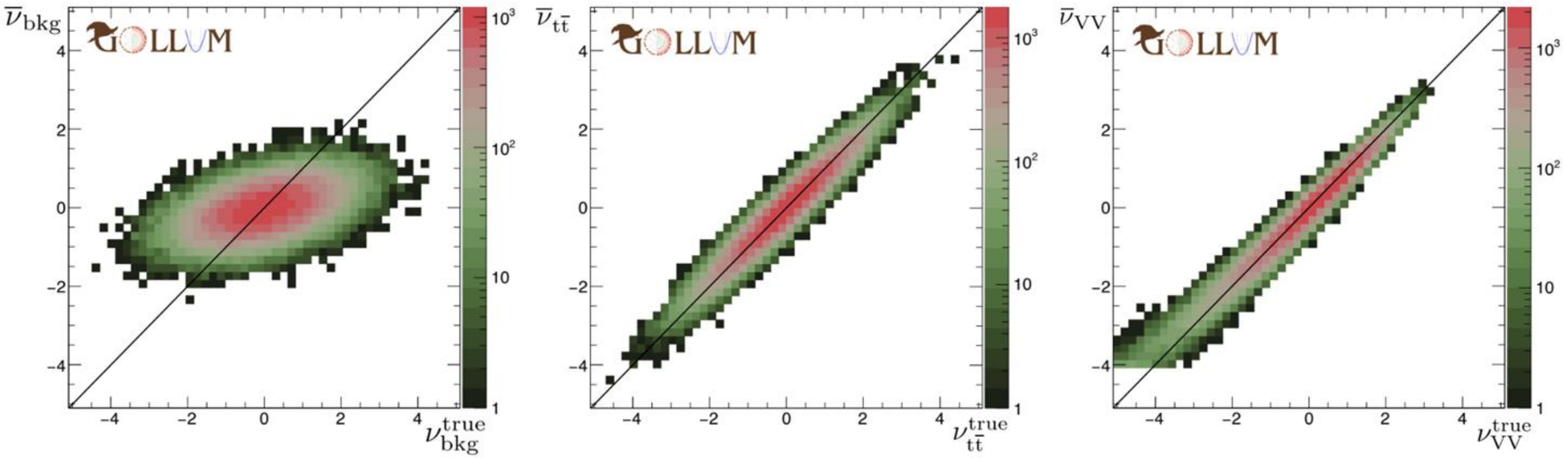
Figure 10: A neural-network classifier tests the surrogate $\hat{S}_{\mathrm{HF}}(\boldsymbol{x}|\nu_{\mathrm{HF}})$ for $\nu_{\mathrm{HF}} = 1$. The black contour shows the distribution of the classifier when the two samples are from the same distribution at $\boldsymbol{\nu}_{\mathrm{HF}} = 0$. The solid orange line indicates the accuracy of a classifier trained with an SM event sample and a sample with $\boldsymbol{\nu}_{\mathrm{HF}} = 1$ that is reweighted back to the SM using the surrogate under test, suggesting no flaw in the performance. The dashed orange line indicates the accuracy of the test trained with two unequal samples at $\boldsymbol{\nu}_{\mathrm{HF}} = 0$ and $\boldsymbol{\nu}_{\mathrm{HF}} = 1$.

Toy studies - II



Worse closure for \nu_MET, but also much less impactful/constrained.

Toy studies - III



Worse closure for \nu_bkg, but also much less impactful/constrained.

Fitting framework

YAML-based configs, defining:

- Processes, samples and regions
- POIs and nuisance parameters (inc. surrogates to use)
- Surrogate information: features, training configurations

Framework builds likelihood from configs and runs fit using Minuit.

Relevant features:

- surrogates evaluated before fit, results cached
- likelihood function JIT-compiled