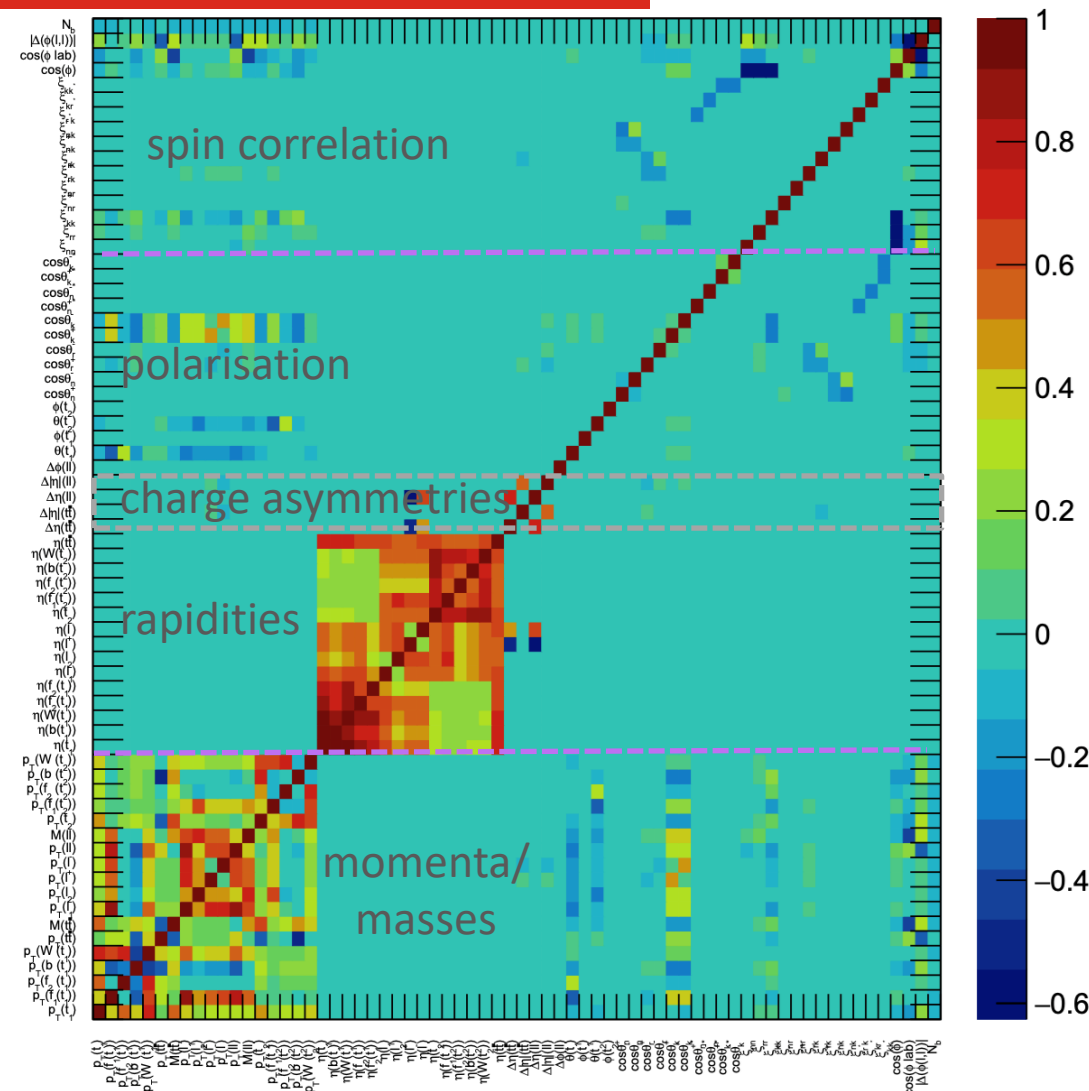
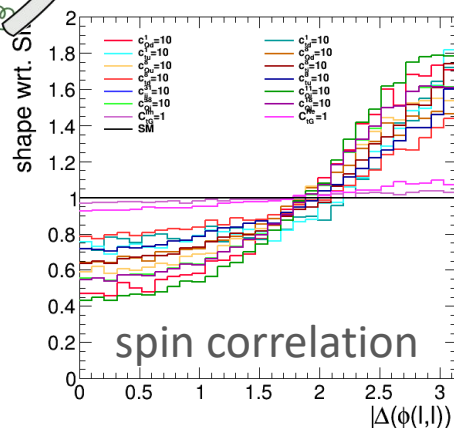
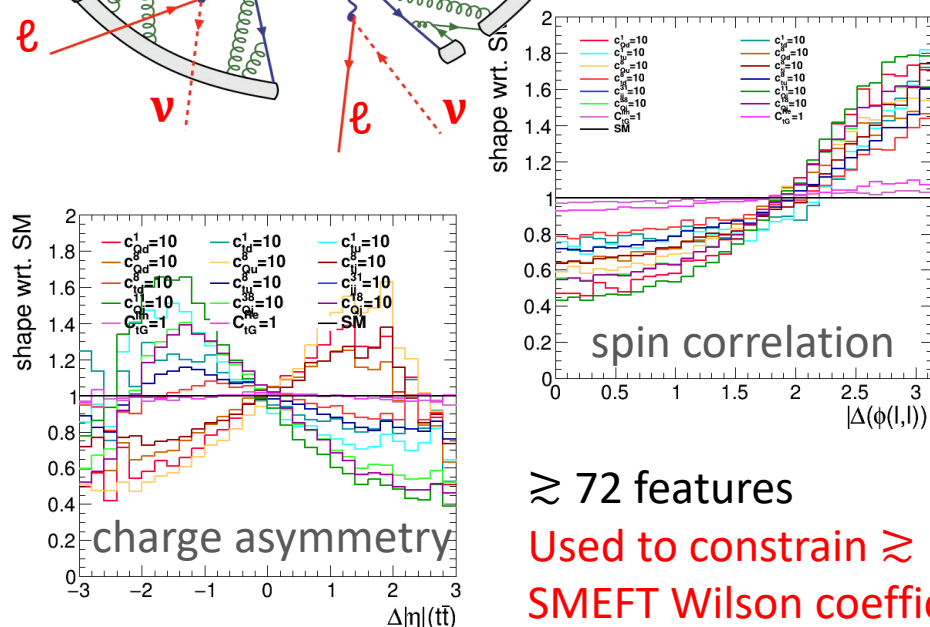
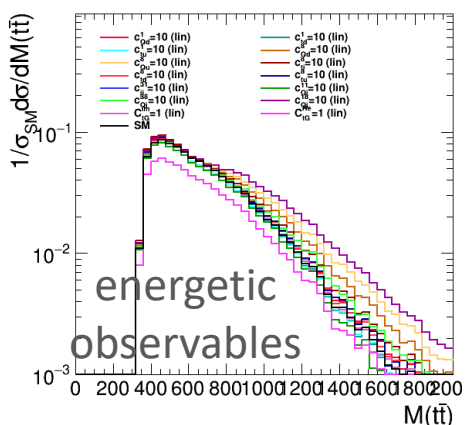
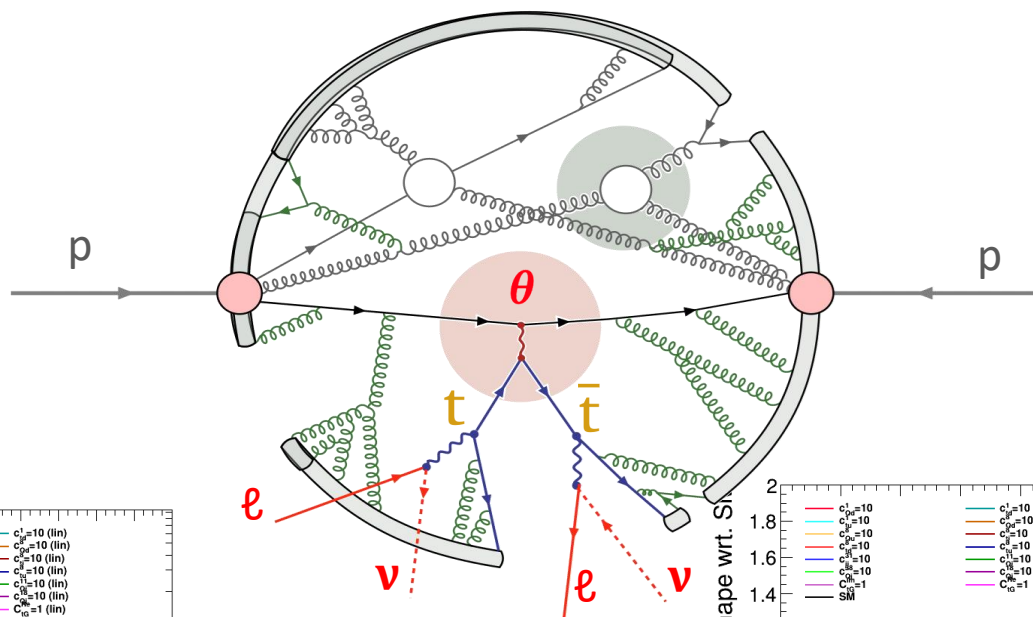


Proton structure with neural simulation based inference (+ implementation in global fits) [[2604.13157](#)]

R. Barrue (MBI, Vienna), L. Benato (MBI, Vienna), A. K. Guven (MBI, Vienna), E. Hammou (Nikhef, Amsterdam), J. t. Hoeve (JCMB, Edinburgh), C. Krause (MBI, Vienna), A. Li (MBI, Vienna), L. Mantani (IFIC, Valencia), J. Rojo (Nikhef, VU, Amsterdam), S.S. Cruz (ICTEA, Oviedo), R. Schoefbeck (MBI, Vienna), M Ubiali (DAMTP, Cambridge), and D. Wang (MBI, Vienna)



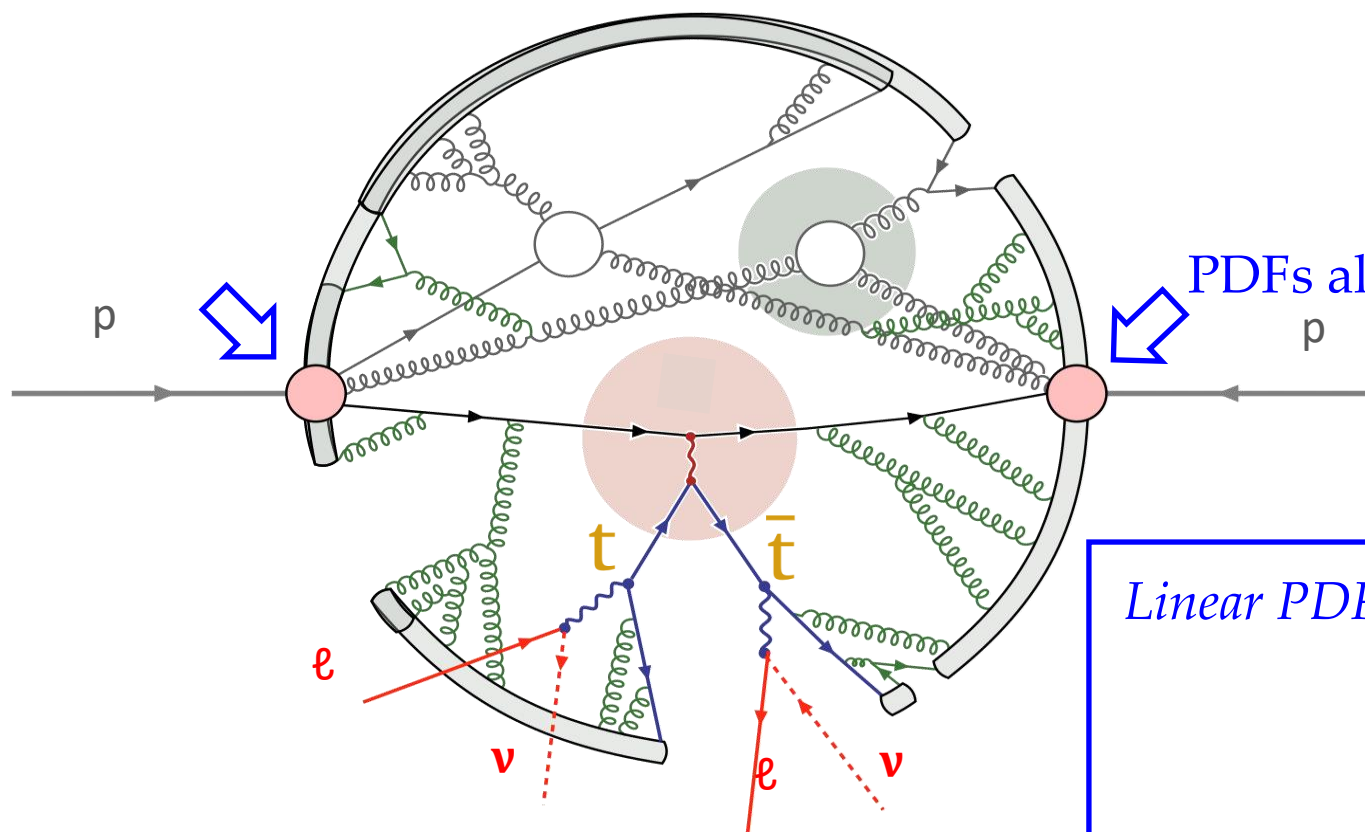
- $t\bar{t}$: Clean probe in the messy LHC environment



≈ 72 features

Used to constrain $\approx 15-25$
SMEFT Wilson coefficients θ

linear feature correlation in $tt(2\ell)$



PDFs also enter quadratically

Linear PDF model [Ubbiali et al. [2507.16913](#)]

$$f_g(x) + \mathbf{c} \cdot \varphi(x)$$

$$d\sigma \propto \mathcal{O}(\mathbf{c}^2)$$

SMEFT: quadratic predictions

$$d\sigma \propto |\mathcal{M}_{\text{SM}} + \boldsymbol{\theta} \mathcal{M}_{\text{EFT}}|^2 \\ \propto \mathcal{O}(\boldsymbol{\theta}^2)$$

Quadratic structure exploited in ML

Madminer [[1805.00013](#)] [[1805.00020](#)] [[1805.12244](#)]

BIT [[2107.10859](#)], [[2205.12976](#)]

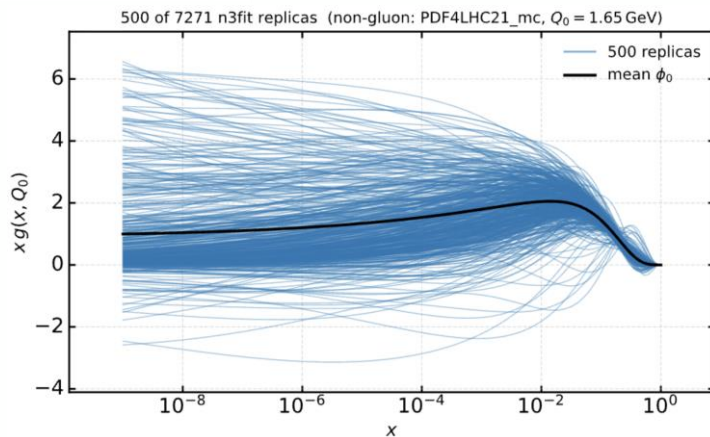
ML4EFT [[2211.02058](#)] "Refinable" [[2406.19076](#)]

$$\left\{ \hat{f}_g^{(m)} \right\}_{m=1}^M$$

1. Define "admissible functions"

- Sample gluon PDF with NNPDF4.0 parametrization

$$x \tilde{f}_g^{(m)}(x, Q_0) = A_m x^{1-\alpha_m} (1-x)^{\beta_m} \text{NN}_m(x; \theta)$$



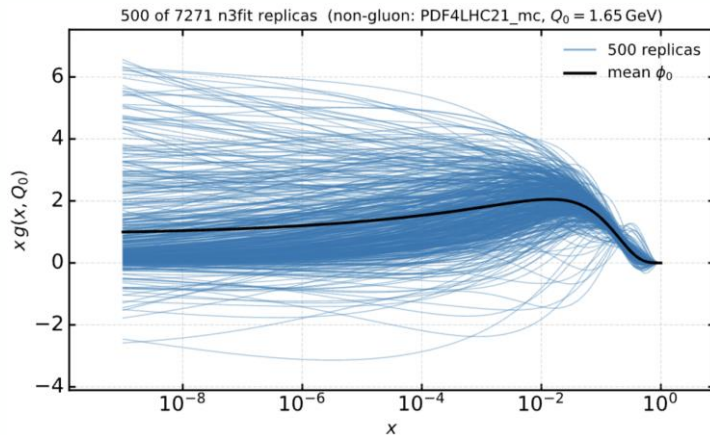
- Keep quark PDF fixed at NNPDF4.0 or PDF4LHC21

$$\left\{ \widehat{f}_g^{(m)} \right\}_{m=1}^M \xrightarrow{\text{POD/SVD}} \left\{ \varphi_g^{(a)} \right\}_{a=1}^N$$

1. Define "admissible functions"

- Sample gluon PDF with NNP4.0 parametrization

$$x \widehat{f}_g^{(m)}(x, Q_0) = A_m x^{1-\alpha_m} (1-x)^{\beta_m} \text{NN}_m(x; \theta)$$

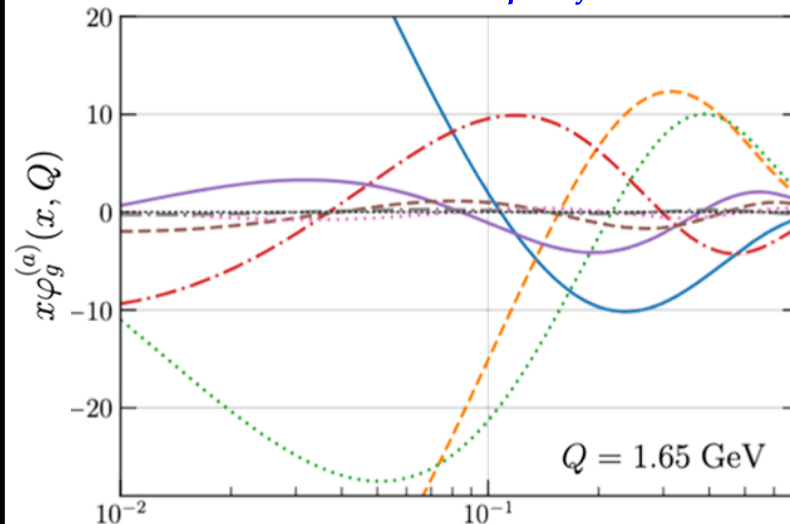


- Keep quark PDF fixed at NNP4.0 or PDF4LHC21

2. Identify leading modes

*"Proper Orthogonal
Decomposition - POD"*

$$\mathcal{L}_N = \frac{1}{M} \sum_{m=1}^M \left\| \widehat{f}_g^{(m)} - \underbrace{\sum_{a=1}^N \left(\varphi_g^{(a)T} \widehat{f}_g^{(m)} \right) \varphi_g^{(a)}}_{\text{projection}} \right\|^2$$

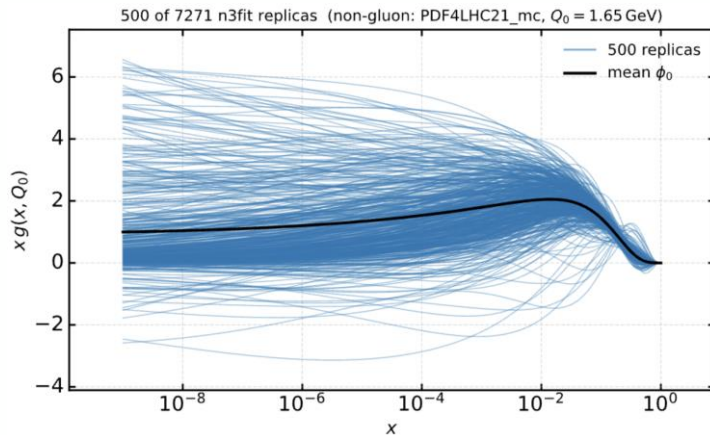


$$\left\{ \widehat{f}_g^{(m)} \right\}_{m=1}^M \xrightarrow{\text{POD/SVD}} \left\{ \varphi_g^{(a)} \right\}_{a=1}^N \xrightarrow{\text{DGLAP}} f_g(x, Q, \mathbf{c})$$

1. Define "admissible functions"

- Sample gluon PDF with NNPDF4.0 parametrization

$$x \widehat{f}_g^{(m)}(x, Q_0) = A_m x^{1-\alpha_m} (1-x)^{\beta_m} \text{NN}_m(x; \boldsymbol{\theta})$$



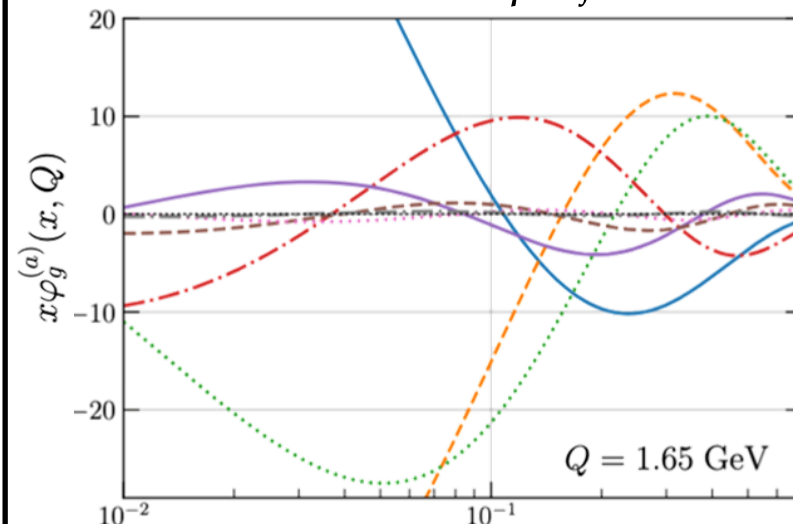
- Keep quark PDF fixed at NNPDF4.0 or PDF4LHC21

2. Identify leading modes

*"Proper Orthogonal
Decomposition - POD"*

$$\mathcal{L}_N = \frac{1}{M} \sum_{m=1}^M \left\| \widehat{\mathbf{f}}_g^{(m)} - \sum_{a=1}^N \underbrace{(\varphi_g^{(a)T} \widehat{\mathbf{f}}_g^{(m)})}_{\text{projection}} \varphi_g^{(a)} \right\|^2$$

projection

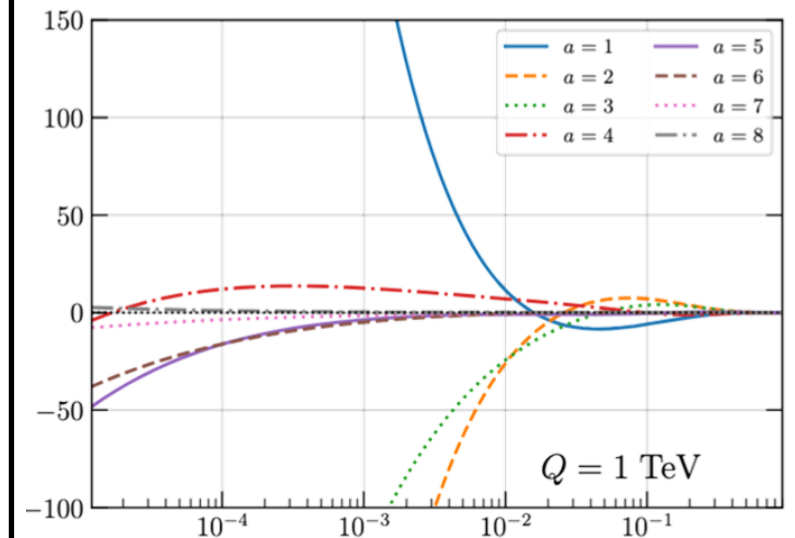


3. DGLAP evolution

NNLO with kernels via EKO

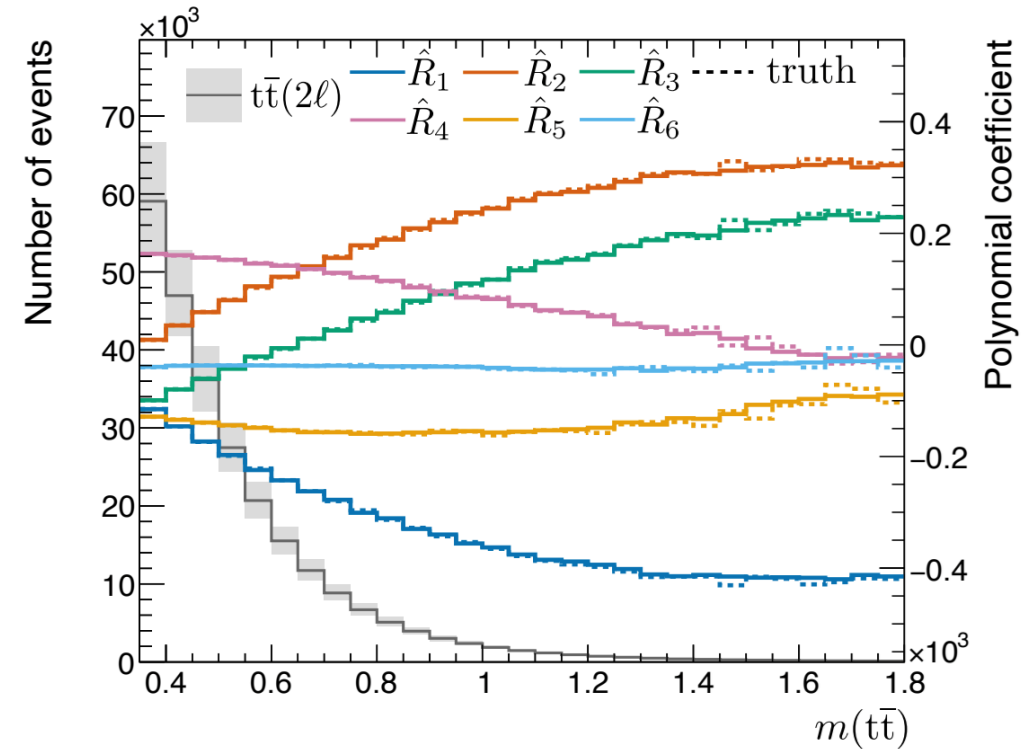
$$f_g(x, Q, \mathbf{c})$$

$$= \Gamma_{gg}(x, Q, Q_0) \otimes \left(\varphi_g^{(0)}(x, Q_0) + \sum_{a=1}^N c_a \varphi_g^{(a)}(x, Q_0) \right) + \Gamma_{g\Sigma}(x, Q, Q_0) \otimes f_\Sigma(x, Q_0)$$



- Top quark pairs $t\bar{t}(2\ell)$ POWHEG+Pythia
 - Integrated luminosity 137/fb
 - Simulated with DELPHES at the CMS detector level
 - Purity >95% → neglect backgrounds
 - Detector-level top quark reconstruction
- PDF affects predictions via parton-level reweighting

$$\omega(\mathbf{z}_i, \mathbf{c}) = \frac{f_{m_i}(x_{1i}; Q_i, \mathbf{c}) f_{n_i}(x_{2i}; Q_i, \mathbf{c})}{f_{m_i}^{(\text{gen})}(x_{1i}; Q_i) f_{n_i}^{(\text{gen})}(x_{2i}; Q_i)} \sim \mathcal{O}(\mathbf{c}^2)$$



- Induces shape- and normalization effects at the detector level

→ *How to constrain coefficients optimally with unbinned inference?*

- Profile log-likelihood test statistic

$$q_{\mathbf{c}}(\mathcal{D}) = -2 \log \frac{\max_{\boldsymbol{\nu}} L(\mathcal{D}|\mathbf{c}, \boldsymbol{\nu})}{\max_{\mathbf{c}, \boldsymbol{\nu}} L(\mathcal{D}|\mathbf{c}, \boldsymbol{\nu})} \rightarrow \text{Profile systematic nuisances } \boldsymbol{\nu}$$

- Extended likelihood

$$L(\mathcal{D}|\mathbf{c}, \boldsymbol{\nu}) = \text{Pois}(N_{\text{obs}} | \mathcal{L}(\boldsymbol{\nu}) \sigma(\mathbf{c}, \boldsymbol{\nu})) \prod_{j=1}^{N_{\text{obs}}} p(\mathbf{x}_j|\mathbf{c}, \boldsymbol{\nu})$$

- Write LL with differential cross section ratio

$$\log \frac{L(\mathcal{D}|\mathbf{c}, \boldsymbol{\nu})}{L(\mathcal{D}|\mathbf{0}, \mathbf{0})} = - \left[\mathcal{L}(\boldsymbol{\nu}) \sigma(\mathbf{c}, \boldsymbol{\nu}) - \mathcal{L}(\mathbf{0}) \sigma(\mathbf{0}, \mathbf{0}) \right] + \sum_{j=1}^{N_{\text{obs}}} \log \left[\frac{\mathcal{L}(\boldsymbol{\nu})}{\mathcal{L}(\mathbf{0})} \frac{d\sigma(\mathbf{x}_j|\mathbf{c}, \boldsymbol{\nu})}{d\sigma(\mathbf{x}_j|\mathbf{0}, \mathbf{0})} \right]$$

- Need to evaluate the differential cross section ratio (DCR) for unseen data ($\mathcal{D}$) and as a function of $\mathbf{c}$ (=POIs) and $\boldsymbol{\nu}$ (=nuisances) $\rightarrow$ ML

- “Likelihood ratio trick”

$$L_{\text{MSE}, \mathbf{c}}[\hat{R}] = \left\langle \hat{R}(\mathbf{x}_i, \mathbf{c}) - \omega(\mathbf{z}_i, \mathbf{c}) \right\rangle^2$$

Training data $\sim \mathcal{O}(\mathbf{c}^2)$

- Parametric ansatz

$$\hat{R}(\mathbf{x}, \mathbf{c}) = 1 + c_a \hat{R}_a(\mathbf{x}) + c_a c_b \hat{R}_{ab}(\mathbf{x})$$

Implement coefficients $R_a(\mathbf{x})$, $R_{ab}(\mathbf{x})$ as trees:

$$\hat{R}_{a(b)}(\mathbf{x}) = \sum_{s=1}^B \eta^{(s)} \sum_{J \in \mathcal{J}^{(s)}} \mathbb{1}_J(\mathbf{x}) \hat{R}_{J,a(b)}^s$$

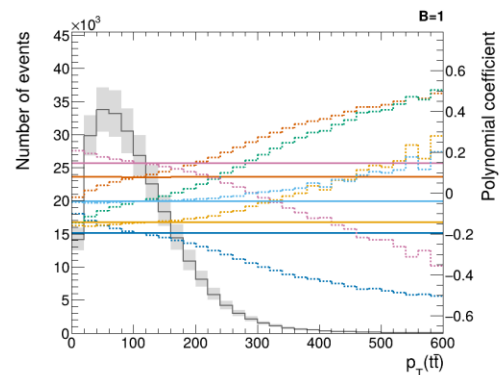
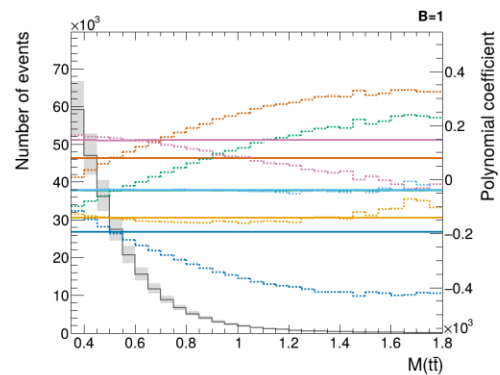
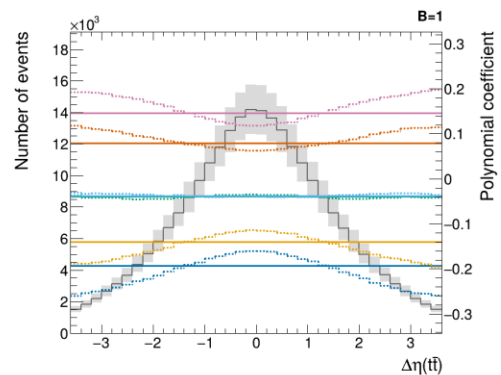
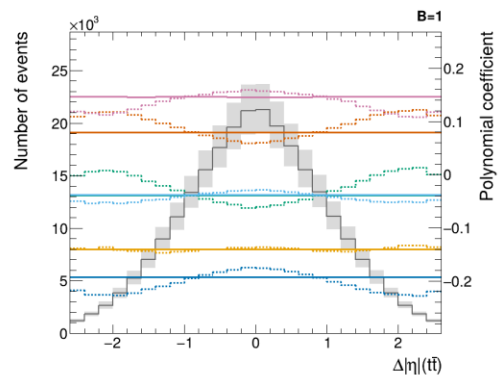
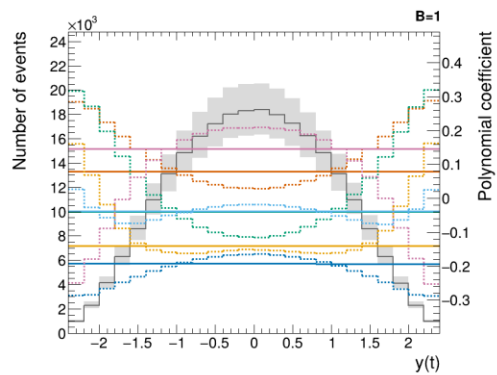
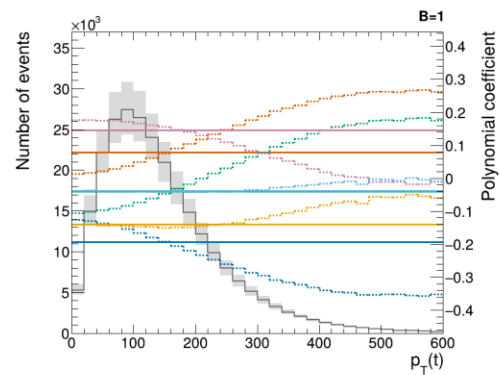
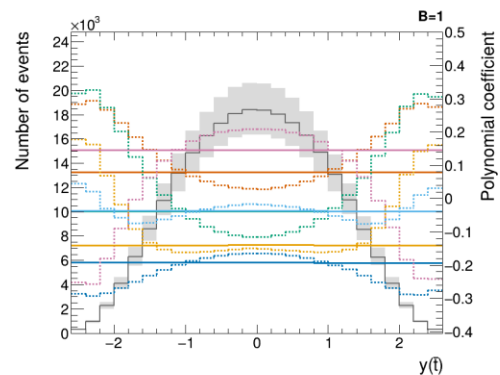
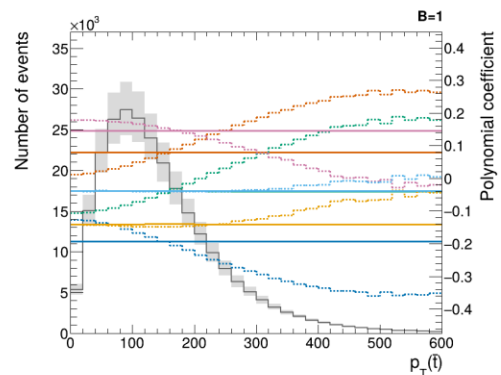
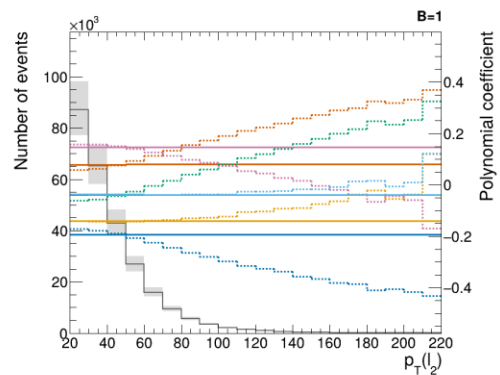
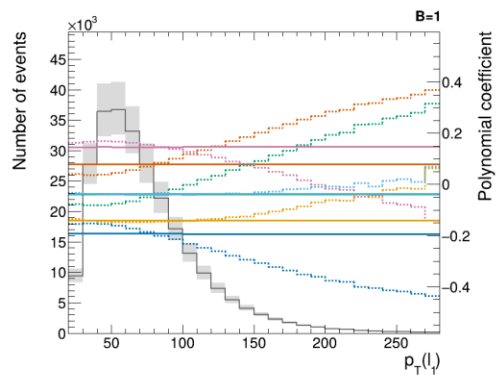
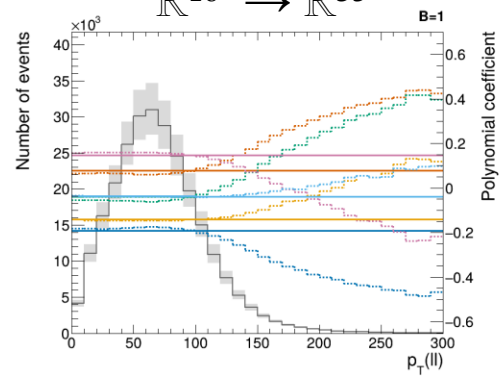
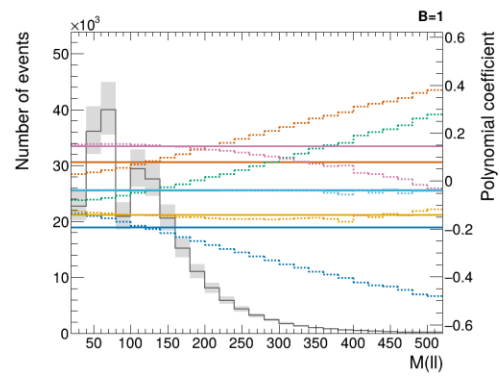
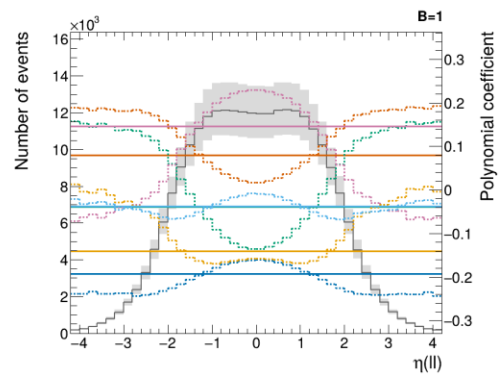
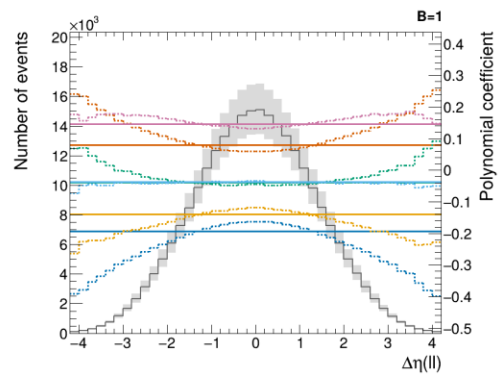
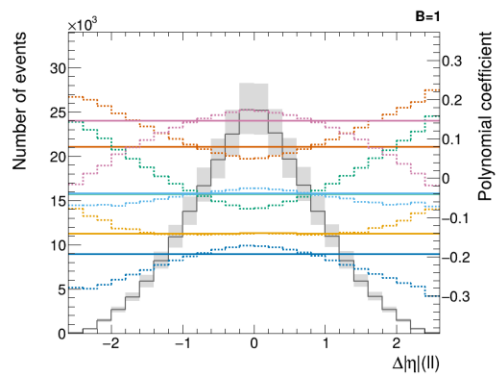
Boosted Information tree

[[2604.13157](#), [2107.10859](#), [2205.12976](#)]

- Sum over base points

$$L_{\text{MSE}}[\hat{R}] = \sum_{\mathbf{c} \in \mathcal{V}} \left\langle \hat{R}(\mathbf{x}_i, \mathbf{c}) - \omega(\mathbf{z}_i, \mathbf{c}) \right\rangle^2 \longrightarrow \hat{R}(\mathbf{x}, \mathbf{c}) \simeq \frac{d\sigma(\mathbf{x}|\mathbf{c}, \mathbf{0})}{d\sigma(\mathbf{x}|\mathbf{0}, \mathbf{0})}$$

$\mathbb{R}^{16} \rightarrow \mathbb{R}^{35}$



$$\frac{d\sigma(\mathbf{x}|\mathbf{c}, \boldsymbol{\nu})}{d\sigma(\mathbf{x}|\mathbf{0}, \mathbf{0})} = \underbrace{\frac{d\sigma(\mathbf{x}|\mathbf{c}, \boldsymbol{\nu})}{d\sigma(\mathbf{x}|\mathbf{c}, \mathbf{0})}}_{\text{Nuis.}} \underbrace{\frac{d\sigma(\mathbf{x}|\mathbf{c}, \mathbf{0})}{d\sigma(\mathbf{x}|\mathbf{0}, \mathbf{0})}}_{\text{POIs}}$$

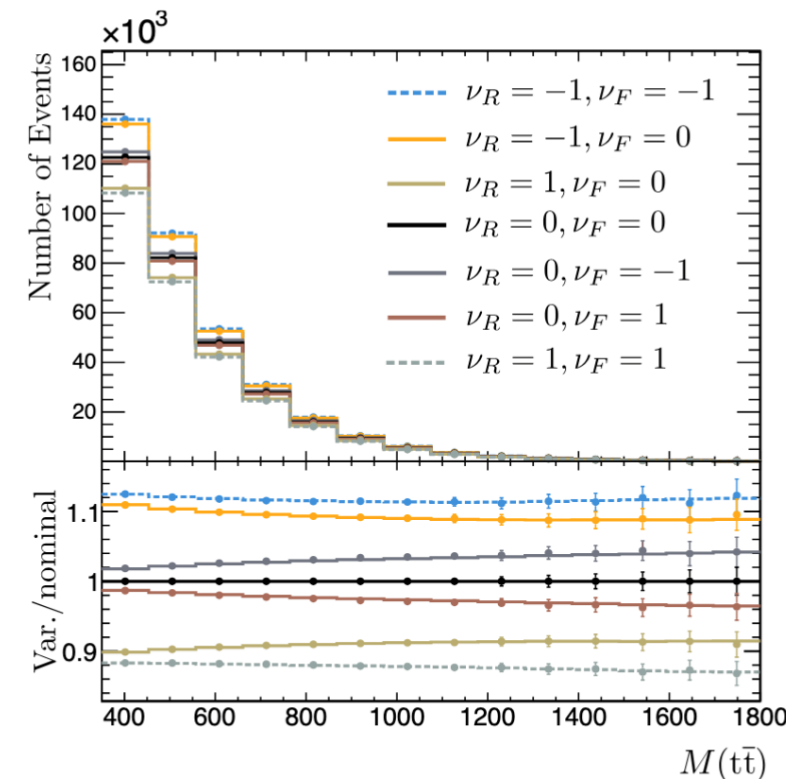
$$\approx S(\mathbf{x}, \boldsymbol{\nu}) R(\mathbf{x}, \mathbf{c})$$

Assume factorization in
nuisances and POIs
[MLST 6 015007 (2025)]

Same as in binned analyses
→ Learn every effect separately

– Systematic variations

- Recipes from CMS open data [<https://cms-opendata-guide.web.cern.ch/>]
 - Ren. & fac. scale (2 nuis.)
 - b-tagging (heavy- and light, 2 nuis.)
 - Jet energy scale (6 nuis.)
 - Jet energy resolution (6 nuis.)
 - Lepton scale factors (2 nuis.)
 - Strong coupling α_s (1 nuis.)
 - Luminosity (1 nuis.)

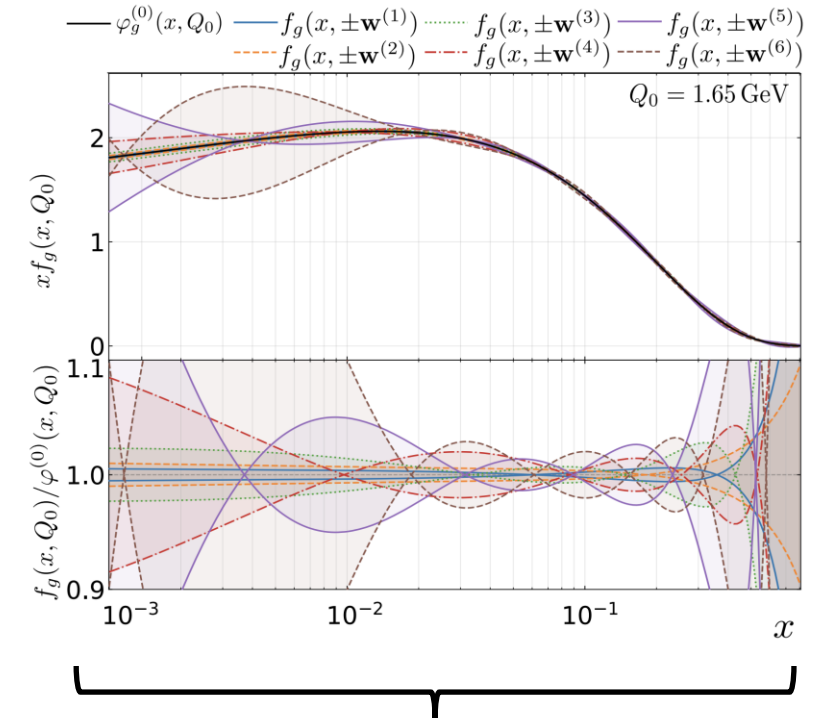
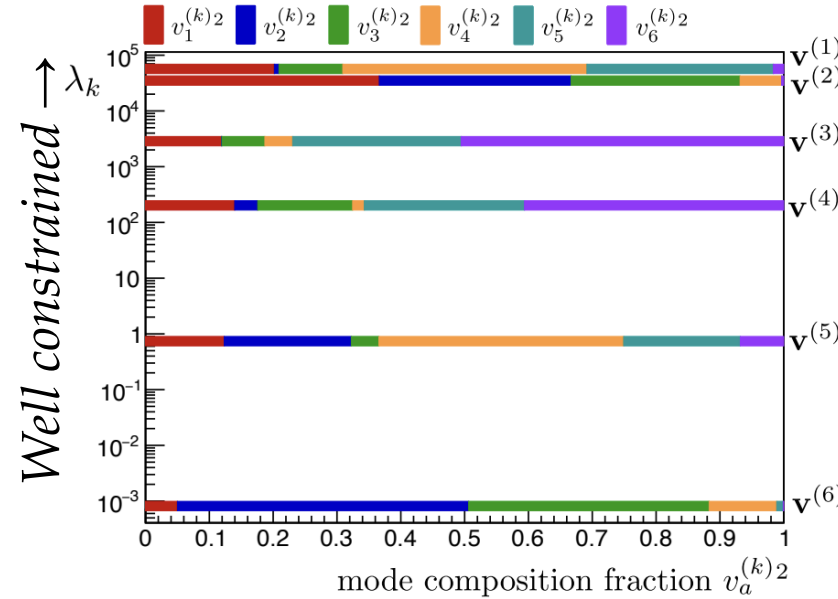
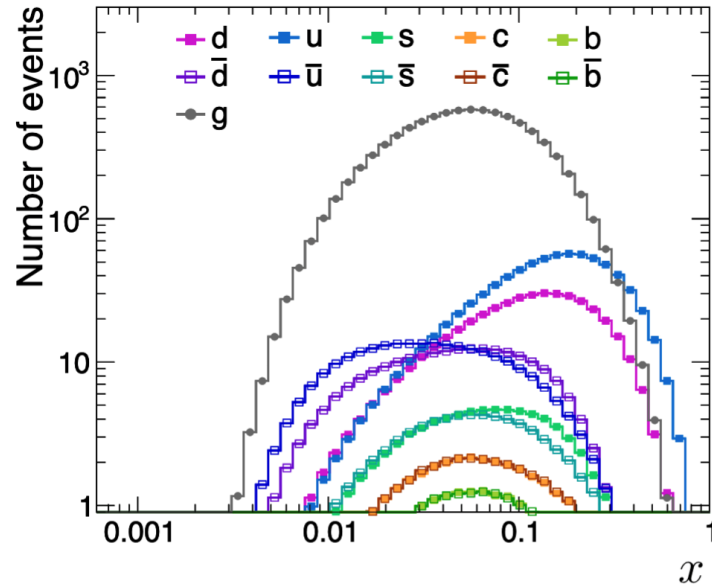


$$\frac{\mathcal{L}(\boldsymbol{\nu})}{\mathcal{L}(\mathbf{0})} \frac{d\sigma(\mathbf{x}|\mathbf{c}, \boldsymbol{\nu})}{d\sigma(\mathbf{x}|\mathbf{0}, \mathbf{0})} \simeq \alpha_{\text{lumi}}^{\nu_{\text{lumi}}} \hat{R}(\mathbf{x}, \mathbf{c}) \hat{S}_{\text{mhou}}(\mathbf{x}, \nu_R, \nu_F) \hat{S}_{\alpha_s}(\mathbf{x}, \nu_{\alpha_s}) \hat{S}_{\text{JES}}(\mathbf{x}, \nu_{\text{JES}}) \hat{S}_{\text{JER}}(\mathbf{x}, \nu_{\text{JER}})$$

$$\times \hat{S}_{\text{HF}}(\mathbf{x}, \nu_{\text{HF}}) \hat{S}_{\text{LF}}(\mathbf{x}, \nu_{\text{LF}}) \hat{S}_{\ell}(\mathbf{x}, \nu_{\mu}, \nu_e)$$

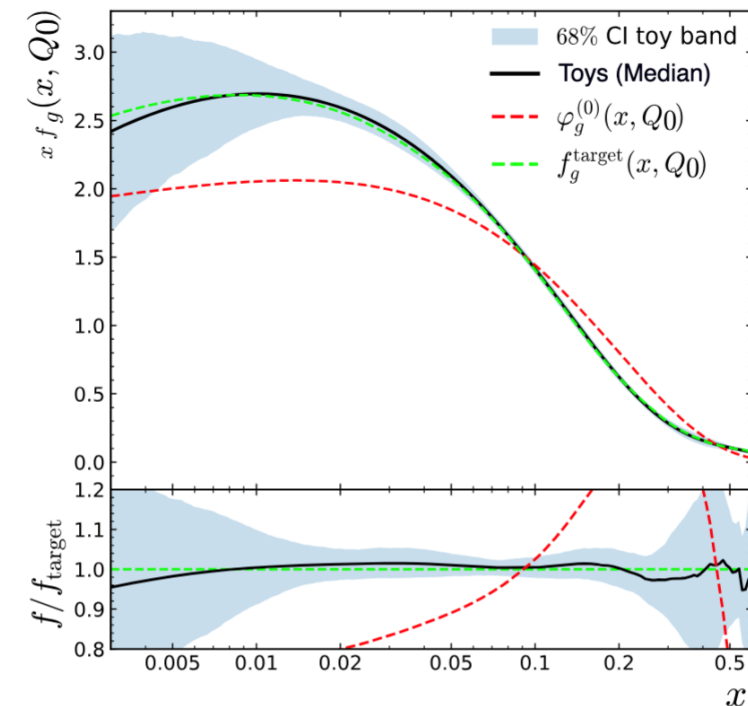
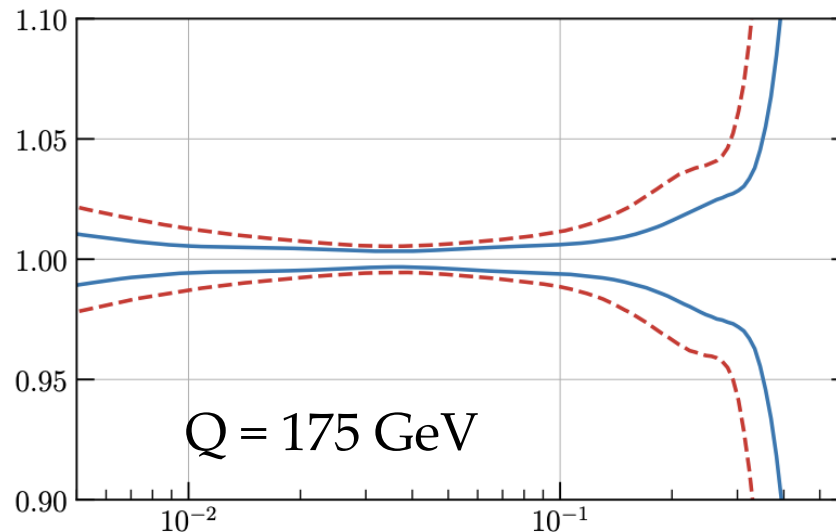
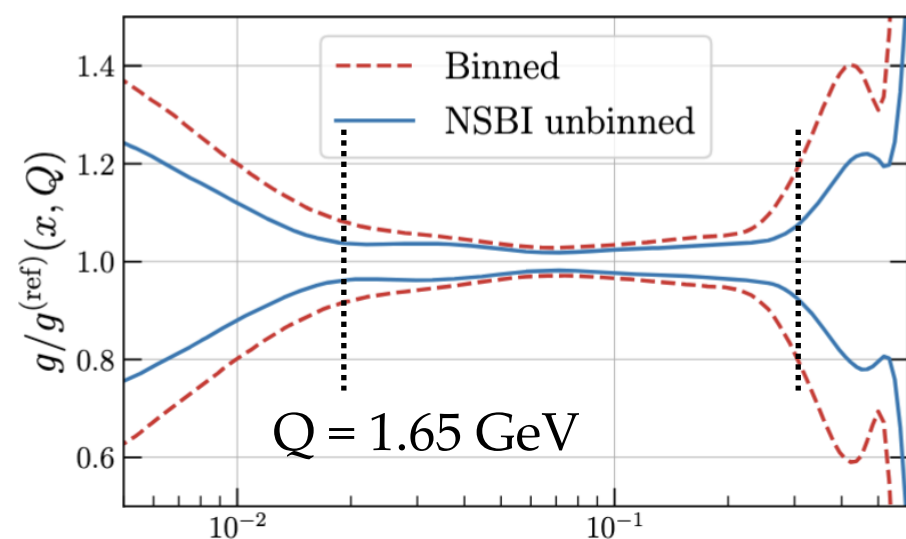
6 POIs, 20 Nuisances

Fit?



- Partonic coverage of $t\bar{t}(2\ell)$: $0.02 < x < 0.3$
 - POD basis has not seen data
 - Some directions in $\mathbf{c}$ -space are not constrained
- With unbinned $L(\mathcal{D}|\mathbf{c}, \mathbf{v})$ defined $\rightarrow$ Expected (Fisher) sensitivity
 - Rank POI directions in expected sensitivity $L(\mathcal{D}_{\text{exp}}|\lambda^{-1/2}) = \pm 1$
 - Rotation in $\mathbf{c}$ -space; Beyond $N=8$, “flat directions” must be removed

*Gluon PDF “wiggles”
in the fit*

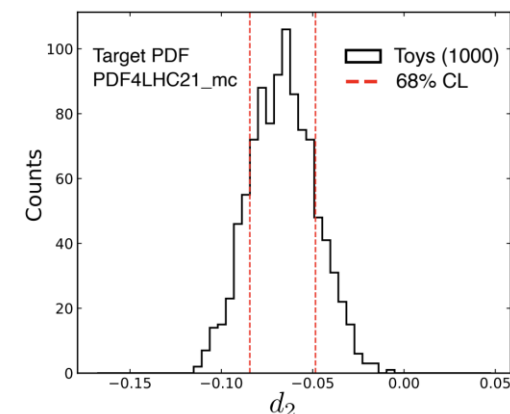
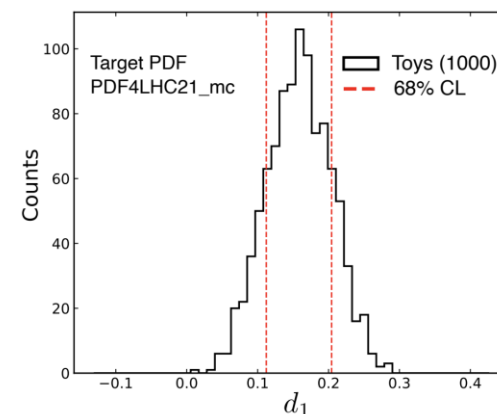


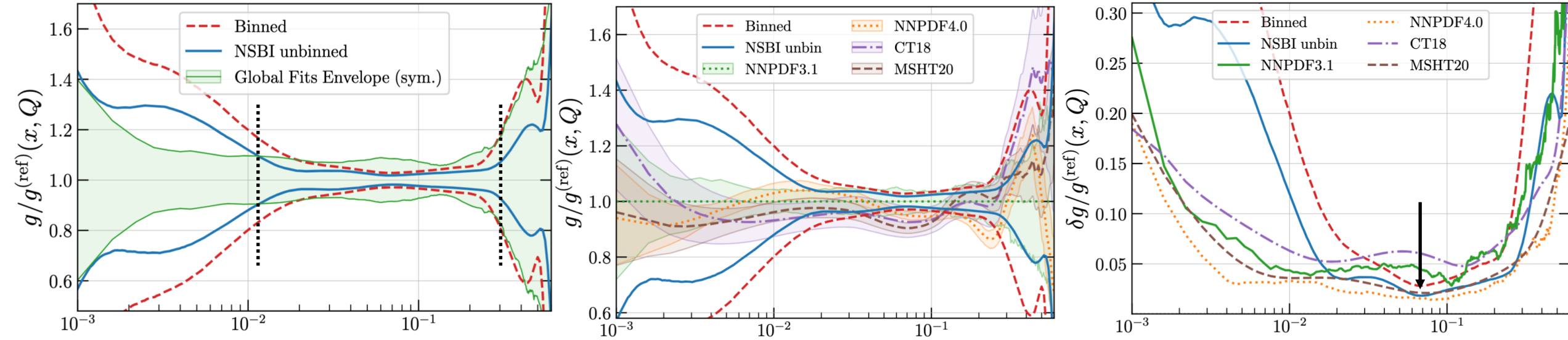
- Fit to toy data sampled from PDF4LHC21
 - Very good agreement! Gaussian toy distributions
- Binned benchmark comparison [\[EPJC 80 \(2020\) 658\]](#)
Use 16 bins in $m(t\bar{t})$ and $y(t\bar{t})$ from CMS meas.

$$m(t\bar{t}) \in [300, 400, 500, 650, 1500] \text{ GeV}$$

$$|y(t\bar{t})| \in [0, 0.4, 0.8, 1.2, 2.5]$$

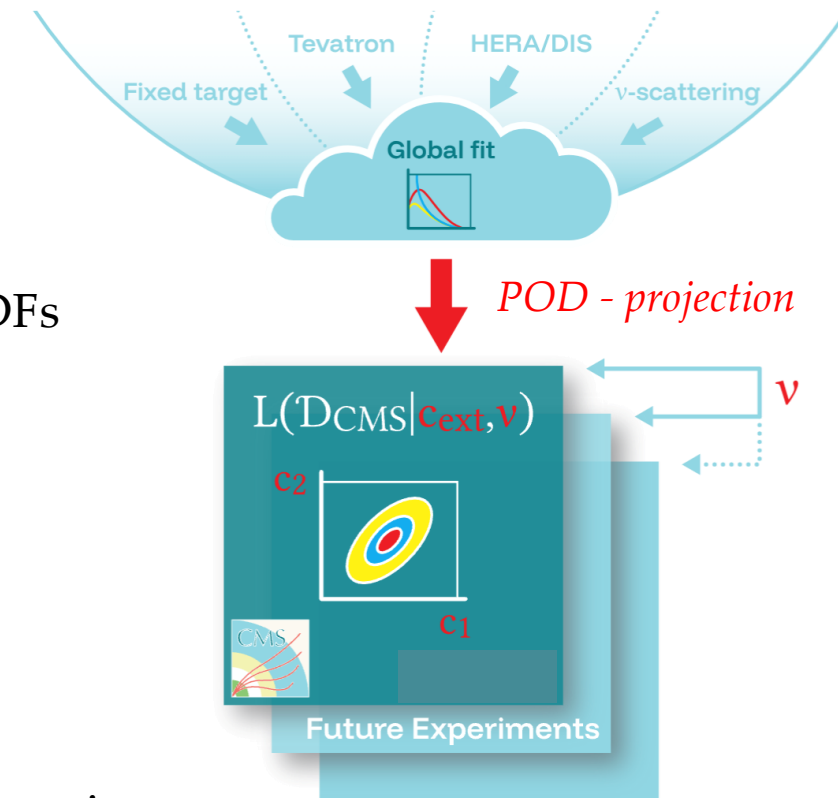
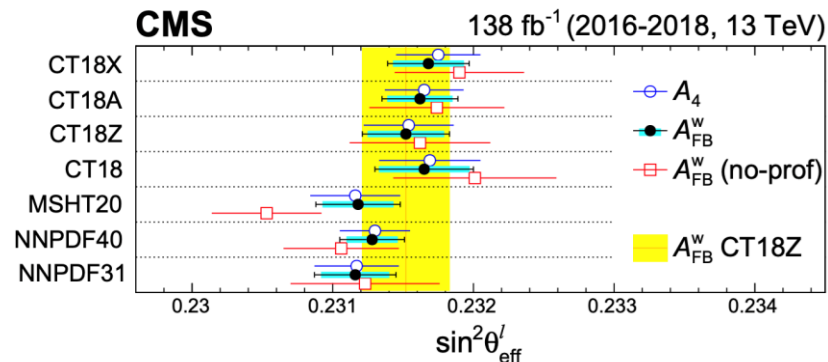
- Significant improvement at high- and low x





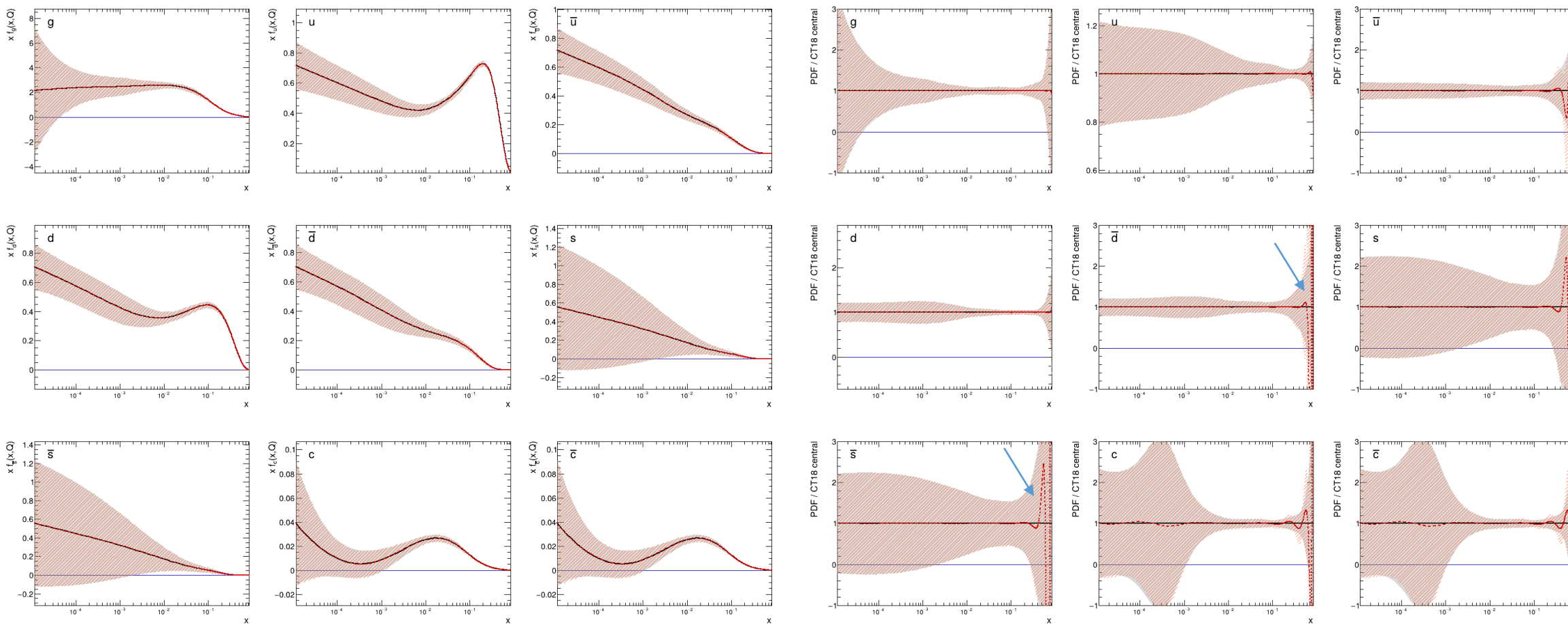
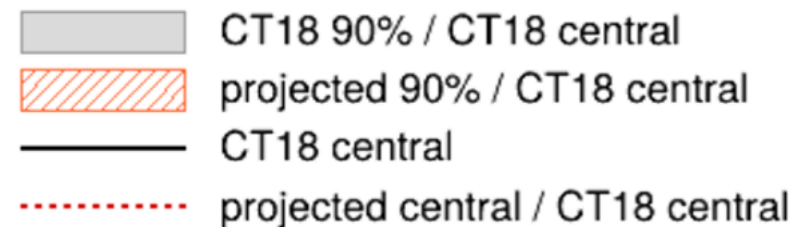
- Compare to NNPDF3.1, NNPDF4.0, CT18, MSHT20
 - Competing with Inclusive Jets, DIS, W/Z, unfolded top quark data, etc.
 - Region $0.02 < x < 0.3$: better than at least one; unbinned better than binned
 - For $x \sim 0.08$: Comparable to NNPDF4.0
- *Unbinned detector-level data may improve PDF inference in a global fit*

- Unbinned measurements provide full $L(D|\mathbf{c}, \mathbf{v})$
 - What to do with $\mathbf{c}$ when every global fit has its own parametrization?
- POD lives in Hilbert space $\rightarrow$ project PDFs
 - An external set maps to $\mathbf{c}, \Sigma$ in a $O(ms)$ projection
 - If a measurement provides a POD result, then $L(D|\mathbf{c}_{ext}[\text{CTEQ/NNPDF/MSHT...}], \mathbf{v})$ is evaluable
 - Bases can be different across measurements – no community challenge
 - This works both ways; experiments can project on *any* external PDFs
 - An external PDF constraint becomes a prior – no PDF bookkeeping



1. Recover detector-level information beyond the Gaussian approximation
2. Experiments don't see the gradient of the global fit
 - "Flavor dependent JEC? b/g corrections? Bias from JER in the crack region?" Or cone-size?
3. Assembly. $N_{bin} \sim O(100)$, covariances scale with $N_{bin}^2 \times N_{nuis}$

- Projecting CT18NNLO onto all-flavor POD $\rightarrow c_{\text{POD}}, \Sigma_{\text{POD}}$
- Excellent closure.



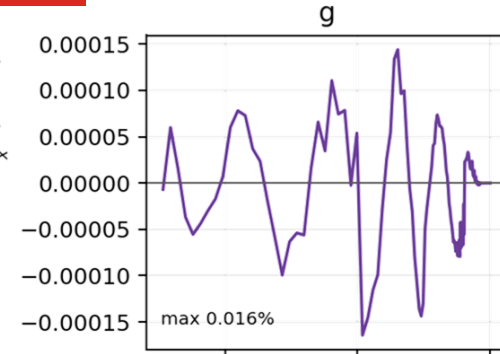
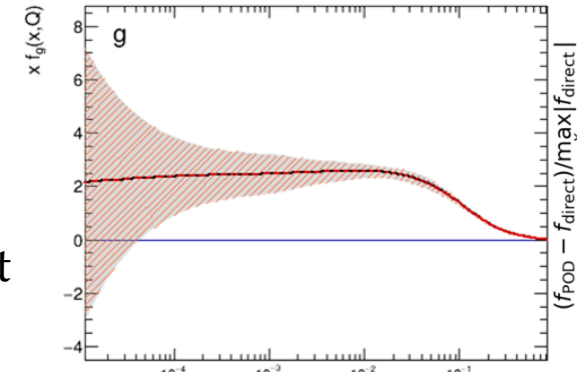
- Example: CMS inclusive jets [[JHEP 12 \(2022\)035](#)] in [xFitter](#)
- Compare:
 1. Analytic model $\rightarrow$ fastNLO $\rightarrow \chi^2 \rightarrow$ fit
 2. Analytic parametrization $\rightarrow$ **POD** $\rightarrow$ fastNLO $\rightarrow \chi^2 \rightarrow$ fit

- Analytic model with 16 parameters

$$\begin{aligned}
 xg &= A_g x^{B_g} (1-x)^{C_g} (1 + D_g x + E_g x^2), \\
 xu_v &= A_{u_v} x^{B_{u_v}} (1-x)^{C_{u_v}} (1 + E_{u_v} x^2), \\
 xd_v &= A_{d_v} x^{B_{d_v}} (1-x)^{C_{d_v}}, \\
 x\bar{U} &= A_{\bar{U}} x^{B_{\bar{D}}} (1-x)^{C_{\bar{U}}} (1 + D_{\bar{U}} x), \\
 x\bar{D} &= A_{\bar{D}} x^{B_{\bar{D}}} (1-x)^{C_{\bar{D}}} (1 + E_{\bar{D}} x^2).
 \end{aligned}$$

- **Fit in POD basis** projection has negligible bias

- Next step: A real analysis



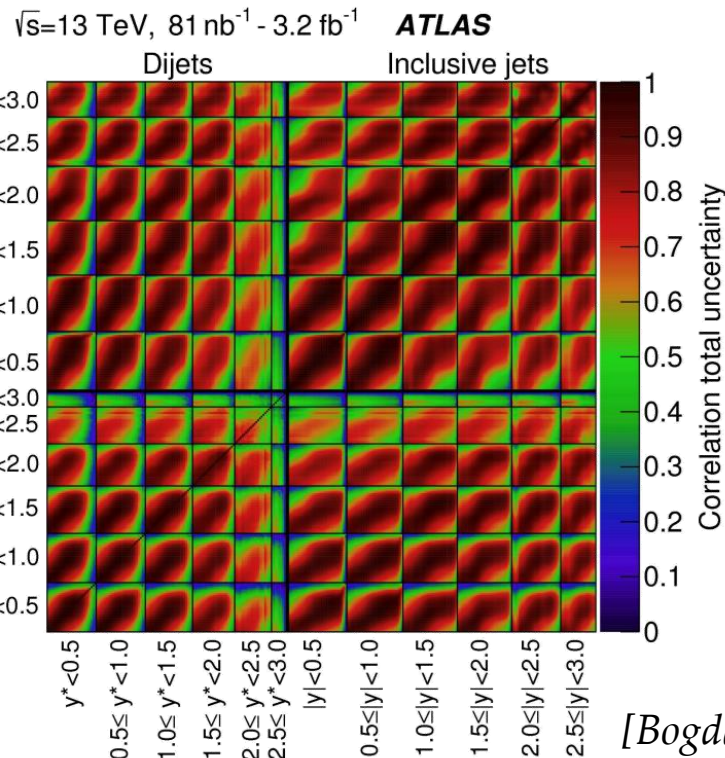
Parameter	Reference	POD-mediated	Shift [(\sigma)]
ADbar	0.267003	0.267843	+0.0536
BDbar	-0.120313	-0.119918	+0.0477
Bdv	0.762795	0.765778	+0.0419
Bg	0.153937	0.153303	-0.0149
Buv	0.714465	0.714054	-0.0159
CDbar	10.668946	10.663250	-0.0060
CUbar	7.471823	7.494759	+0.0348
Cdv	3.743693	3.757982	+0.0404

Full list in backup

- Unbinned top quark data has extra PDF constraining power [[2604.13157](#)]
 - Strong case (and vehicle) for publishing finer non-profiled experimental likelihoods
 - Demonstrated simulation-based PDF inference is feasible
 - Author team Members from NNPDF + Experimental top (CMS) + Machine learning/SBI
- Workhorses
 - Linear PDF model [[2507.16913](#)] & workflows
 - Boosted Information tree [[2604.13157](#), [2107.10859](#), [2205.12976](#)]
 - “Refinable” modeling of systematics [[2406.19076](#)]
- POD reduces the need for community wide organization of PDF likelihoods
 - Would like to begin set it up

Backup

- How can we incorporate unbinned detector-level likelihood $L(D|\mathbf{c}, \mathbf{v})$ in global fits?
- Staging: First, **take unfolded, binned χ^2 -based measurement** → **Recast with POD** → global fits
Followed by, possibly → Binned full likelihood with POD
 - Example: CMS inclusive jets (13 TeV) [[JHEP 12 \(2022\)035](#)]
 - NNLO theory predictions are binned. No point for NSBI. Jet-based measurements are indispensable for f_g



- Why could the full likelihood be useful on nuisance side?

1. Recover detector-level information in $L(D|\mathbf{c}, \mathbf{v})$

“When jets and top quarks don’t agree [in a global PDF fit], what do you do?” (June 19th)
→ What’s the postfit result on $\mathbf{v}$?

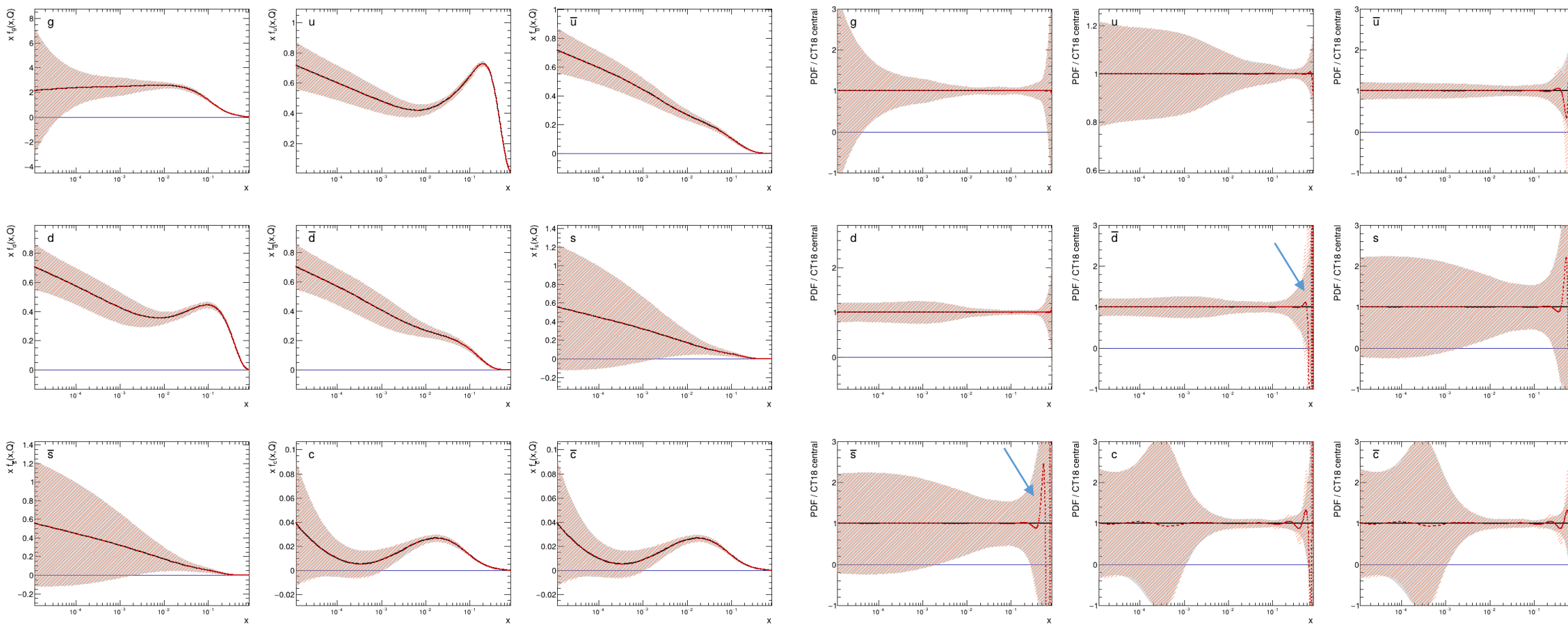
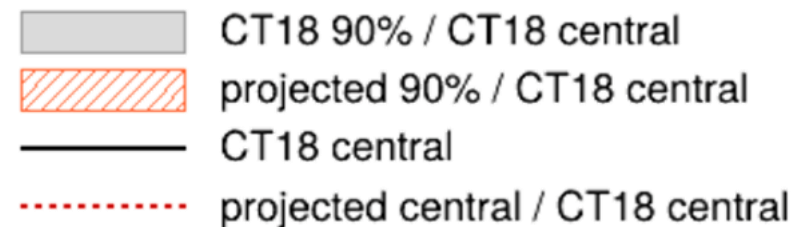
2. Experiments don’t see the gradient of the global fit

- “Flavor dependent JEC? b/g corrections?
Bias from JER in the crack region?” Or cone-size?
- We compress the full information when results are made public

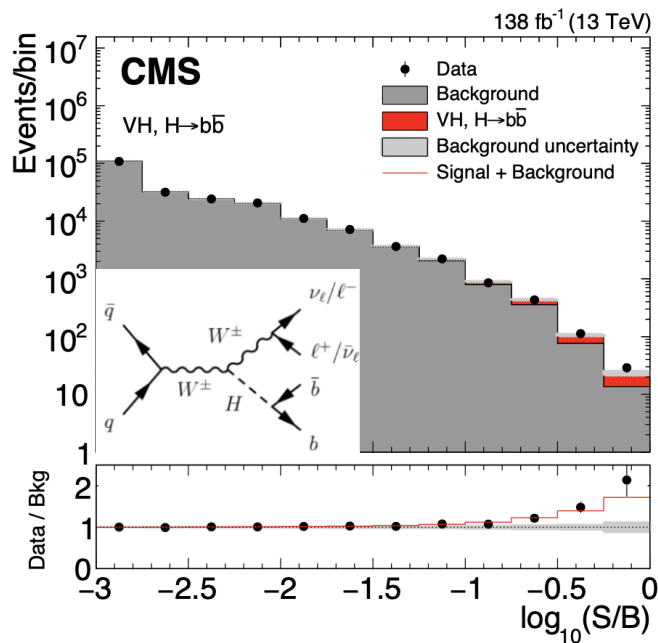
3. Assembly. $N_{\text{bin}} \sim O(100)$, covariances scale with $N_{\text{bin}}^2 \times N_{\text{nuis}}$

- POD lives in Hilbert space → PDF *projections* can be exploited

- Projecting CT18NNLO onto all-flavor POD $\rightarrow c_{\text{POD}}, \Sigma_{\text{POD}}$
- Excellent closure. Apply within experiment *as a prior*.

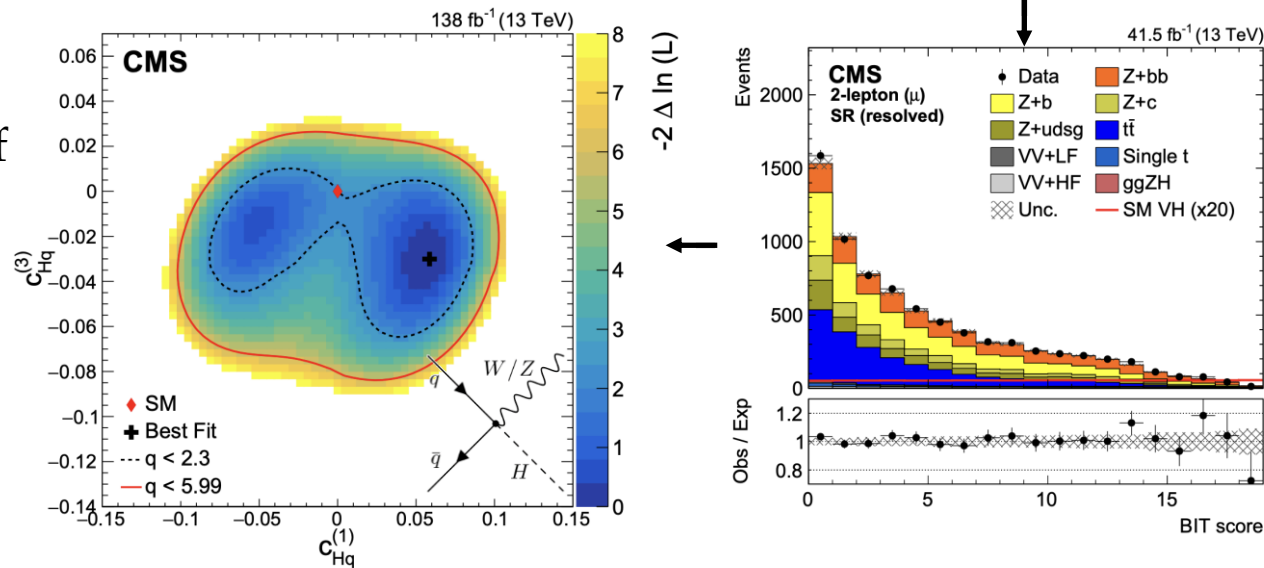
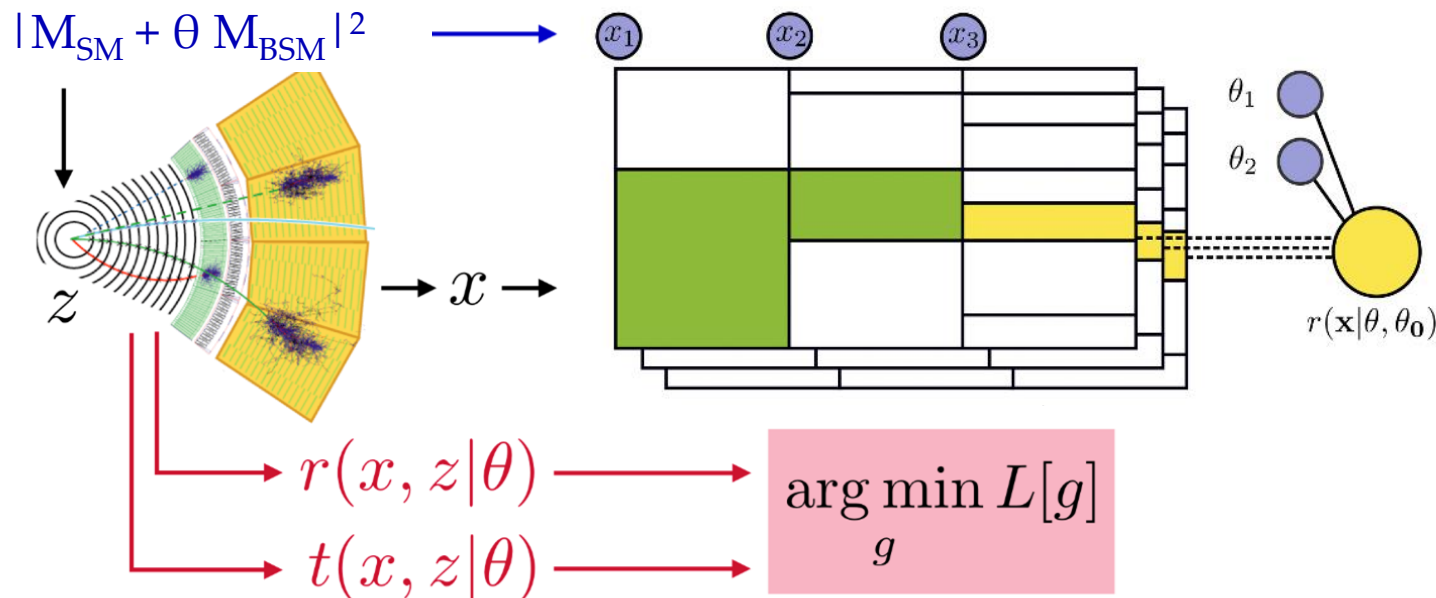


- Unbinned top quark data has extra PDF constraining power [[2604.13157](#)]
 - Strong case (and vehicle) for publishing finer non-profiled experimental likelihoods
 - Demonstrated simulation-based PDF inference is feasible
 - Author team Members from NNPDF + Experimental top (CMS) + Machine learning/SBI
- Workhorses
 - Linear PDF model [[2507.16913](#)] & workflows
 - Boosted Information tree [[2604.13157](#), [2107.10859](#), [2205.12976](#)]
 - “Refinable” modeling of systematics [[2406.19076](#)]
- POD reduces the need for community wide organization of PDF likelihoods
 - Would like to begin set it up



H unitarizes SM at high energy

- cancellations over many orders of magnitude $\rightarrow$ BSM probes
- Six CP-even/odd effects modify angular distributions & induce E-growth
- Developed ML algorithm fully aware of $|M_{SM} + \theta M_{BSM}|^2$
[[MLST 2025](#), [2205.12976](#), [CPC 2022](#)]



“ML optimal observable”

0 ℓ /1 ℓ /2 ℓ

high dim!

- 48 features
- 6 BSM effects

up to x8 sensitivity
[[2411.16907](#)]

1. ``Refinable modeling for unbinned SMEFT analyses`` [MLST 6 015007 (2025)]

- Stage-wise unbinned model building for SMEFT
- Statistical footing of “systematics recipes”
- Dissect the effect of (groups of) systematics & POIs
- Machine-learn the components (with trees)

2. Contribution to [FAIR-HUC Challenge] preprint [arXiv:2505.05544]

- Cross section measurement of $H \rightarrow \tau\tau$ in DELPHES simulation

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

- Our approach
 - Classic unbinned SBI
 - Build all known dependencies into the profiled test statistic
 - Arbitrarily accurate models for systematics; fully refinable

Refinable modeling for unbinned SMEFT analyses

Robert Schöfbeck

*Institute for High Energy Physics, Austrian Academy of Sciences,
Georg-Coch-Platz 2, 1010 Vienna, Austria*

Abstract

We present techniques for estimating the effects of systematic uncertainties in unbinned data analyses at the LHC. Our primary focus is constraining the Wilson coefficients in the standard model effective field theory (SMEFT), but the methodology applies to broader parametric models of phenomena beyond the standard model (BSM). We elevate the well-established procedures for binned Poisson counting experiments to the unbinned case by utilizing machine-learned surrogates of the likelihood ratio. This approach can be applied to various theoretical, modeling, and experimental uncertainties. By establishing a common statistical framework for BSM and systematic effects, we lay the groundwork for future unbinned analyses at the LHC. Additionally, we introduce a novel tree-boosting algorithm capable of learning highly accurate parameterizations of systematic effects. This algorithm extends the existing toolkit with a versatile and robust alternative. We demonstrate our approach using the example of an SMEFT interpretation of highly energetic top quark pair production in proton-proton collisions.

[RS, [Mach. Learn. Sci. Technol. 6 015007 \(2025\)](#)]

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

Lisa Benato, Cristina Giordano, Claudius Krause, Ang Li,

Robert Schöfbeck, Dennis Schwarz, Maryam Shooshtari, and Daohan Wang

Institute for High Energy Physics, Austrian Academy of Sciences Dominikanerbastei 16, 1010 Vienna, Austria

(Dated: May 12, 2025)

We introduce a novel methodology for addressing systematic uncertainties in unbinned inclusive cross-section measurements and related collider-based inference problems. Our approach incorporates known analytic dependencies on parameters of interest, including signal strengths and nuisance parameters. When these dependencies are unknown, as is frequently the case for systematic uncertainties, dedicated neural network parametrizations provide an approximation that is trained on simulated data. The resulting machine-learned surrogate captures the complete parameter dependence of the likelihood ratio, providing a near-optimal test statistic. As a case study, we perform a first-principles inclusive cross-section measurement of $H \rightarrow \tau\tau$ in the single-lepton channel, utilizing simulated data from the FAIR Universe Higgs Uncertainty Challenge. Results in Asimov data, from large-scale toy studies, and using the Fisher information demonstrate significant improvements over traditional binned methods. Our computer code “Guaranteed Optimal Log-Likelihood-based Unbinned Method” (GOLLUM) for machine-learning and inference is publicly available.

[RS et al., [arXiv:2505.05544](#)]

1. $d\sigma(x|\mathbf{c}, \mathbf{v})$ encodes the field's knowledge about how phenomena depend on model parameters.
 - *Extract from the bottom of a river (of wikis)*
2. Wield its power to our advantage: Expand $d\sigma(x|\mathbf{c}, \mathbf{v})$ into sub-processes and encode analytic dependencies, *as if you had the power to evaluate*
 - Can't keep it, though.
3. Get rid of $d\sigma(x|\mathbf{c}, \mathbf{v})$ by forming ratios, which we *do* evaluate with ML surrogates



What is $d\sigma(x|c, v)$?

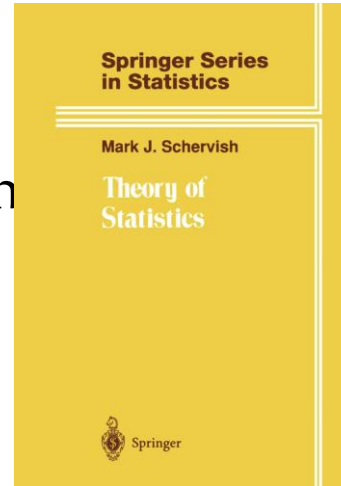
1. **Physics:** $d\sigma(x|c, v) dx$

Unbinned differential cross section
(detector-level, fiducial, total)

$c \dots POI \in \mathbb{R}^N$

$v \dots \text{nuisance parameters} \in \mathbb{R}^M$

$x \dots \text{observation (1 event)}$



Access to $d\sigma(x|c, v)$ via expensive sampling

2. **Machine-learning:** $d\sigma(x|c, v)$ defines a pdf
– The inclusive model is *not* particularly useful.

$$d\sigma(x|\mu, \nu) = \sigma(\mu, \nu)p(x|\mu, \nu)dx$$

3. **Mathematical statistics:** $d\sigma(x|\mu, \nu)$

A **measure** on a **probability space**

- Random variable x takes values in a “set”
- $d\sigma$ maps unions and intersections of subsets of this space (**σ -field**) to $\mathbb{R}$
- DCR = Radon-Nikodym derivative
- Technically relevant for overflow bins, Fisher information, etc.
 - And when you ask yourself “can I do this?”

Algorithm 1 Boosting parametric tree for learning of systematics

Require: base points $\nu \in \mathcal{V}$, sample $\mathcal{D}_0$ and $\mathcal{D}_\nu$ for all $\nu \in \mathcal{V}$,
boosting iterations B , learning rates $0 \leq \eta^{(b)} \leq 1$ for $b = 1, \dots, B$.

Ensure: $\sum_{\nu \in \mathcal{V}} \nu_A \nu_B$ has full rank

$$G_\nu^{(0)*}(\mathbf{x}) \leftarrow 0$$

$$\hat{g}_\nu^{(0)}(\mathbf{x}) \leftarrow 0$$

$$\mathcal{D}_\nu^{(0)} \leftarrow \mathcal{D}_\nu \text{ for all } \nu \in \mathcal{V}$$

for $b = 1, \dots, B$ **do**

$$\mathcal{D}_\nu^{(b)} \leftarrow \left\{ w_i^{(b)} \leftarrow \exp(-\eta^{(b-1)} G_\nu^{(b-1)*}(\mathbf{x}_i)) w_i^{(b-1)}, \mathbf{x}_i, \mathbf{z}_i \right\} \text{ for all } \{w^{(b-1)}, \mathbf{x}_i, \mathbf{z}_i\} \in \mathcal{D}_\nu^{(b-1)}$$

$$\mathcal{J}^{(b)} \leftarrow \arg \min_{\mathcal{J}} L[\mathcal{J}] \text{ with } \mathcal{D}_0^{(b)} \text{ and } \mathcal{D}_\nu^{(b)} \text{ using CART or TAO}$$

for all $j \in \mathcal{J}^{(b)}$ **do**

$$\lambda_{j,0} \leftarrow \sum_{(\mathbf{x}_i, w_i) \in \mathcal{D}_0 \cap j} w_i$$

$$\lambda_{j,\nu} \leftarrow \sum_{(\mathbf{x}_i, w_i) \in \mathcal{D}_\nu^{(b)} \cap j} w_i \text{ for all } \nu \in \mathcal{V}$$

$$\hat{\Delta}_{A,j}^{(b)} \leftarrow \left[\sum_{\nu \in \mathcal{V}} \nu \nu^T \right]_{AB}^{-1} \left[\sum_{\nu \in \mathcal{V}} \nu \log \frac{\lambda_{j,\nu}}{\lambda_{j,0}} \right]_B$$

end for

$$G_\nu^{(b)*}(\mathbf{x}) \leftarrow \sum_{j \in \mathcal{J}^{(b)}} \mathbb{1}_j(\mathbf{x}) \left(\nu_A \hat{\Delta}_{A,j}^{(b)} \right)$$

$$\hat{g}_\nu^{(b)}(\mathbf{x}) \leftarrow \hat{g}_\nu^{(b-1)}(\mathbf{x}) + \eta^{(b)} G_\nu^{(b)*}(\mathbf{x})$$

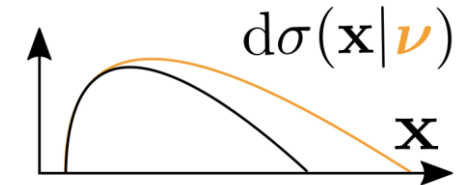
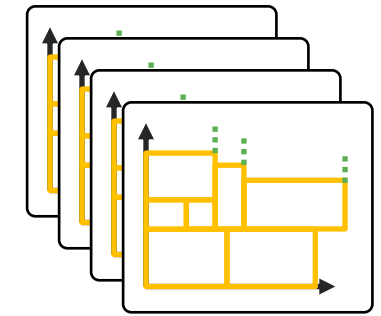
end for

$$\text{return } \hat{g}_\nu^{(B)}(\mathbf{x}) = \sum_{b=1}^B \eta^{(b)} \sum_{j \in \mathcal{J}^{(b)}} \mathbb{1}_j(\mathbf{x}) \nu_A \hat{\Delta}_{A,j}^{(b)}$$

A tree node is a bin.

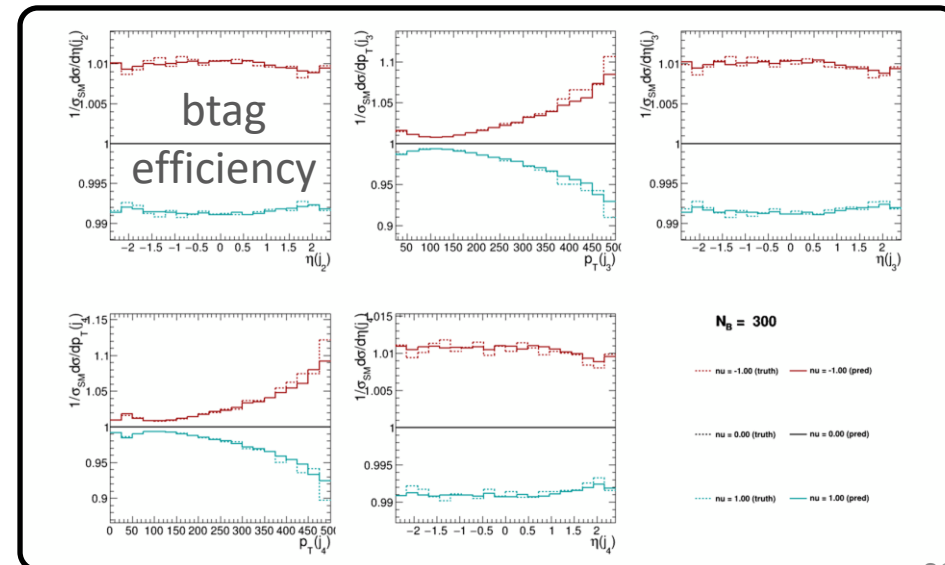
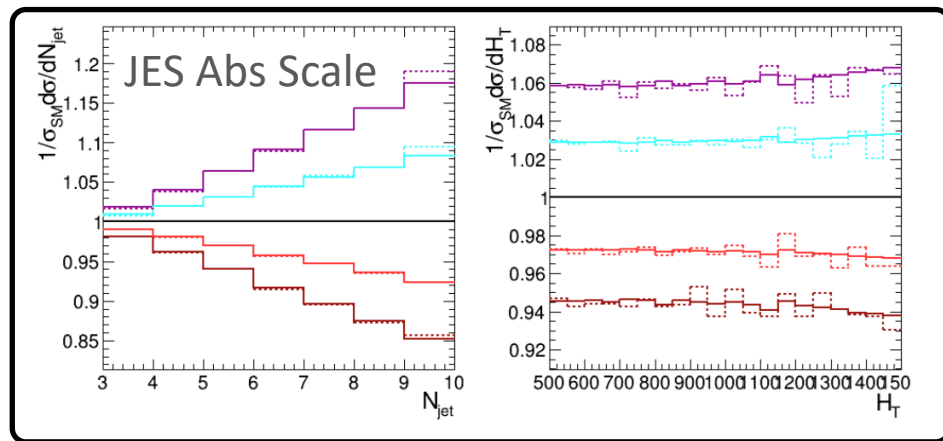
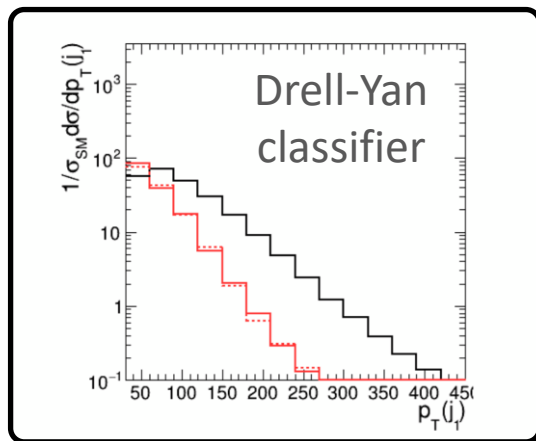
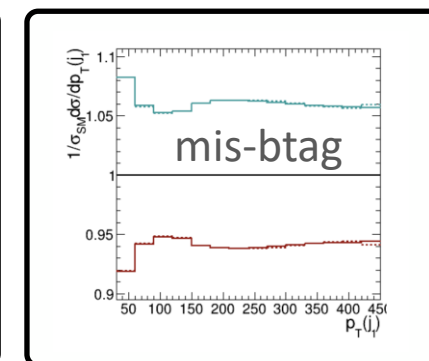
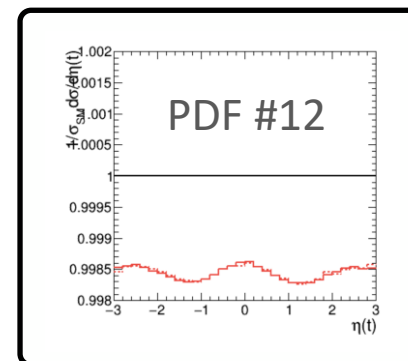
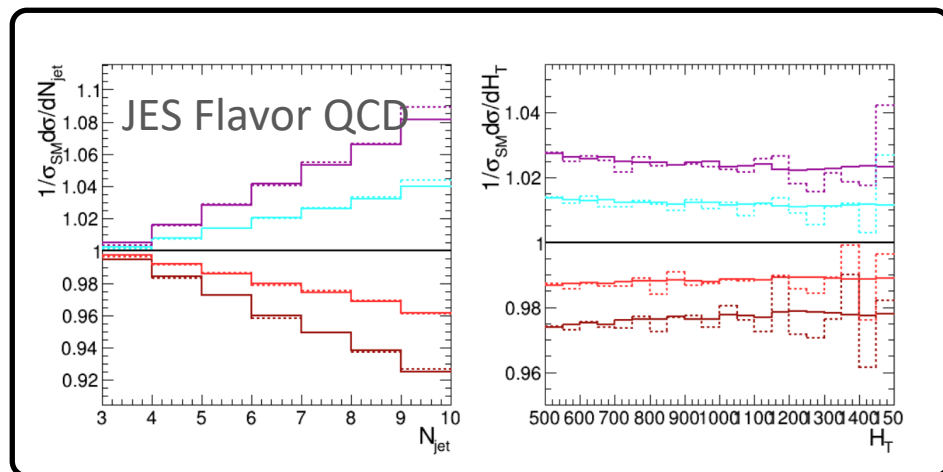
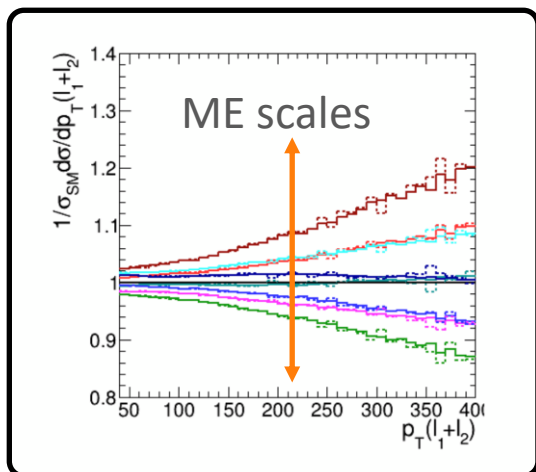
Same parametrization
as in [\[combine paper\]](#)

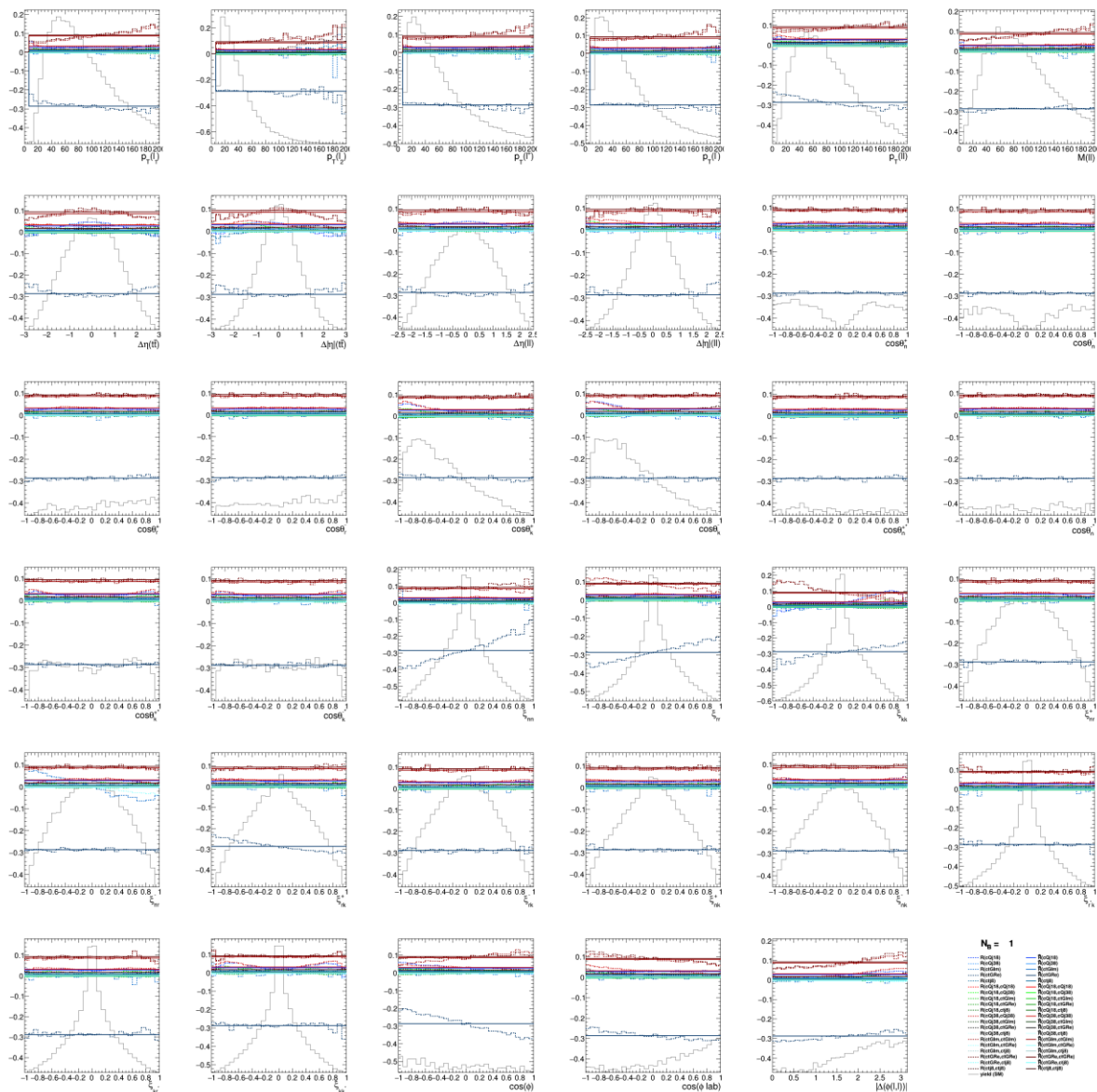
Boosted *Parametric* Tree



implements
x-dependent model
of cross section ratio

- 1D projections for a few values of $\mathbf{v}$. Available parametrically as $g_{\mathbf{v}}(\mathbf{x})$.
- Semi-realistic toy study (Delphes) with publicly available information





- From the playbook:

$$\frac{d\Sigma(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\nu})}{d\Sigma(\mathbf{x}|\text{SM})} \approx \frac{\hat{R}_{\text{EFT}}(\mathbf{x}|\boldsymbol{\theta}) \hat{S}_{\text{EFT}}(\mathbf{x}|\boldsymbol{\nu}) + \sum_{p=1}^{N_{\text{bkg}}} \alpha_p^{\nu_p} \hat{S}_p(\mathbf{x}|\boldsymbol{\nu}) \hat{g}_p(\mathbf{x})}{1 + \sum_{p=1}^{N_{\text{bkg}}} \hat{g}_p(\mathbf{x})}$$

↑
Quadratic
SMEFT model

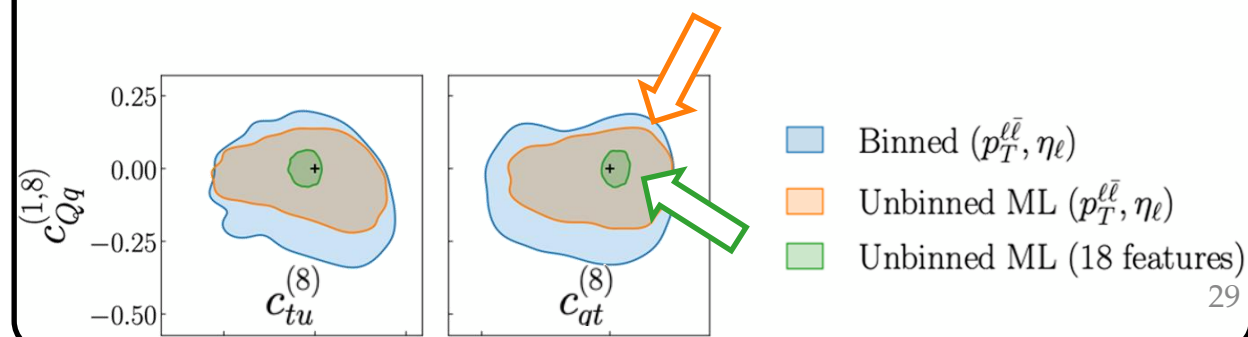
- Learn with “Boosted Information Tree (BIT)”

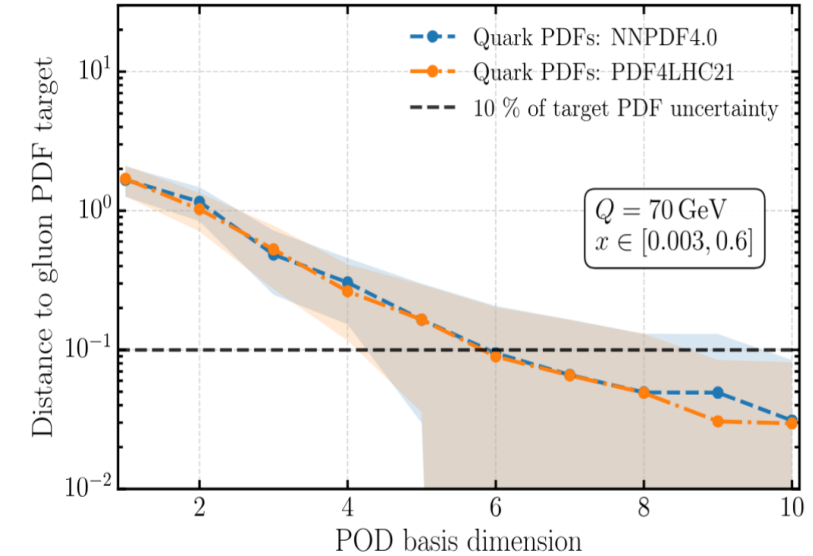
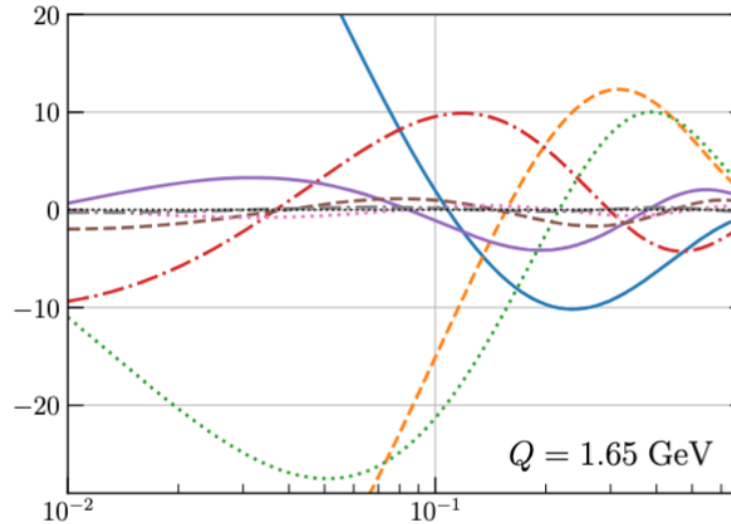
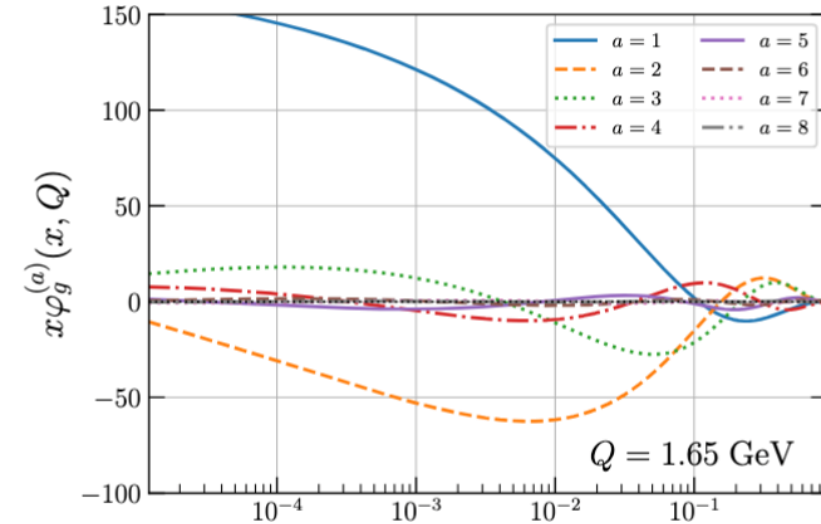
[2107.10859], [2205.12976] [MLST 6 015007 (2025)]

- 5 POIs, 20 functions simultaneously learned
- 35 features top quark pairs (2ℓ)
- 300 trees, D=5, ~9 hrs of training

- (Systematics-free) Unbinned analysis shows ~x5 sensitivity gain in profiled limits

ML4EFT R. Ambrosio, J. Hoeve, M. Madigan, J. Rojo, V. Sanz [2211.02058]





- Resulting basis elements define “wiggles” in PDF space
 - Orthogonal in *function space*
- With 6 basis elements, achieve sufficient precision
 - 10% of PDF4LHC21 uncertainty
- Quark PDFs: PDF4LHC21 at $Q_0=1.65$ GeV
 - DGLAP at NNLO with EKO [[Eur.Phys.J.C 82 \(2022\) 10, 976](#)] and $m_{b/c}$ as in NNPDF4.0
- How can we constrain the coefficients in the linear basis?

- “Likelihood ratio trick”
[PRD112 (2025) 5, 052006]

$$L_{\boldsymbol{\nu}, \text{CE}}[\hat{f}] = -\langle \log \hat{f}(\mathbf{x}) \rangle_{\mathbf{x}, \mathbf{z}, \text{SM}} - \langle \log(1 - \hat{f}(\mathbf{x})) \rangle_{\mathbf{x}, \mathbf{z} | \boldsymbol{\nu}}$$

$$f_{\text{CE}}^*(\mathbf{x}) = \frac{1}{1 + \frac{d\sigma(\mathbf{x} | \boldsymbol{\nu})}{d\sigma(\mathbf{x} | \text{SM})}}$$

- Parametric ansatz:

$$\hat{f}_{\boldsymbol{\nu}}(\mathbf{x}) = \frac{1}{1 + \exp(\hat{g}_{\boldsymbol{\nu}}(\mathbf{x}))}$$

$$\hat{g}_{\boldsymbol{\nu}}(\mathbf{x}; \hat{\Delta}_1, \hat{\Delta}_2, \dots) = \boldsymbol{\nu}^T \hat{\Delta}_1(\mathbf{x}) + \boldsymbol{\nu}^T \hat{\Delta}_2(\mathbf{x}) \boldsymbol{\nu} + \dots$$

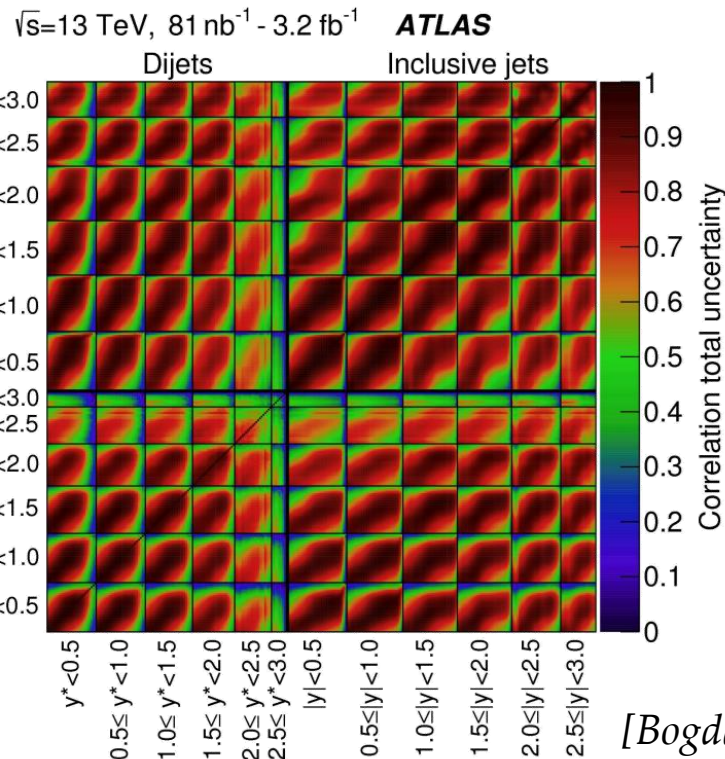
$\hat{\Delta}_{1,2,\dots}(\mathbf{x})$

Implement coefficient functions $\Delta(\mathbf{x})$ as NNs:

- Sufficiently many synthetic data sets

$$L[\hat{\Delta}] = \sum_{\boldsymbol{\nu} \in \mathcal{V}} L_{\boldsymbol{\nu}, \text{CE}}[\hat{f}_{\boldsymbol{\nu}}] \longrightarrow \exp(g_{\boldsymbol{\nu}}^*(\mathbf{x})) \approx \frac{d\sigma(\mathbf{x} | \boldsymbol{\nu})}{d\sigma(\mathbf{x} | \text{SM})}$$

- How can we incorporate unbinned detector-level likelihood $L(D|\mathbf{c}, \mathbf{v})$ in global fits?
- Staging: First, **take unfolded, binned χ^2 -based measurement** → **Recast with POD** → global fits
Followed by, possibly → Binned full likelihood with POD
 - Example: CMS inclusive jets (13 TeV) [[JHEP 12 \(2022\)035](#)]
 - NNLO theory predictions are binned. No point for NSBI. Jet-based measurements are indispensable for f_g



- Why could the full likelihood be useful on nuisance side?

1. Recover detector-level information in $L(D|\mathbf{c}, \mathbf{v})$

“When jets and top quarks don’t agree [in a global PDF fit], what do you do?” (June 19th)
 → What’s the postfit result on $\mathbf{v}$?

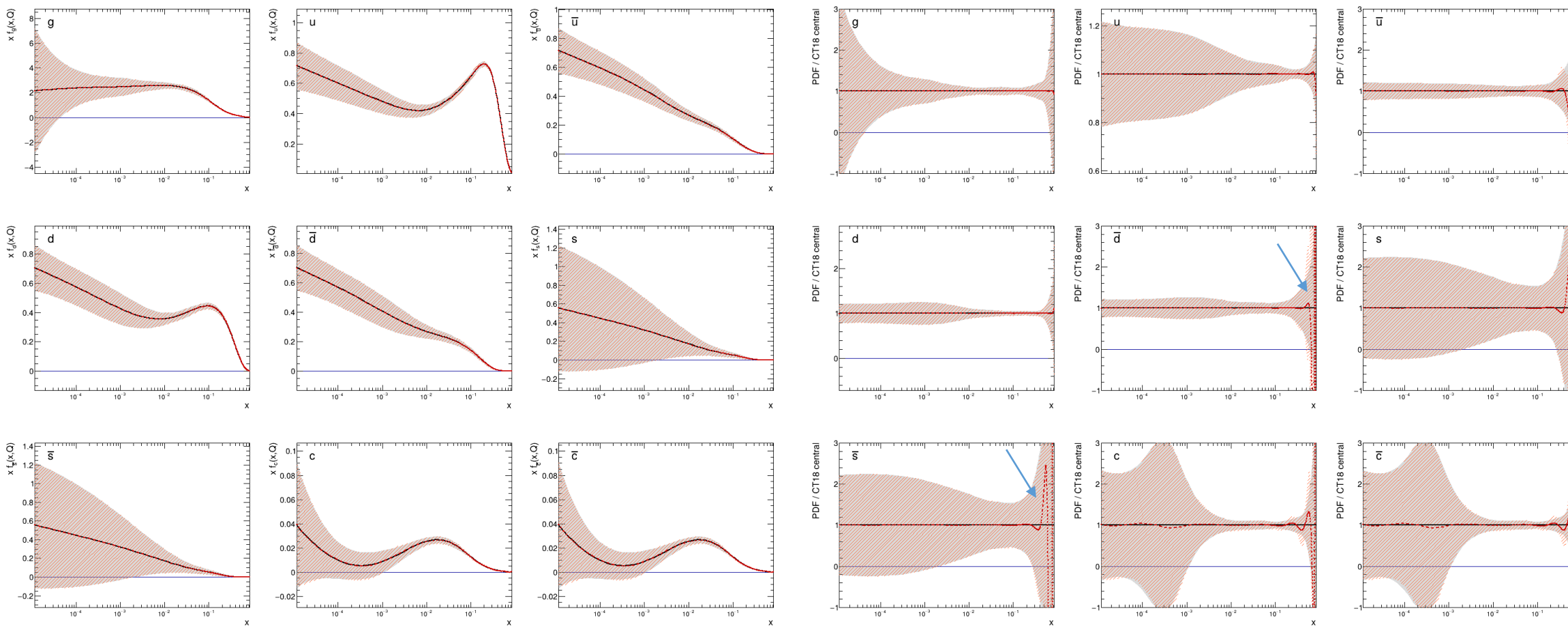
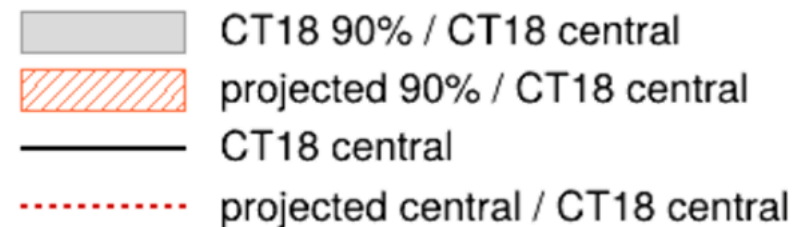
2. Experiments don’t see the gradient of the global fit

- “Flavor dependent JEC? b/g corrections?
Bias from JER in the crack region?” Or cone-size?
- We compress the full information when results are made public

3. Assembly. $N_{\text{bin}} \sim O(100)$, covariances scale with $N_{\text{bin}}^2 \times N_{\text{nuis}}$

- POD lives in Hilbert space → PDF *projections* can be exploited

- Projecting CT18NNLO onto all-flavor POD $\rightarrow c_{\text{POD}}, \Sigma_{\text{POD}}$
- Excellent closure. Apply within experiment *as a prior*.



- CMS 13 TeV (36fb⁻¹) and HERA DIS data

- χ^2 approx. is in [xFitter](#), public

- Parametrization (16 parameters)

$$xg = A_g x^{B_g} (1-x)^{C_g} (1 + D_g x + E_g x^2),$$

$$xu_v = A_{u_v} x^{B_{u_v}} (1-x)^{C_{u_v}} (1 + E_{u_v} x^2),$$

$$xd_v = A_{d_v} x^{B_{d_v}} (1-x)^{C_{d_v}},$$

$$x\bar{U} = A_{\bar{U}} x^{B_{\bar{U}}} (1-x)^{C_{\bar{U}}} (1 + D_{\bar{U}} x),$$

$$x\bar{D} = A_{\bar{D}} x^{B_{\bar{D}}} (1-x)^{C_{\bar{D}}} (1 + E_{\bar{D}} x^2).$$

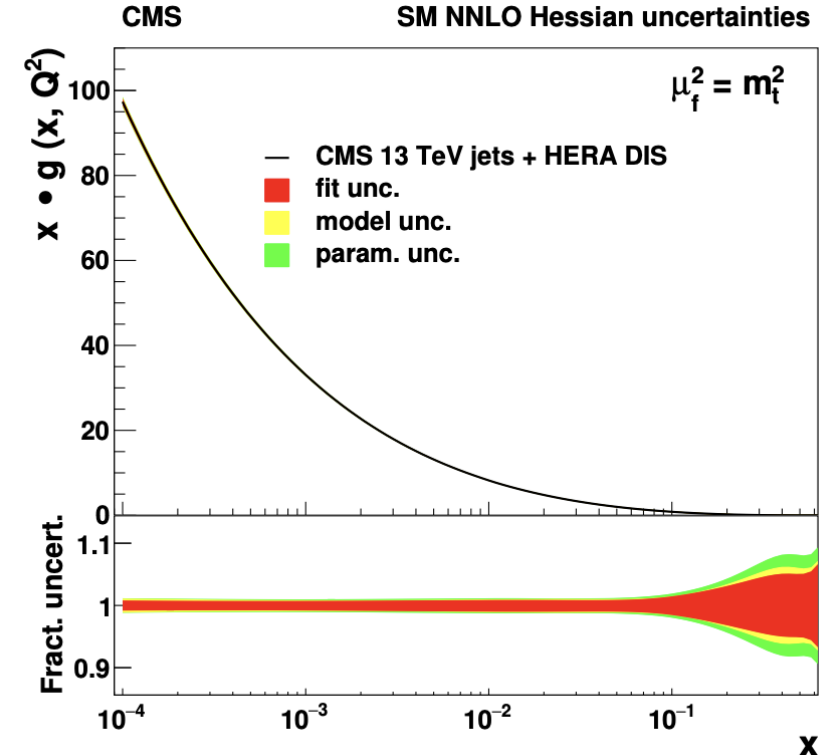
- Reproduce paper result to rounding error

- Implemented POD basis in xFitter [[repo](#)]

- *Exposes systematics*; useful testbed; In HEPData, therefore in CT18 [[Phys.Rev.D 111 \(2025\) 3, 036033](#)]

- Compare:

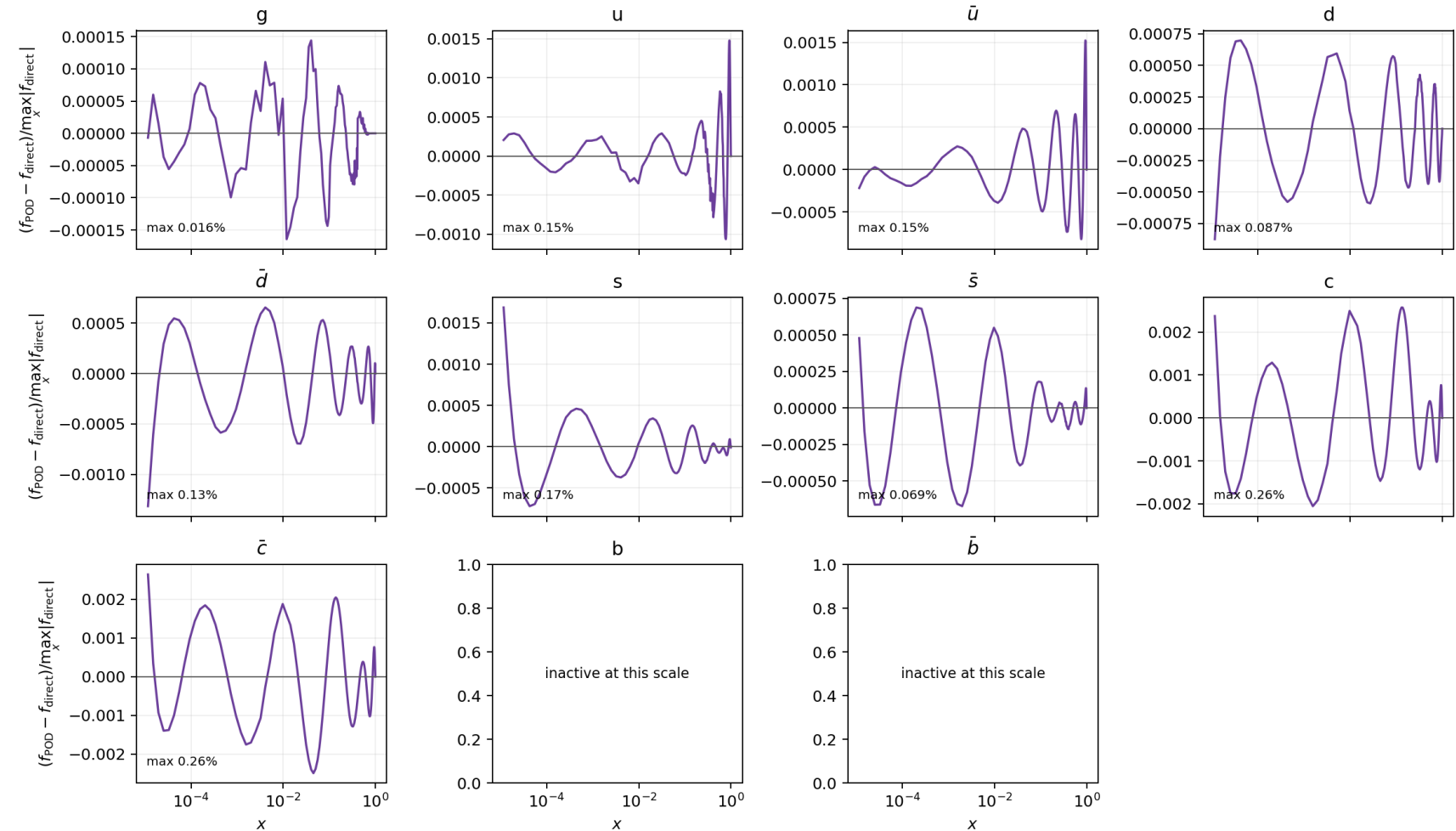
1. Analytic parametrization $\rightarrow$ fastNLO $\rightarrow \chi^2 \rightarrow$ fit
2. Analytic parametrization $\rightarrow$ **POD** $\rightarrow$ **fastNLO** $\rightarrow \chi^2 \rightarrow$ fit

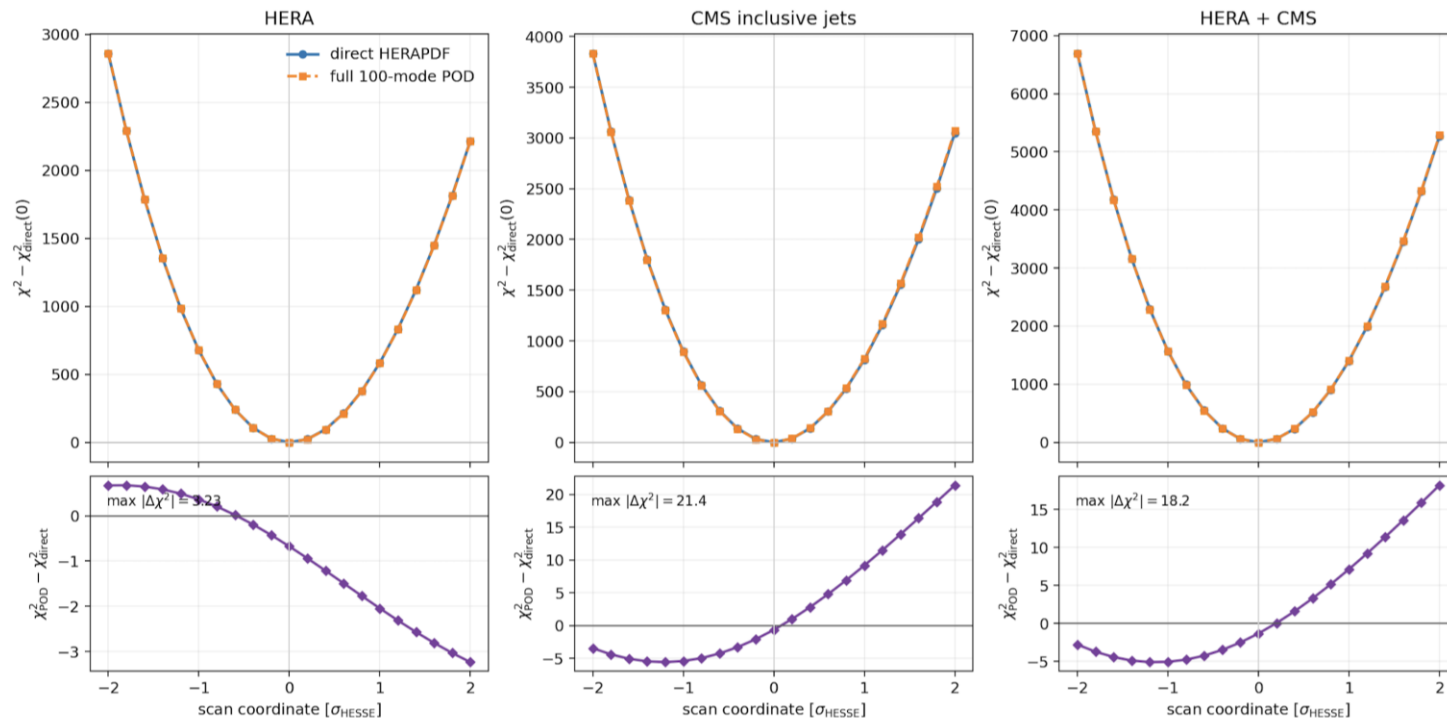


CMS uncertainties

lumi, JER, PU, miss, fake, AbsoluteScale, AbsoluteMPFBias, Fragmentation, SinglePionECAL, SinglePionHCAL, FlavorQCD, TimePtEta, RelativeJEREC1, RelativeJEREC2, RelativePtBB, RelativePtEC1, RelativePtEC2, RelativeBal, RelativeSample, RelativeFSR, RelativeStatFSR, RelativeStatEC, PileUpDataMC, PileUpPtRef, PileUpPtBB, PileUpPtEC1, PileUpPtEC2, model, NP + *HERA uncertainties (162)*

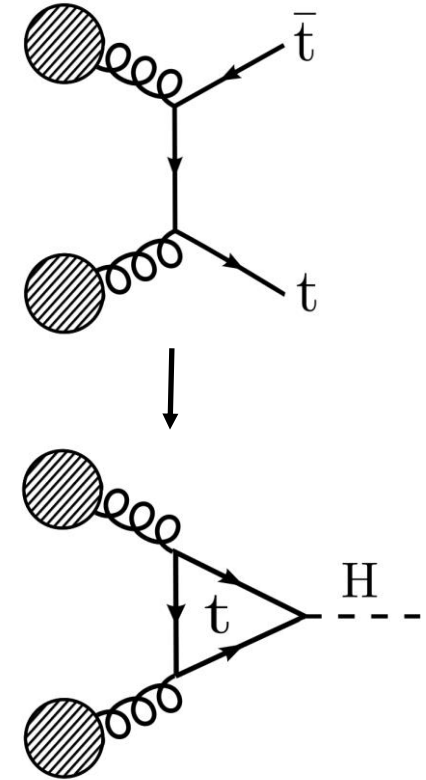
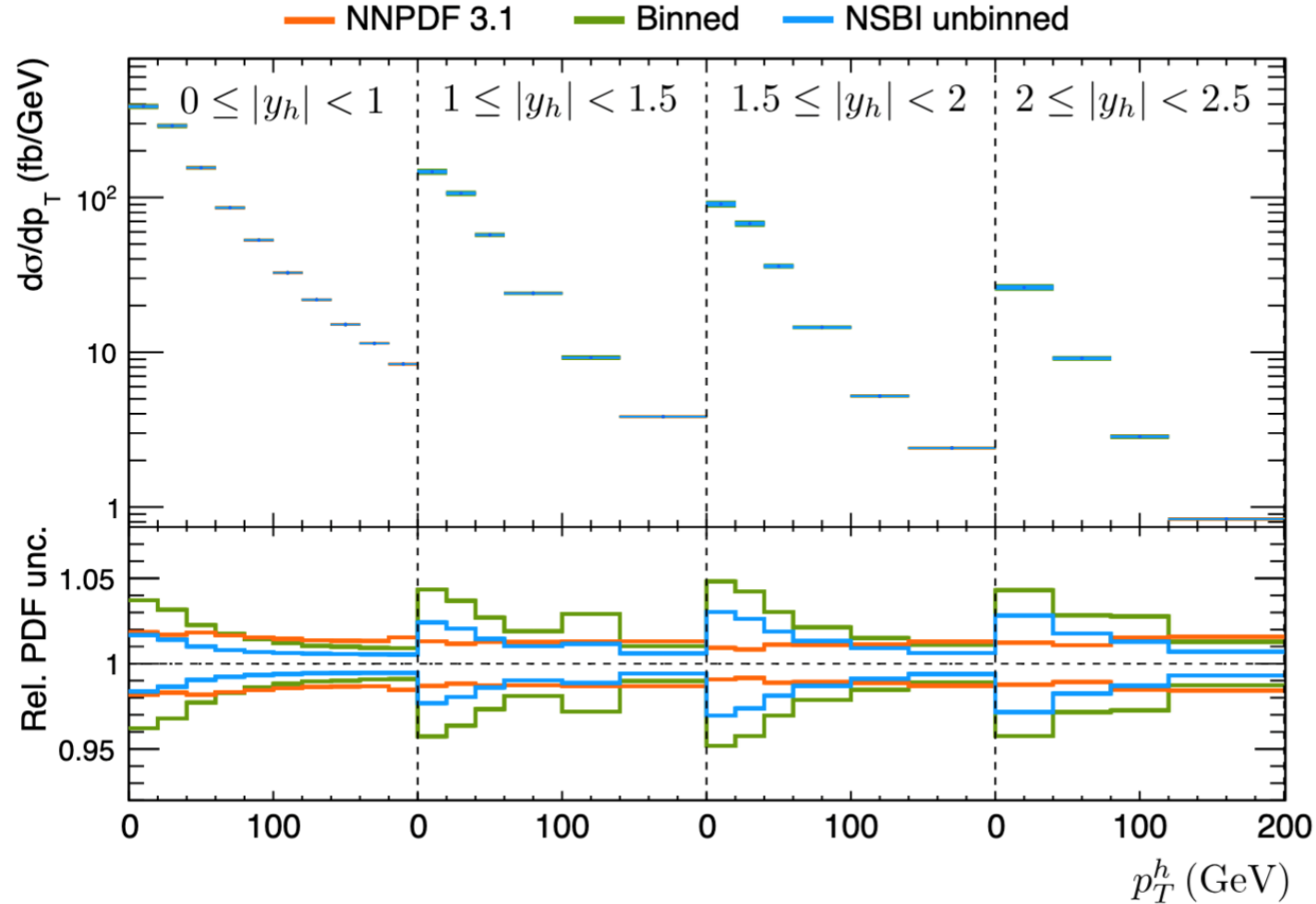
- Project analytic model on the POD
- Needed to specify x-range, scale of the projection, active flavors, DGLAP, etc.
- Needed to upweight up high-x
- Bias of POD projection $\sim 10^{-3} - 10^{-4}$ on xf
- Sub-% level differences in predicted cross-sections





- Scan likelihood and fit
 - Analytic parametrization $\rightarrow$ POD $\rightarrow$ fastNLO $\rightarrow \chi^2 \rightarrow$ fit
 - POD bias $< 0.05 \sigma$

Parameter	Reference	POD-mediated	Shift $[(\sigma)]$
ADbar	0.267003	0.267843	+0.0536
BDbar	-0.120313	-0.119918	+0.0477
Bdv	0.762795	0.765778	+0.0419
Bg	0.153937	0.153303	-0.0149
Buv	0.714465	0.714054	-0.0159
CDbar	10.668946	10.663250	-0.0060
CUbar	7.471823	7.494759	+0.0348
Cdv	3.743693	3.757982	+0.0404
Cg	9.691203	9.696871	+0.0112
Cuv	4.901321	4.901681	+0.0046
DUbar	4.145963	4.128594	-0.0121
Dg	-4.525369	-4.510517	+0.0264
EDbar	23.988433	24.170836	+0.0101
Eg	32.131128	32.167993	+0.0105
Euv	13.770841	13.822875	+0.0374
(\alpha_s(M_Z))	0.116638	0.11664050	+0.0016



- For central rapidity, top quarks pair production gives better precision than NNPDF 3.1

$$\left\{ \tilde{f}_g^{(m)} \right\}_{m=1}^M \xrightarrow{\text{subtract mean}} \left\{ \hat{f}_g^{(m)} \right\}_{m=1}^M \xrightarrow{\text{POD/SVD}} \left\{ \varphi_g^{(a)} \right\}_{a=1}^N \xrightarrow{\text{DGLAP}} f_g(x, Q, \mathbf{c})$$

1. Make the gluon PDF variable & *sample*

$$x \tilde{f}_g^{(m)}(x, Q_0) = A_m x^{1-\alpha_m} (1-x)^{\beta_m} \text{NN}_m(x; \boldsymbol{\theta})$$

NNs with hidden layers 25/20

[JHEP 04 (2026) 068]

2×10^4 random samples

A_m frozen by momentum sum rule

$\alpha, \beta \dots$ respect High-/low- x asymptotic

Quark PDFs fixed NNPDF 3.1

[J.Phys.G 49 (2022) 8]

2. Subtract the mean

$$x \hat{f}_g^{(m)}(x, Q_0) = x \tilde{f}_g^{(m)}(x, Q_0) - x \varphi_g^{(0)}(x, Q_0)$$

$$\varphi_g^{(0)}(x, Q_0) = \frac{1}{M} \sum_{m=1}^M \tilde{f}_g^{(m)}(x, Q_0)$$

3. Assume a grid on x and orthonormal modes

$$\mathcal{L}_N = \frac{1}{M} \sum_{m=1}^M \left\| \hat{\mathbf{f}}_g^{(m)} - \sum_{a=1}^N \underbrace{\left(\varphi_g^{(a)T} \hat{\mathbf{f}}_g^{(m)} \right)}_{\text{projection}} \varphi_g^{(a)} \right\|^2$$

Many variations $M = 2 \times 10^4$

Few basis elements

$a = 1 \dots N \ll M$

Equivalent to PCA of grid x grid matrix

$$A = \sum_{m=1}^M \hat{\mathbf{f}}_g^{(m)} \hat{\mathbf{f}}_g^{(m)T}$$

4. DGLAP of $f_g(x, Q_0, \mathbf{c}) \equiv \varphi_g^{(0)}(x, Q_0) + \sum_{a=1}^N \mathbf{c}_a \varphi_g^{(a)}(x, Q_0)$

$$f_g(x, Q) = \Gamma_{gg}(x, Q, Q_0) \otimes f_g(x, Q_0) + \Gamma_{g\Sigma}(x, Q, Q_0) \otimes f_\Sigma(x, Q_0)$$

$$f_\Sigma(x, Q) = \Gamma_{\Sigma g}(x, Q, Q_0) \otimes f_g(x, Q_0) + \Gamma_{\Sigma\Sigma}(x, Q, Q_0) \otimes f_\Sigma(x, Q_0)$$