MSHT20: Updates

Lucian Harland-Lang, University College London

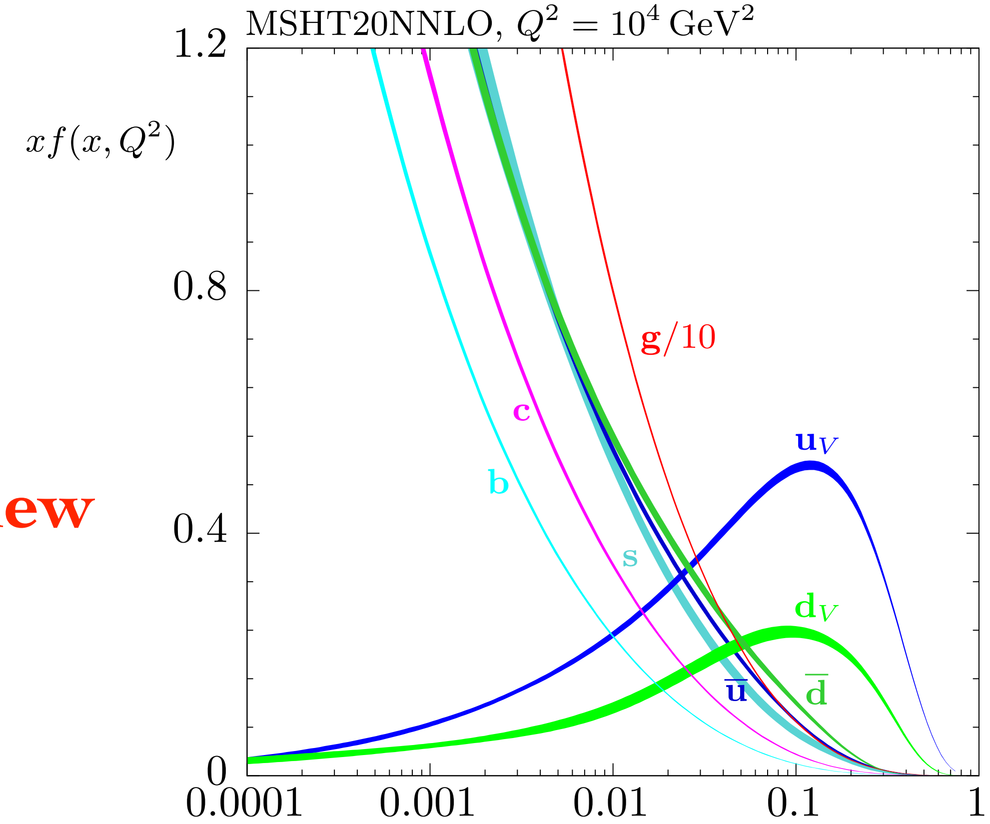
DIS 2023, MSU, 28 March 2023

In collaboration with Tom Cridge, Jamie McGowan and Robert Thorne

UCL

Outline

- The 'Post-Run I' set from the MSTW, MMHT… group: **MSHT20.**
- Focus on including significant amount of **new data**, higher **precision theory** and on **methodological improvements**.



See talk by T. Cridge

- More recent major update: extended to (approximate) N3LO order.
- Will discuss here on a selection of follow up studies, at both NNLO and aN3LO.
- **Main Focus**: analysis of jet and dijet data at NNLO and aN3LO.

Jets and Dijets in MSHT20

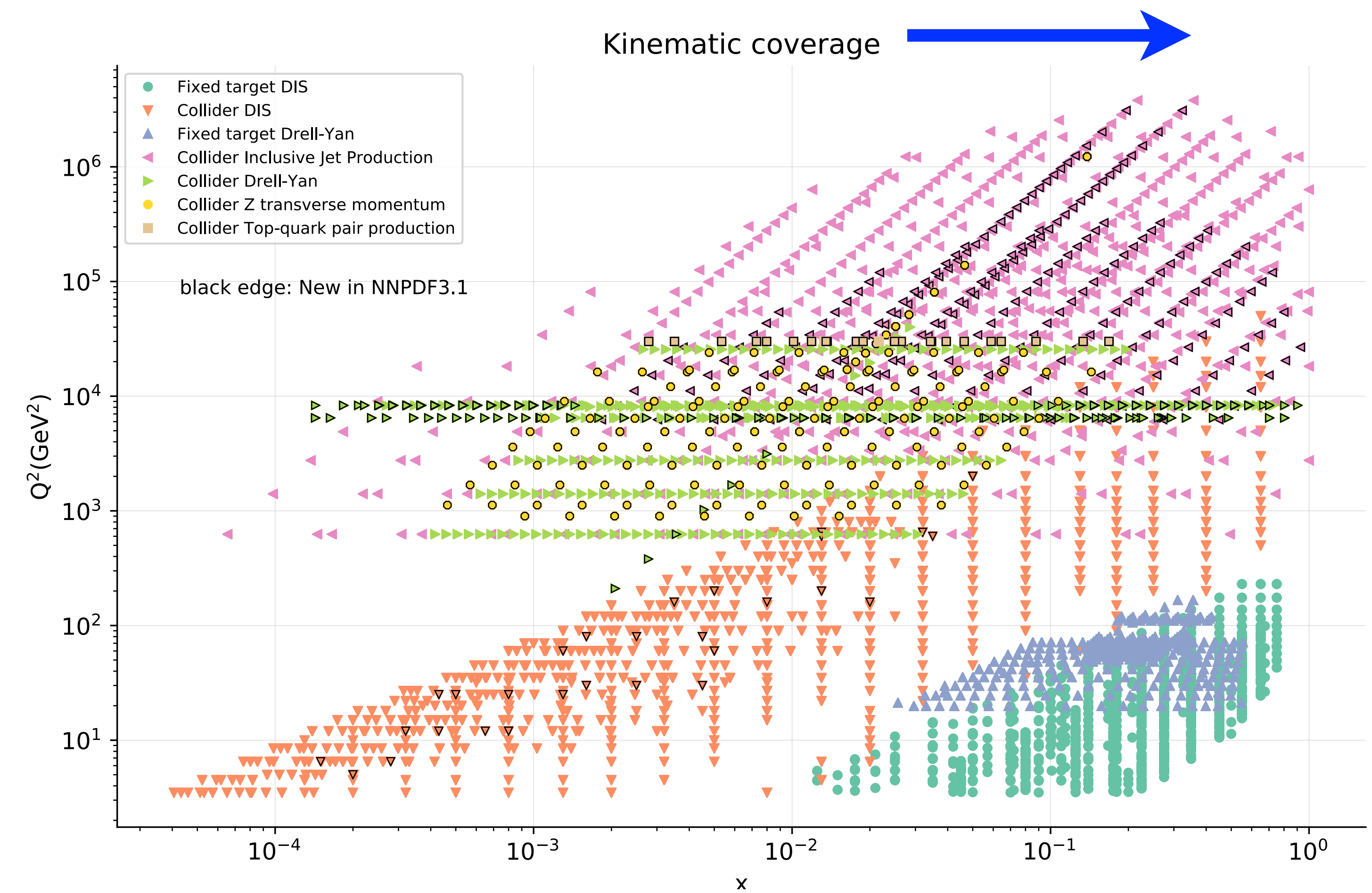
Jets for PDF fits

- Jet production a key ingredient in modern PDF fits.
- By pushing to larger jet $p_\perp$ (dijet m_{jj}) go to larger x.
- Quark-initiated contribution tends to better constrained $\longrightarrow$ particularly relevant for gluon at high x.

High $p_\perp$, m_{jj}

- NNLO QCD (and NLO EW) theory available for both inclusive and dijet data.

- In addition, high precision LHC data available, spanning large range of kinematic space.



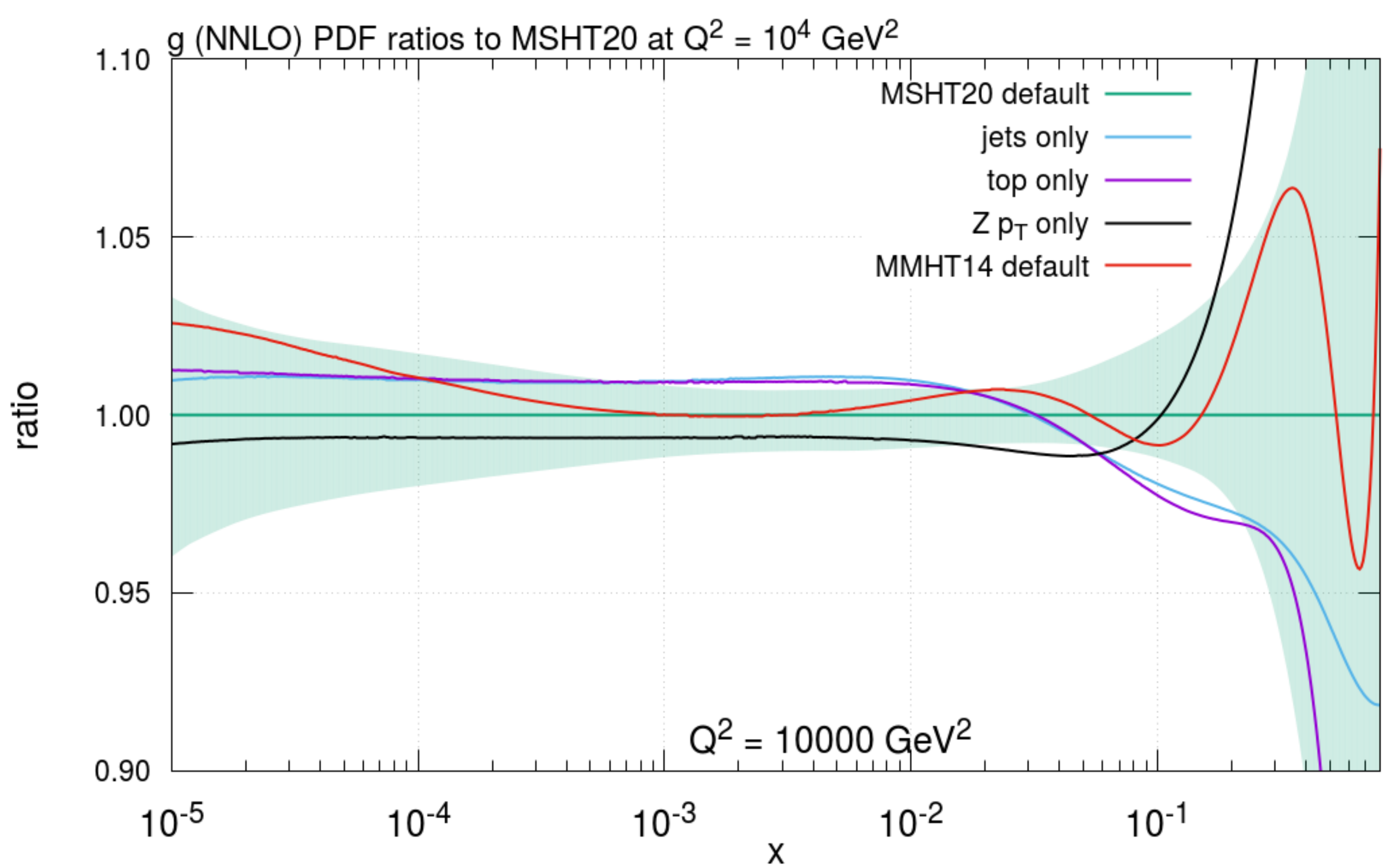
Jets in MSHT20

- Range of inclusive LHC jet dat fit:

	NNLO, χ^2/N_{pt}
ATLAS 7 TeV jets [18]	221.6/140
CMS 2.76 TeV jet [107]	102.9/81
CMS 7 TeV jets [100]	175.8/158
CMS 8 TeV jets [101]	261.3/174

- Fit quality acceptable. N.B. For ATLAS data smooth decorrelation of systematic errors applied.

- PDF impact tied up with other high x gluon sensitive data….



MSHT20 updates: Jet data

- Focussing on Run-I data (i.e. current PDF fits):

$\mathrm{d}^2\sigma/\mathrm{d}p_\perp \mathrm{d}y$

- Inclusive jets:

$0.0 < |y| < 2.5 - 3.0$

★ CMS 2.76 TeV: 81 points – $5.43\ \mathrm{pb}^{-1}$ – $74 < p_\perp < 592\,\mathrm{GeV}$

★ CMS 7 TeV: 158 points – $5.0\ \mathrm{fb}^{-1}$ – $74 < p_\perp < 2500\,\mathrm{GeV}$

★ CMS 8 TeV: 174 points – $19.7\ \mathrm{fb}^{-1}$ – $60 < p_\perp < 1300\,\mathrm{GeV}$

★ ATLAS 7 TeV: 140 points – $4.5\ \mathrm{fb}^{-1}$ – $100 < p_\perp < 2000\,\mathrm{GeV}$

★ **ATLAS 8 TeV**: 171 points – $20.2\ \mathrm{fb}^{-1}$ – $70 < p_\perp < 2500\,\mathrm{GeV}$

⟶ ***724*** points in total, v.s. ~ 4500 in global MSHT fit (inc.).

- We take the larger of the jet radii available in both cases, i.e. R=0.6/0.7.

- Dijets:

  ⋆ ATLAS 7 TeV: 90 points – $4.5\ \text{fb}^{-1}$ – $\mathrm{d}^2\sigma/\mathrm{d}m_{jj}\mathrm{d}|y_{\max}|$ $0.26 < m_{jj} < 5.04\,\text{TeV}$

  ⋆ CMS 7 TeV: 54 points – $5.0\ \text{fb}^{-1}$ – $\mathrm{d}^2\sigma/\mathrm{d}m_{jj}\mathrm{d}|y^*|$ $0.25 < m_{jj} < 4.48\,\text{TeV}$

  ⋆ CMS 8 TeV: 122 points – $19.7\ \text{fb}^{-1}$ – $\mathrm{d}^3\sigma/\mathrm{d}p_{\perp,avg}\mathrm{d}y_b\mathrm{d}y^*$ $143 < p_{\perp,avg} < 1638\,\text{GeV}$

$\longrightarrow$ **266** points in total, v.s. ~ 4000 in global MSHT fit (inc.).

- Again take the larger of the jet radii available in both cases, i.e. R=0.6/0.7.

- **CMS 8 TeV** data the only cases where this is **triple differential**. Only case where LO kinematics specified $\Rightarrow$ higher impact (backup).

Fit Quality

All results preliminary

- Consider impact of both inclusive or dijet data at NNLO and aN3LO in the MSHT20 fit.

Jet fit:

	$N_{\rm pts}$	NNLO	aN^{3}LO
ATLAS 7 TeV jets	140	1.54	1.46
CMS 7 TeV jets	158	1.29	1.32
ATLAS 8 TeV jets	171	1.96	1.90
CMS 8 TeV jets*	174	1.83	1.80
Total Jets	643	1.67	1.63

NB: smooth decorrelation of systematics applied for ATLAS inclusive jet data.

$\chi^2/N_{\rm pts}$

Dijet fit:

	$N_{\rm pts}$	NNLO	aN^{3}LO
ATLAS 7 TeV dijets	90	1.06	1.12
CMS 7 TeV dijets	54	1.43	1.39
CMS 8 TeV dijets	122	1.05	0.82
Total Dijets	266	1.13	1.04

★ NNLO: Fit quality to dijet data very good (1.13), clearly worse for jets (1.67).

★ aN3LO: Some improvement in both cases (1.13, 1.63 for jets, dijets) but inclusive jet remains a rather bad fit!

Dijets ✓ **Jets** ✗

*NB we use stat. correlations here. Not included by other groups, and leads to deterioration in fit quality.

- What about interplay with other gluon sensitive data?

Jet fit:

	$N_{\rm pts}$	NNLO	aN^{3}LO
ATLAS $Z\ p_\perp$	104	1.89	1.03
Diff. top	54	1.10	1.06
7 + 8 TeV dijets	266	[1.30]	[1.10]
7 + 8 TeV jets	643	1.67	1.63

Dijet fit:

	$N_{\rm pts}$	NNLO	aN^{3}LO
ATLAS $Z\ p_\perp$	104	1.66	1.05
Diff. top	54	1.26	1.09
7 + 8 TeV jets	643	[1.75]	[1.65]
7 + 8 TeV dijets	266	1.13	1.04

★ Jet data: no signs of significant inconsistency in fit vs. prediction though some difference in pull implied.

★ NNLO: Fit quality to top ($Z\ p_\perp$) data better in jet (dijet) fit. Latter particularly notable $\Rightarrow$ overall tension less in dijet fit.

★ aN3LO: tensions reduced in all cases. No clear difference between jet/dijets.

★ (Not shown) - fit quality to other data in global fit v. similar.

Dijets ✔ **Jets** ✘

With some preference for aN3LO

Impact of EW corrections

Jet fit:

NNLO: χ^2 (no EW) $\to \chi^2$ (EW) : $1.57 \to 1.67$

aN3LO: χ^2 (no EW) $\to \chi^2$ (EW): $1.59 \to 1.63$

Dijet fit:

NNLO: χ^2 (no EW) $\to \chi^2$ (EW) : $1.37 \to 1.13$

aN3LO: χ^2 (no EW) $\to \chi^2$ (EW) : $1.27 \to 1.04$

- ★ Significant improvement in dijet fit upon including EW corrections. However trend is opposite for inclusive jets (!). Given these are there:

  Dijets ✓ **Jets** ✗

  indeed even absent EW correction dijet fit quality is better.

- ★ Remains true at aN3LO. Deterioration in fit quality for no EW fit somewhat improved but not entirely $\Rightarrow$ not true that freedom in aN3LO K-factors can (fully) absorb other theoretical deficiencies.

Impact on PDFs small: Backup

Inclusive Jets: scale choice

J. Currie et al., *JHEP* 10 (2018) 155

- Default inclusive fits taken with $\mu = p_\perp^j$ scale choice. However some indication that $\mu = \hat{H}_\perp$ may be preferable.

$$\hat{H}_\perp = \sum_i p_{i\perp}$$

- What does global fit say?

NLO: $$\chi^2(p_\perp^j) \to \chi^2(\hat{H}_\perp) : 1.68 \to 1.60$$

NNLO: $$\chi^2(p_\perp^j) \to \chi^2(\hat{H}_\perp) : 1.64 \to 1.65$$

aN3LO: $$\chi^2(p_\perp^j) \to \chi^2(\hat{H}_\perp) : 1.58 \to 1.60$$

★ NLO fit quality better with $\mu = \hat{H}_\perp$ but difference marginal at NNLO/aN3LO.

★ Trend for improved description with order not present with $\mu = \hat{H}_\perp$.

$\longrightarrow$ Scale choice does not appear to play significant role at NNLO and beyond.

Impact on PDFs small: Backup

NB. Cuts for CMS 7 jet data slightly different here

Taking step back: pQCD working?

- Worth taking a look at **NLO** fit quality…

Jets fit:

	$N_{\rm pts}$	NLO	NNLO	aN^{3}LO
ATLAS 7 TeV jets	140	1.61	1.54	1.46
CMS 7 TeV jets	158	1.37	1.29	1.32
ATLAS 8 TeV jets	171	2.24	1.96	1.90
CMS 8 TeV jets	174	1.66	1.83	1.80
Total Jets	643	1.73	1.67	1.63

Dijets fit:

	$N_{\rm pts}$	NLO	NNLO	aN^{3}LO
ATLAS 7 TeV dijets	90	1.12	1.06	1.12
CMS 7 TeV dijets	54	1.70	1.43	1.39
CMS 8 TeV dijets	122	5.27	1.05	0.82
Total Dijets	266	3.14	1.13	1.04

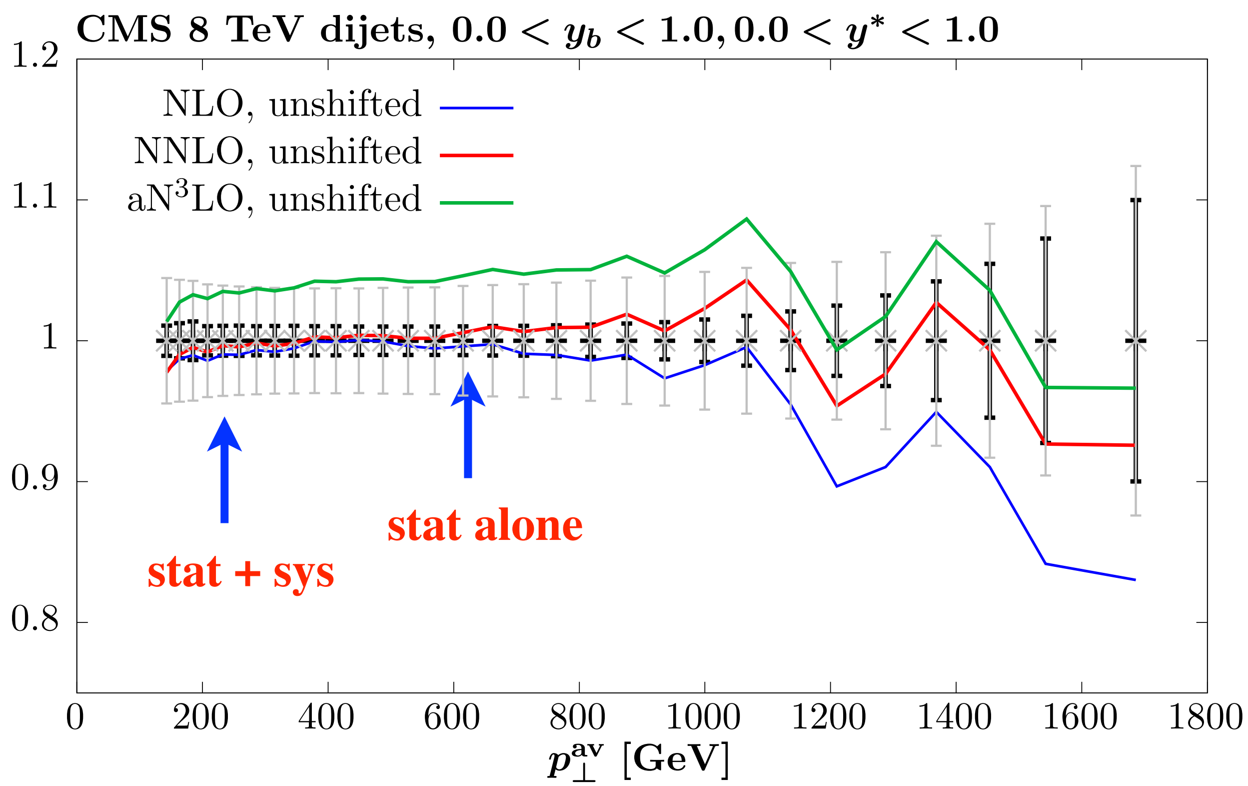
Not a typo!

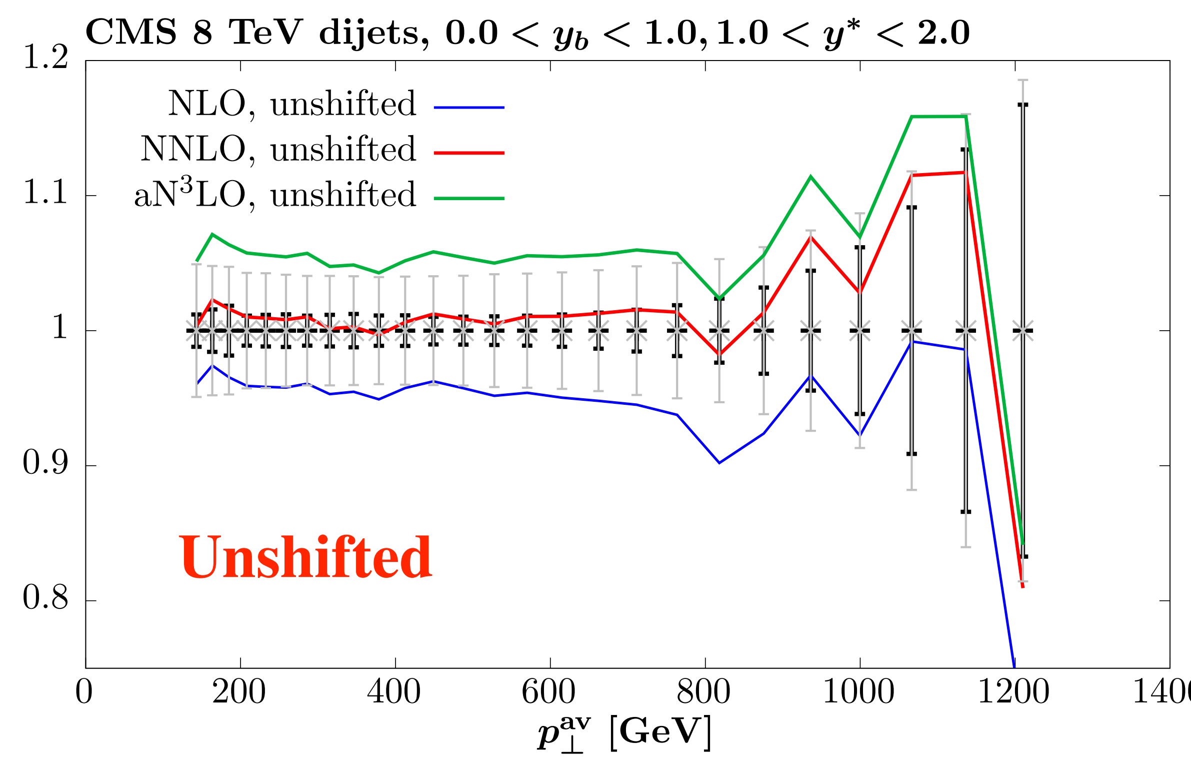
★ Clear trend in both cases for QCD corrections to improve fit quality. pQCD working as it should!

★ Improvement in CMS 8 TeV dijets particularly **remarkable**. Clear need for NNLO QCD at high precision + multi-differential LHC. In more detail…

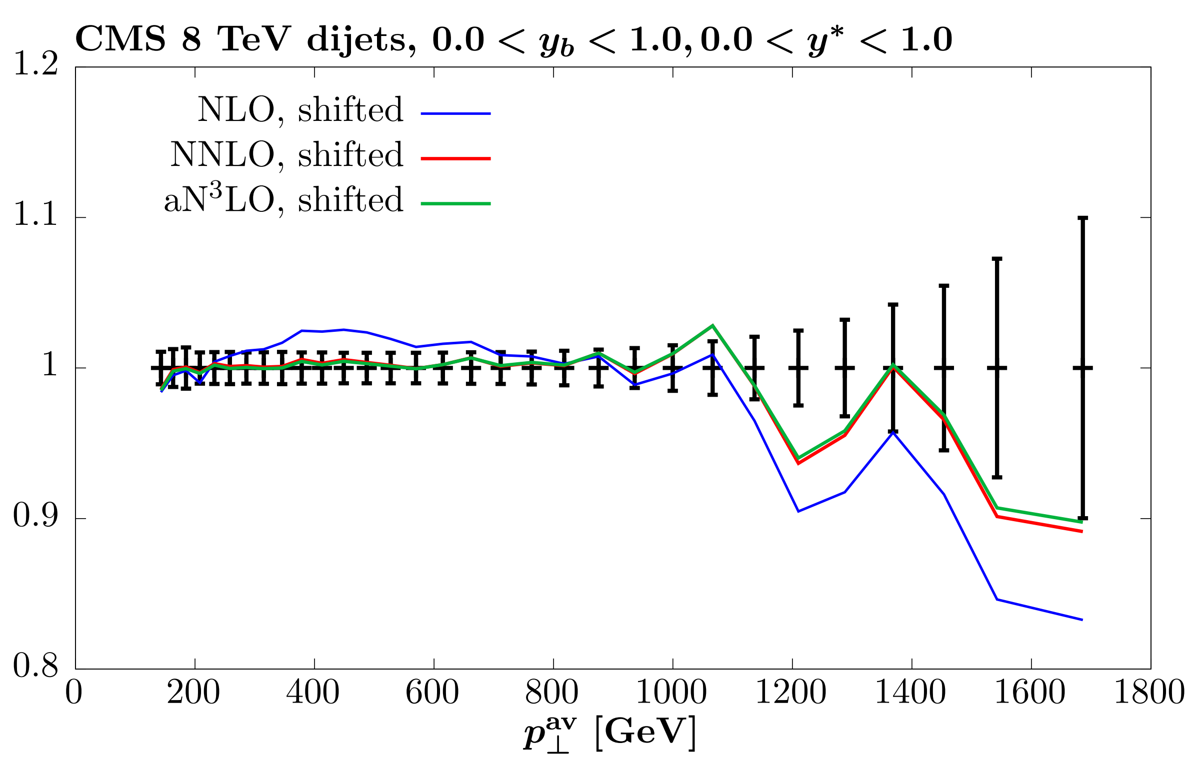
See also: ATLAS high precision W,Z

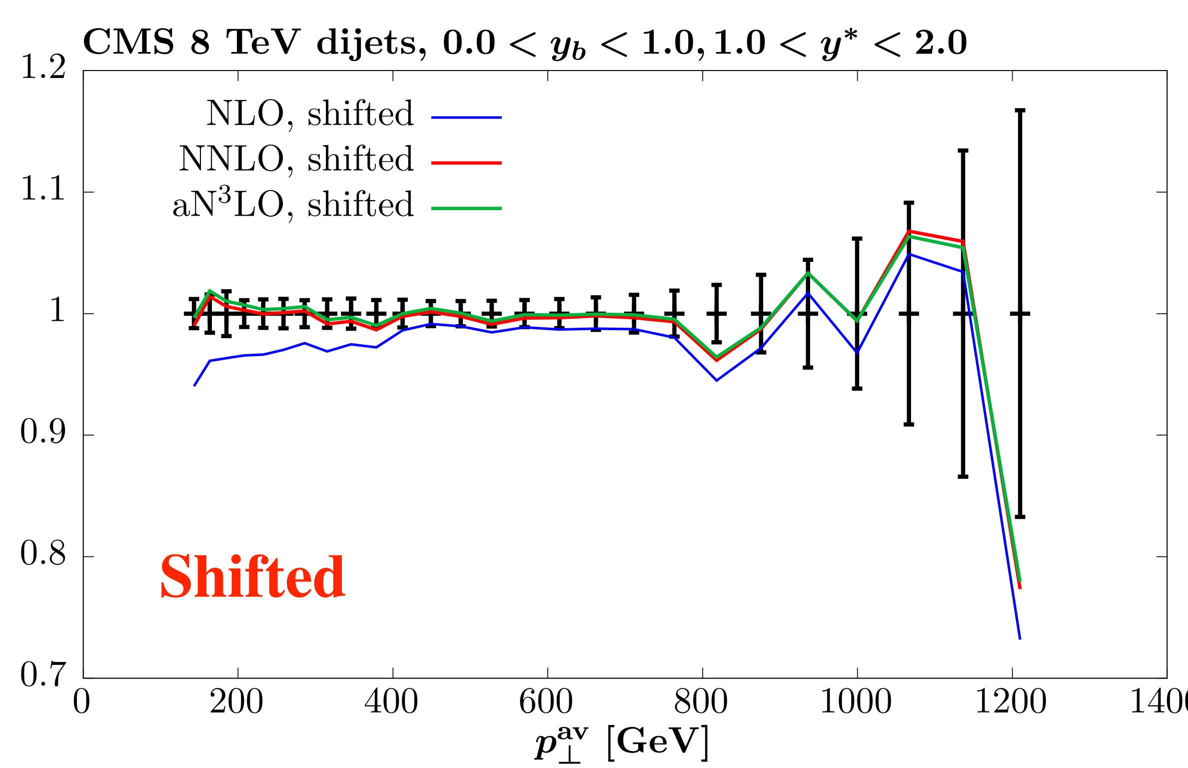
- No clear, by eye, trend for better description at NNLO, aN3LO.



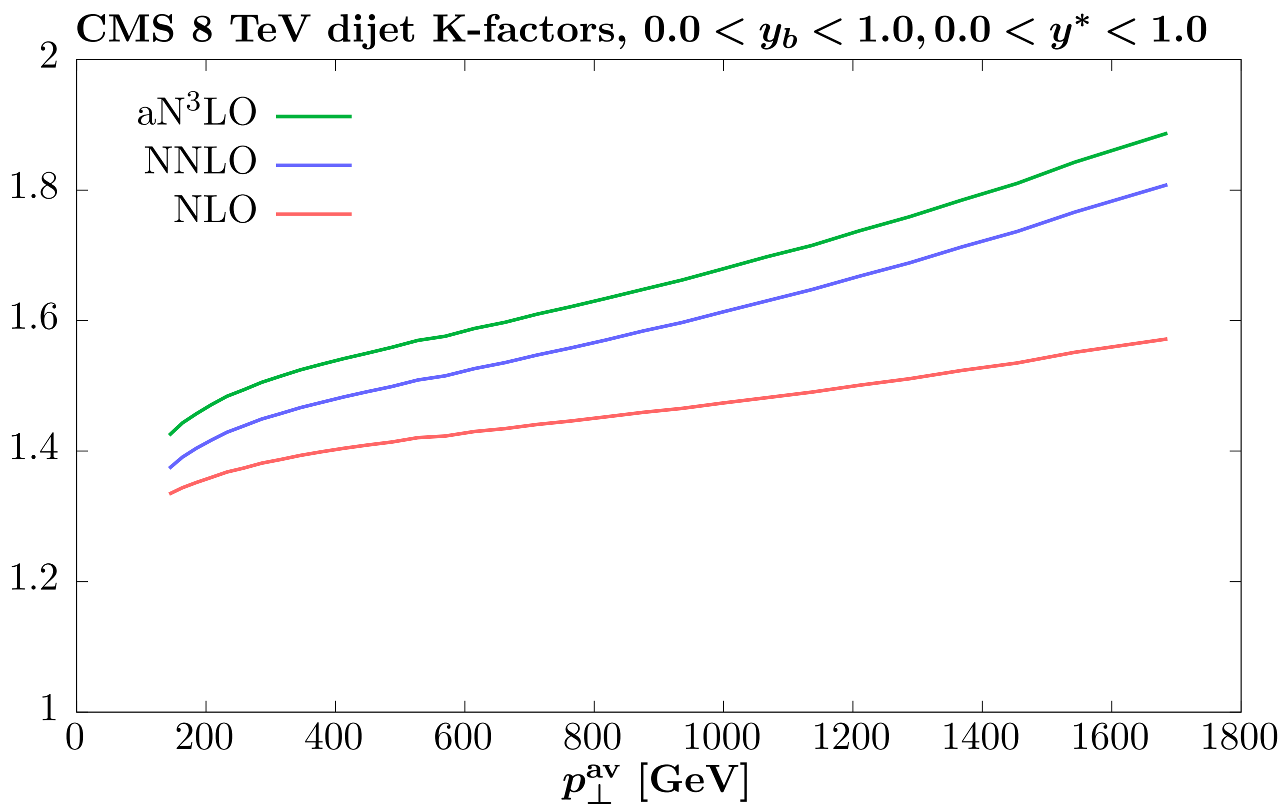


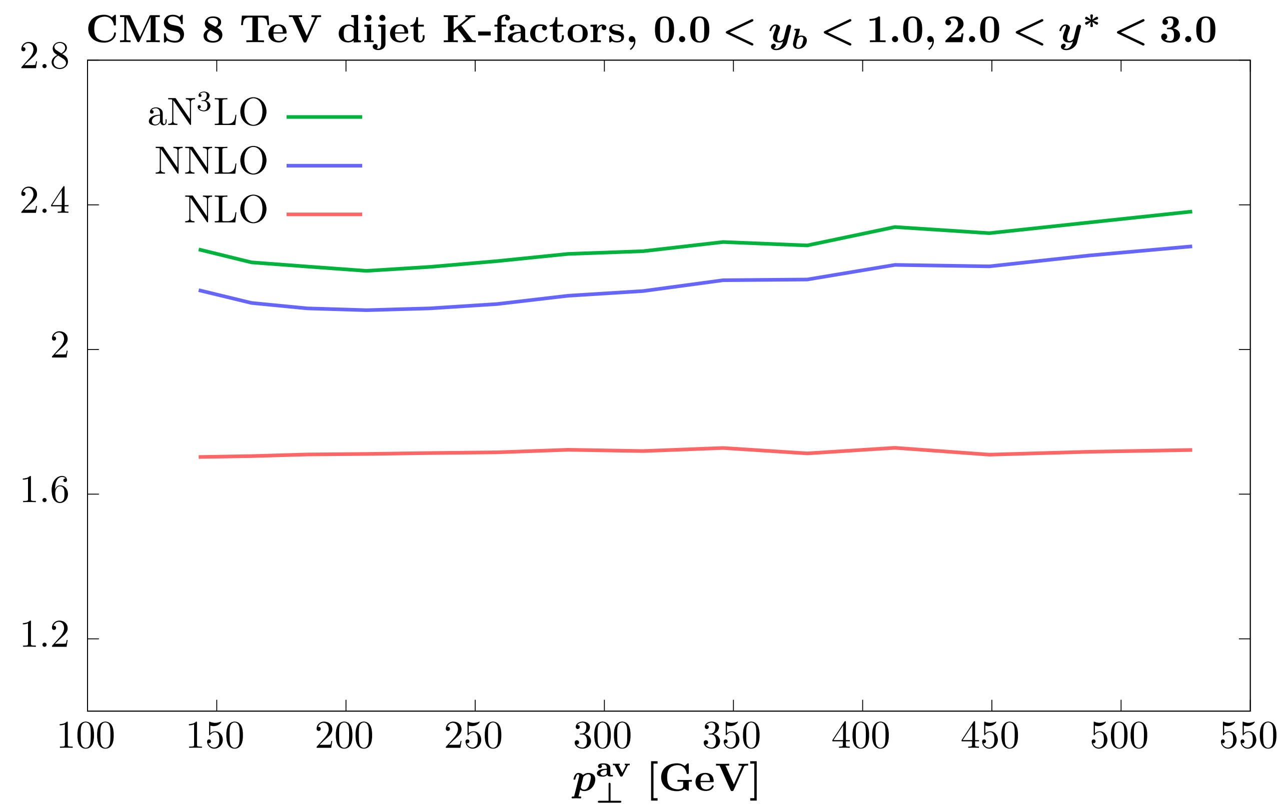
- However this is **before** shifting by correlated systematics.





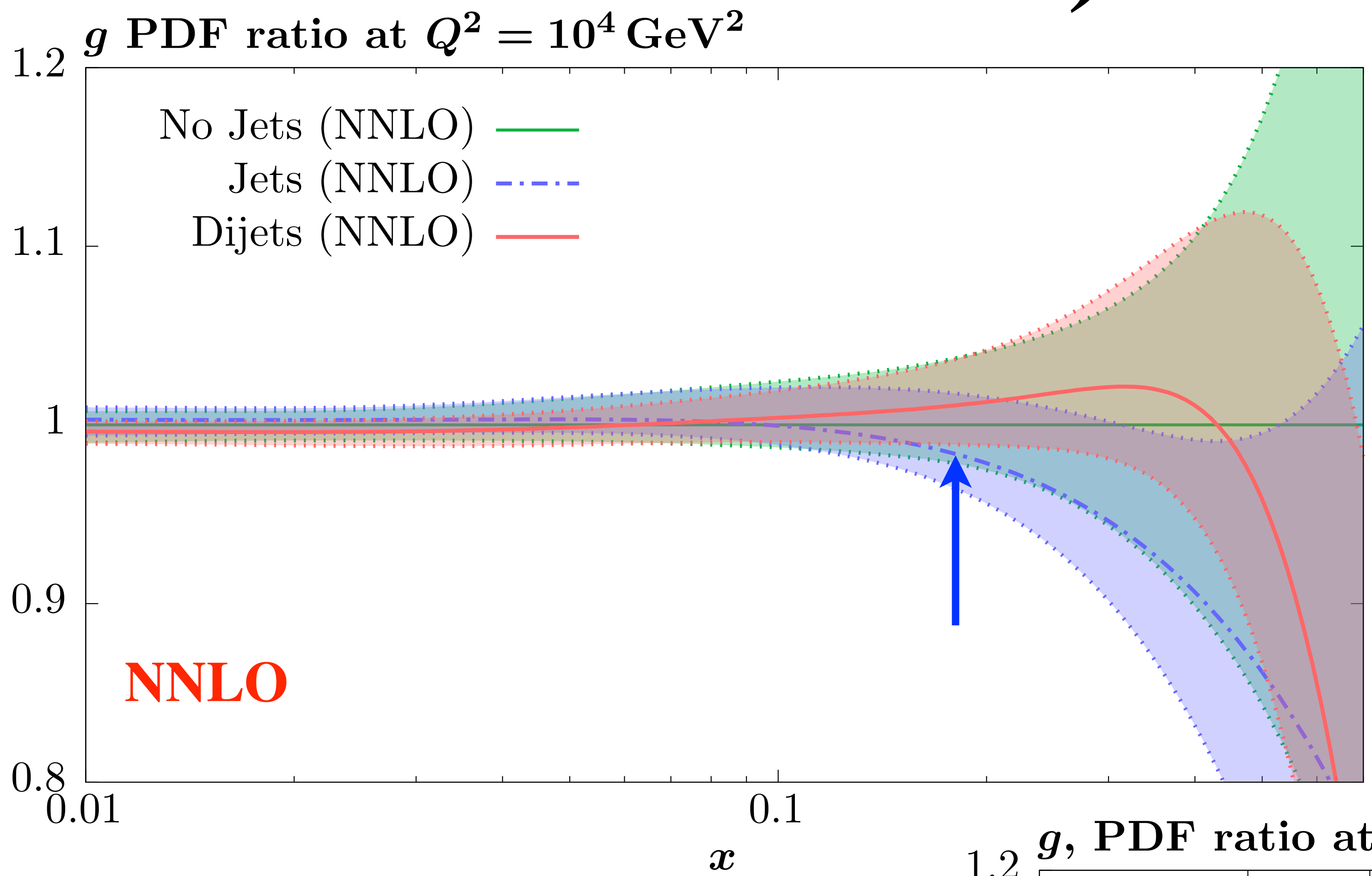
- Impact on shape of distributions in 3D kinematic space and interplay with correlated systematics drives this.
- However some clue from looking at K-factors:





- ★ NNLO corrections reasonable large, in particular in some regions of phase space.
- ★ Also shown are the aN3LO K-factors preferred by the fit: nice trend for perturbative stability, in line with lower orders.
- ★ Similar stability in inclusive jet case (backup).

PDFs: dijets vs. Jets

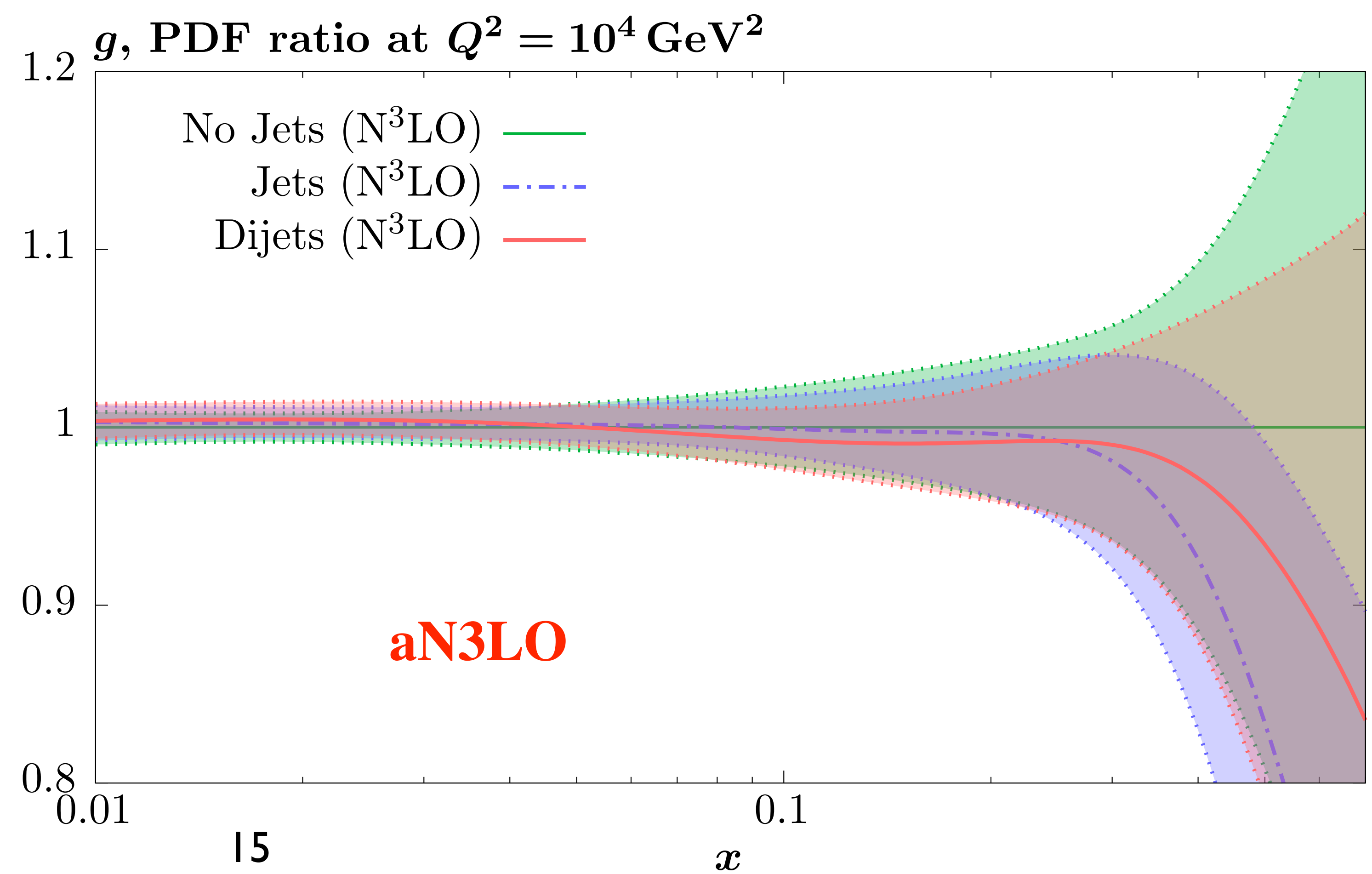


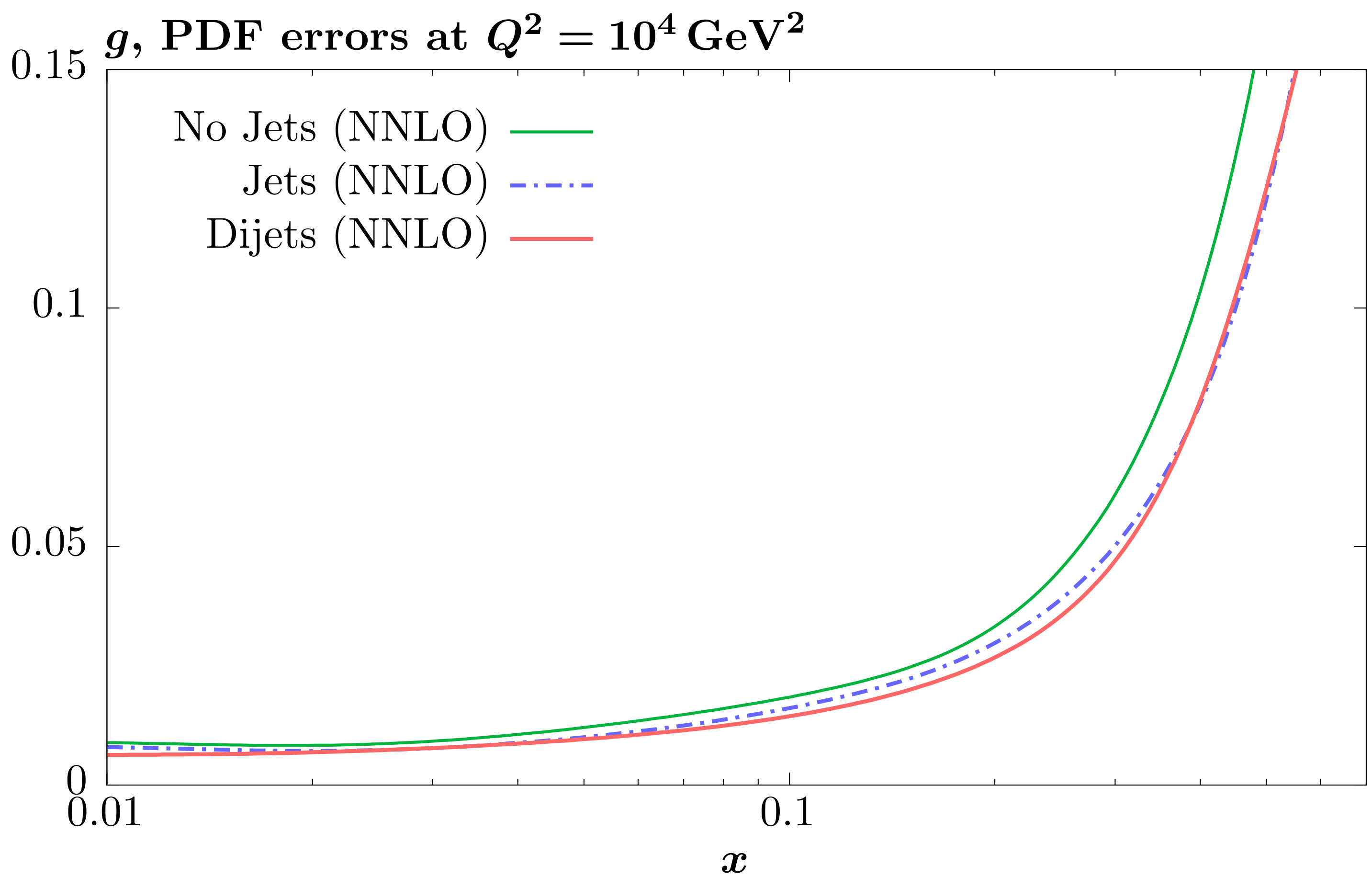
★ Focus on gluon: largest expected impact.

★ Overall consistency between two cases…

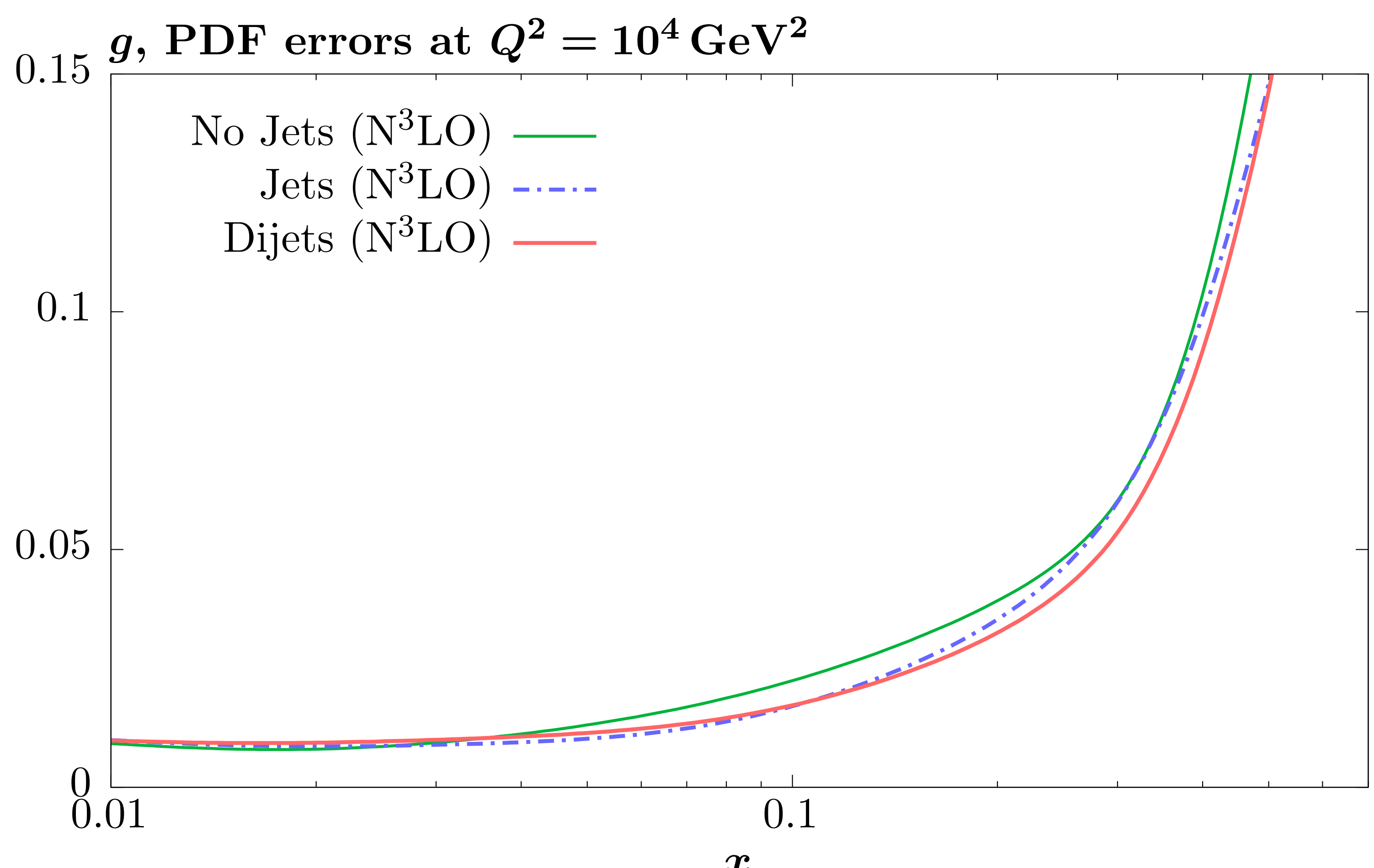
★ But some difference in pull observed between jets/dijets at NNLO.

★ At aN3LO pulls rather similar.





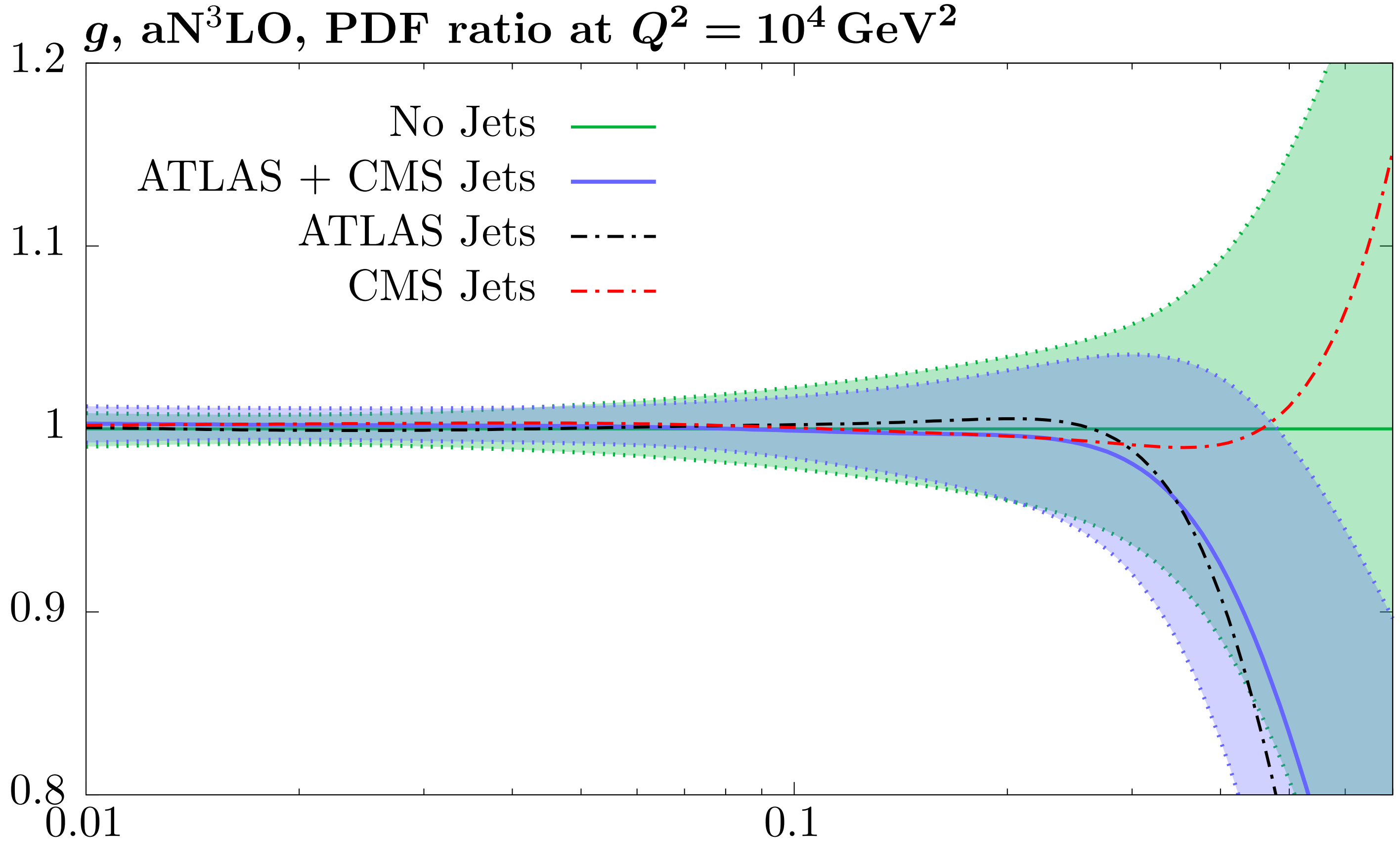
★ Clear reduction in uncertainty in both cases and at both orders.



★ Marginally more significant for dijets.

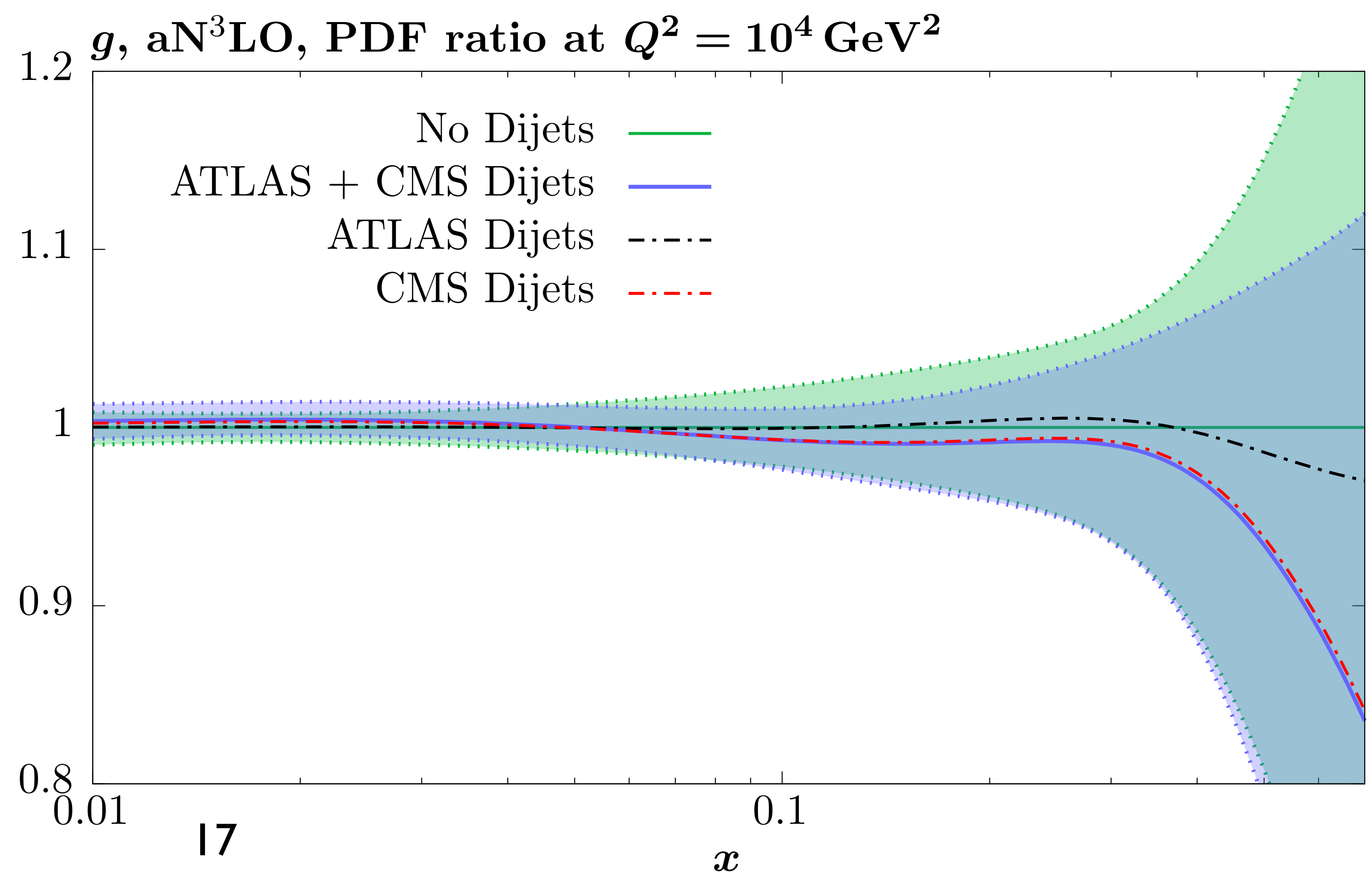
★ Slightly less significant at aN3LO.

Consistency within datasets



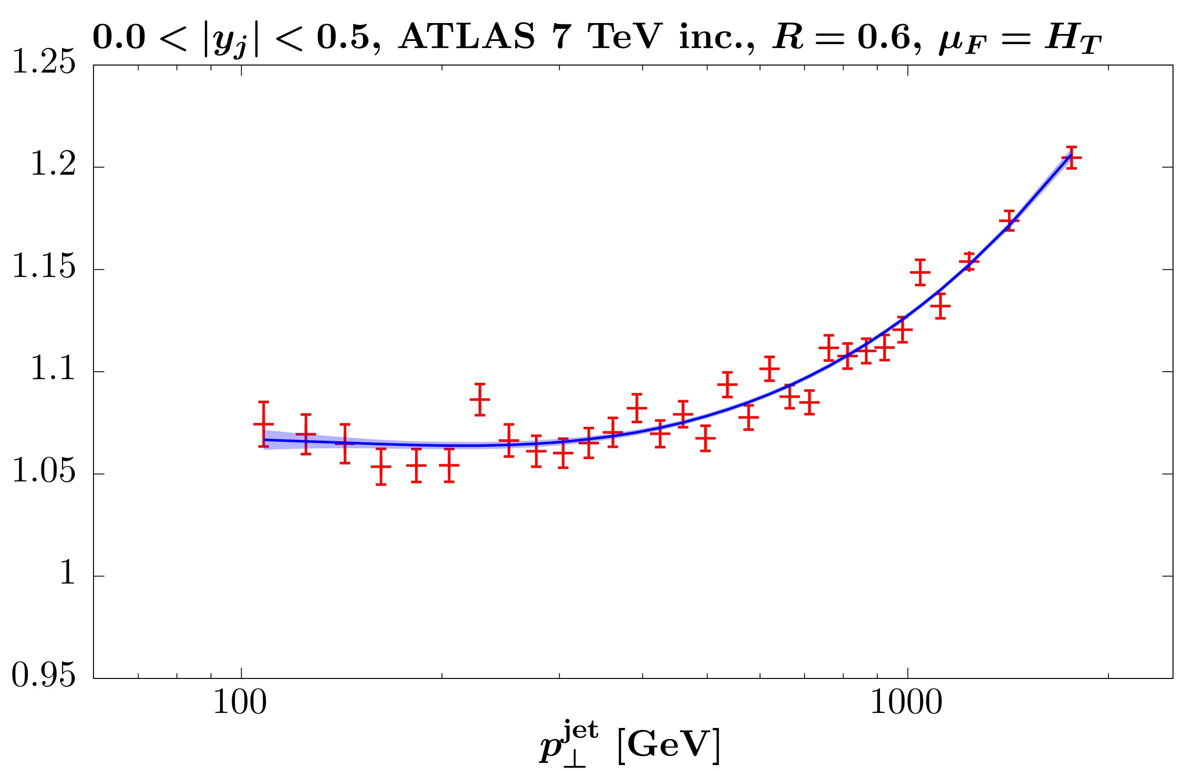
- At higher x clear difference between pulls of ATLAS and CMS (also seen in MSHT20).
- Final result compromise between these.

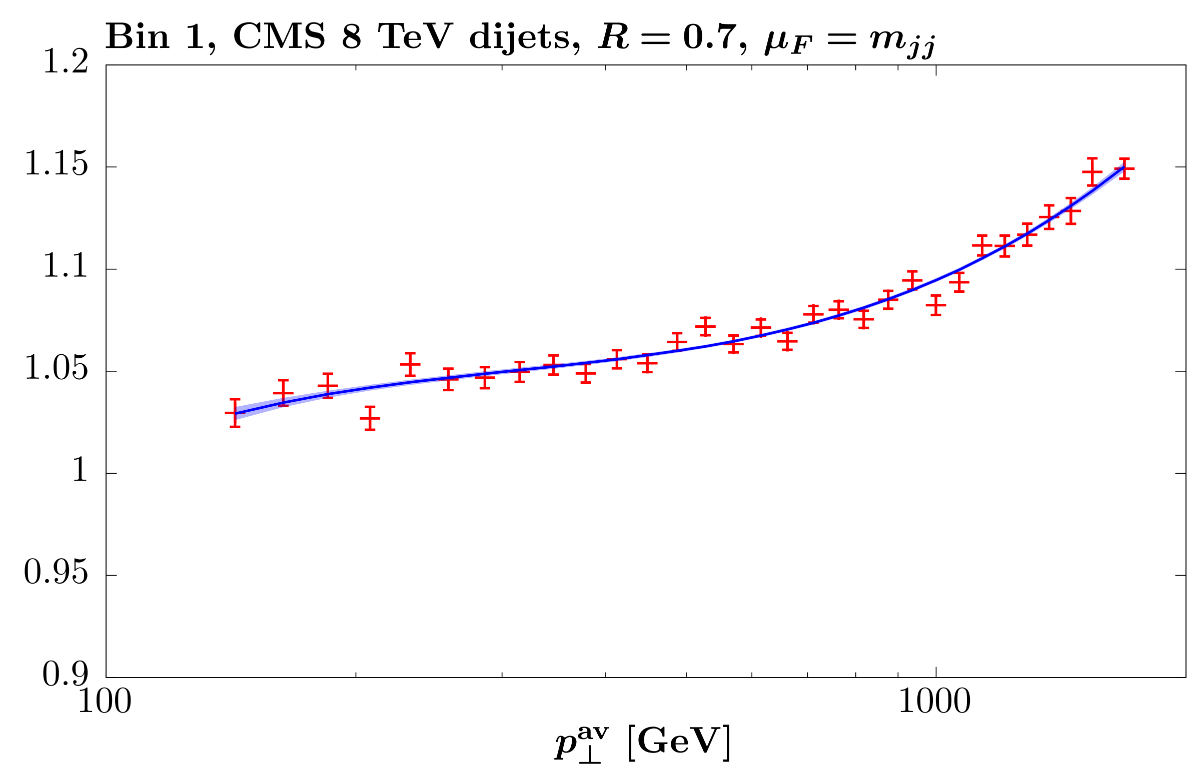
- Consistency between CMS and ATLAS, but latter has very little impact alone.
- Again CMS 8 TeV driving fit.
- Again similar story at NNLO (not shown).



Technical aside (1) - K-factors

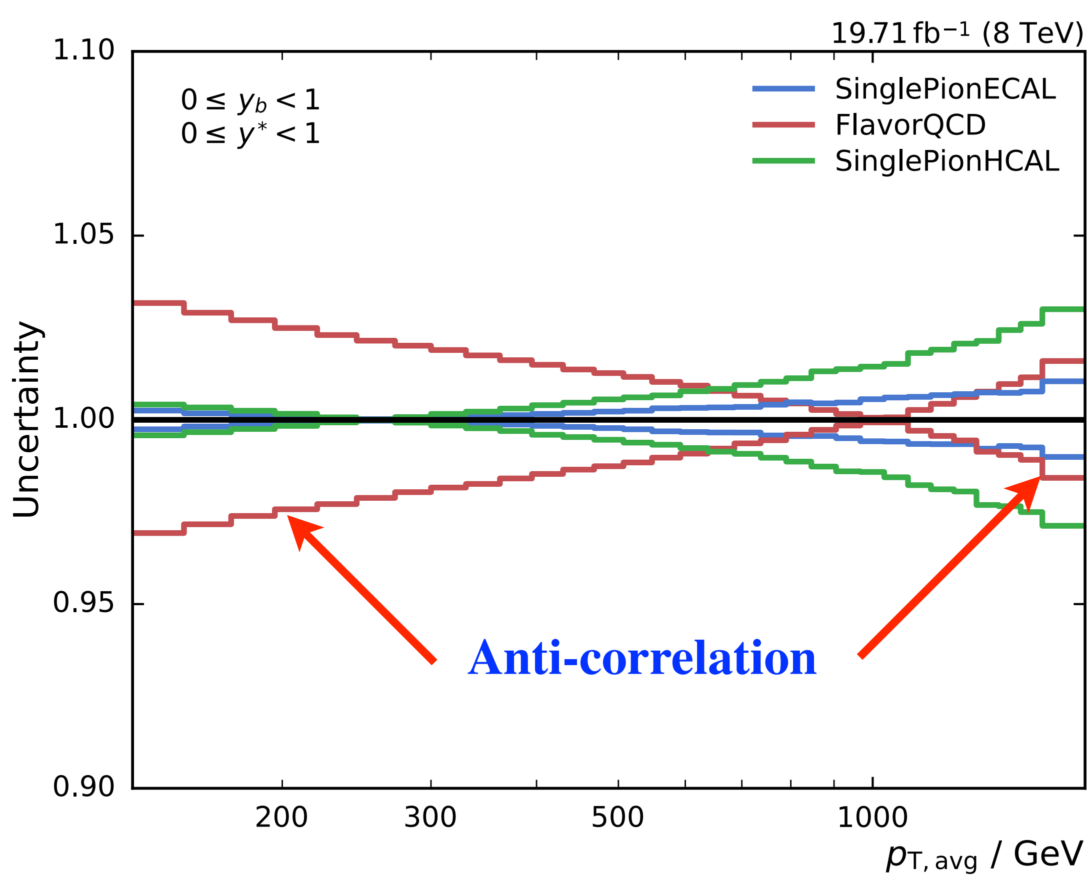
- NNLO QCD corrections included via K-factors. MC uncertainties on these not negligible.
- We argue better to fit these to smooth functions. Can impact on fit quality at the ~ 0.1-0.2 per point level, though PDFs very stable.
- Provides cleaner idea of improvements from NLO to NNLO etc. Find that interpretation can be washed out somewhat otherwise.

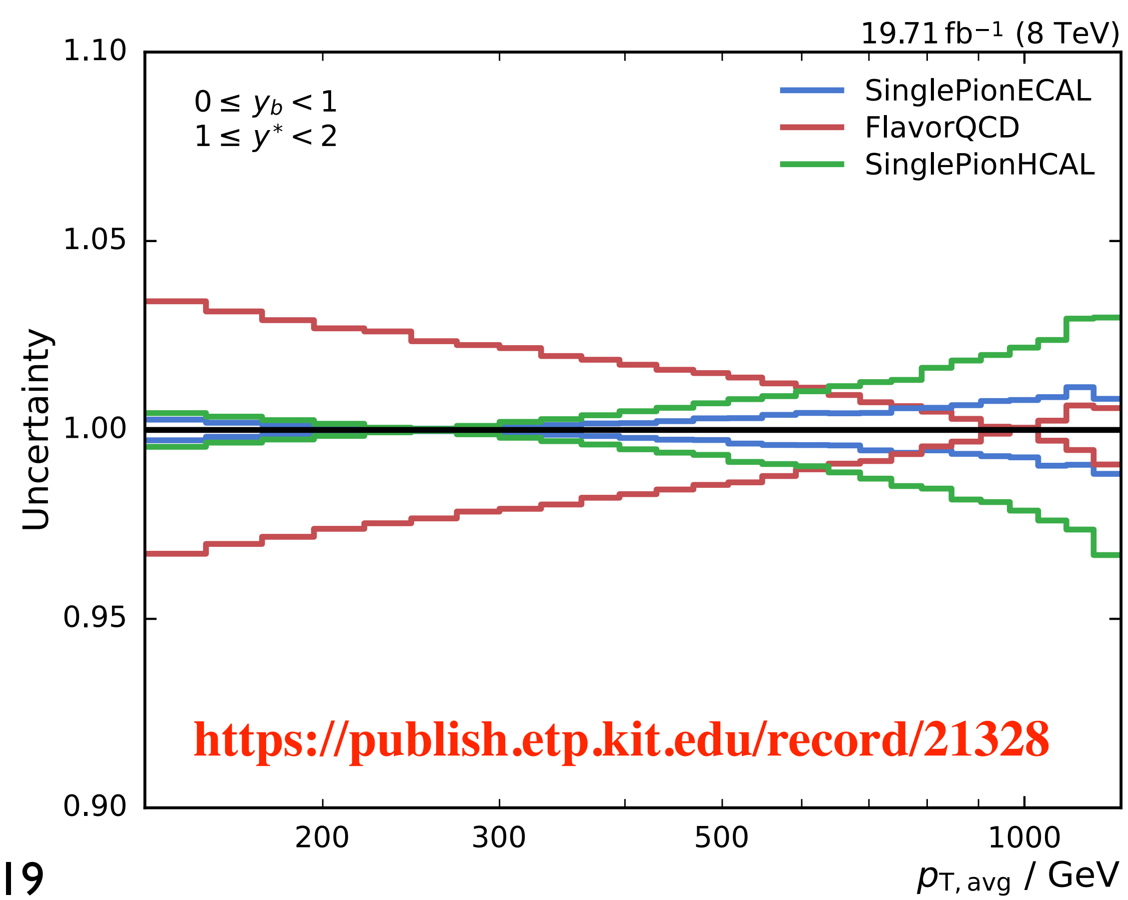




Technical aside (2) - CMS 8 dijets

- Systematic uncertainties related to jet calibration correlated across kinematic (rapidity/$p_\perp$) space. Shape of these indicates anti-correlation between certain regions. However hepdata entries entirely positive.
- Through discussion with CMS colleagues have changed sign to more 'natural' (anti)-correlation.
- In the end this makes very little difference: improves χ^2 by ~ 1-2 points and gluon very stable. But more by chance than design.
- Detailed understanding/bookkeeping of systematic correlations key.



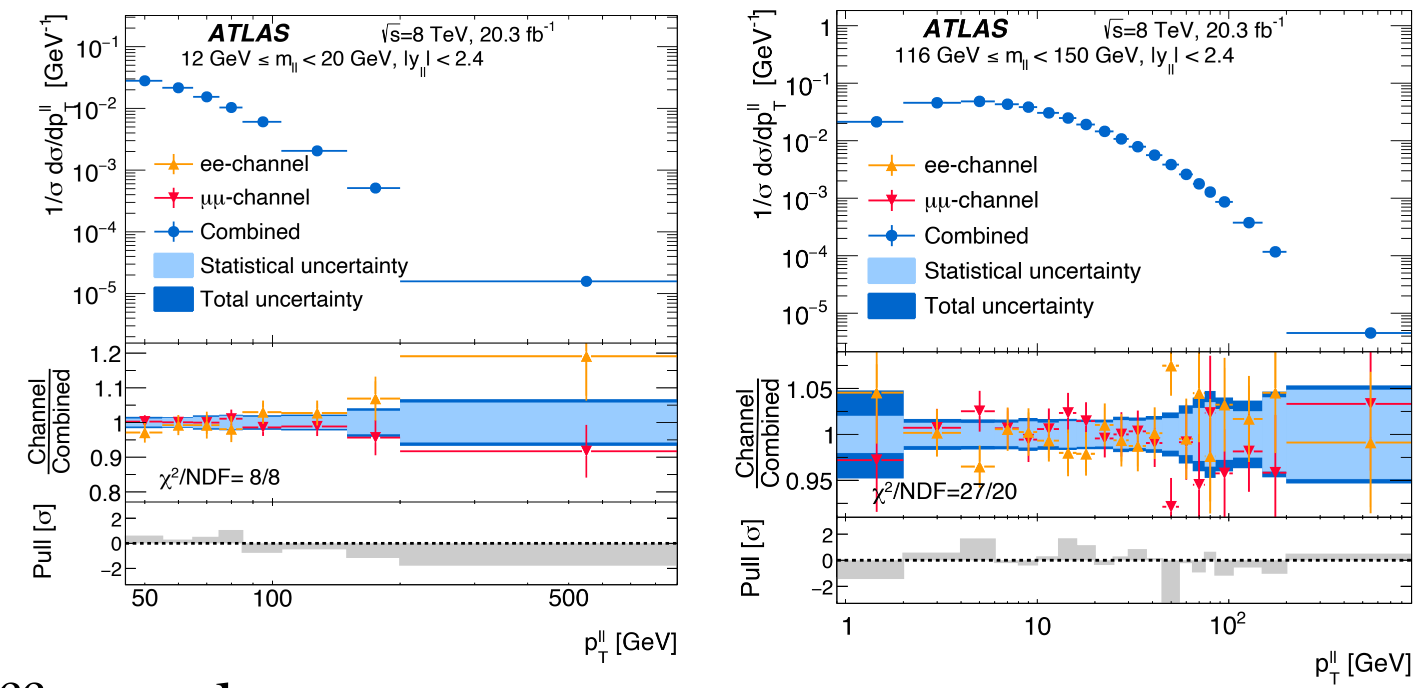


ATLAS $Z\, p_\perp$ data: a closer look

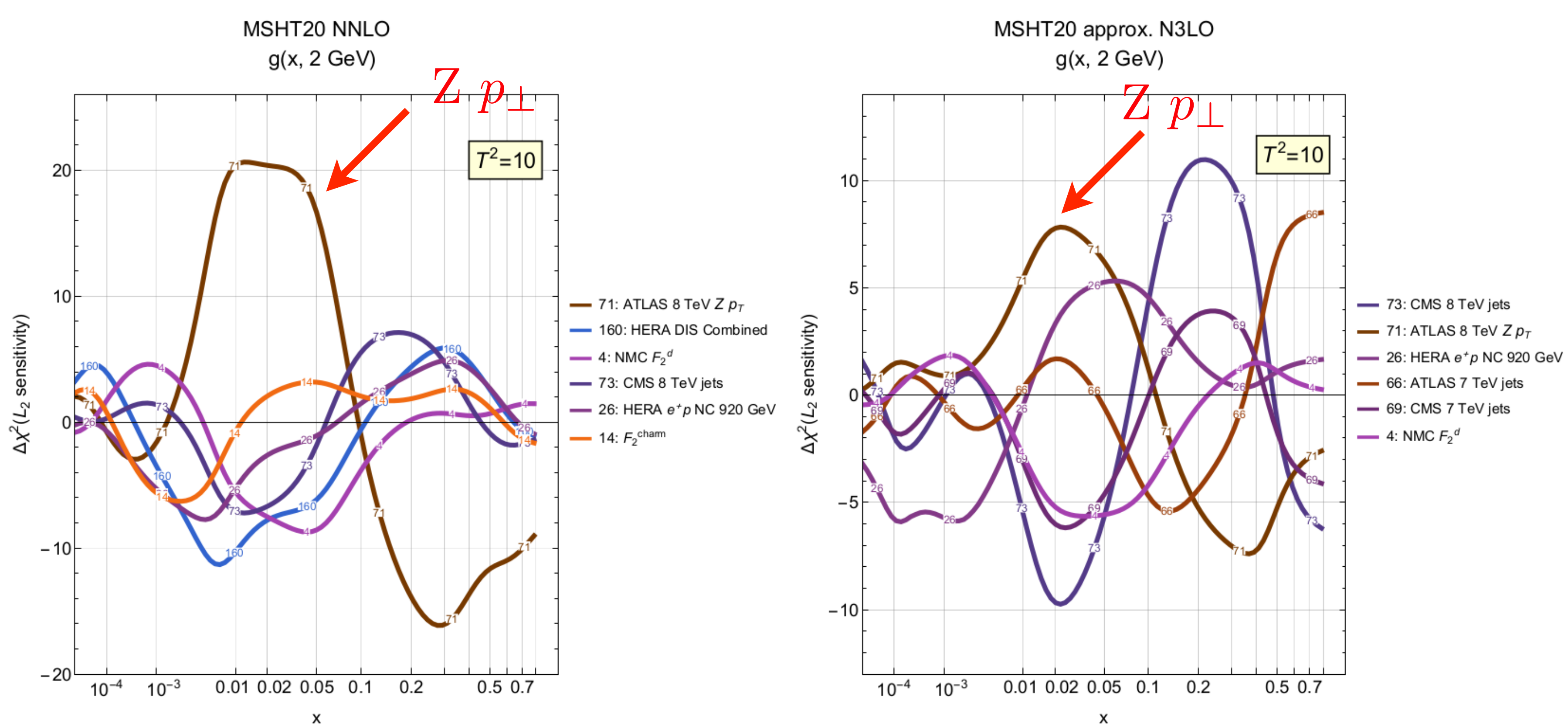
ATLAS, G. Aad et al., Eur. Phys. J. C 76, 291 (2016)

- ATLAS $Z\,p_\perp$ (more properly dilepton $p_\perp$) data presented double differentially in $m_{ll}, p_\perp^{ll}$

$12 < m_{ll} < 150\,\mathrm{GeV} \quad p_\perp^{ll} > 30\,\mathrm{GeV}$



- Treatment of this dataset rather different between groups.
- Fit quality v. poor in default NNLO fit, with dramatic improvement at aN3LO (1.86 vs. 1.04), and highly sensitive to other data in fit (jets vs. dijets).
- Reduced **tension** at aN3LO also backed up by L2 sensitivities (reduced scale).



Preliminary

$\longrightarrow$ Worth revisiting, and considering impact of data selection/ treatment.

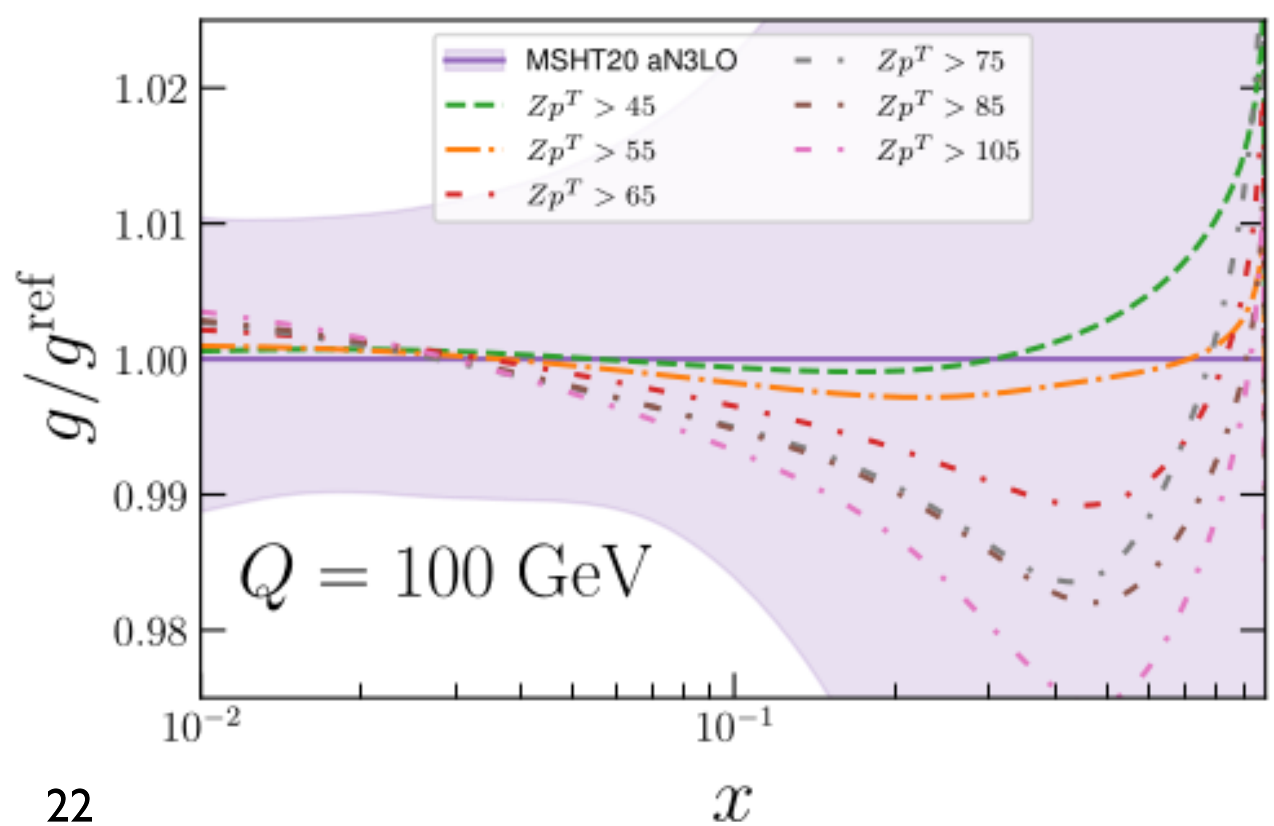
- **First step**: consider impact of raising $p_\perp^{ll}$ cut.

$\chi^2/N_{\rm pts}$

Order of fit/p_T^{cut} (GeV)	30 (default)	45	55	65	75	85	105
NNLO	1.86	1.68	1.67	1.42	1.39	1.42	1.21
aN3LO	1.04	0.95	1.01	0.84	0.86	0.87	0.81
N_{pts}	104	88	77	66	55	44	33

- Fit quality improves slowly as amount of data is reduced.
- Effect larger at NNLO, but NNLO always worse.

- No obvious sign of issue with particular $p_\perp^{ll}$ region.
- Next steps: impact of m_{ll} selection, interplay with other datasets…

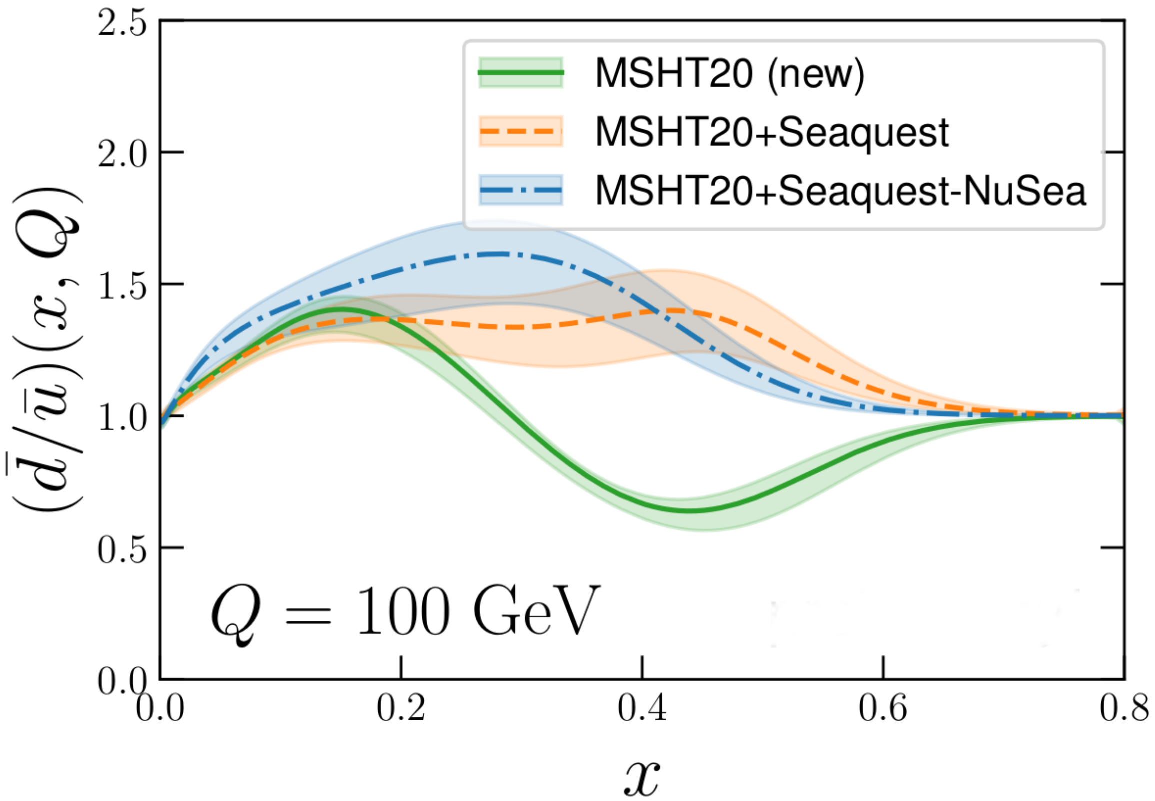


Impact of SeaQuest data

New data - Seaquest (NNLO)

Preliminary!

- Seaquest (E906) fixed target DY data - sensitivity to high x q, $\bar{q}$:
  $\Rightarrow \sigma_D/\sigma_H \sim 1 + \bar{d}/\bar{u}$. Direct measurement of $\bar{d}/\bar{u}$ at high x.
- Various models for $\bar{d}/\bar{u}$ at high x: Pauli blocking, pion cloud, etc.
- Previous questions of NuSea (E866) data preferring $\bar{d} < \bar{u}$ at $x \approx 0.4$.
- Clearly raises high x $\bar{d}/\bar{u}$. Tension with NuSea which pulls it down.



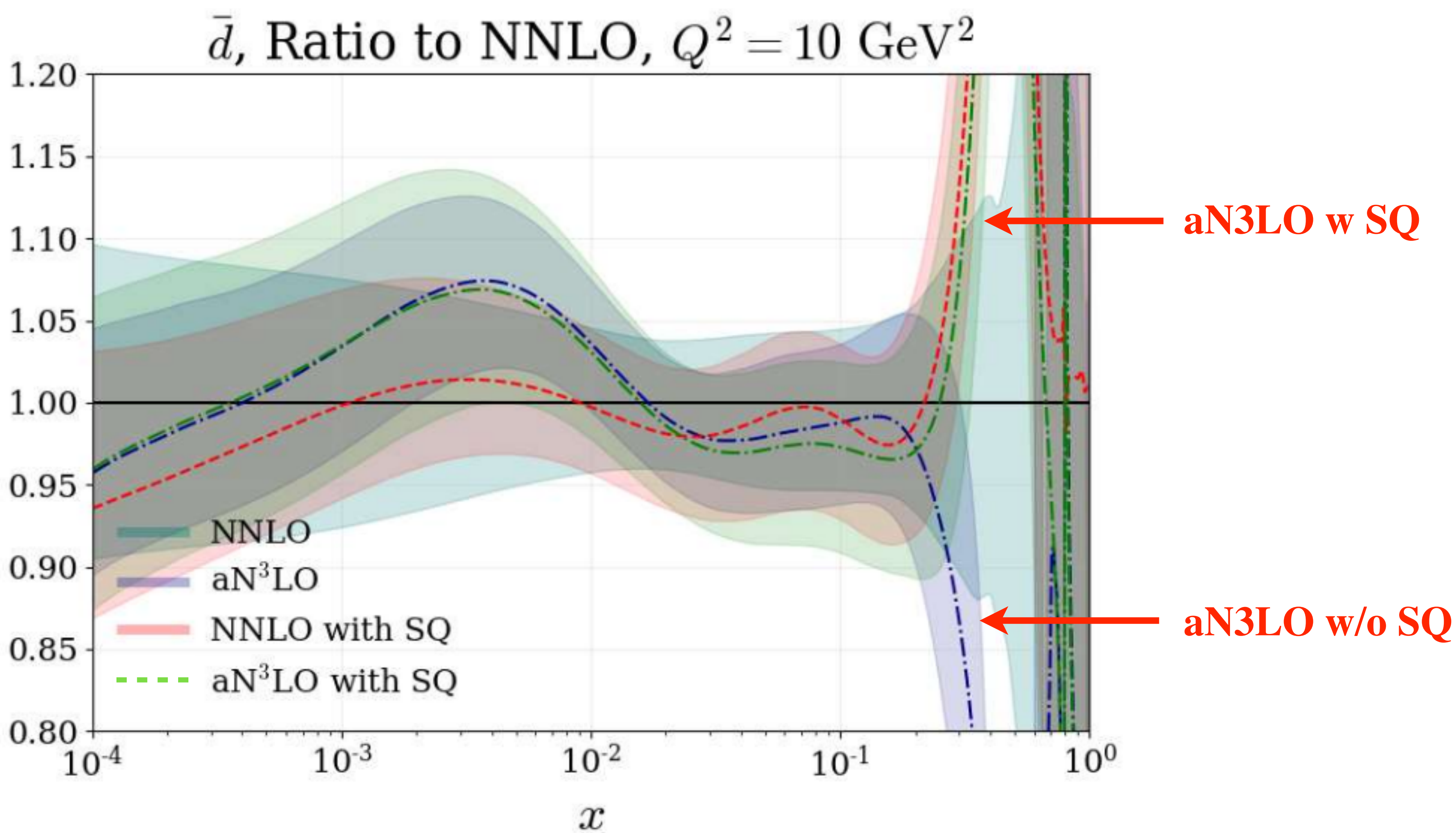
Dataset	$N_{\rm pts}$	MSHT20	New
Seaquest	6	-	8.2
NuSea	15	9.8	19.0
Total (without Seaquest or NuSea)	4348	5102.3	5112.1

- NuSea $\chi^2/N_{\rm pts}$: $0.65 \rightarrow 1.27$, when Seaquest added.

- Rest of data also worsens in χ^2 by 9 points, with 4.5 in E866 absolute DY (rather than ratio), 4.4 in NMC n/p, 4.3 in DØ W asymmetry.

Preliminary!

- At aN3LO, the $\bar{d}$ become negative above $x \sim 0.5$ with a minimum at $x \sim 0.6$. Nonetheless remains positive within uncertainties.
- Like at NNLO, adding the Seaquest data raises the $\bar{d}/\bar{u}$.
- Adding Seaquest $\Rightarrow$ NNLO and aN3LO $\bar{d}$, $\bar{u}$ again very similar.
- Effect on fit quality of adding Seaquest similar to NNLO, $\Delta\chi^2 = +6$ in rest of data, NuSea χ^2/N doubles from ~ 0.6 to ~ 1.3.



Summary

★ MSHT group busy working on range of follow up studies, making use of NNLO and new aN3LO machinery.

★ Jets/Dijets:

- Jet fit quality relatively poor, remains so in aN3LO fit.
- Dijet fit quality good, and with improvement at aN3LO in line with expectations.
- Scale choice does not play big role in inclusive, EW corrections make fit quality worse (!).

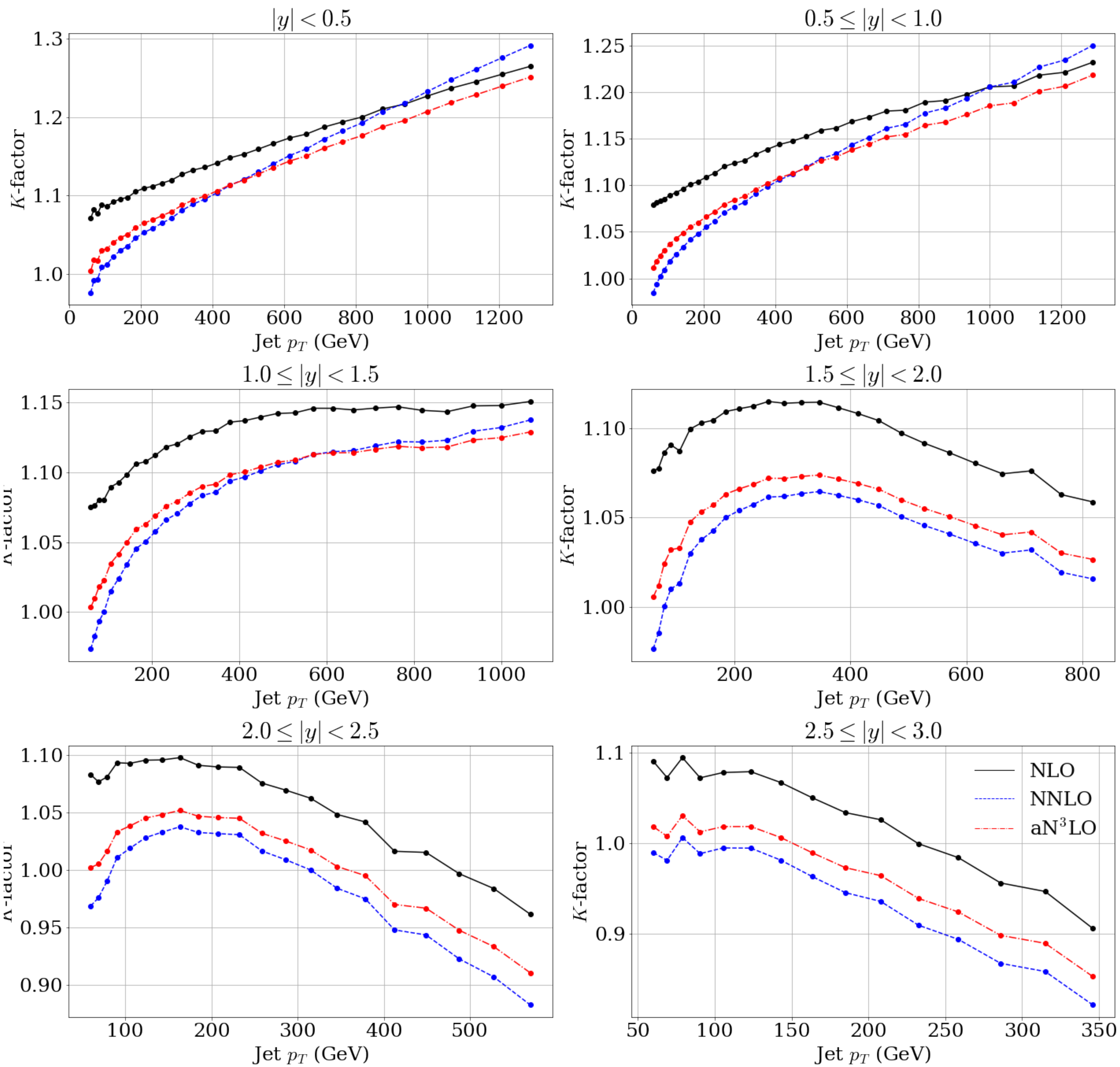
★ All indicates that dijet data may be preferable.

★ Working ongoing to understand these questions, and connected ones related to high x ($Z\,p_\perp$, Seaquest) at NNLO and aN3LO.

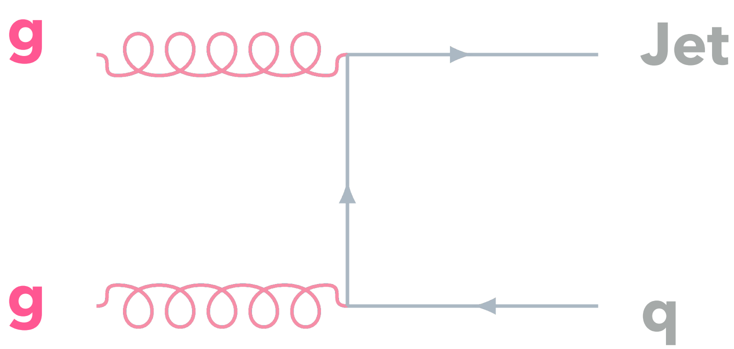
Thank you for listening!

Backup

Inclusive Jet K-factors



Jet Kinematics: Inclusive



- Inclusive jets measured in terms of jet $p_\perp$ and y_j.
- Schematically, **LO** relationship to high x parton:

$$x = \frac{p_\perp}{\sqrt{s}} \left(e^{y_j} + e^{y_{j'}} \right)$$

Observed Jet j $(y_j > 0)$

'Unobserved' Jet j'

$\longrightarrow$ Need 3 kinematic inputs to uniquely determine x.

- Inclusive jets: effectively integrate over $x \gtrsim \frac{p_\perp}{\sqrt{s}} e^{y_j}$.
- So certainly sensitive to high x region, but washed out somewhat.

Jet Kinematics: Dijets

- For dijets, both jets measured. Same schematic **LO** relationship:

$$x_{1,2} = \frac{p_\perp}{\sqrt{s}} \left(e^{\pm y_j} + e^{\pm y_{j'}}\right)$$

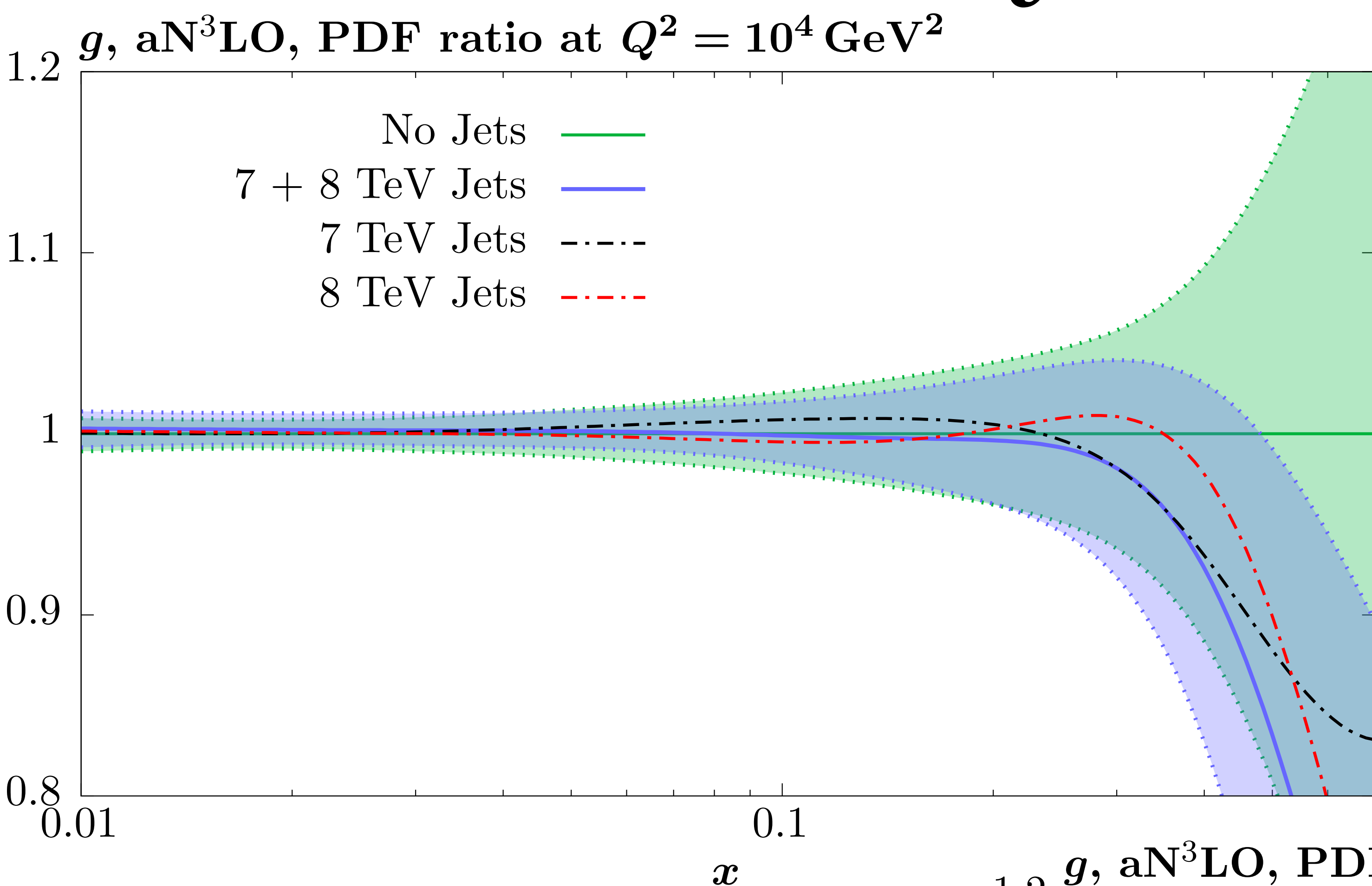
- Double differential measurements in terms of m_{jj} and $y^*/y_{\rm max}$: not sufficient to uniquely pin down LO x.

- That is, some washing out (though precise effect depends on choice of variable).

- However, also possible to measure triple differentially - expect to provide stronger, more direct constraints.

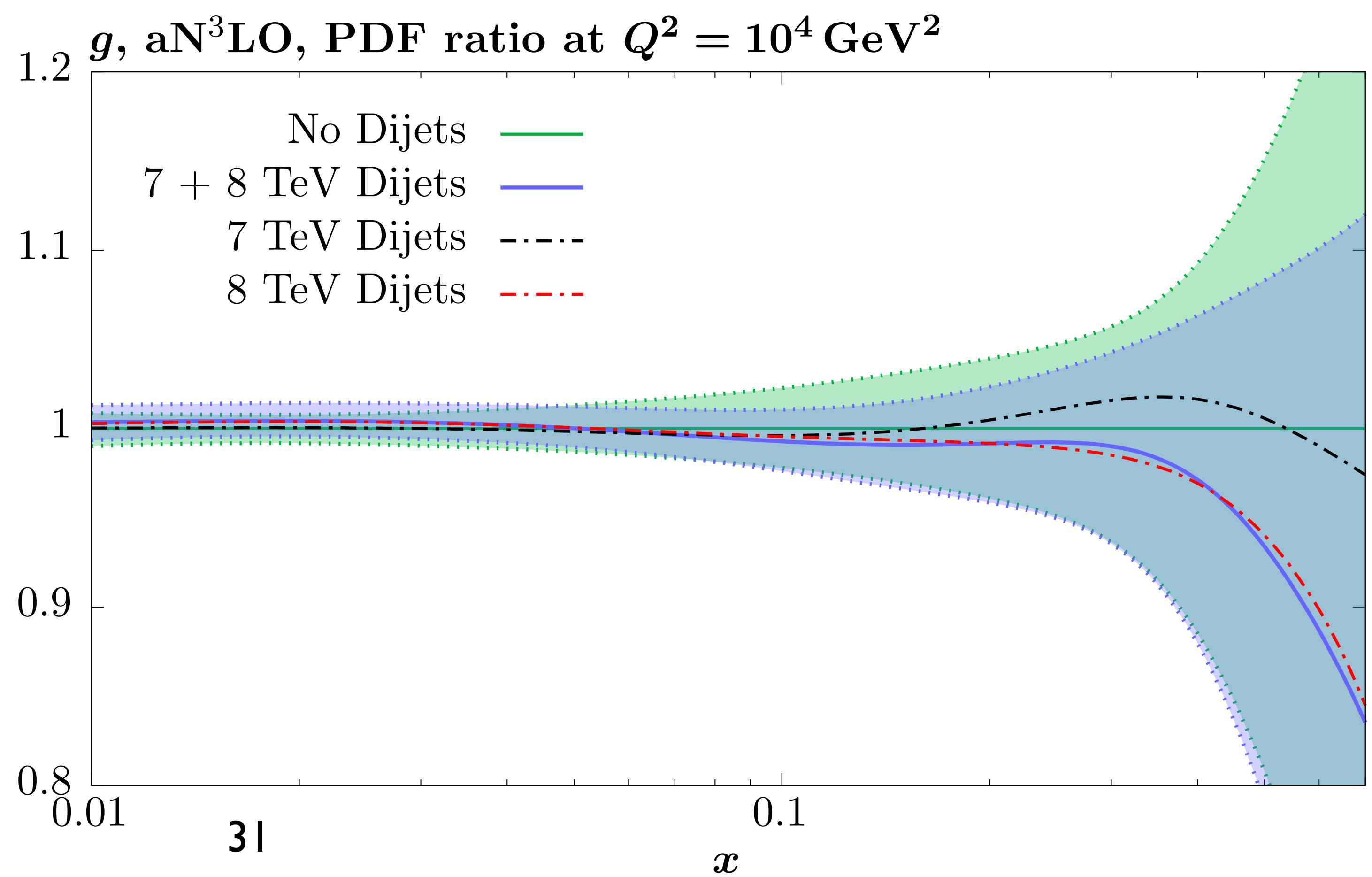
$$\mathrm{d}^3\sigma/\mathrm{d}p_{\perp,avg}\mathrm{d}y_b\mathrm{d}y^*$$

Consistency within datasets

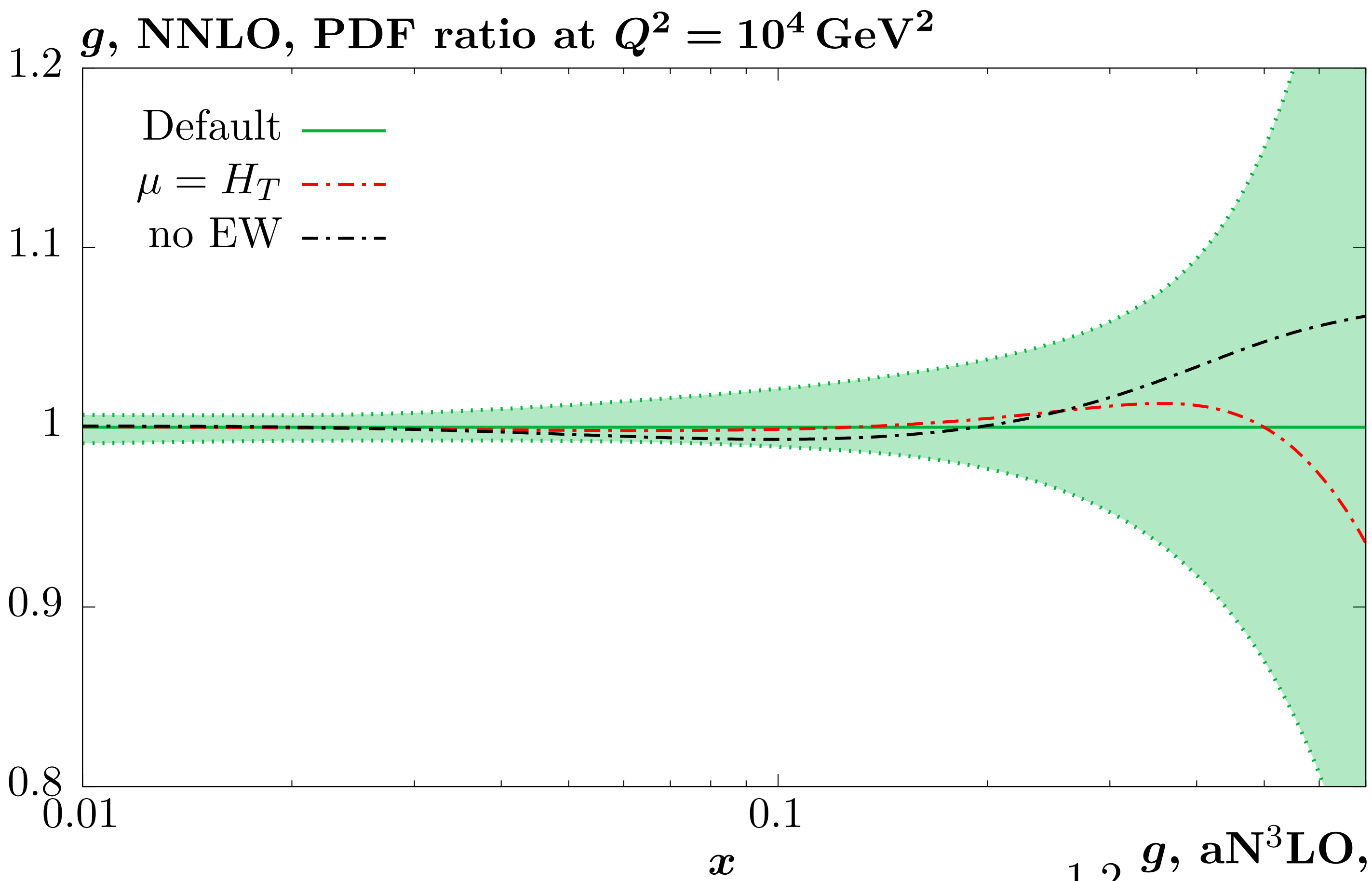


- *7* & 8 TeV data ~ consistent pulls inclusive jets.
- Similar for NNLO (not shown).

- *7* & 8 TeV data consistent for dijets, but this is due to broader result.
- All dijet fits completely driven by CMS 8 TeV data
- Similar for NNLO (not shown).



PDFs: EW corrections/scale choice



★Impact of these on gluon small, though not completely negligible.