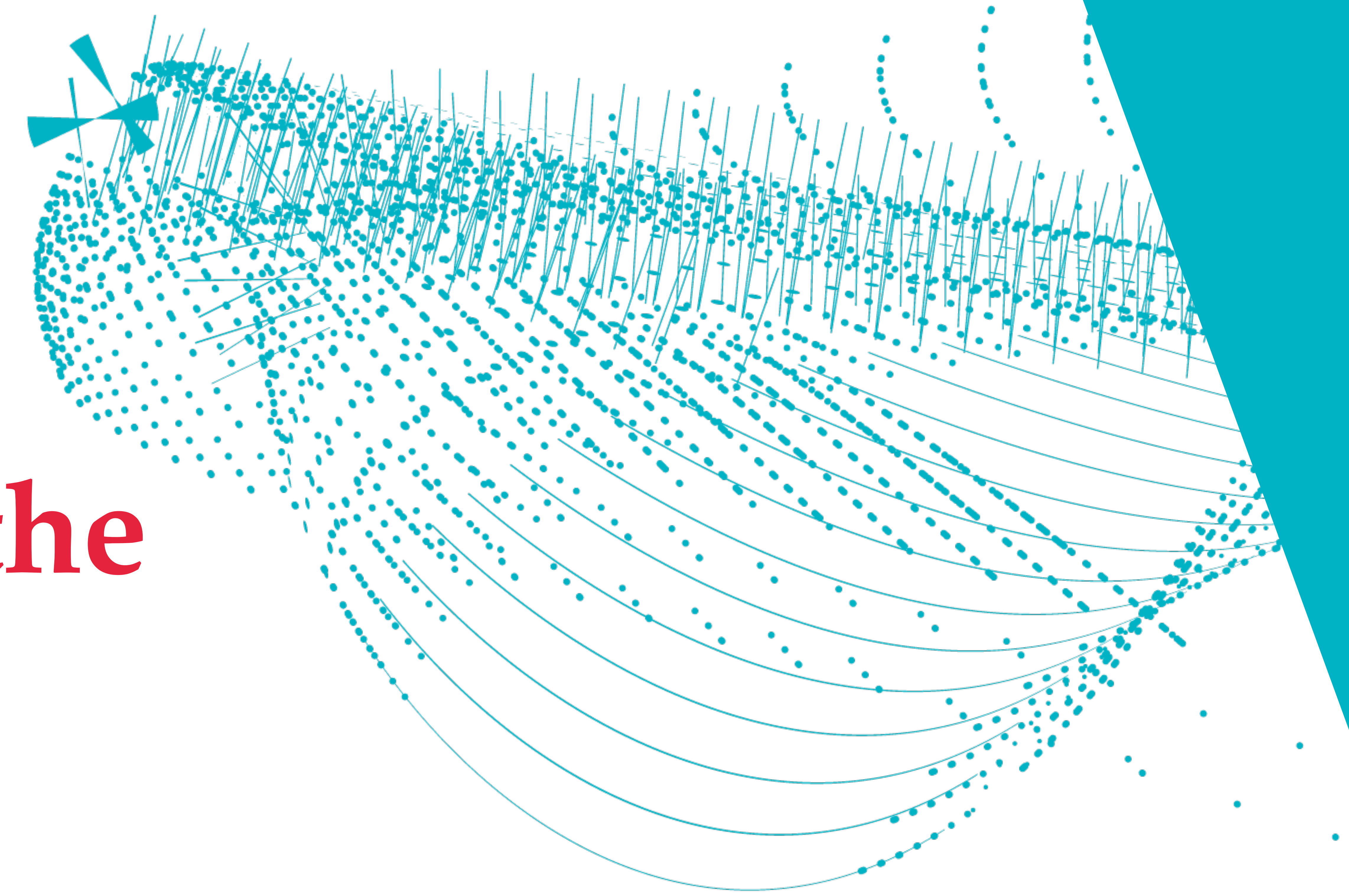
PDF constraints from the Top sector

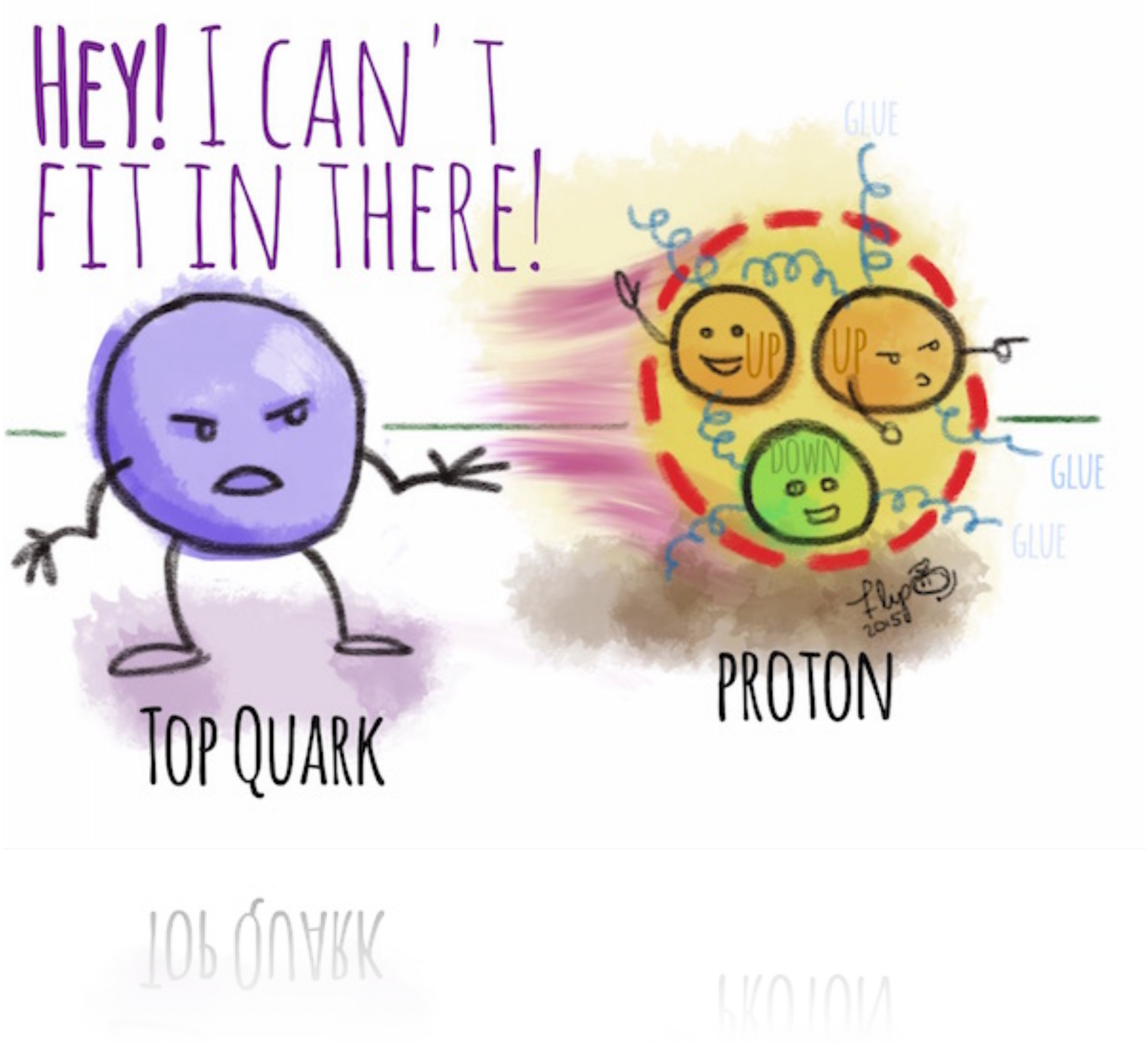


Tanjona R. Rabemananjara on behalf of the NNPDF collaboration
LHCP2025, May 8th 2025
Taipei, Taiwan

Outline of the Talk

Introduction & Motivations

- Top quark productions at the LHC complements (di-)jet measurements in providing a clean probe for PDFs at intermediate and large momentum fraction
- Processes such as top-pair productions are relevant in extracting the value of the strong coupling through the strong correlation between α_s and the gluon PDF at large partonic momentum
- Single top and top-pair productions are among the processes that can be computed at high perturbative accuracy in QCD (with electroweak corrections)



This Talk:

- Review the impacts of top measurements in current PDF fits
- Quantify the goodness of current available PDFs in describing new precision top data measurements from Run I/II
- Provide preliminary quantifications of their impacts in constraining PDFs
- Top data towards NNPDF4.1

Outline of the Talk

	Introduction & Motivations	
Part I	Top Data in NNPDF4.0	[arXiv:2109.02653; arXiv:2109.02671]
Part II	PDFs confront Top Data: A Quantitative Appraisal	[arXiv:2501.10359]
Part III	Impacts of new Top data on PDFs & NNPDF4.1	
	Conclusions & Outlook	

PDF Determination: Formalism & Ingredients

Collinear Factorisation enables the separation between short-range and long-range physics:

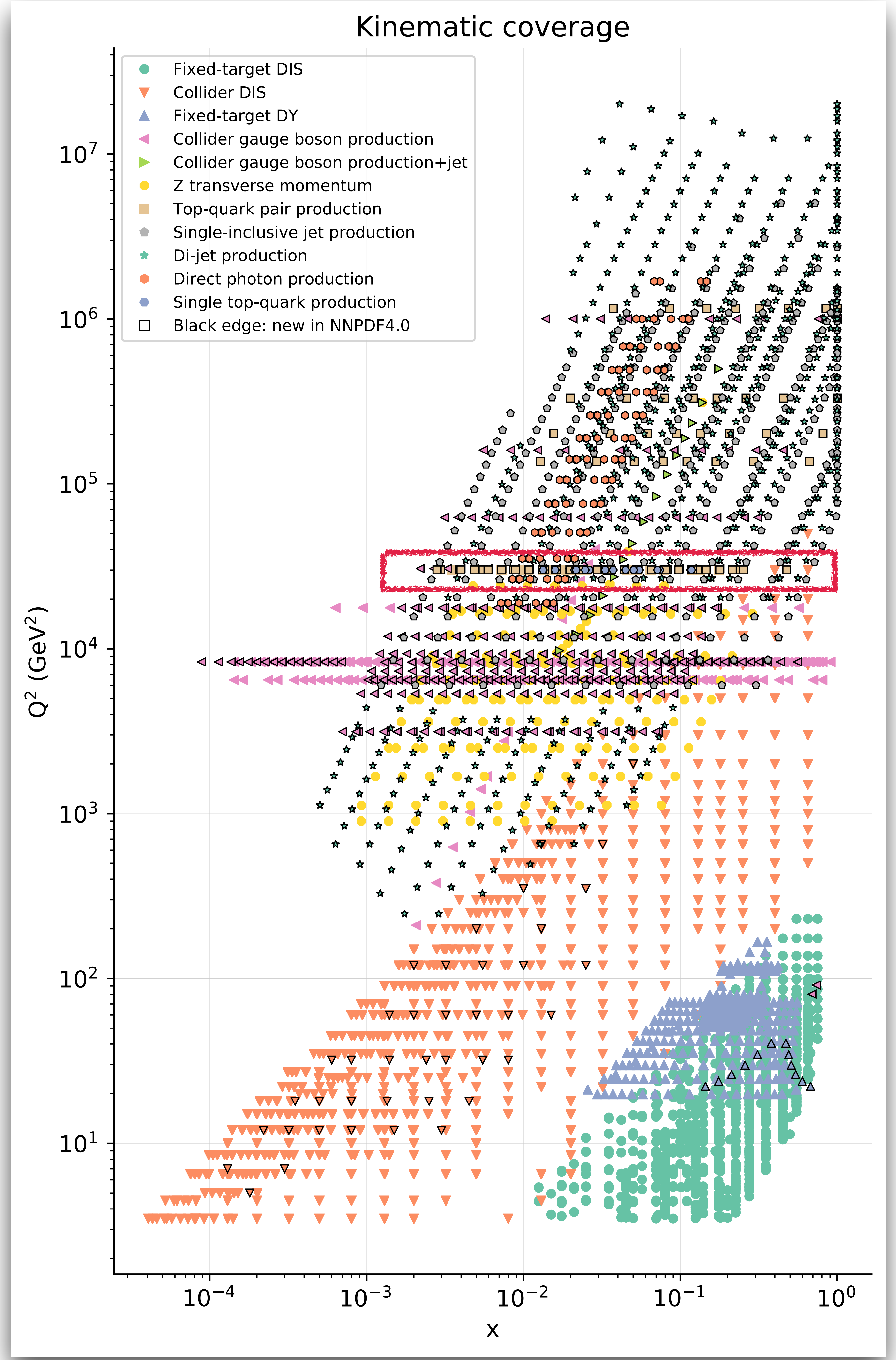
$$\mathcal{O}_{\mathrm{LHC}}(M,s) \propto \sum_{ij} \int_{M^2}^{s} d\hat{s}\, \mathcal{L}_{ij}(\hat{s},s)\hat{\mathcal{O}}_{ij}\left(\hat{s},\alpha_s(M)\right)$$

Where the Partonic luminosity directly relates to the PDFs:

$$\mathcal{L}_{ij}(Q,s) = \frac{1}{s}\int_{Q^2/s}^{1} \frac{dx}{x} f_i\left(\frac{Q^2}{sx},Q\right) f_j(x,Q)$$

PDFs are universal and their determinations involve **three main ingredients**:

- **Experimental data** — contains $\sim \mathcal{O}(4000)$ datapoints, probe different processes and channels, include LHC Run II, etc.
- **Theory predictions** — Heavy quark schemes, NNLO (aN3LO) accuracy, MHOUs, Photon PDFs, intrinsic charm, etc.
- **Methodology** — Neural Network (NN) parametrisation, closure & future tests, uncertainty propagation, etc.



Top-pair Productions

- NNPDF4.0 includes Top data in the form of single-top and $t\bar{t}$ pair production
- Top datasets significantly extended in NNPDF4.0: **~30 datapoints more than NNPDF3.1**

Dataset	Ref.	$N_{\rm dat}$	Kin$_1$	Kin$_2$ [GeV]	Theory
CMS $\sigma_{tt}^{\rm tot}$ 5 TeV **(*)**	[88]	1 (1/1)	—	$Q = m_t$	MCFM+top++
ATLAS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV	[65]	2 (2/2)	—	$Q = m_t$	MCFM+top++
CMS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV	[146]	2 (2/2)	—	$Q = m_t$	MCFM+top++
ATLAS $\sigma_{tt}^{\rm tot}$ 13 TeV ($\mathcal{L}$=139 fb^{-1}) **(*)**	[134]	1 (1/1)	—	$Q = m_t$	MCFM+top++
CMS $\sigma_{tt}^{\rm tot}$ 13 TeV	[69]	1 (1/1)	—	$Q = m_t$	MCFM+top++
[ATLAS $t\bar{t}$ ℓ+jets 8 TeV ($1/\sigma d\sigma/dp_T^t$)]	[67]	8 (—/8)	$0 \leq p_T^t \leq 500$ GeV	$Q = m_t$	Sherpa+NNLO
ATLAS $t\bar{t}$ ℓ+jets 8 TeV ($1/\sigma d\sigma/dy_t$)	[67]	5 (4/4)	$\lvert y_t\rvert < 2.5$	$Q = m_t$	Sherpa+NNLO
ATLAS $t\bar{t}$ ℓ+jets 8 TeV ($1/\sigma d\sigma/dy_{t\bar{t}}$)	[67]	5 (4/4)	$\lvert y_{t\bar{t}}\rvert < 2.5$	$Q = m_t$	Sherpa+NNLO
[ATLAS $t\bar{t}$ ℓ+jets 8 TeV ($1/\sigma d\sigma/dm_{t\bar{t}}$)]	[67]	7 (—/7)	$345 \leq m_{t\bar{t}} \leq 1600$ GeV	$Q = m_t$	Sherpa+NNLO
ATLAS $t\bar{t}$ 2ℓ 8 TeV ($1/\sigma d\sigma/dy_{t\bar{t}}$) **(*)**	[89]	5 (5/5)	$\lvert y_{t\bar{t}}\rvert < 2.8$	$Q = m_t$	mg5_aMC+NNLO
CMS $t\bar{t}$ ℓ+jets 8 TeV ($1/\sigma d\sigma/dy_{t\bar{t}}$)	[70]	10 (9/9)	$-2.5 < y_{t\bar{t}} < 2.5$	$Q = m_t$	Sherpa+NNLO
CMS $t\bar{t}$ 2D 2ℓ 8 TeV ($1/\sigma d\sigma/dy_t dm_{t\bar{t}}$) **(*)**	[90]	16 (16/16)	$\lvert y_t\rvert < 2.5$	$340 \leq m_t \leq 1500$	mg5_aMC+NNLO
CMS $t\bar{t}$ ℓ+jet 13 TeV ($d\sigma/dy_t$) **(*)**	[91]	10 (10/10)	$\lvert y_t\rvert < 2.5$	$Q = m_t$	mg5_aMC+NNLO
CMS $t\bar{t}$ 2ℓ 13 TeV ($d\sigma/dy_t$) **(*)**	[92]	11 (11/11)	$\lvert y_t\rvert < 2.5$	$Q = m_t$	mg5_aMC+NNLO

- not included in the fit
- New in NNPDF4.0

Single-Top Production

- NNPDF4.0 includes Top data in the form of single-top and $t\bar{t}$ pair production
- Top datasets significantly extended in NNPDF4.0: **~30 datapoints more than NNPDF3.1**

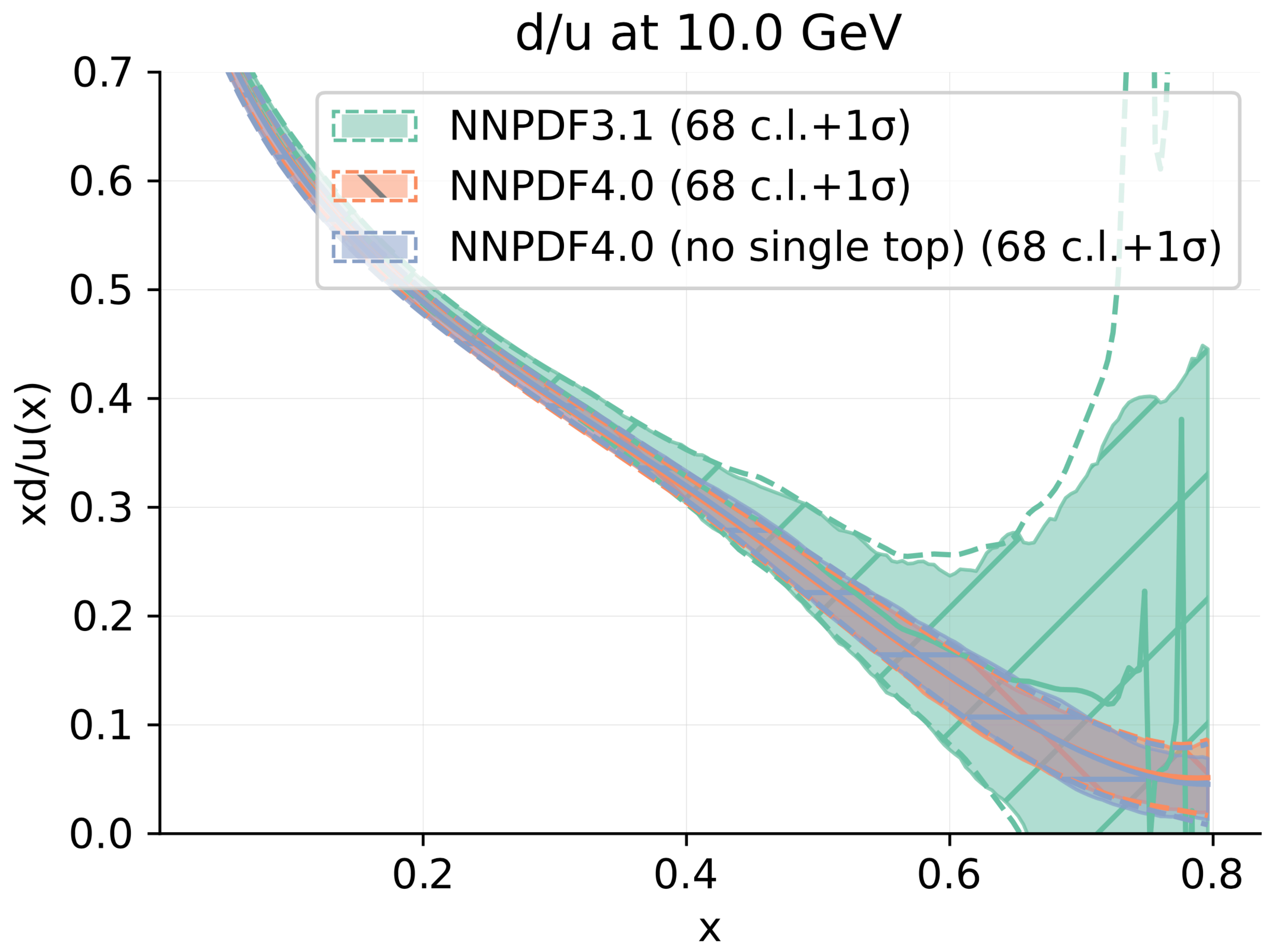
Dataset	Ref.	$N_{\rm dat}$	Kin$_1$	Kin$_2$ [GeV]	Theory
ATLAS single t R_t 7 TeV **(*)**	[94]	1 (1/1)	—	$Q = m_t$	mg5_aMC+NNLO
CMS single t $\sigma_t + \sigma_{\bar{t}}$ 7 TeV **(*)**	[95]	1 (1/1)	—	$Q = m_t$	mg5_aMC+NNLO
ATLAS single t R_t 8 TeV **(*)**	[96]	1 (1/1)	—	$Q = m_t$	mg5_aMC+NNLO
CMS single t R_t 8 TeV **(*)**	[97]	1 (1/1)	—	$Q = m_t$	mg5_aMC+NNLO
ATLAS single t R_t 13 TeV **(*)**	[98]	1 (1/1)	—	$Q = m_t$	mg5_aMC+NNLO
CMS single t R_t 13 TeV **(*)**	[99]	1 (1/1)	—	$Q = m_t$	mg5_aMC+NNLO
ATLAS single t 7 TeV $(1/\sigma d\sigma/dy_t)$ **(*)**	[94]	4 (3/3)	$\lvert y_t\rvert < 3.0$	$Q = m_t$	mg5_aMC+NNLO
ATLAS single t 7 TeV $(1/\sigma d\sigma/dy_{\bar{t}})$ **(*)**	[94]	4 (3/3)	$\lvert y_{\bar{t}}\rvert < 3.0$	$Q = m_t$	mg5_aMC+NNLO
ATLAS single t 8 TeV $(1/\sigma d\sigma/dy_t)$ **(*)**	[96]	4 (3/3)	$\lvert y_t\rvert < 2.2$	$Q = m_t$	mg5_aMC+NNLO
ATLAS single t 8 TeV $(1/\sigma d\sigma/dy_{\bar{t}})$ **(*)**	[96]	4 (3/3)	$\lvert y_{\bar{t}}\rvert < 2.2$	$Q = m_t$	mg5_aMC+NNLO

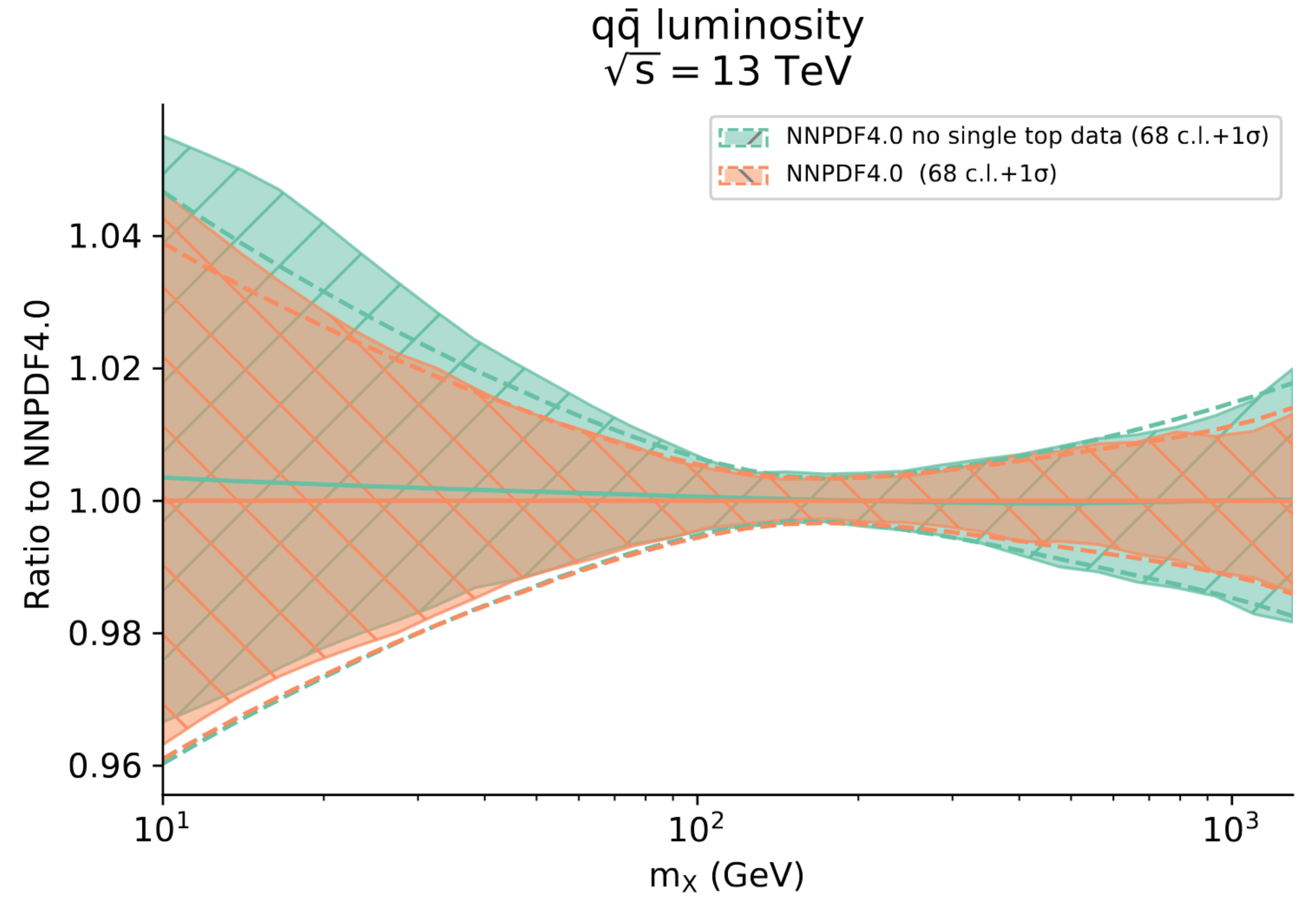
- not included in the fit
- New in NNPDF4.0

Impacts of LHC Single-Top Datasets

- Single-top datasets in NNPDF4.0 include:
  - normalised $y_t/y_{\bar{t}}$ distributions from ATLAS at 7 and 8 TeV
  - total/ratio cross-sections from ATLAS at 7 and 13 TeV
  - total/ratio cross-sections from CMS at 7 and 13 TeV
- **Quark-flavour separation at large-x (d/u and $\bar{d}/\bar{u}$)** are in principle provided by single-top quark production (included for the first time in NNPDF4.0 along with fixed-target DY from **SeaQuest**)
- Single-top data have no impact on the d/u ratio and more generally on the whole PDF determination $\Longleftrightarrow$ **due to the relatively large experimental uncertainties of the corresponding measurements**
- Reduction of uncertainties from NNPDF3.1 to NNPDF4.0 is purely of methodological origin
- Removing single-top production data only results to a **slight increase of the uncertainties** in the small-invariant mass of the Luminosity

	NNPDF4.0	**NNPDF4.0 w/o single-top**
Chi2	1.16022	1.16278

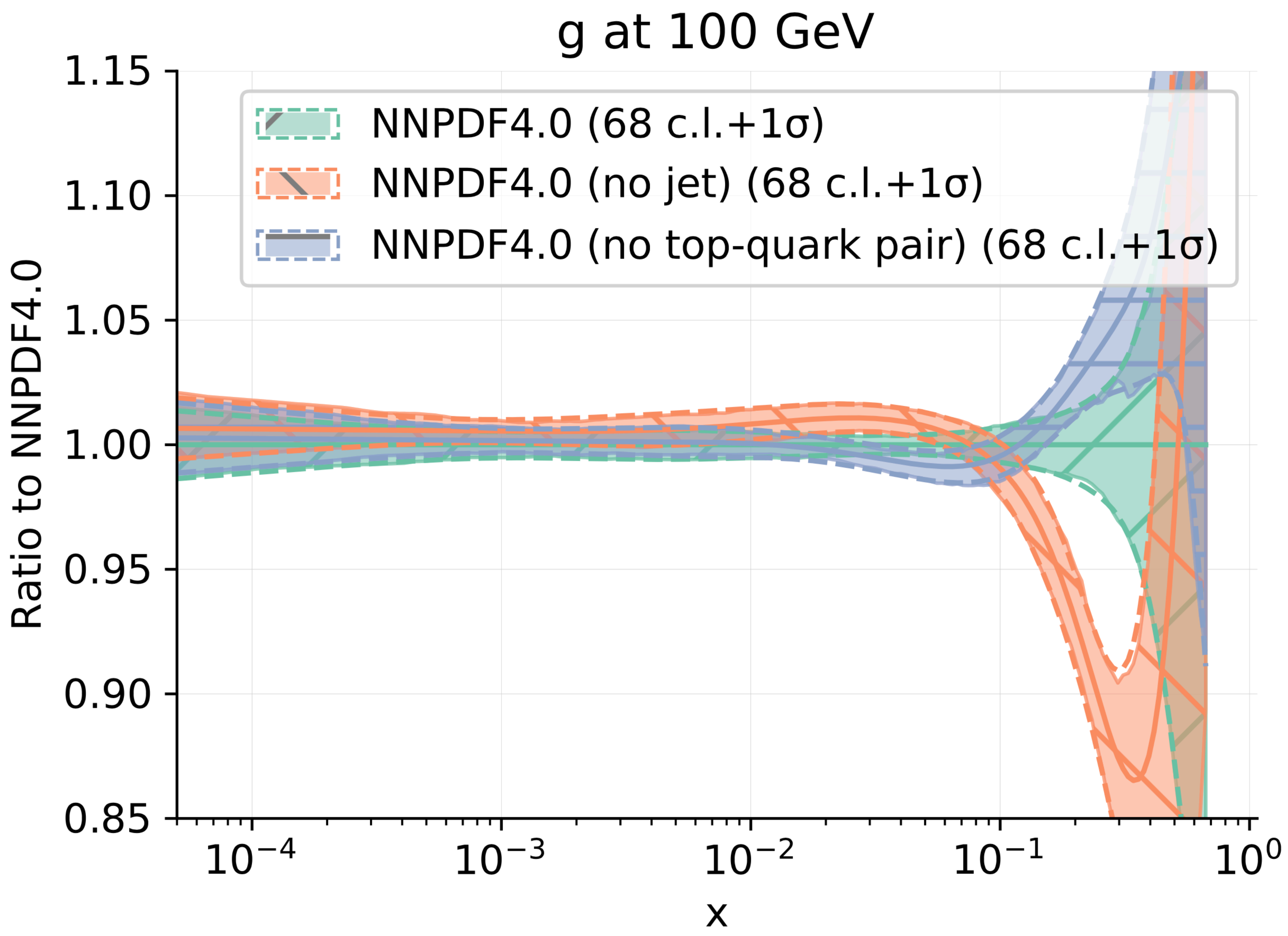


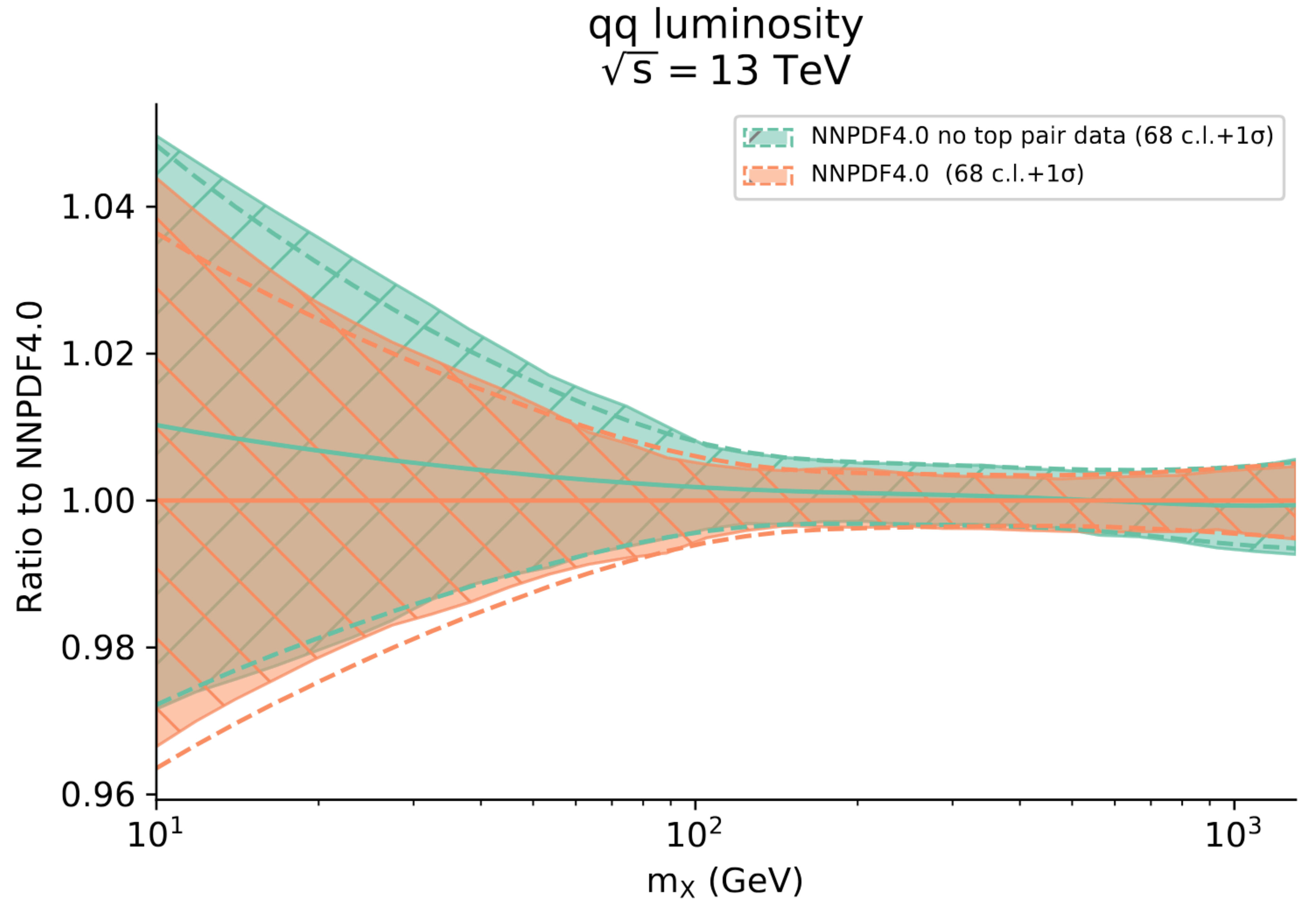


Impacts of LHC top-quark pair

- Top-pair production datasets in NNPDF4.0 include:
    - Total cross-sections from ATLAS at 7, 8, and 13 TeV
    - Total cross-sections from CMS at 5, 7, 8, and 13 TeV
    - normalised $y_t/y_{\bar{t}}$ distributions from ATLAS at 8 and 13 TeV (W/ leptons+jets in the final-state)
    - y_t and 2D $y_t m_{t\bar{t}}$ distributions from CMS at 8 and 13 TeV
- Top-pair production, which is initiated by gg-scattering, **primarily probes gluon PDF at large-x**
- **Removing jet** data **strongly suppresses (~15%) the gluon at medium-x** $\Longleftrightarrow$ pulls from other datasets such as top-quark pair production
- **single-inclusive jet** and **di-jet** production data, which are the most abundant and precise, **drive the features of the gluon** in the global fit
- Top-pair data have really moderate impact $\Longleftrightarrow$ slight enhancement of the gluon PDF for $x \geq 0.1$

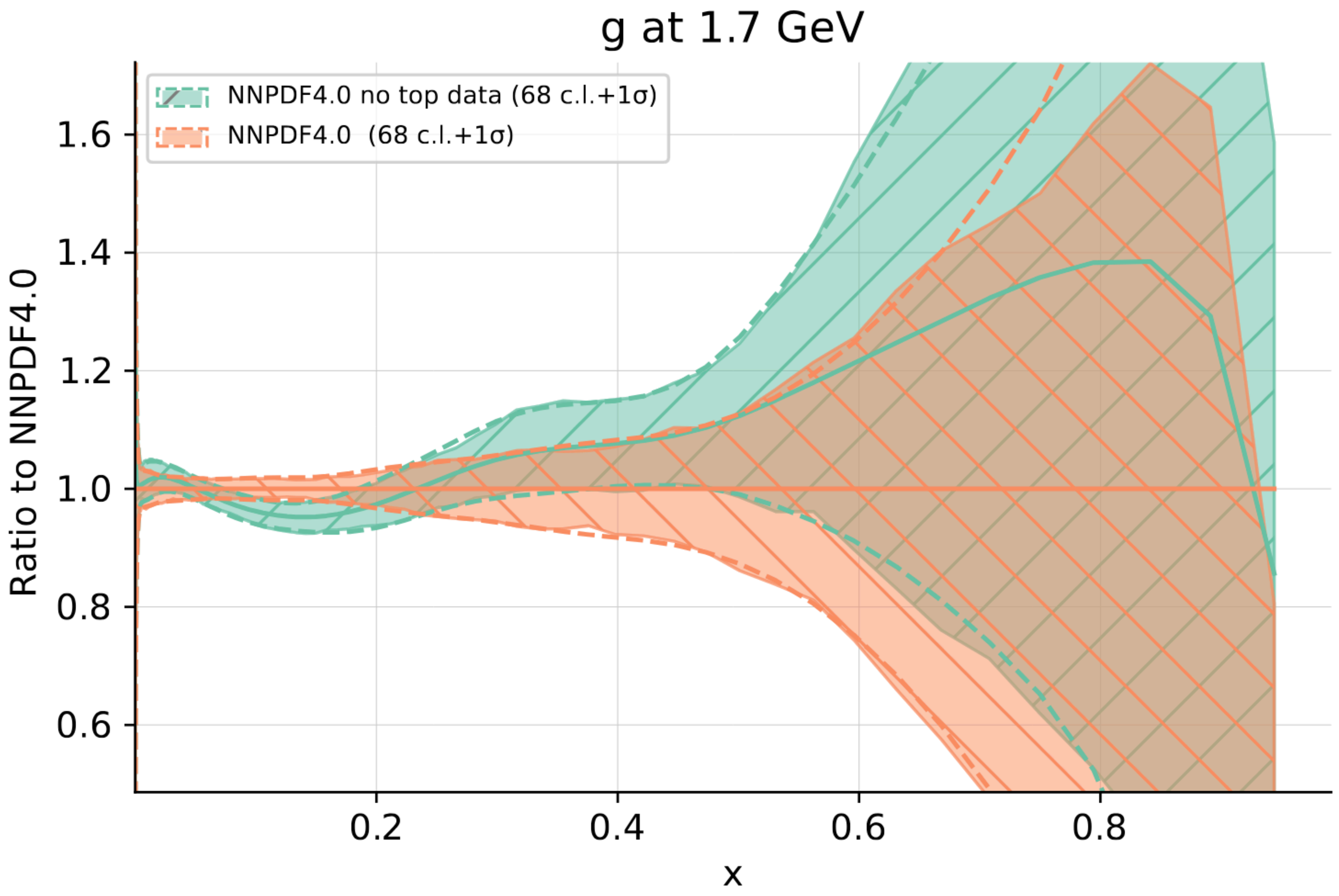
	NNPDF4.0	NNPDF4.0 w/o top-pair
Chi2	1.16022	1.16256

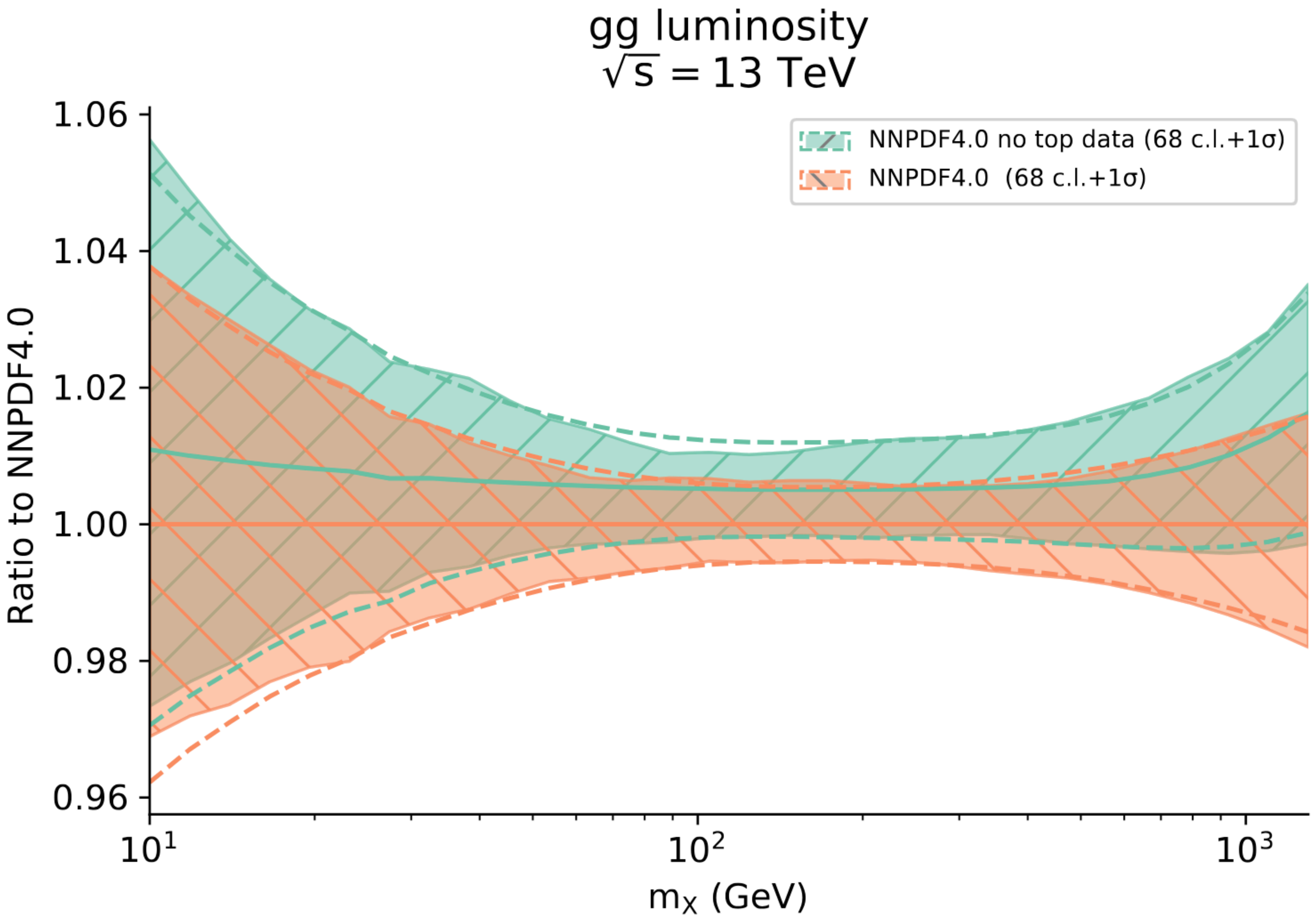




Impacts of removing all Top data

- ❖ Removing altogether all Top data results in the **suppression of the gluon PDF at medium-x (~0.1)** and a stronger enhancement at large-x — Major changes are only seen in the gluon while the quark PDFs remain similar
- ❖ This results in some enhancements in the gg-luminosity with a slight increase of the uncertainties





Outline of the Talk

How well do PDFs accommodate new Top data?

- Test PDF sets against new precise measurements from Run I/II using NNLO theories (w/o K-factors)
- Aim to assess how well PDF sets describe unseen data and whether these data will have effects on fits
- Agreement between data and theoretical predictions are quantified using the χ^2 definition:

$$\chi^2 = \frac{1}{n_{\rm dat}} \sum_{i,j=1}^{n_{\rm dat}} \left(T_i^{(0)} - D_i\right) \left(\mathrm{cov}^{-1}\right)_{ij} \left(T_j^{(0)} - D_j\right)$$

- Contributions to the covariance matrix include all sources of theoretical uncertainties:

$$\mathrm{cov}_{ij} = \left(\mathrm{cov}_{\rm exp}\right)_{ij} + \left(\mathrm{cov}_{\rm th}\right)_{ij}$$

$$\left(\mathrm{cov}_{\rm th}\right)_{ij} = \left(\mathrm{cov}_{\rm mho}\right)_{ij} + \left(\mathrm{cov}_{\rm pdf}\right)_{ij} + \left(\mathrm{cov}_{\rm as}\right)_{ij}$$

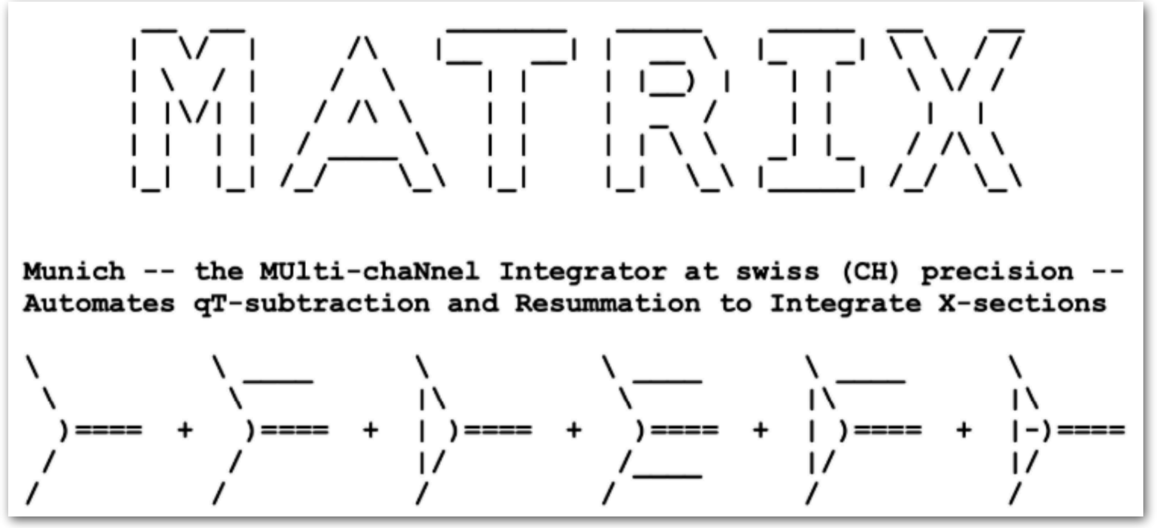
$$\left(\mathrm{cov}_{\rm pdf}^{\rm HES}\right)_{ij} = \sum_{k=1}^{n_{\rm eig}} \left(T_i^{(k)} - T_i^{(0)}\right)\left(T_j^{(k)} - T_j^{(0)}\right)$$

$$\left(\mathrm{cov}_{\rm pdf}^{\rm MC}\right)_{ij} = \frac{1}{n_{\rm rep}} \sum_{k=1}^{n_{\rm rep}} \left(T_i^{(k)} - \langle T_i \rangle_{\rm rep}\right)\left(T_j^{(k)} - \langle T_j \rangle_{\rm rep}\right)$$

PineAPPL

https://github.com/NNPDF/pineappl

NNLOJET



https://matrix.hepforge.org/

https://ploughshare.web.cern.ch/ploughshare/

$$\left(\mathrm{cov}_{\rm mho}\right)_{ij} = \frac{1}{3}\left\{\Delta_i^{+0}\Delta_j^{+0} + \Delta_i^{-0}\Delta_j^{-0} + \Delta_i^{0+}\Delta_j^{0+} + \Delta_i^{0-}\Delta_j^{0-} + \Delta_i^{++}\Delta_j^{++} + \Delta_i^{--}\Delta_j^{--}\right\}$$

$$\Delta_i\left(\kappa_R, \kappa_F\right) = T_i\left(\mu_R = \kappa_R \mu_R^{(0)}, \mu_F = \kappa_R \mu_F^{(0)}\right) - T_i\left(\mu_R^{(0)}, \mu_F^{(0)}\right)$$

$$\Delta_i^{+0} = \Delta_i(2,1), \quad \Delta_i^{-0} = \Delta_i(1/2,1), \quad \Delta_i^{0+} = \Delta_i(1,1/2)$$
$$\Delta_i^{0-} = \Delta_i(1,1/2), \quad \Delta_i^{++} = \Delta_i(2,2), \quad \Delta_i^{--} = \Delta_i(1/2,1/2)$$

$$\left(\mathrm{cov}_{\rm as}\right)_{ij} = \frac{1}{2}\left\{\Delta_{i,\alpha_s}^{+}\Delta_{j,\alpha_s}^{+} + \Delta_{i,\alpha_s}^{-}\Delta_{j,\alpha_s}^{-}\right\}$$

$$\Delta_{i,\alpha_s}^{+} \equiv T_i\left(\alpha_s = 0.119\right) - T_i\left(\alpha_s = 0.118\right)$$

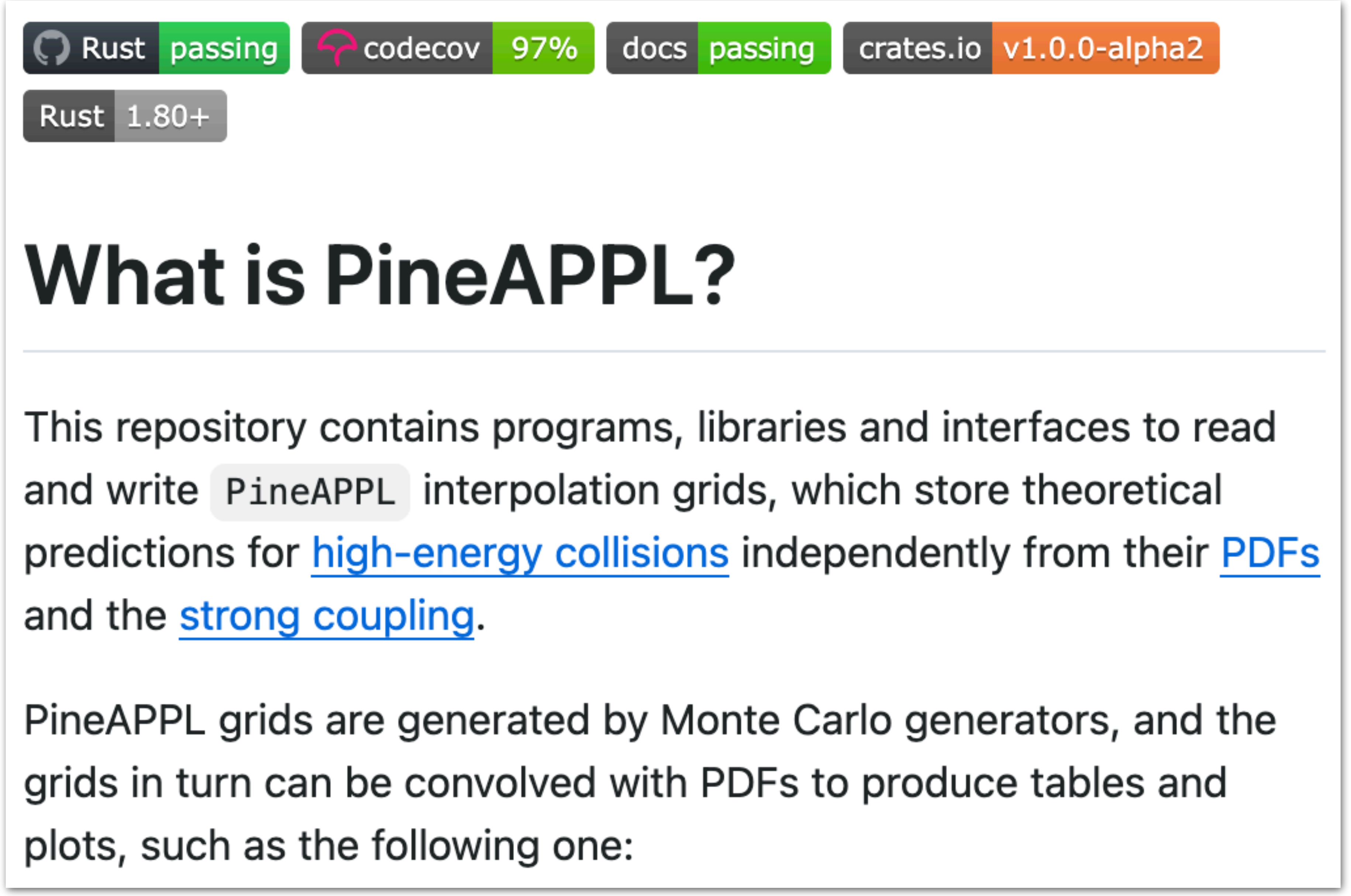
$$\Delta_{i,\alpha_s}^{-} \equiv T_i\left(\alpha_s = 0.118\right) - T_i\left(\alpha_s = 0.117\right)$$

Theoretical Predictions

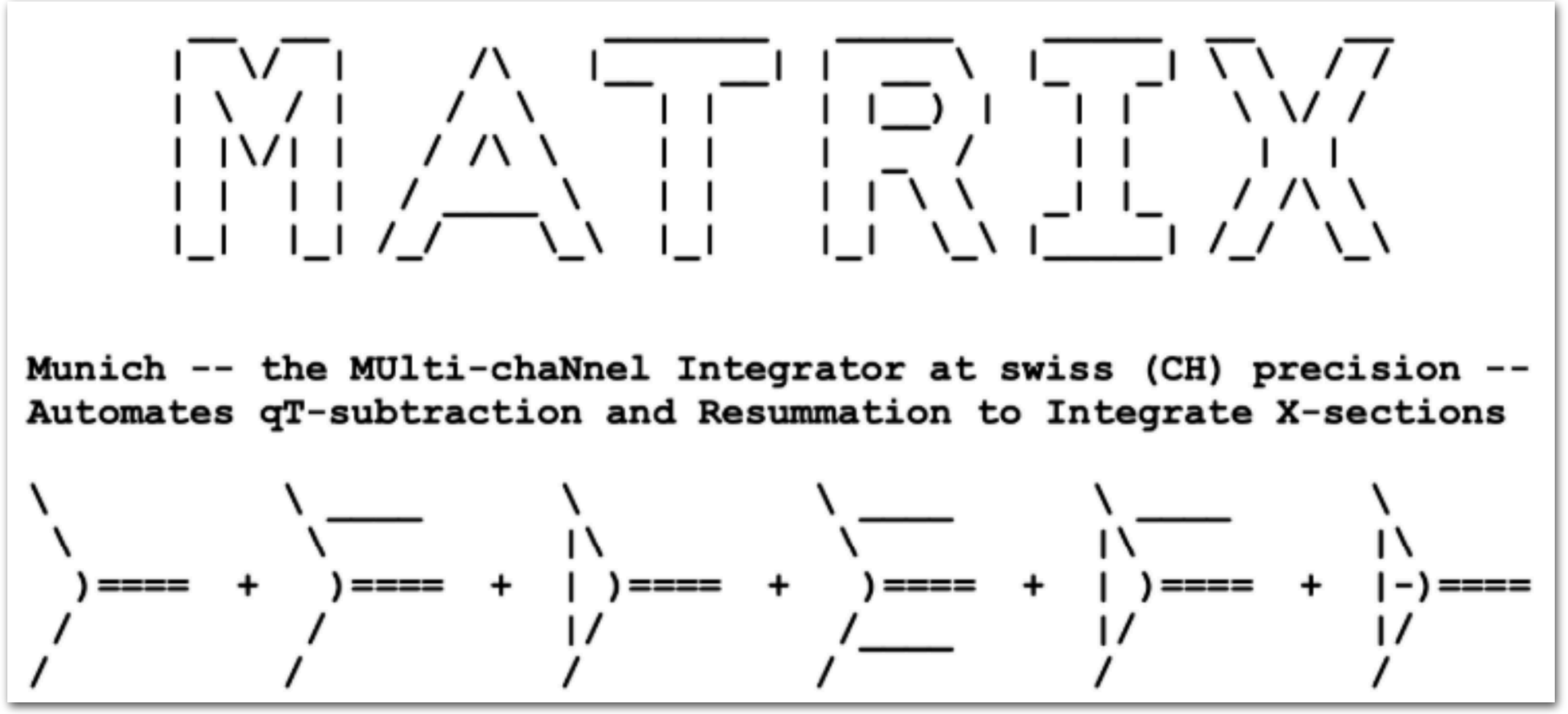
- Theory predictions are **accurate to NNLO QCD** $\Longleftrightarrow$ precision measurements require precise theory predictions
- Predictions are computed using MATRIX interfaced to PineAPPL fast interpolation grids
- $m_t^{\text{pole}} = 172.5$ GeV is used for the top-quark pole mass (consistent with latest PDG average)
- Computations have been benchmarked whenever possible with FastNLO tables

$$\mu_R = \mu_F = \frac{H_T}{2} = \frac{\sqrt{m_t^2 + (p_T^t)^2}}{2}$$

- Electroweak, QED, and photon-induced cross-sections are **not included**



https://github.com/NNPDF/pineappl

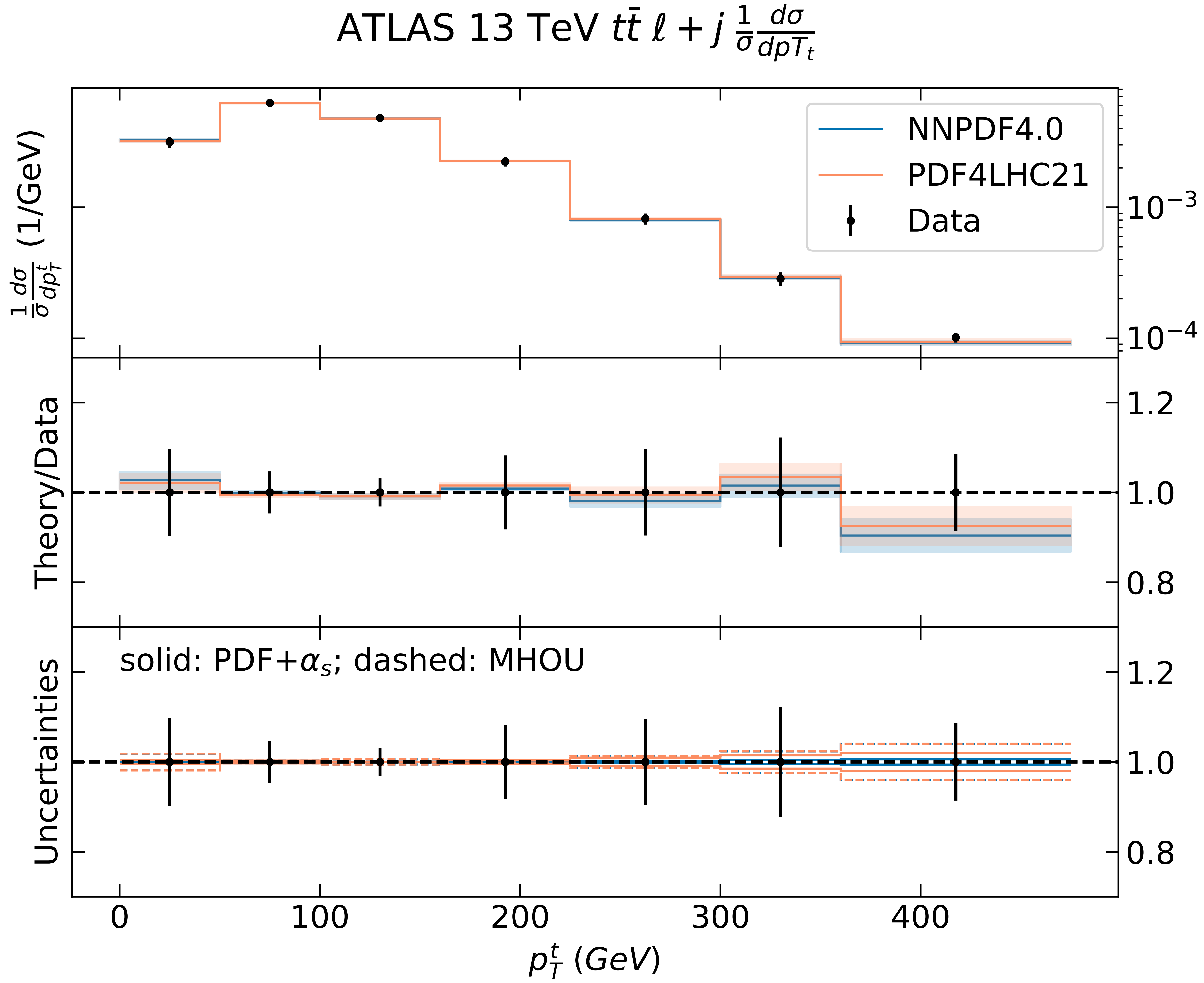


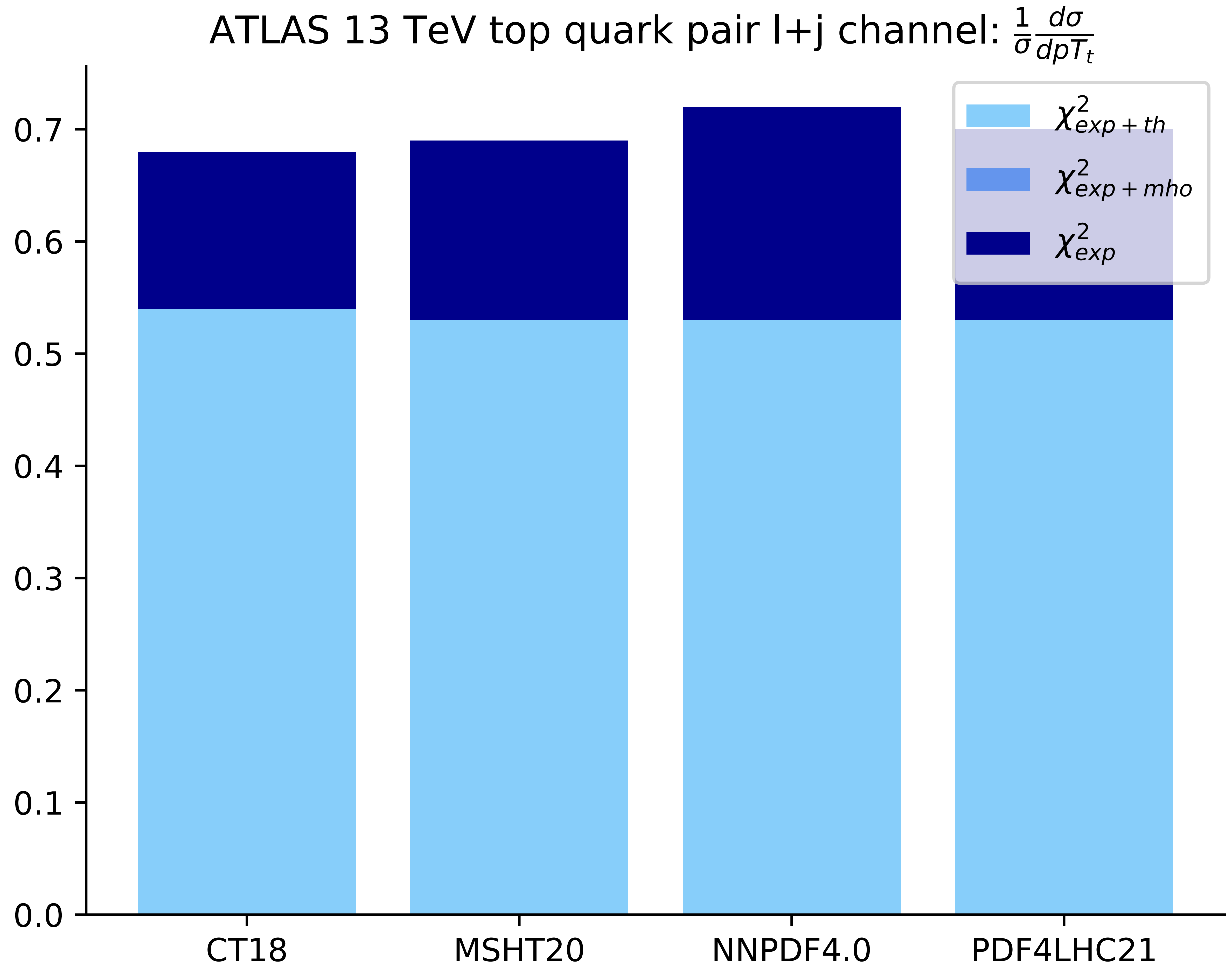
Top data included in the study

- **Dataset Criteria:** Not included in NNPDF4.0 and Publicly available on HepData
- Provide info on PDFs of # partons & computable @ NNLO interfaced to PineAPPL fast interpolation grids

Process	Experiment	Final State	Observable	$\sqrt{s}$ (TeV)	$\mathcal{L}$ (fb^{-1})	$n_{\rm dat}$	Ref.
top-pair	ATLAS	all-hadronic	$\left(\frac{1}{\sigma}\right)\frac{d\sigma}{dm_{t\bar{t}}}$	13	36.1	9	[53]
			$\left(\frac{1}{\sigma}\right)\frac{d\sigma}{d\|y_{t\bar{t}}\|}$	13	36.1	12	[53]
			$\left(\frac{1}{\sigma}\right)\frac{d^2\sigma}{d\|y_{t\bar{t}}\|dm_{t\bar{t}}}$	13	36.1	11	[53]
	ATLAS	ℓ+jets	$\left(\frac{1}{\sigma}\right)\frac{d\sigma}{dm_{t\bar{t}}}$	13	36.1	9	[54]
			$\left(\frac{1}{\sigma}\right)\frac{d\sigma}{dp_T^t}$	13	36.1	8	[54]
			$\left(\frac{1}{\sigma}\right)\frac{d\sigma}{d\|y_t\|}$	13	36.1	5	[54]
			$\left(\frac{1}{\sigma}\right)\frac{d\sigma}{d\|y_{t\bar{t}}\|}$	13	36.1	7	[54]
	CMS	ℓ+jets	$\left(\frac{1}{\sigma}\right)\frac{d\sigma}{dm_{t\bar{t}}}$	13	137	15	[55]
			$\left(\frac{1}{\sigma}\right)\frac{d\sigma}{dp_T^t}$	13	137	16	[55]
			$\left(\frac{1}{\sigma}\right)\frac{d\sigma}{d\|y_{t\bar{t}}\|}$	13	137	10	[55]
			$\left(\frac{1}{\sigma}\right)\frac{d\sigma}{d\|y_t\|}$	13	137	11	[55]
			$\left(\frac{1}{\sigma}\right)\frac{d^2\sigma}{d\|y_{t\bar{t}}\|dm_{t\bar{t}}}$	13	137	35	[55]

Normalised p_T distributions at 13 TeV with $(\ell + j)$ from ATLAS





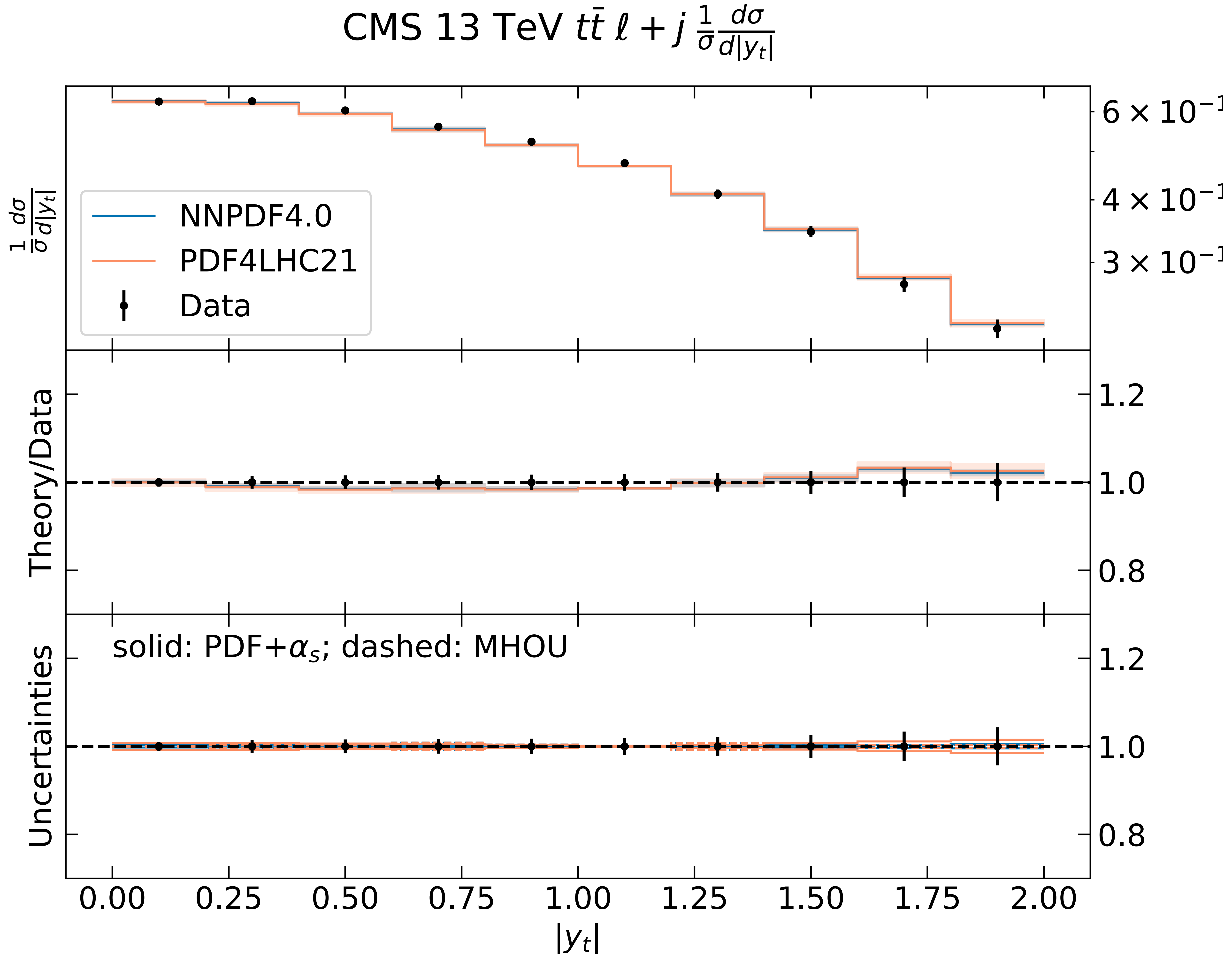
All PDF sets have similar predictive power

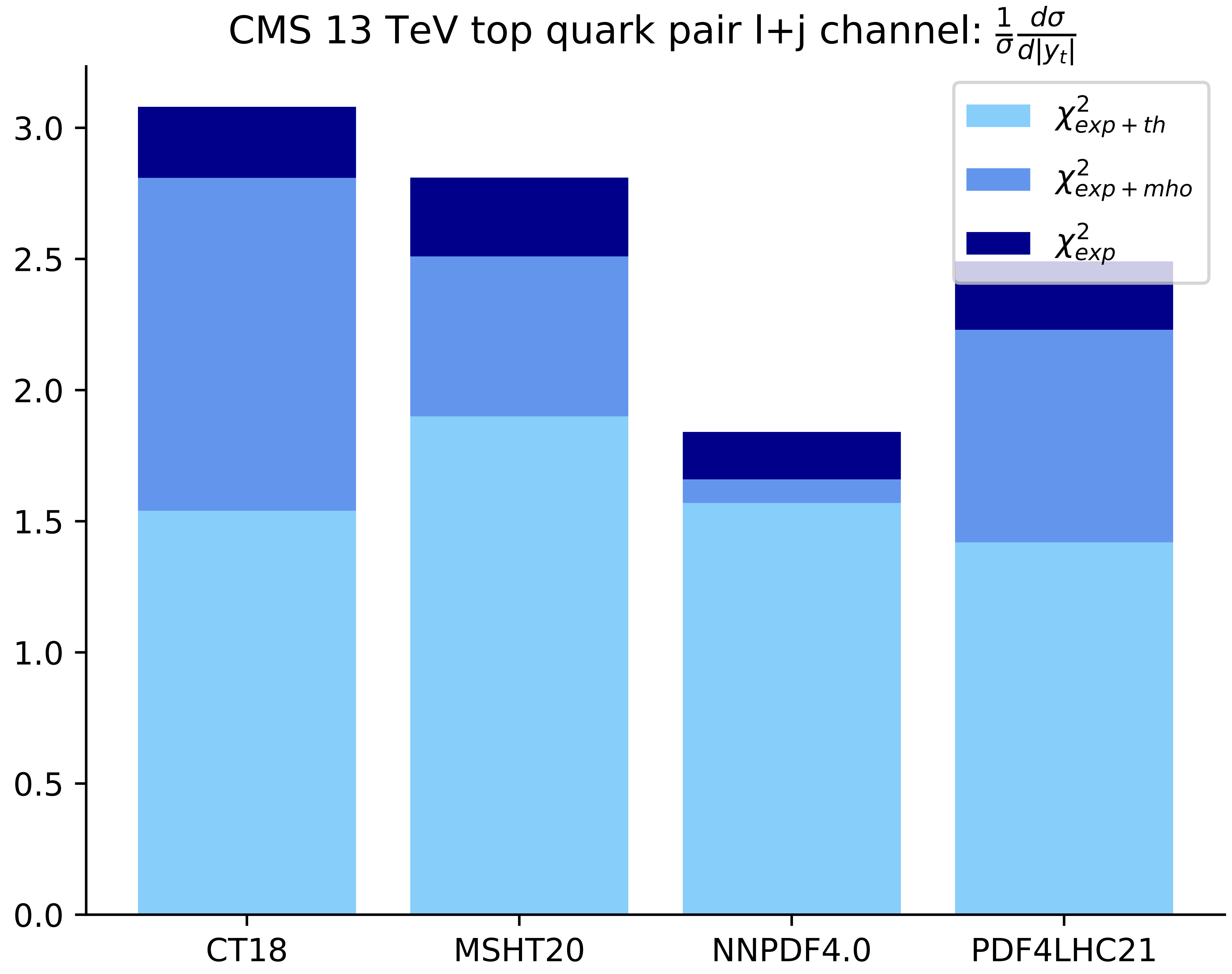
The inclusion of MHO uncertainties can have some impacts on the description of data

By taking into account all possible sources of uncertainties (MHO, PDF, α_s), the differences at the cross section level dissipate

LHC measurements do not strongly discriminate among PDF

Normalised y_t distributions at 13 TeV with $(\ell + j)$ from CMS





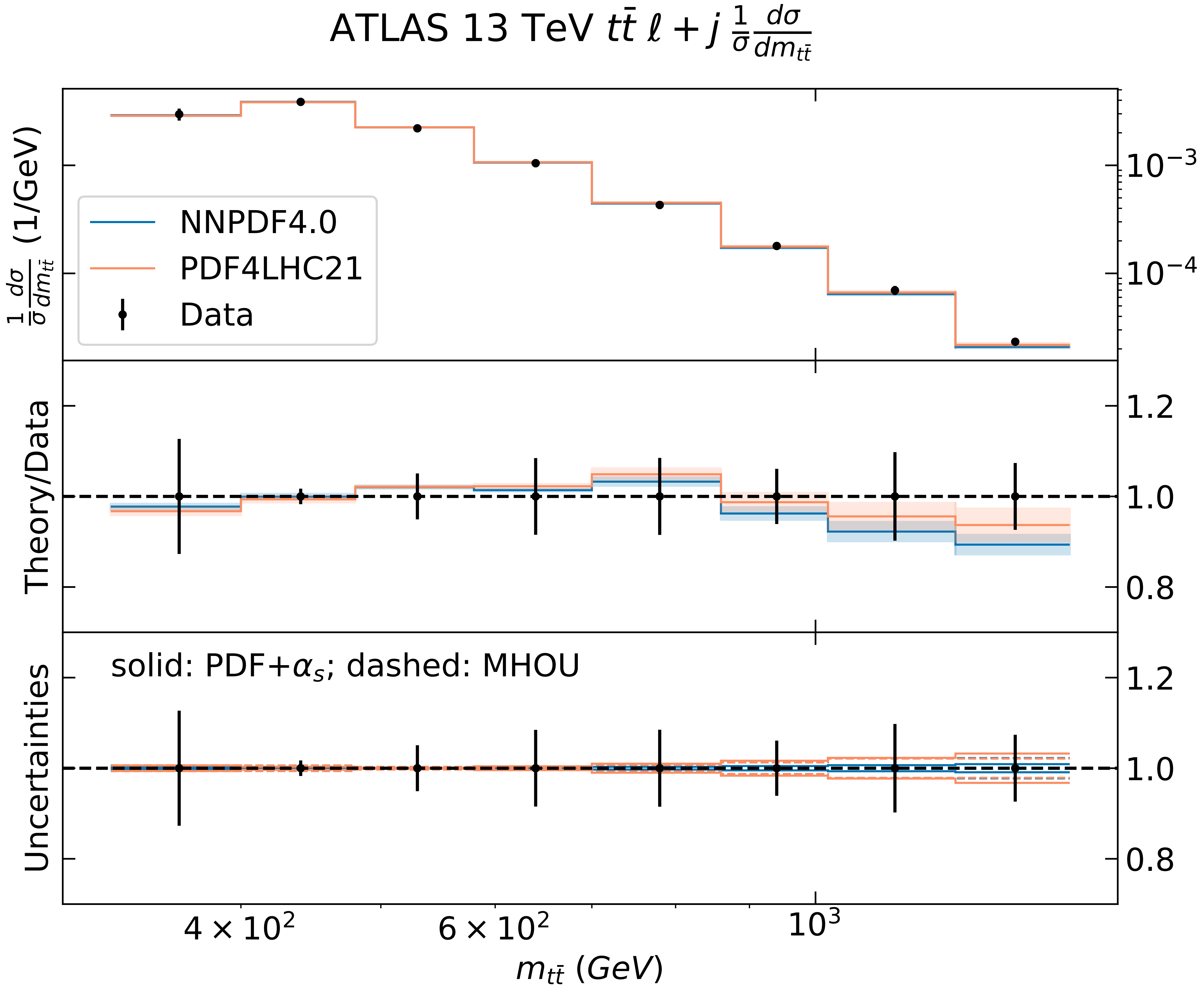
All PDF sets have similar predictive power

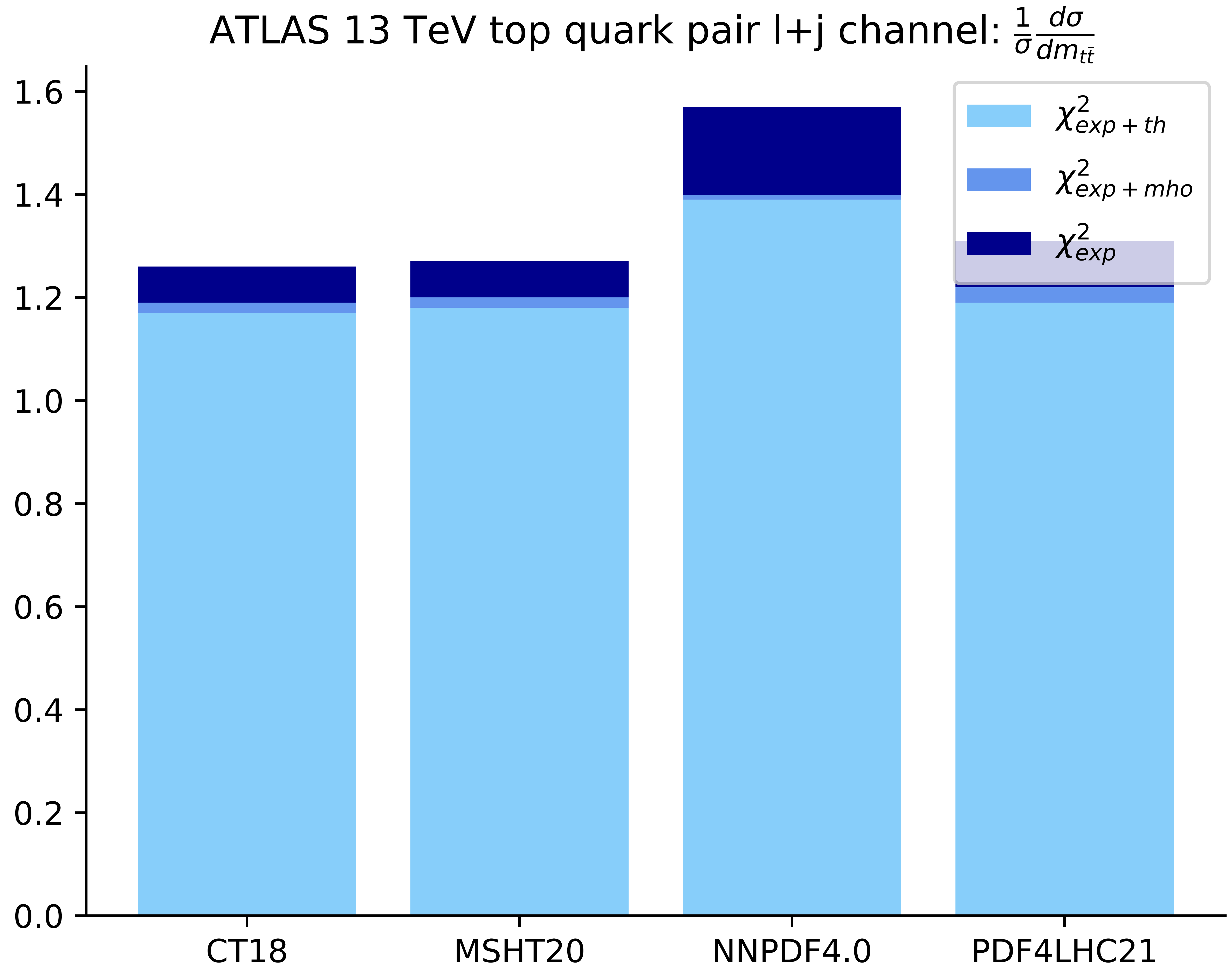
The inclusion of MHO uncertainties can have some impacts on the description of data

By taking into account all possible sources of uncertainties (MHO, PDF, α_s), the differences at the cross section level dissipate

LHC measurements do not strongly discriminate among PDF

Normalised $m_{t\bar{t}}$ distributions at 13 TeV with $(\ell + j)$ from ATLAS





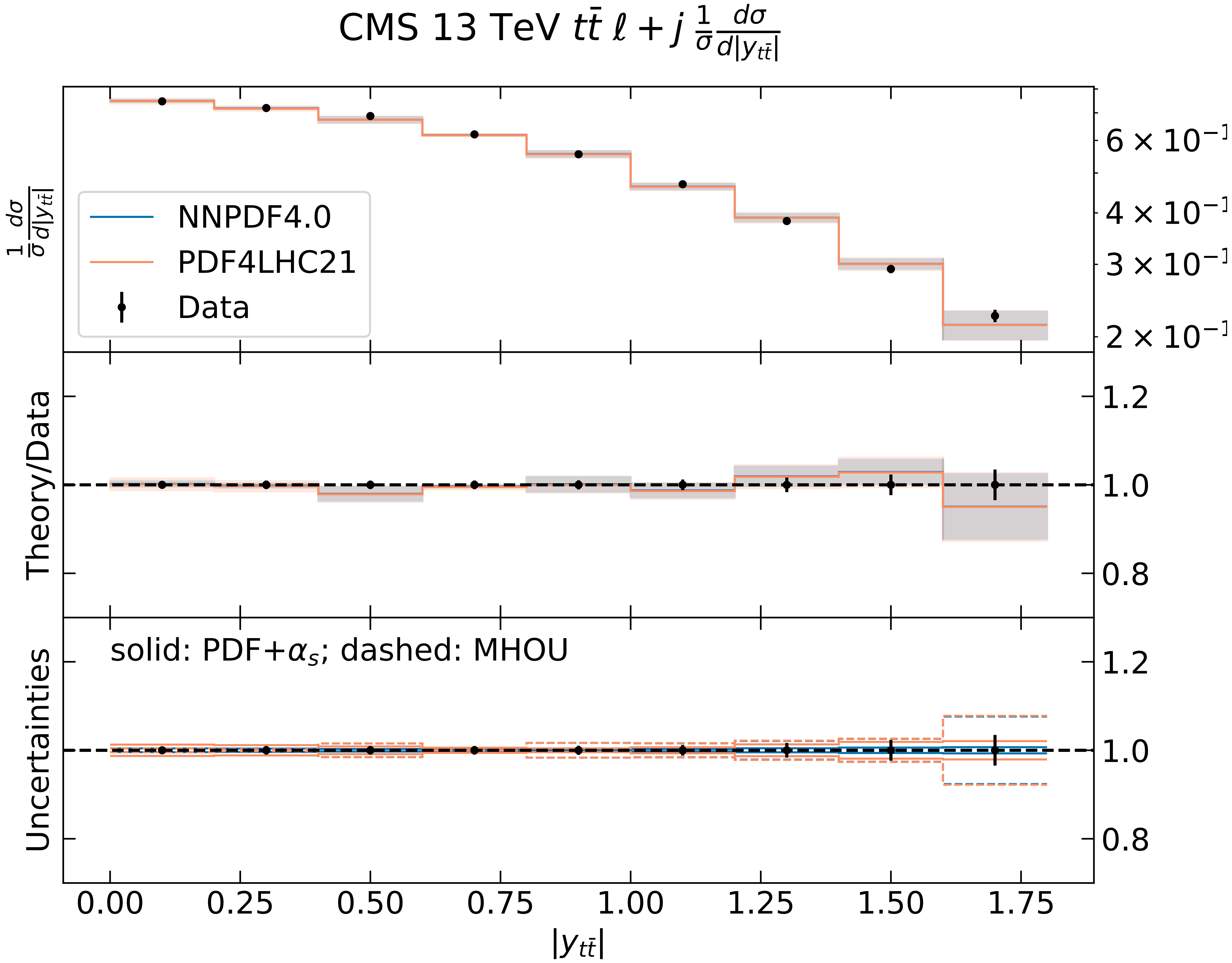
All PDF sets have similar predictive power

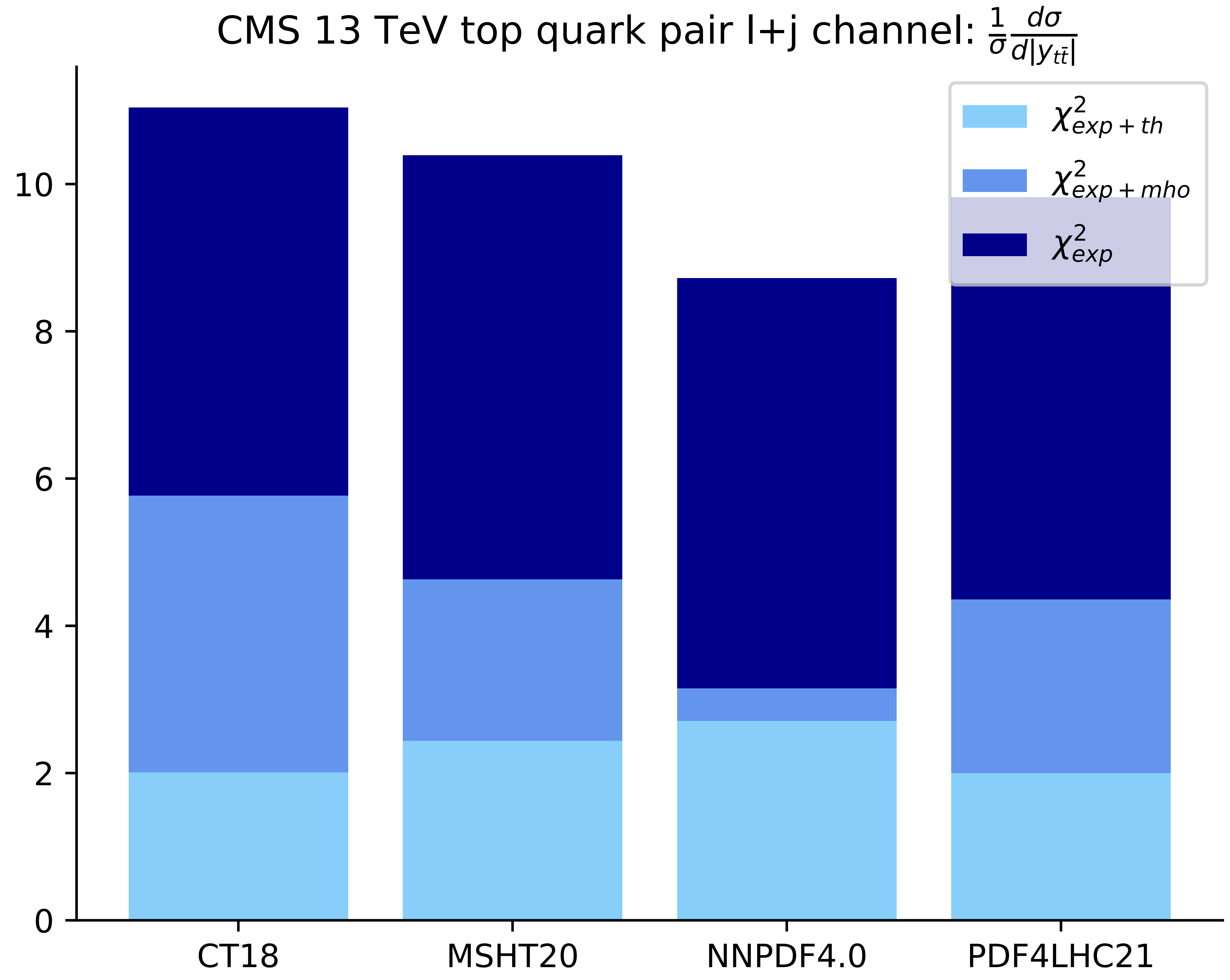
The inclusion of MHO uncertainties can have some impacts on the description of data

By taking into account all possible sources of uncertainties (MHO, PDF, α_s), the differences at the cross section level dissipate

LHC measurements do not strongly discriminate among PDF

Normalised $y_{t\bar{t}}$ distributions at 13 TeV with $(\ell + j)$ from CMS



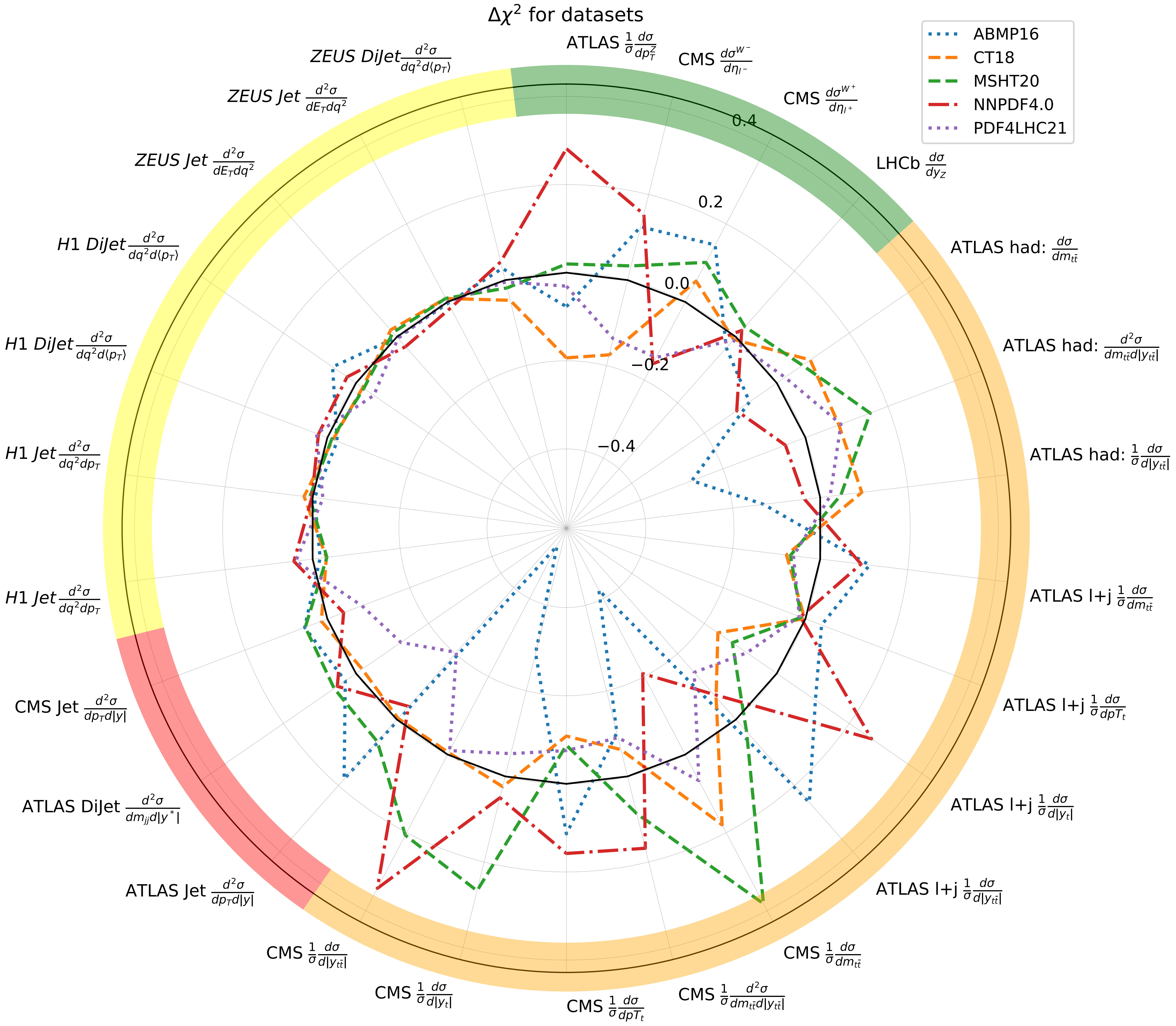


All PDF sets have similar predictive power

The inclusion of MHO uncertainties can have some impacts on the description of data

By taking into account all possible sources of uncertainties (MHO, PDF, α_s), the differences at the cross section level dissipate

LHC measurements do not strongly discriminate among PDF



$\Delta\chi^2$ Results

❖ Relative change in the total χ^2 due to a change in the input PDF

$$\Delta\chi^{2(i)} = \frac{\chi^{2(i)}_{\text{exp+th}} - \left\langle \chi^2_{\text{exp+th}} \right\rangle_{\text{pdfs}}}{\left\langle \chi^2_{\text{exp+th}} \right\rangle_{\text{pdfs}}}$$

where

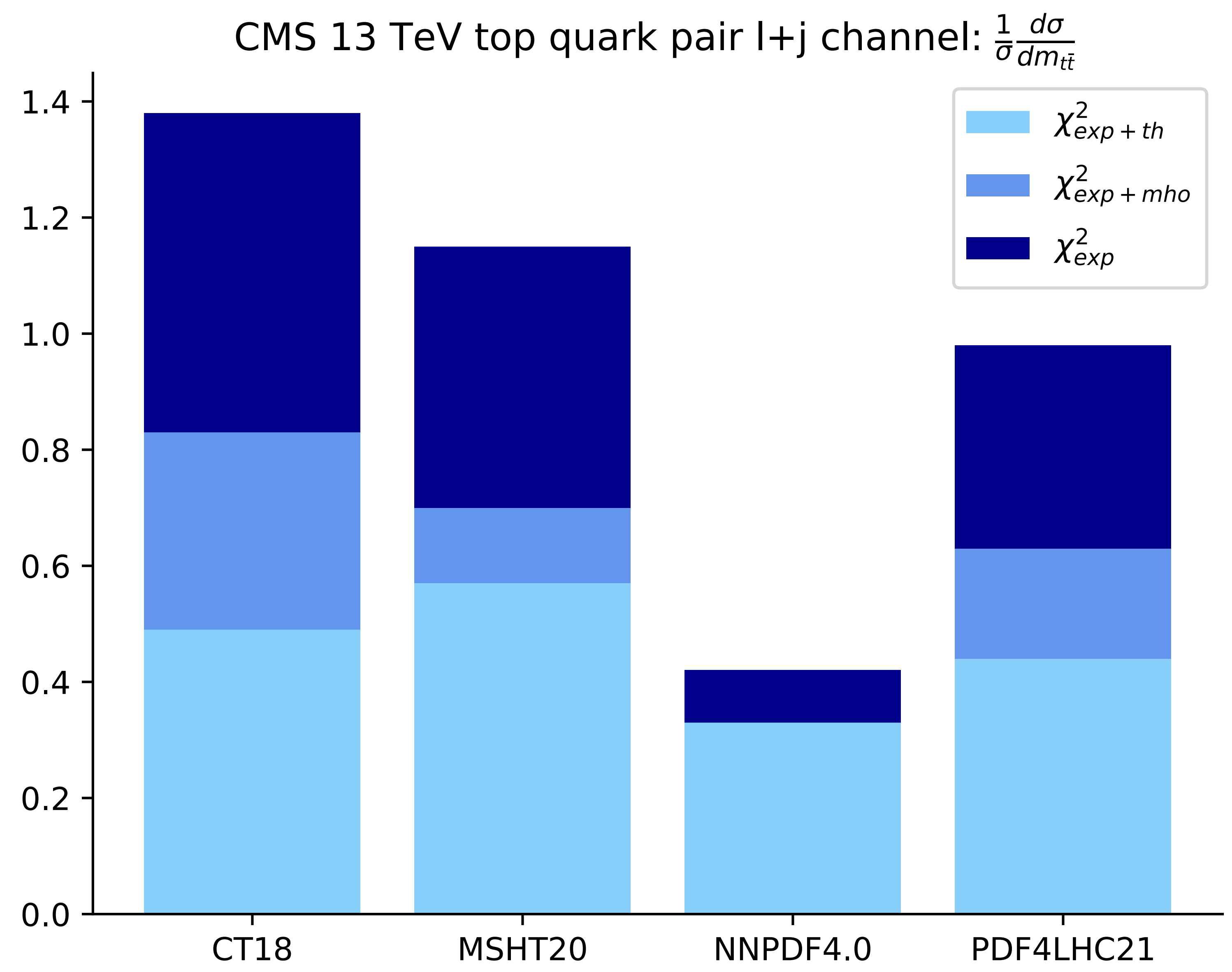
$$\left\langle \chi^2_{\text{exp+th}} \right\rangle_{\text{pdfs}} = \frac{1}{n_{\text{pdfs}}} \sum_{i=1}^{n_{\text{pdfs}}} \chi^{2(i)}_{\text{exp+th}}$$

❖ **No systematic outlier** seen in the data description despite noticeable differences at the level of PDF

❖ As anticipated, PDF4LHC21 represents the **average** (with $\Delta\chi^2 \sim 0$)

Outline of the Talk

Impacts of new Top data on PDFs



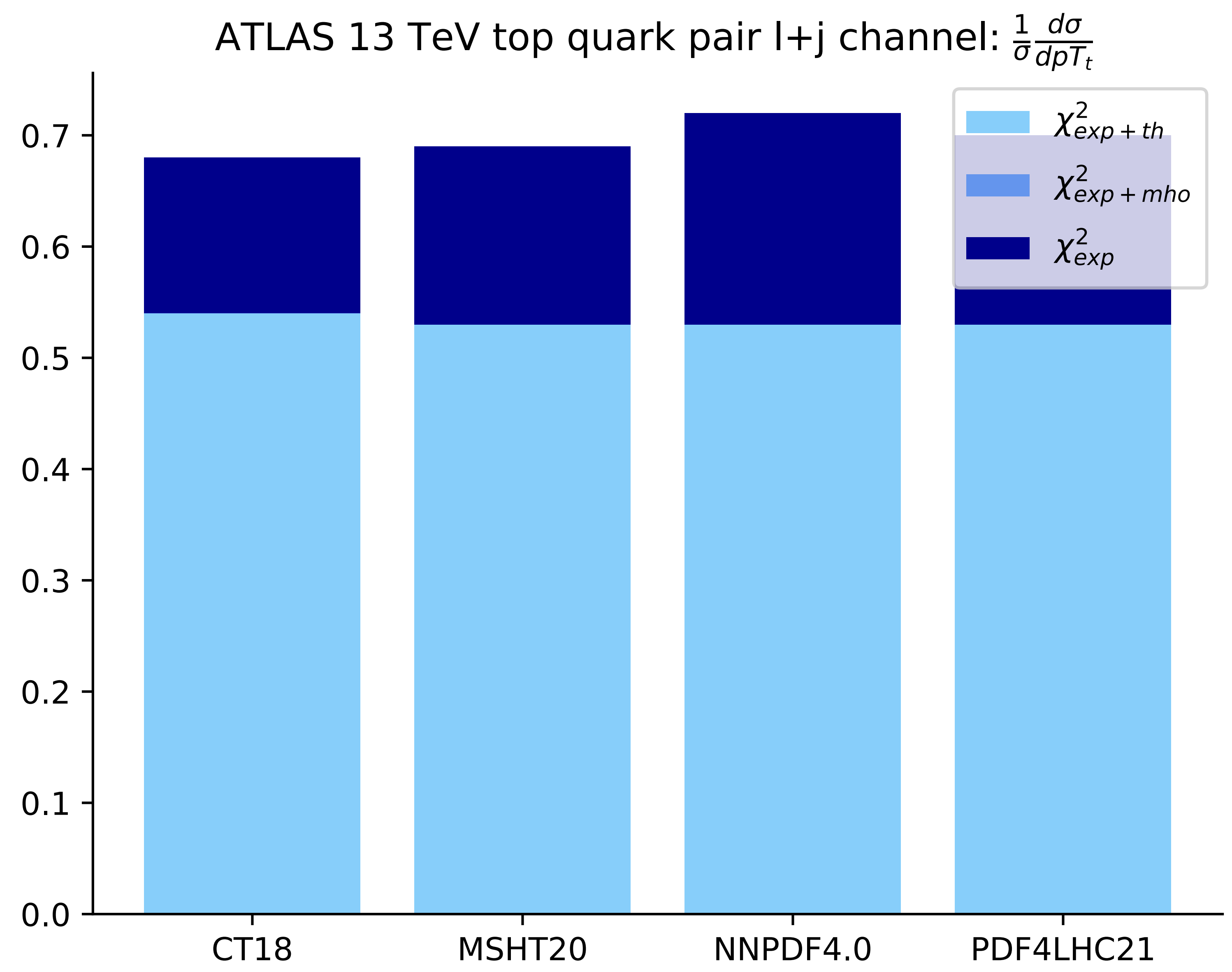
$1/\sigma\, d\sigma/dm_{t\bar{t}}$ — CMS at $\sqrt{s} = 13$ TeV

- $\chi^2(\text{NNPDF4.0} + 1/\sigma\, d\sigma/dm_{t\bar{t}}^{\text{CMS}}) = 1.12838$ vs. $\chi^2(\text{NNPDF4.0}) = 1.12832$
- $\chi^2(\text{data}) \;/\; N_{\text{dat}} = 0.4796 \iff$ consistent with the phenomenology predictions

Impacts of new Top data on PDFs

$1/\sigma\, d\sigma/dp_T$ — ATLAS at $\sqrt{s} = 13$ TeV

- $\chi^2(\text{NNPDF4.0} + 1/\sigma\, d\sigma/dp_T^{\text{ATLAS}}) = 1.12757$ vs. $\chi^2(\text{NNPDF4.0}) = 1.12832$
- $\chi^2(\text{data}) \,/\, N_{\text{dat}} = 0.7822 \iff$ consistent with the phenomenology predictions



Top Datasets under consideration for NNPDF4.1 (else?)

ATLAS			
Description	**c.m.e. (GeV)**	**L (fb^{-1})**	**Publication**
W, Z total cross section	13.6	0.03	2403.12902
Z pT differential cross section	13	36.1	1912.02844
top-pair differential distributions (all hadronic channel)	13	36.1	2006.09274
top-pair differential distributions (lepton+jets)	13	36.1	1908.07305
top-pair total cross section	13.6	29.0	2308.09529
top-pair total cross section	5.02	0.257	2207.01354
single-inclusive jet differential cross sections	13	3.2	1711.02692
di-jet differentail cross sections	13	3.2	1711.02692
W+c differential cross sections	13	139	2302.00336
prompt-photon differential cross sections	13	139	2302.00510
CMS			
Description	**c.m.e. (GeV)**	**L (fb^{-1})**	**Publication**
W differential cross sections	13	35.9	2008.04174
forward-backward asymmetry	13	139	2202.12327
Z pT differential cross section	13	35.9	2205.02872
top-pair differential cross sections (lepton+jets)	13	137	2108.02803
top-pair differential cross sections (dilepton)	13	137	2402.08486
top-pair total cross section	13	137	1812.10505
single-inclusive jet differential cross sections	13	36.3	2111.10431
di-jet differentail cross sections	13	36.3	2312.16669

LHCb			
Description	**c.m.e. (GeV)**	**L (fb^{-1})**	**Publication**
Z boson production differential cross section	13	5.1	2112.07458
top-pair inclusive cross sections	13	1.93	1803.05188
H1, ZEUS			
Description	**c.m.e. (GeV)**	**L (fb^{-1})**	**Publication**
H1 single-inclusive jet cross sections (low Q)	0.319	0.290	1611.03421
H1 single-inclusive jet cross sections (high Q)	0.319	0.351	1406.4709
ZEUS single-inclusive jet cross sections	0.300	0.038	hep-ex/0208037
ZEUS single-inclusive jet cross sections	0.319	0.082	hep-ex/0608048
H1 di-jet cross sections (low Q)	0.319	0.290	1611.03421
H1 di-jet cross sections (high Q)	0.319	0.351	1406.4709
ZEUS di-jet cross sections	0.319	0.374	1010.6167

Outline of the Talk

	Introduction & Motivations	
Part I	PDF Determination: the NNPDF Methodology	[arXiv:2109.02653; arXiv:2109.02671]
Part II	Top Data in NNPDF4.0	[arXiv:2109.02653; arXiv:2109.02671]
Part III	PDFs confront Top Data: A Quantitative Appraisal	[arXiv:2501.10359]
Part III	Impacts of new Top data on PDFs & NNPDF4.1	
	Conclusions & Outlook	

Conclusions & Outlook

- The precision era at the LHC requires precise & accurate PDFs determination and the top sector plays a major role
- (Di-)jets still place stronger constraints on the gluon PDF distribution
- A quantitative appraisal of PDF fits using precision Top measurements show that **all PDF sets have similar predictive power** despite significant differences at the PDF level
- Preliminary look at the impacts of new top data on PDFs shows consistent results as in the appraisal studies $\Longleftrightarrow$ no major impacts at the level of the PDFs
- Various (top) datasets are currently investigation for the new NNPDF4.1 release

THANKS FOR YOUR ATTENTION

“Wanderer above the Sea of Fog” by Caspar David Friedrich