MSHT20: Review and Updates

Lucian Harland-Lang, University of Oxford

DIS 2022, 4 May 2022

T. Cridge et al., *Eur.Phys.J.C* 81 (2021) 4, 341
T. Cridge et al., *Eur.Phys.J.C* 81 (2021) 744
T. Cridge et al., *Eur.Phys.J.C* 82 (2022) 1, 90
T. Cridge et al., *in preparation*.

UNIVERSITY OF OXFORD

Outline

- Since the release of **MSHT20** we have been busy performing a number of follow up studies and extensions.
- I will cover these in different levels of detail:

  ★ **MSHT20** - recap.

  ★ Dedicated study on **strong coupling/heavy quark** masses.

  ★ Including **dijet** production in MSHT20.

  ★ Fitting **SeaQuest** data: first look.

  ★ **MSHT20qed** - including QED corrections.

  ★ Theoretical uncertainties: approximate higher order corrections.

- In the latter cases stay tuned for details talks today + tomorrow!

MSHT20

MSHT20: Recap

Parton distributions from LHC, HERA, Tevatron and fixed target data:

MSHT20 PDFs

S. Bailey[a], T. Cridge[b], L. A. Harland-Lang[a], A. D. Martin[c], and R.S. Thorne[b]

[a] Rudolf Peierls Centre, Beecroft Building, Parks Road, Oxford, OX1 3PU
[b] Department of Physics and Astronomy, University College London, London, WC1E 6BT, UK
[c] Institute for Particle Physics Phenomenology, Durham University, Durham, DH1 3LE, UK

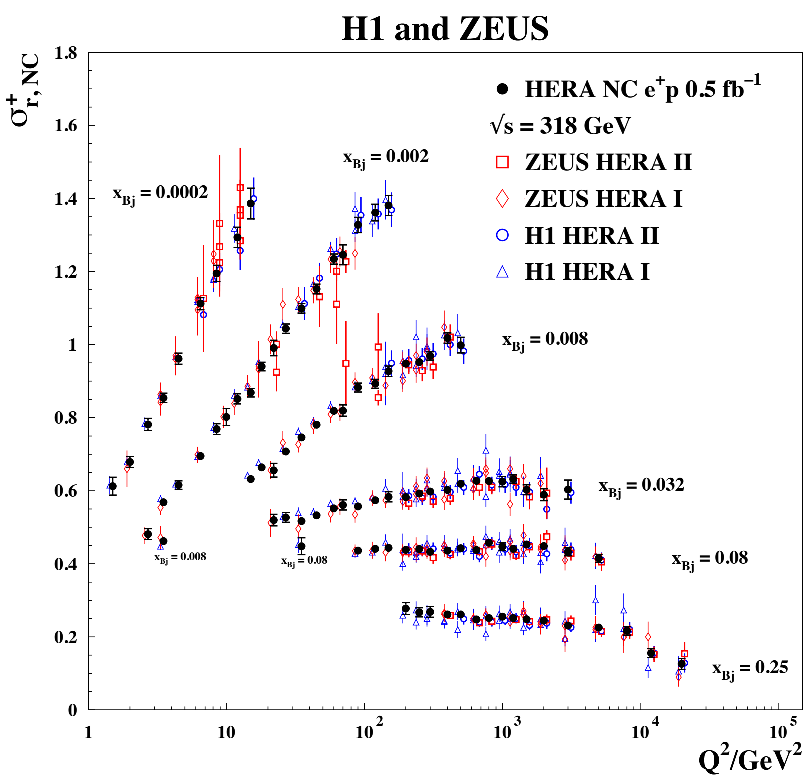
Abstract

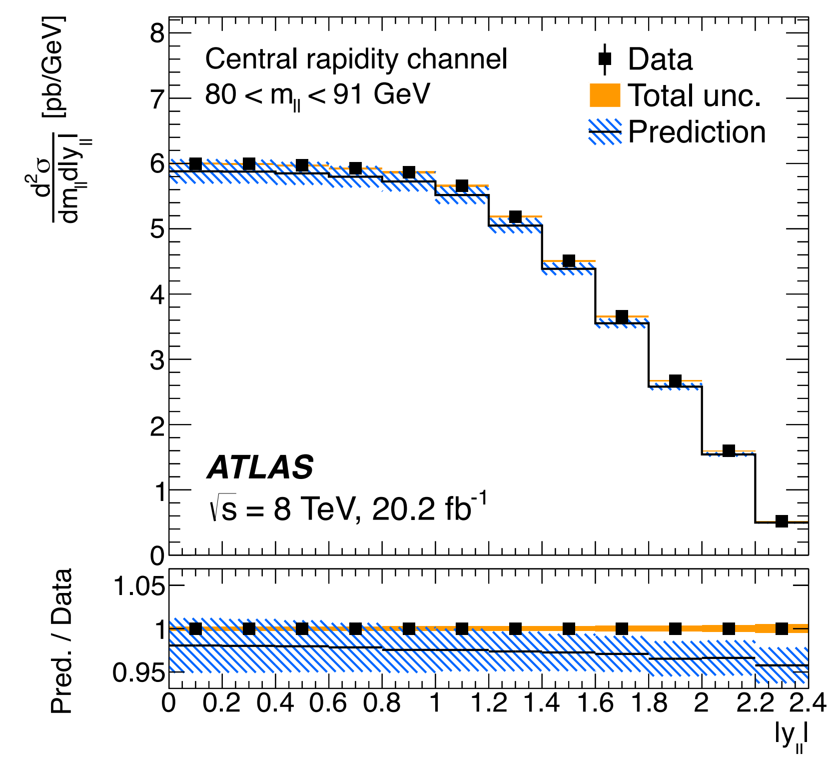
We present the new MSHT20 set of parton distribution functions (PDFs) of the proton, determined from global analyses of the available hard scattering data. The PDFs are

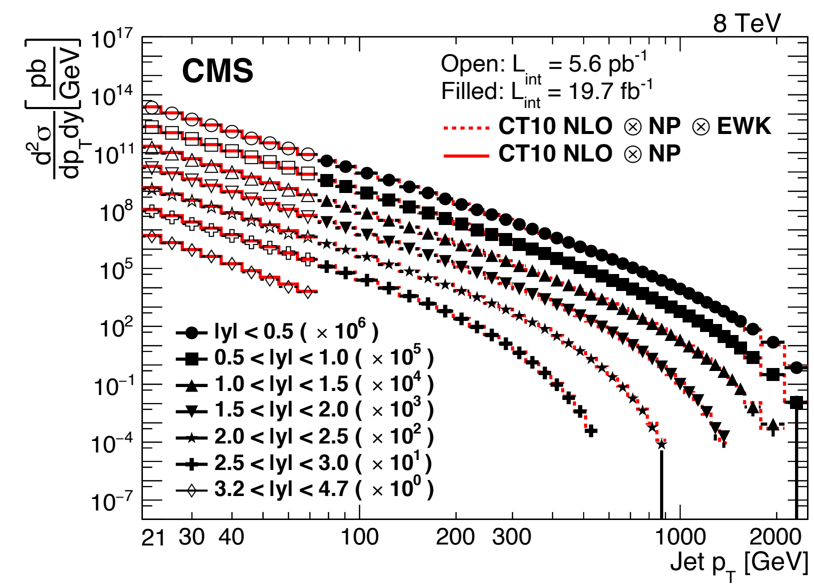
- The ‘Post-Run I’ set from the MSTW, MMHT… group: **MSHT20**.
- Focus on including significant amount of **new data**, higher **precision theory** and on **methodological improvements**.

New Data

- Can divide into 3 broad (non-exhaustive) categories:







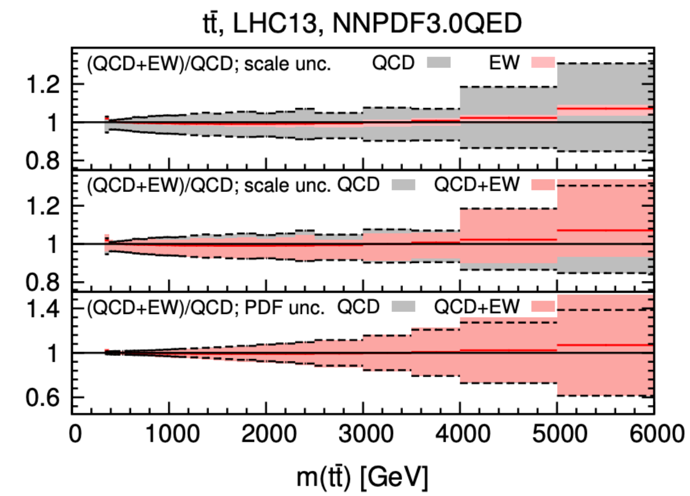
★ Final HERA H1 + ZEUS combination data on inclusive and heavy flavour **DIS**.

★ High precision multi-differential **DY** data. Flavour decomposition.

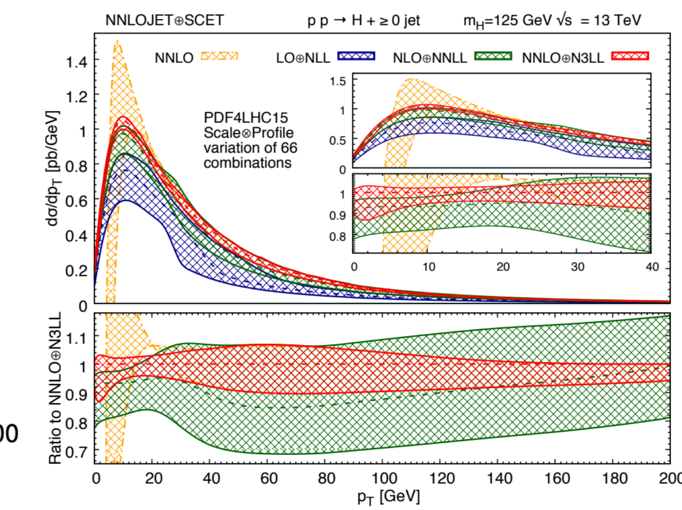
★ Inclusive jet, $Z\,p_\perp$, differential $t\bar{t}$. High x partons.

Precision Theory

- Vast majority of processes included in fits have full **NNLO** QCD theory (+ **NLO** EW where relevant) available and included.

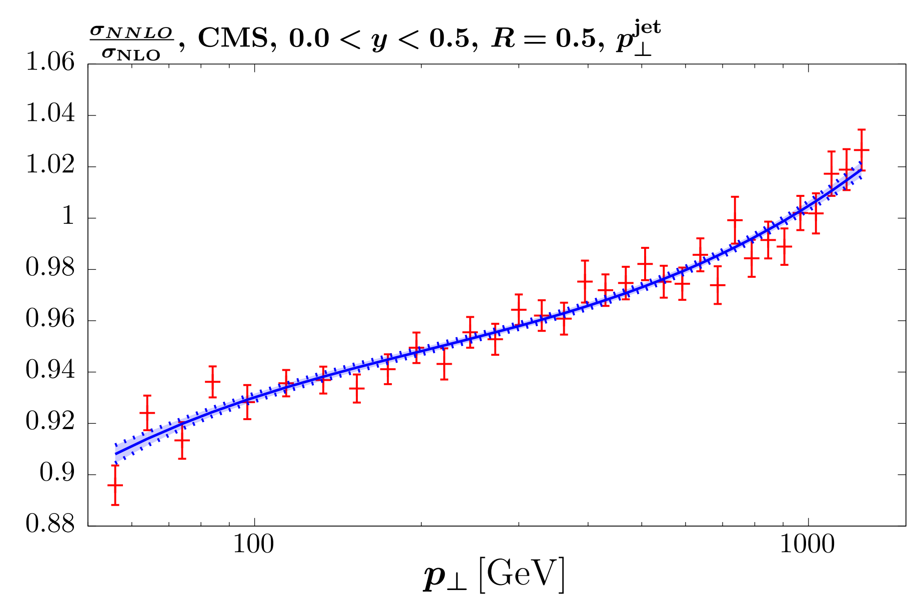


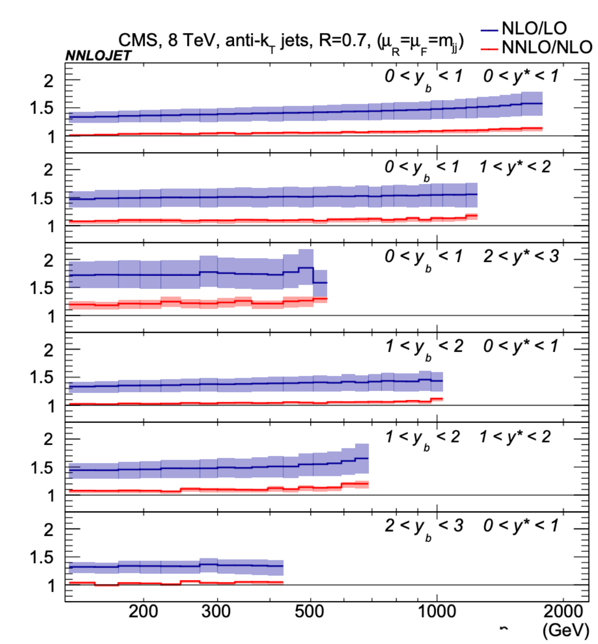
Top quarks - single/ double differential



W, Z transverse momentum distributions

- LHC processes: NLO implemented with **Fastnlo/ Applgrids**.
- NNLO included via K-factors (exception of $t\bar{t}$). Smoothed/with full account of MC error.





Inclusive jets/dijets

Parameterisation Flexibility

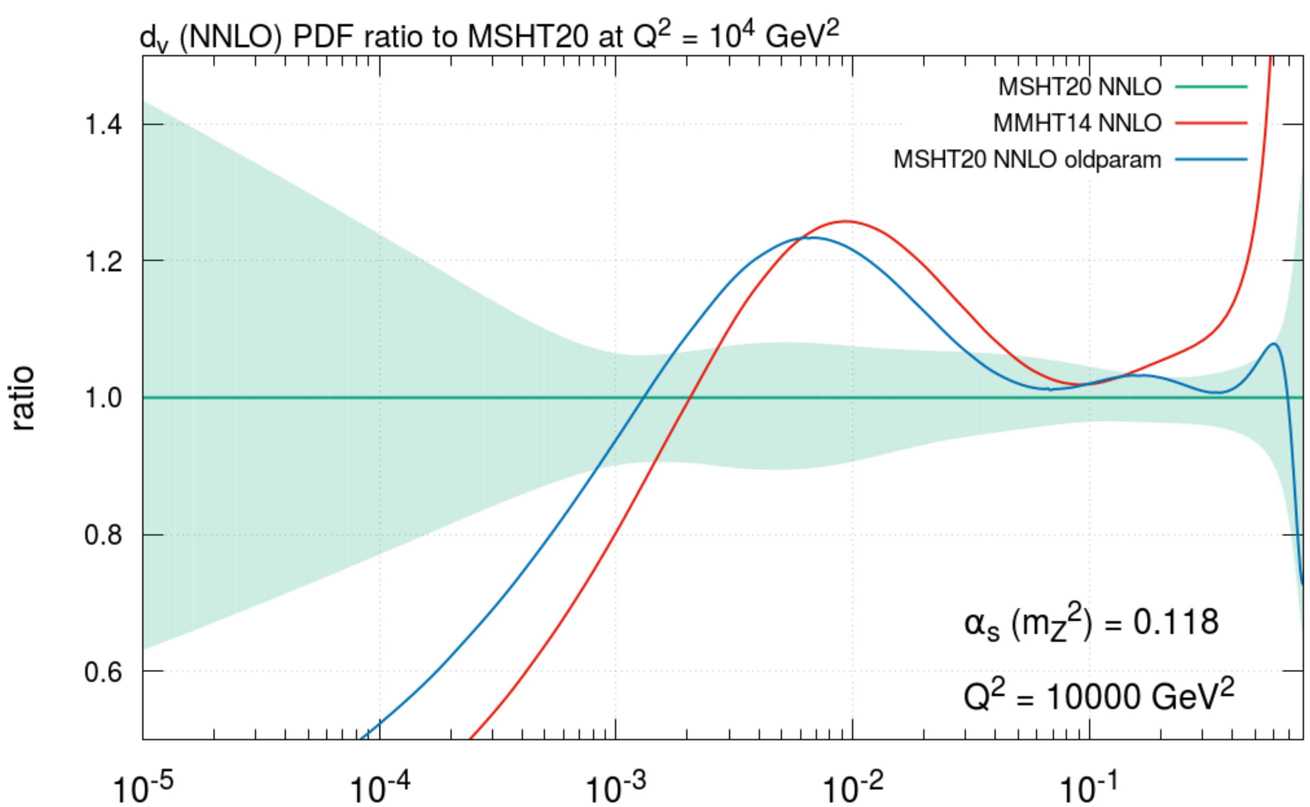
- Necessary to continually assess PDF parameterisation to account for increasingly **precise** data.
- MSHT20 - based on **Chebyshev polynomials** as in MMHT14:

$$xf(x, Q_0^2) \; = \; A(1-x)^\eta x^\delta \left(1 + \sum_{i=1}^{n} a_i T_i^{\mathrm{Ch}}(y(x))\right)$$

- In MMHT14 we generally took $n = 4$ Chebyshevs. Now take $n = 6$ (and $\bar{d} - \bar{u} \rightarrow \bar{d}/\bar{u}$).
- Gives some improvement in fit quality:

$$\Delta\chi^2/N_{\mathrm{pts}} \sim -0.02$$

- And impact on PDFs. Dominantly in region where constraints limited.



Results: Fit Quality

Data set	NLO	NNLO
BCDMS μp F_2 [49]	169.4/163	180.2/163
BCDMS μd F_2 [49]	135.0/151	146.0/151
NMC μp F_2 [50]	142.9/123	124.1/123
NMC μd F_2 [50]	128.2/123	112.4/123
NMC $\mu n/\mu p$ [51]	127.8/148	130.8/148
E665 μp F_2 [52]	59.5/53	64.7/53
E665 μd F_2 [52]	50.3/53	59.7/53
SLAC ep F_2 [53,54]	29.4/37	32.0/37
SLAC ed F_2 [53,54]	37.4/38	23.0/38
NMC/BCDMS/SLAC/HERA F_L [49,50,54,146–148]	79.4/57	68.4/57
E866/NuSea pp DY [149]	216.2/184	225.1/184
E866/NuSea pd/pp DY [150]	10.6/15	10.4/15
NuTeV νN F_2 [55]	43.7/53	38.3/53
CHORUS νN F_2 [56]	27.8/42	30.2/42
NuTeV νN xF_3 [55]	37.8/42	30.7/42
CHORUS νN xF_3 [56]	22.0/28	18.4/28
CCFR $\nu N \to \mu\mu X$ [57]	73.2/86	67.7/86
NuTeV $\nu N \to \mu\mu X$ [57]	41.0/84	58.4/84
HERA e^+p CC [84]	54.3/39	52.0/39
HERA e^-p CC [84]	80.4/42	70.2/42
HERA e^+p NC 820 GeV [84]	91.6/75	89.8/75
HERA e^+p NC 920 GeV [84]	553.9/402	512.7/402
HERA e^-p NC 460 GeV [84]	253.3/209	248.3/209
HERA e^-p NC 575 GeV [84]	268.1/259	263.0/259
HERA e^-p NC 920 GeV [84]	252.3/159	244.4/159
HERA ep $F_2^{\rm charm}$ [26]	125.6/79	132.3/79
DØ II $p\bar{p}$ incl. jets [125]	117.2/110	120.2/110
CDF II $p\bar{p}$ incl. jets [124]	70.4/76	60.4/76
CDF II W asym. [90]	19.1/13	19.0/13
DØ II $W \to \nu e$ asym. [151]	44.4/12	33.9/12
DØ II $W \to \nu\mu$ asym. [152]	13.9/10	17.3/10
DØ II Z rap. [153]	15.9/28	16.4/28
CDF II Z rap. [154]	36.9/28	37.1/28
DØ W asym. [21]	13.1/14	12.0/14

Data set	NLO	NNLO
ATLAS W^+, W^-, Z [119]	34.7/30	29.9/30
CMS W asym. $p_T > 35$ GeV [155]	11.8/11	7.8/11
CMS asym. $p_T > 25, 30$ GeV [156]	11.8/24	7.4/24
LHCb $Z \to e^+e^-$ [157]	14.1/9	22.7/9
LHCb W asym. $p_T > 20$ GeV [158]	10.5/10	12.5/10
CMS $Z \to e^+e^-$ [159]	18.9/35	17.9/35
ATLAS High-mass Drell-Yan [160]	20.7/13	18.9/13
CMS double diff. Drell-Yan [72]	222.2/132	144.5/132
Tevatron, ATLAS, CMS $\sigma_{t\bar{t}}$ [93]- [94]	22.8/17	14.5/17
LHCb 2015 W, Z [95,96]	114.4/67	99.4/67
LHCb 8 TeV $Z \to ee$ [97]	39.0/17	26.2/17
CMS 8 TeV W [98]	23.2/22	12.7/22
ATLAS 7 TeV jets [18]	226.2/140	221.6/140
CMS 7 TeV $W + c$ [99]	8.2/10	8.6/10
ATLAS 7 TeV high precision W, Z [20]	304.7/61	116.6/61
CMS 7 TeV jets [100]	200.6/158	175.8/158
CMS 8 TeV jets [101]	285.7/174	261.3/174
CMS 2.76 TeV jet [107]	124.2/81	102.9/81
ATLAS 8 TeV Z p_T [75]	235.0/104	188.5/104
ATLAS 8 TeV single diff $t\bar{t}$ [102]	39.1/25	25.6/25
ATLAS 8 TeV single diff $t\bar{t}$ dilepton [103]	4.7/5	3.4/5
CMS 8 TeV double differential $t\bar{t}$ [105]	32.8/15	22.5/15
CMS 8 TeV single differential $t\bar{t}$ [108]	12.9/9	13.2/9
ATLAS 8 TeV High-mass Drell-Yan [73]	85.8/48	56.7/48
ATLAS 8 TeV W [106]	84.6/22	57.4/22
ATLAS 8 TeV W + jets [104]	33.9/30	18.1/30
ATLAS 8 TeV double differential Z [74]	157.4/59	85.6/59

→

	NLO	**NNLO**
Total, LHC data in MSHT20	1.79	1.33
Total, non-LHC data in MSHT20	1.13	1.10
Total, all data	1.33	1.17

~700 points in χ^2 $(\sim 7\sigma)$

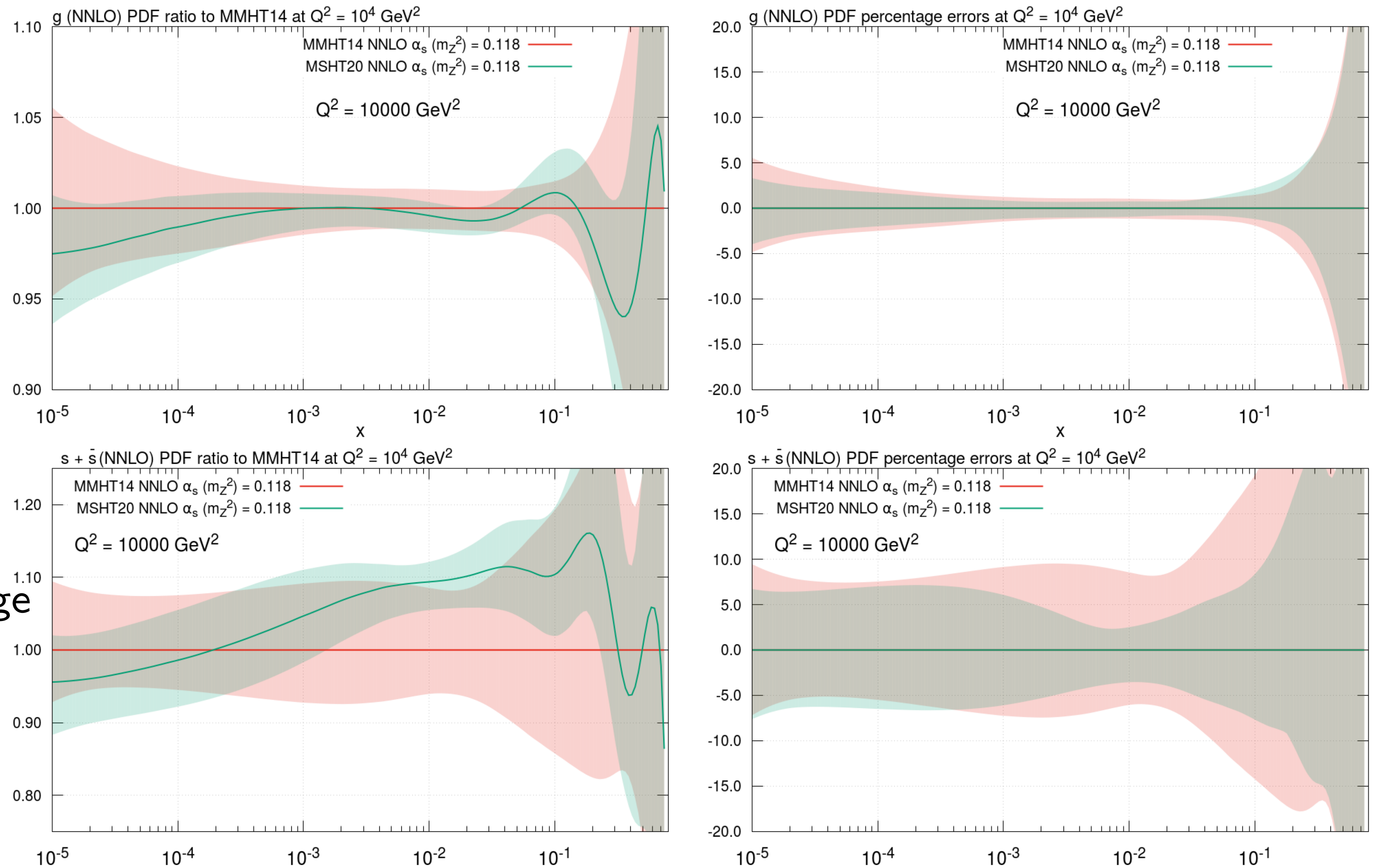
- Global **fit quality** very good at NNLO. NLO clearly worse.
- Mainly driven by new high precision **LHC data** in fit, where even NNLO fit quality gets worse.
- We have produced a LO fit for completeness, but the fit quality is now extremely poor.

$\longrightarrow$ Importance of NNLO theory very clear, but also points the way to including **theory uncertainties** (N3LO…) in future.

Results: PDFs

Gluon

Strange
-ness



Changes in high x gluon

← Reduced uncertainty

More data here -

Jets, top, Zp_T.

Increased Strangeness

← Reduced uncertainty

Both due to new

ATLAS 7, 8 TeV

W, Z data.

- Nice **reduction in uncertainties** for gluon & light sea. Central values generally within errors. Similar effect seen in benchmark cross sections.

Backup

Strong coupling + heavy quarks

An investigation of the α_S and heavy quark mass dependence in the MSHT20 global PDF analysis

T. Cridge[a], L. A. Harland-Lang[b], A. D. Martin[c], and R.S. Thorne[a]

[a] Department of Physics and Astronomy, University College London, London, WC1E 6BT, UK

[b] Rudolf Peierls Centre, Beecroft Building, Parks Road, Oxford, OX1 3PU

[c] Institute for Particle Physics Phenomenology, Durham University, Durham, DH1 3LE, UK

Abstract

We investigate the MSHT20 global PDF sets, demonstrating the effects of varying the strong coupling $\alpha_S(M_Z^2)$ and the masses of the charm and bottom quarks. We determine

T. Cridge et al., *Eur.Phys.J.C* 81 (2021) 744

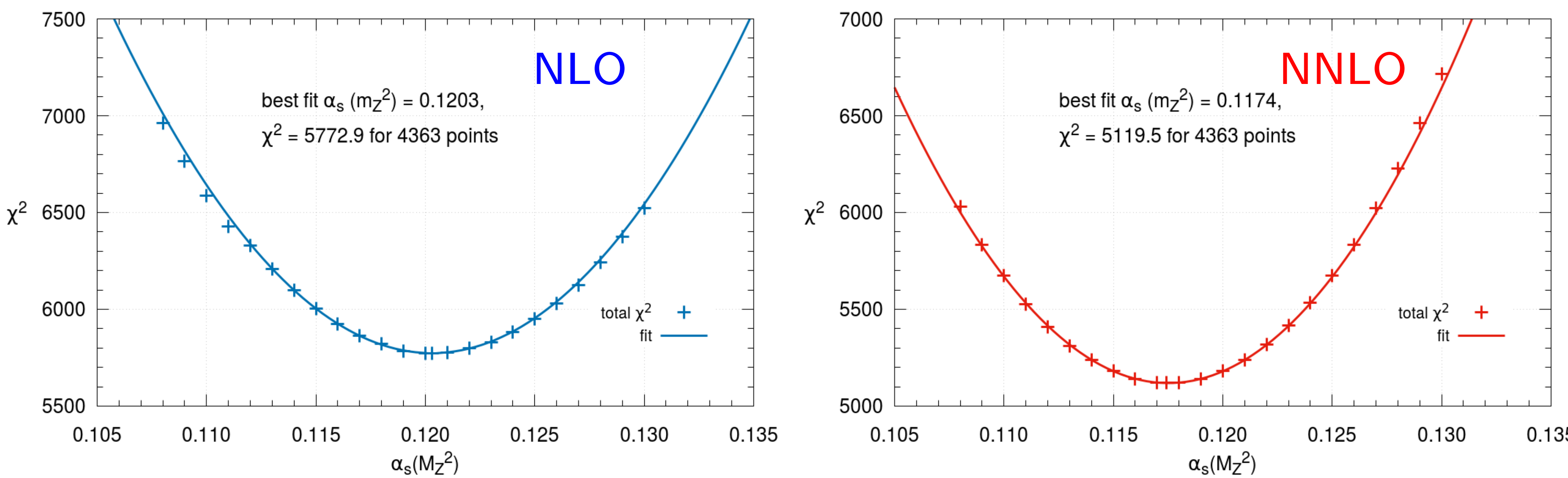
- Have followed up baseline fit with dedicated study on strong coupling and heavy quark mass dependence.

MSHT20: Strong Coupling

- Default PDFs provided at standard fixed value of $\alpha_S(M_Z^2) = 0.118$.
- Can also allow α_S to be a free parameter in the fit.
- Global fit nature of PDFs $\Rightarrow$ can provide a precise, accurate determination of α_S.
- Individual datasets have different dependences on α_S, but robust determination utilising all datasets.
- The best fit values are found to be:

$\alpha_{S,\mathrm{NNLO}}(M_Z^2) < \alpha_{S,\mathrm{NLO}}(M_Z^2)$ as NNLO corrections +ve, so fitting same data $\Rightarrow$ lower α_S.

$$\alpha_{S,\mathrm{NLO}}(M_Z^2) = 0.1203 \qquad \alpha_{S,\mathrm{NNLO}}(M_Z^2) = 0.1174$$



Nice Quadratic χ^2 profile ✓

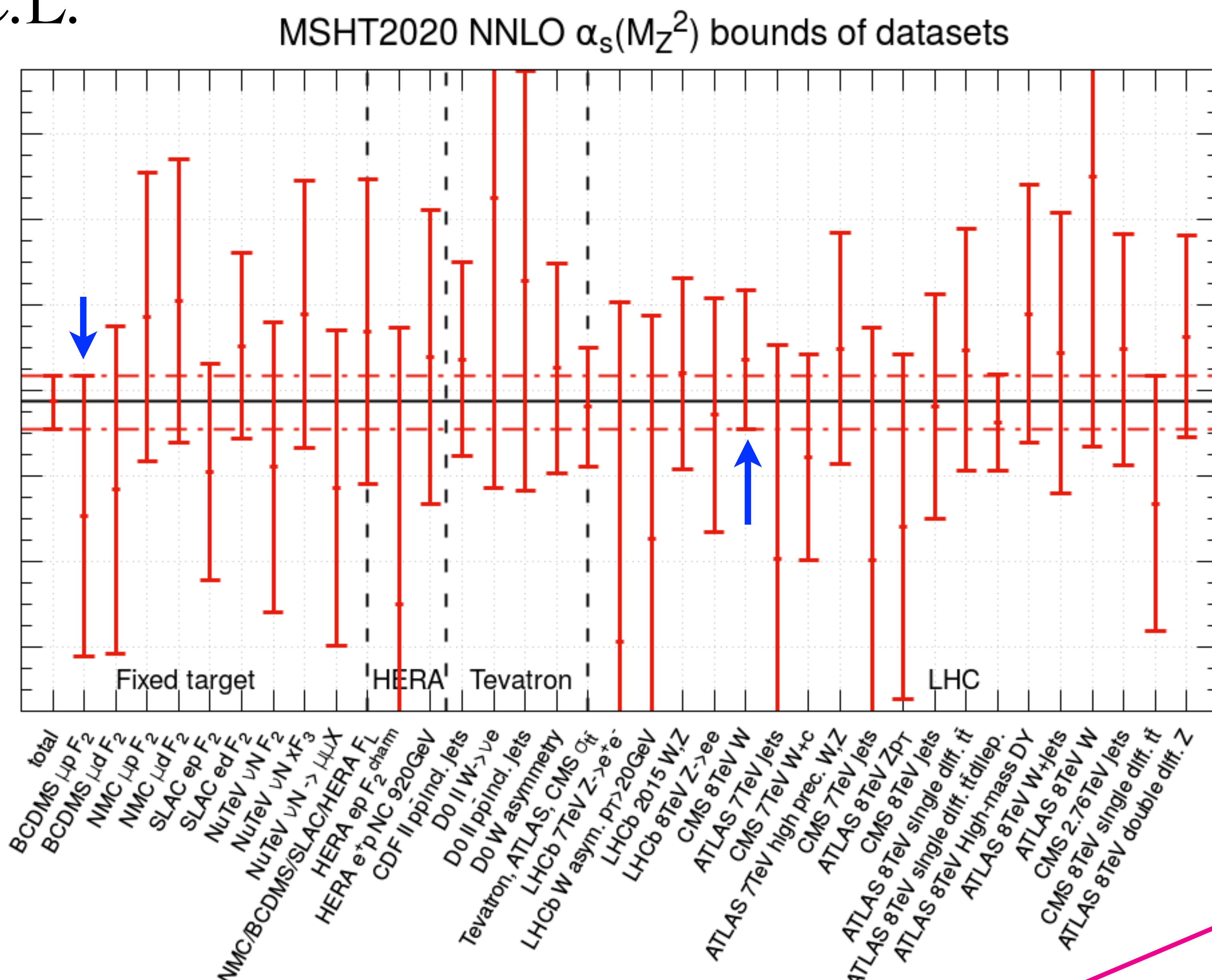
NNLO Strong Coupling Determination

- Set using dynamical tolerance: point in $\Delta\chi_i^2$ for each dataset that exceeds 68% C.L.

BCDMSp data strongest constraint upwards: $\Delta\alpha_S(M_Z^2) = +0.0012$.

SLACp data provides upwards bound of: $\Delta\alpha_S(M_Z^2) = +0.0018$.

CMS and ATLAS (dilepton) $t\bar{t}$ single diff. prefer lower α_S but bound not used.



CMS 8 TeV *W* data gives lower bound: $\Delta\alpha_S(M_Z^2) = -0.0013$.

ATLAS 8 TeV *Z*, SLACd, ATLAS 8 TeV High-mass give lower bounds of $\Delta\alpha_S(M_Z^2) = -0.0017, -0.001$ -0.0019

Corresponds to: $\Delta\chi^2_{\rm global} = 17$

- Therefore upper and lower bounds are +0.0012 and -0.0013.

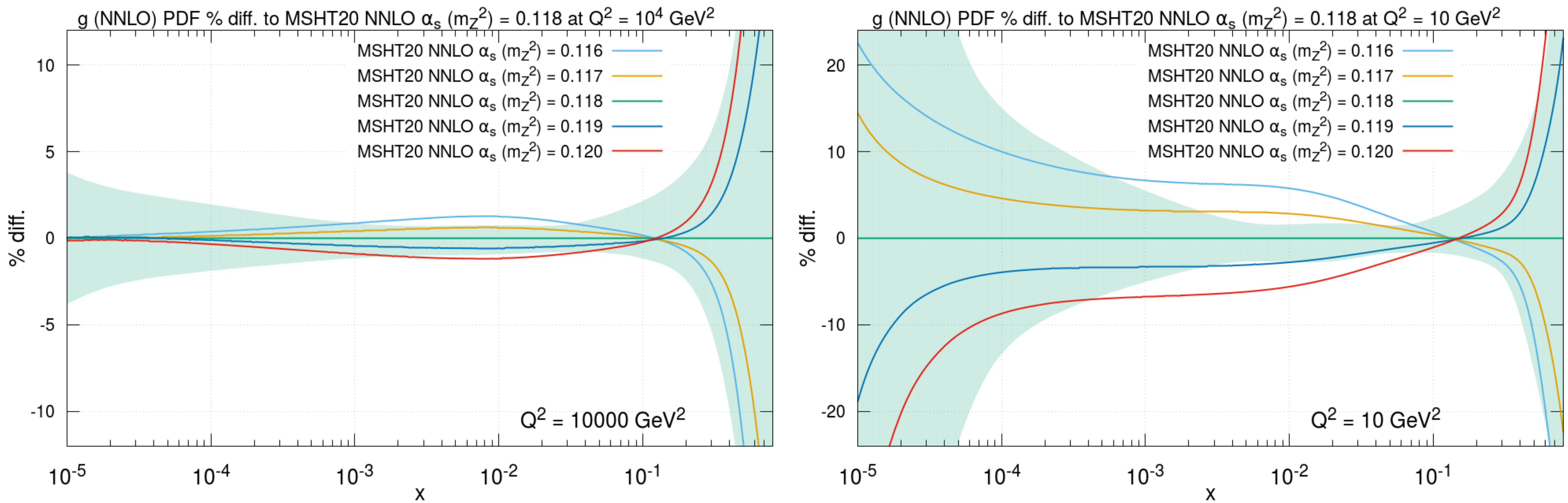
$$\alpha_{S,\rm NNLO}(M_Z^2) = 0.1174 \pm 0.0013$$

Consistent with World Average of 0.1179 ± 0.0009.

Impact on PDFs Gluon

- Correlations between PDFs and α_S.

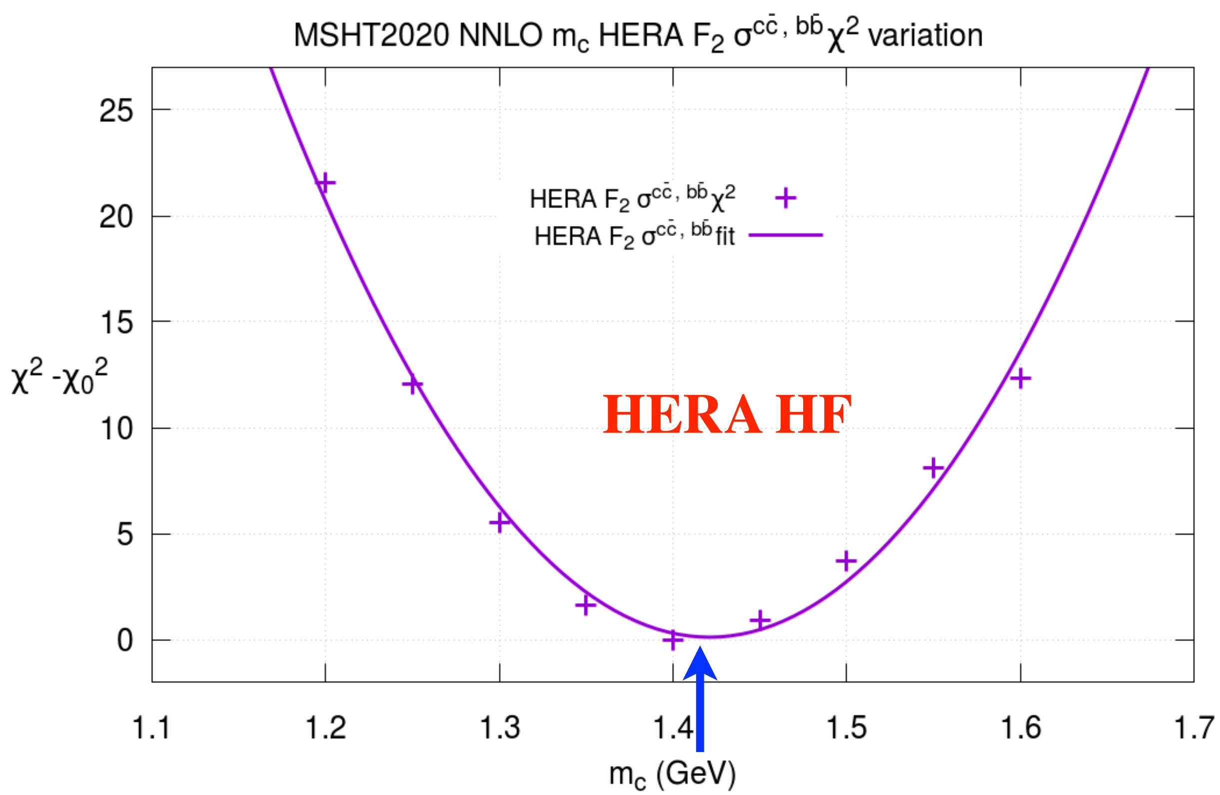
Changes of PDFs generally within PDF uncertainties, certainly at larger scales for $\Delta\alpha_S(M_Z^2) = \pm 0.001$.

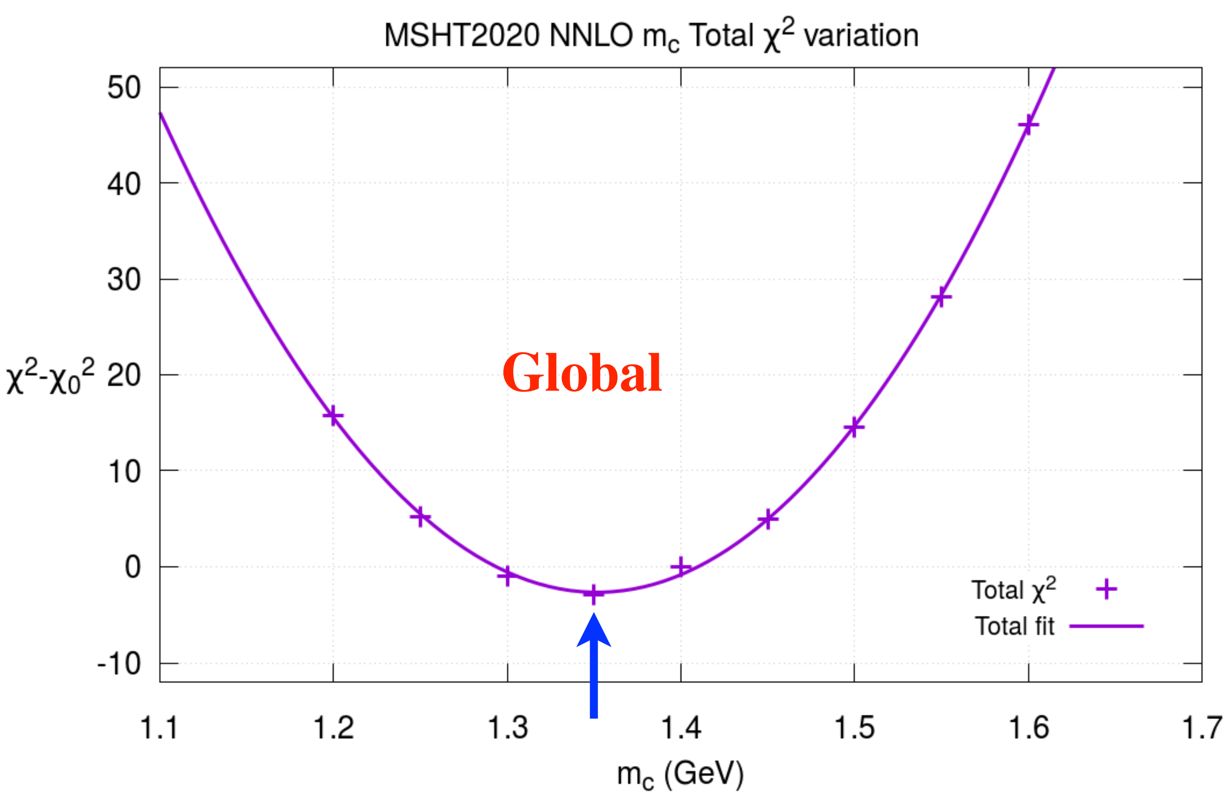


- Gluon anti-correlated with $\alpha_S(M_Z^2)$ for $x \lesssim 0.1$ as maintains product $\alpha_S g$ for structure functions.
- Gluon therefore correlated with $\alpha_S(M_Z^2)$ at high $x \gtrsim 0.1$ due to momentum sum rule.

(For quarks see backup)

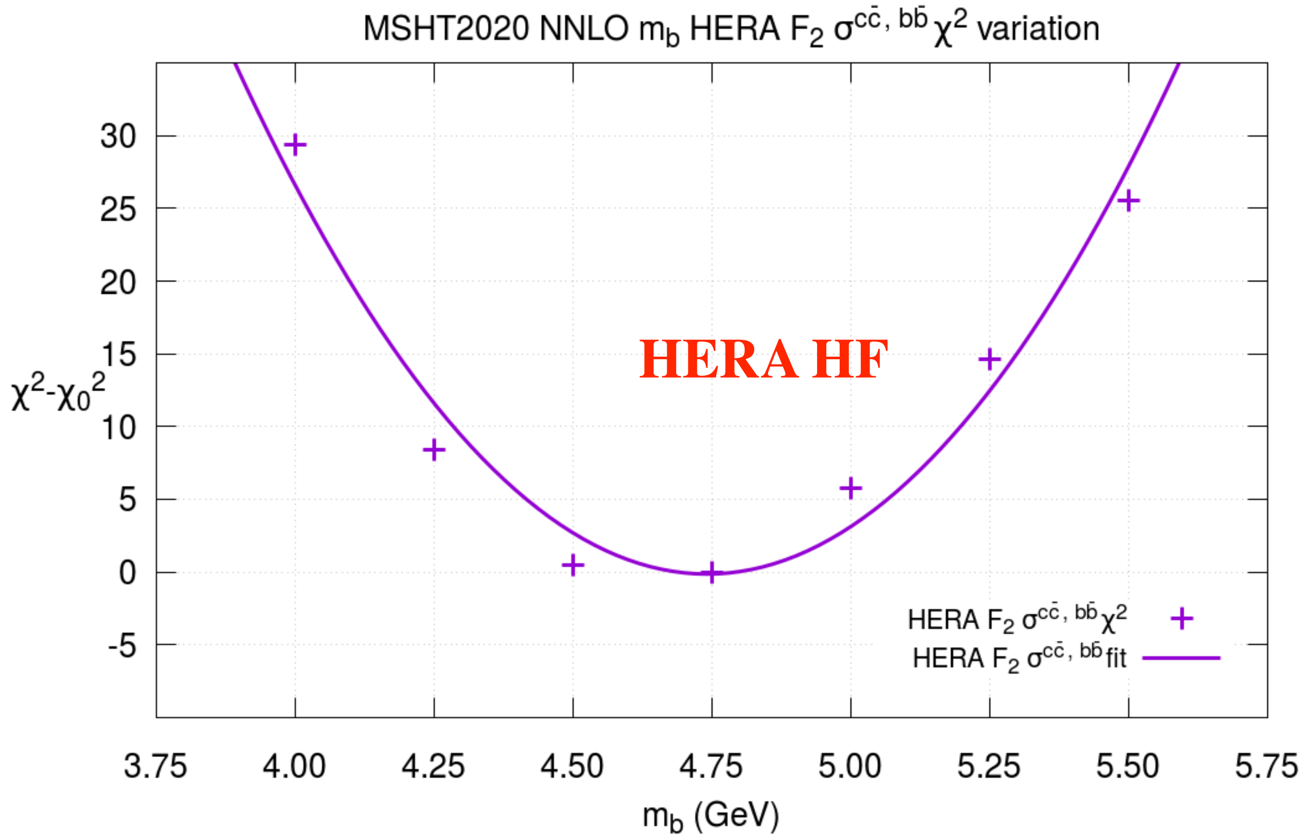
Heavy Quarks: Charm

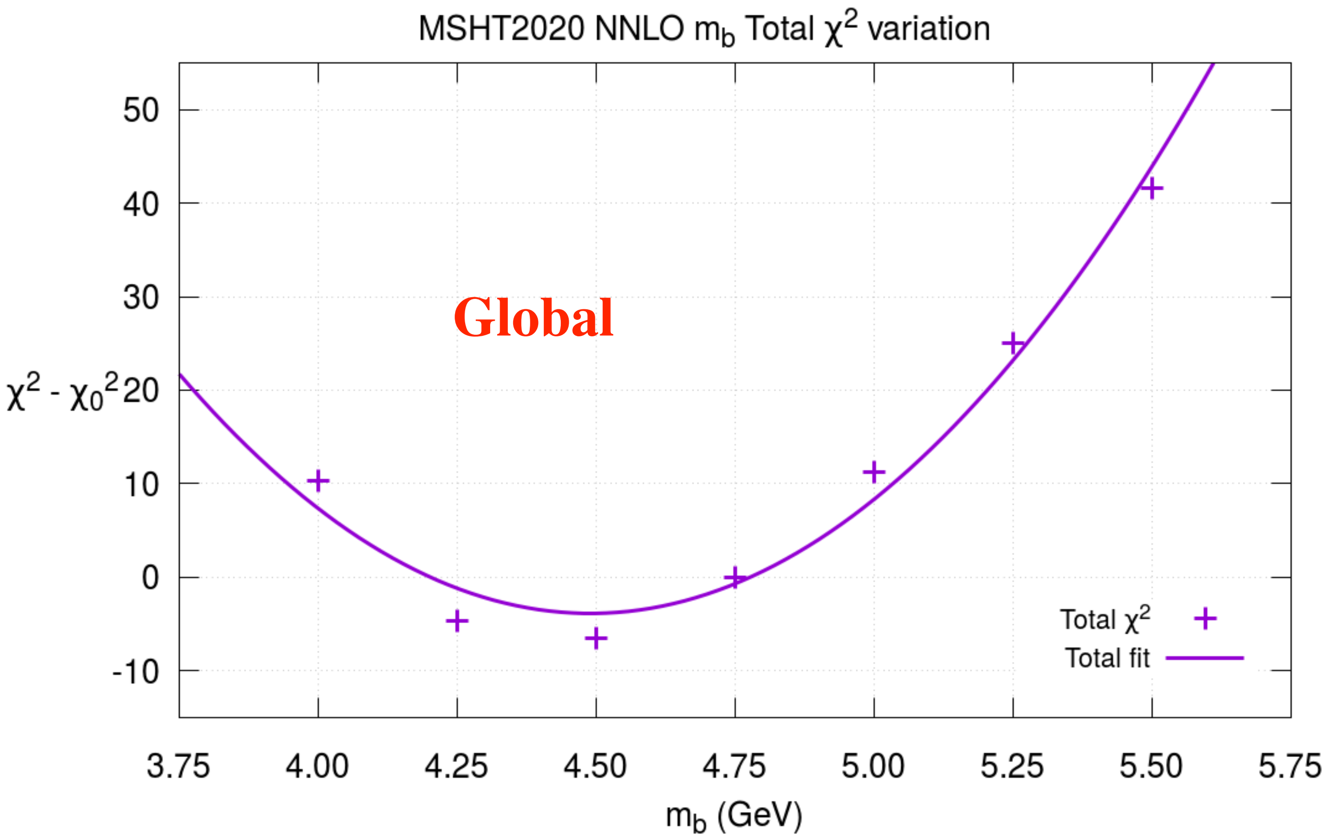




- Similar procedure to determine dependence on heavy quark pole masses.
- Notable addition v.s MSHT14: final combined HERA charm + bottom data.
- Broad consistency between this + global profiles, with some mild tension.
- We do not make a detailed determination (other methods to do that), but not these are ~ consistent with our default value of $m_c = 1.4\,\text{GeV}$.

Heavy Quarks: Bottom





- Again broad consistency between global and HERA profiles, with latter preferring slightly larger values. Less well determined.
- Consistent with our default of $m_b = 4.75\,\text{GeV}$.

Benchmarks

- Evaluate uncertainty on benchmark cross section, including full PDF dependence: $\alpha_s \pm 0.001$ $m_{c,b} \pm \sim 10\%$

	σ	PDF unc.	m_c var.	m_b var.
W Tevatron (1.96 TeV)	2.705	$^{+0.054}_{-0.057}$ $\left(^{+2.0\%}_{-2.1\%}\right)$	$^{+0.010}_{-0.013}$ $\left(^{+0.37\%}_{-0.47\%}\right)$	$^{-0.0079}_{+0.0029}$ $\left(^{-0.29\%}_{+0.11\%}\right)$
$t\bar{t}$ LHC (13 TeV)	796.8	$^{+16.0}_{-10.6}$ $\left(^{+2.0\%}_{-1.3\%}\right)$	$^{+2.9}_{-2.6}$ $\left(^{+0.36\%}_{-0.33\%}\right)$	$^{+2.0}_{-2.2}$ $\left(^{+0.25\%}_{-0.27\%}\right)$
Higgs LHC (13 TeV)	42.13	$^{+0.47}_{-0.51}$ $\left(^{+1.1\%}_{-1.2\%}\right)$	$^{+0.27}_{-0.23}$ $\left(^{+0.64\%}_{-0.57\%}\right)$	$^{+0.14}_{-0.16}$ $\left(^{+0.32\%}_{-0.38\%}\right)$

	σ	PDF unc.	α_S unc.
W Tevatron (1.96 TeV)	2.705	$^{+0.054}_{-0.057}$ $\left(^{+2.0\%}_{-2.1\%}\right)$	$^{+0.018}_{-0.017}$ $\left(^{+0.66\%}_{-0.61\%}\right)$
$t\bar{t}$ LHC (13 TeV)	796.8	$^{+16.0}_{-10.6}$ $\left(^{+2.0\%}_{-1.3\%}\right)$	$^{+12}_{-13}$ $\left(^{+1.5\%}_{-1.6\%}\right)$
Higgs LHC (13 TeV)	42.13	$^{+0.47}_{-0.51}$ $\left(^{+1.1\%}_{-1.2\%}\right)$	$^{+0.64}_{-0.65}$ $\left(^{+1.5\%}_{-1.5\%}\right)$

- Uncertainty due to strong coupling **similar to/larger than** PDF uncertainty. Quark mass dependence lower.

Impact of dijet data

Preliminary

Dijet Data at the LHC

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

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

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

- In what follows will also compare to ATLAS/CMS data on inclusive jets at 7 and 8 TeV.

- At NNLO, we find:

Dijet fit:

	N_{pts}	χ^2/N_{pt}
ATLAS 7 TeV dijets	90	1.05
CMS 7 TeV dijets	54	1.43
CMS 8 TeV dijets	122	1.04
Total Dijets	266	1.12

	N_{pts}	χ^2/N_{pt}
ATLAS Z $p_\perp$	104	1.65
Diff. top	54	1.24
7 + 8 TeV Jets	643	[1.62]

Prediction

Jet fit:

	N_{pts}	χ^2/N_{pt}
ATLAS 7 TeV jets	140	1.53
ATLAS 8 TeV jets	171	1.45
CMS 7 TeV jets	158	1.22
CMS 8 TeV jets*	174	1.80
Total Jets	643	1.50

	N_{pts}	χ^2/N_{pt}
ATLAS Z $p_\perp$	104	1.85
Diff. top	54	1.12
7 + 8 TeV Dijets	643	[1.32]

★ Fit quality to dijet data very good (1.12), clearly worse for jets (1.50).

★ No signs of significant inconsistency in fit vs. predicted χ^2, though some difference in pull implied.

★ Fit quality to top (Z $p_\perp$) data better in jet (dijet) fit. Latter particularly notable.

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

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

★ **NNLO QCD** corrections.

Jets fit:

	N_{pts}	NLO	NNLO
ATLAS 7 TeV jets	140	1.69	1.53
ATLAS 8 TeV jets	171	2.37	1.45
CMS 7 TeV jets	158	1.38	1.22
CMS 8 TeV jets	174	1.65	1.80
Total Jets	643	1.78	1.50

Dijet fit:

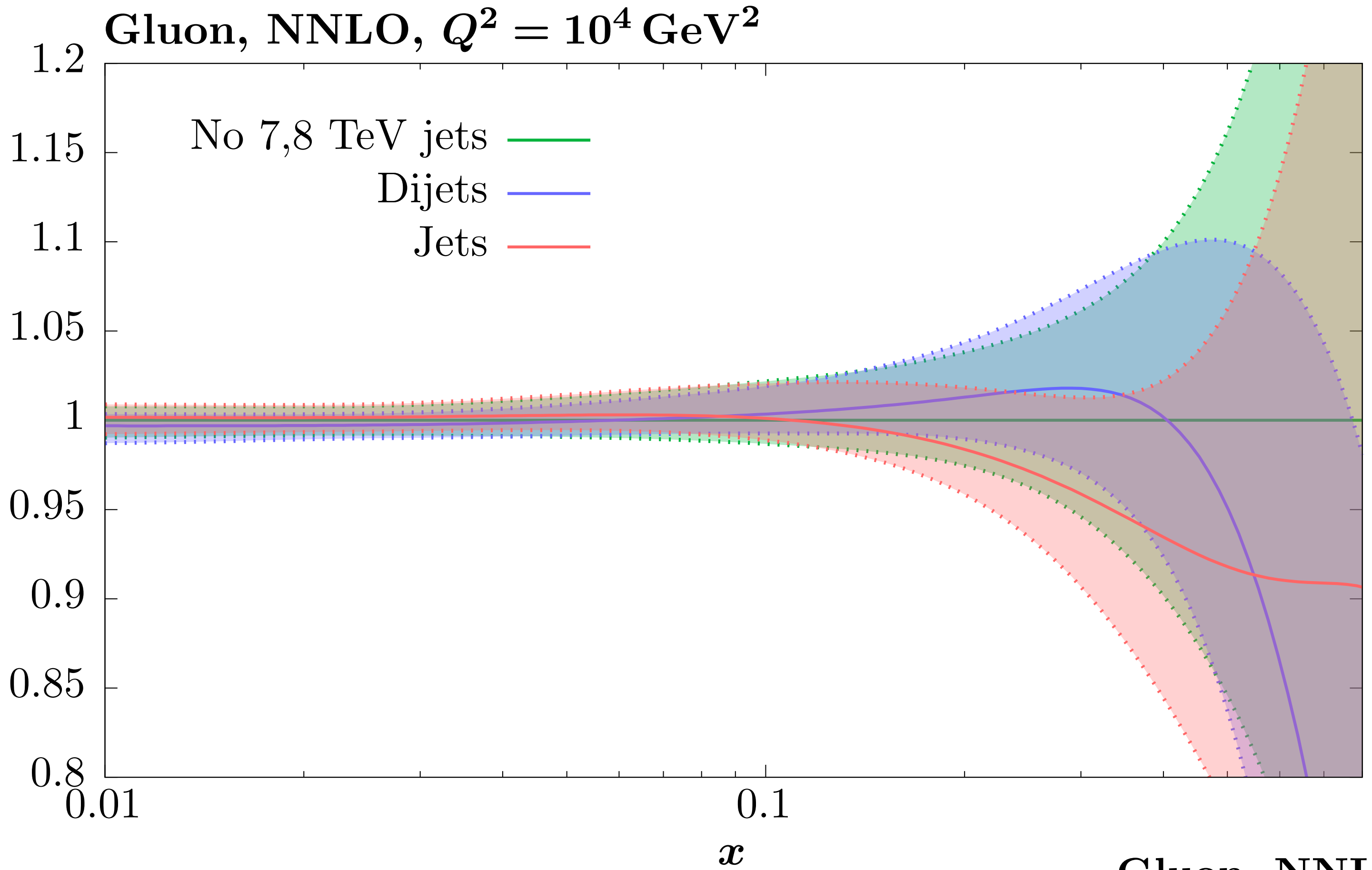
	N_{pts}	NLO	NNLO
ATLAS 7 TeV dijets	90	1.10	1.05
CMS 7 TeV dijets	54	1.71	1.43
CMS 8 TeV dijets	122	5.30	1.04
Total Dijets	266	3.15	1.12

Not a typo!

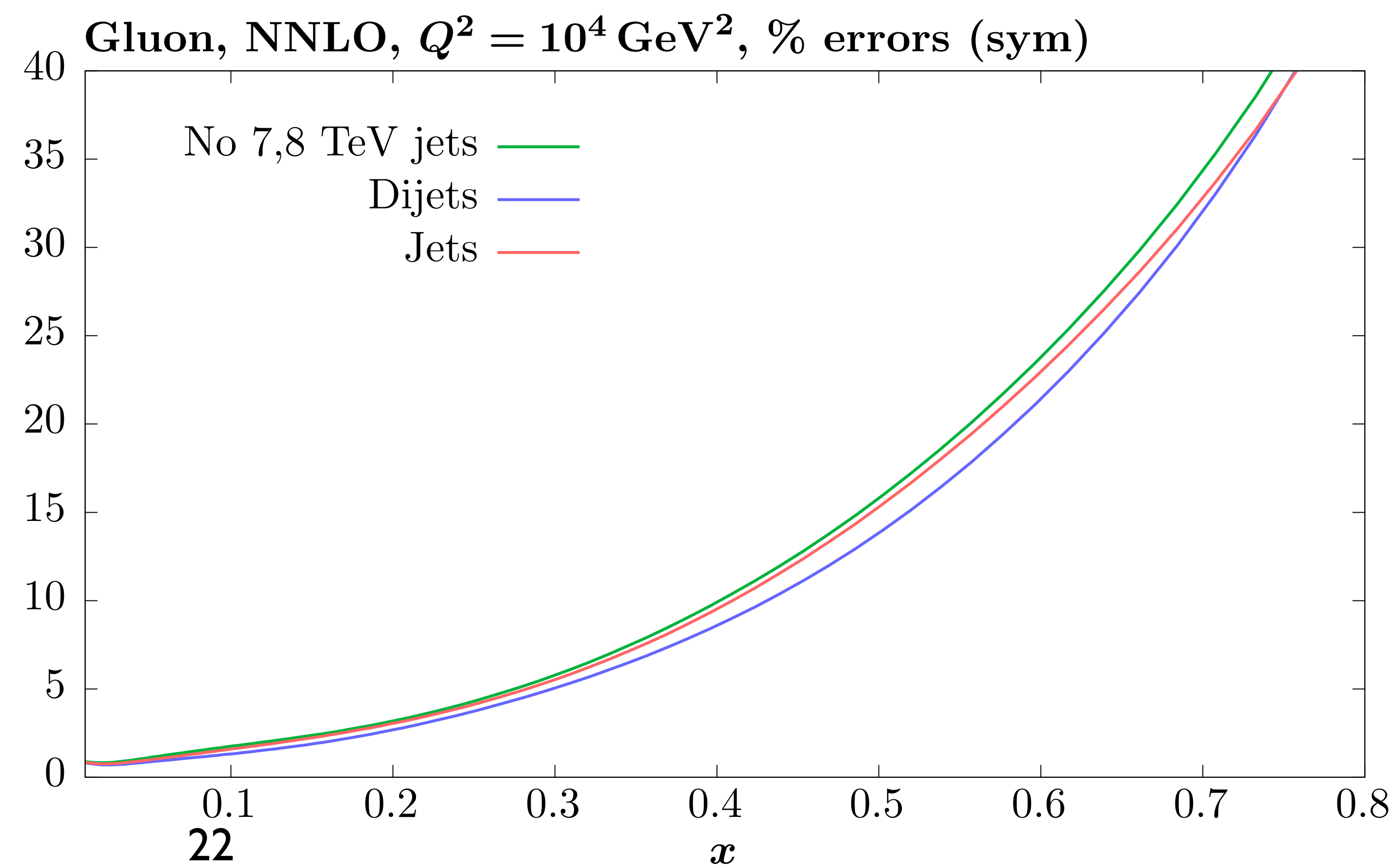
- ★ Clear trend in both cases for QCD corrections to improve fit quality!
- ★ For jets, this is different to arXiv:2005.11327 trend (though same as in MSHT20), but note scale different ($p_\perp^j$ rather than H_T).
- ★ Improvement in CMS 8 TeV dijets remarkable. Not particularly evident by just looking at data/theory. Depends sensitively on correlated systematics.
- ★ Impact of full colour will be very interesting to see, in light of this.
- ★ Impact of EW corrections more mixed (backup).

X. Chen et al., 2204.10173

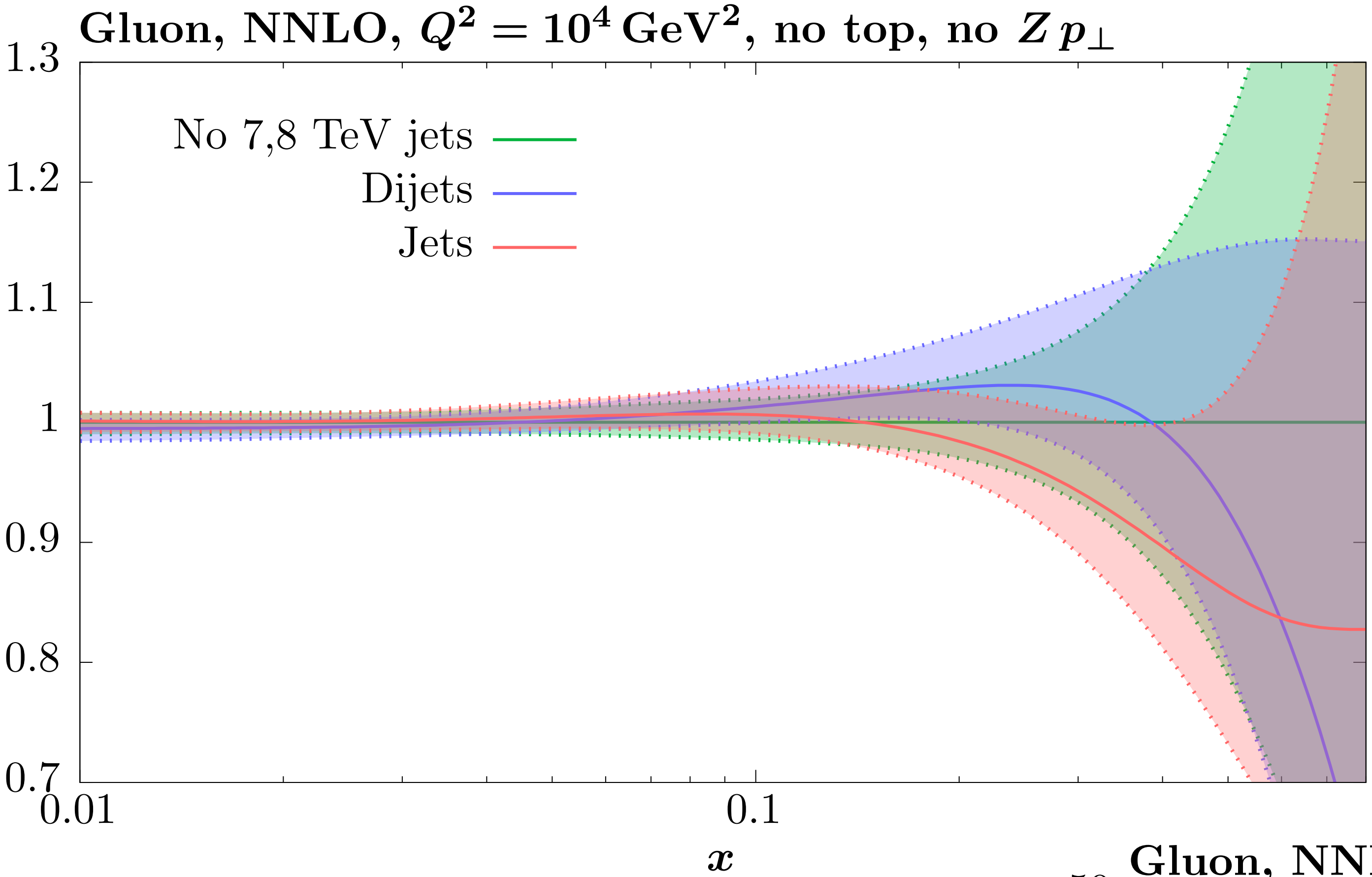
PDFs: Dijets vs. Jets



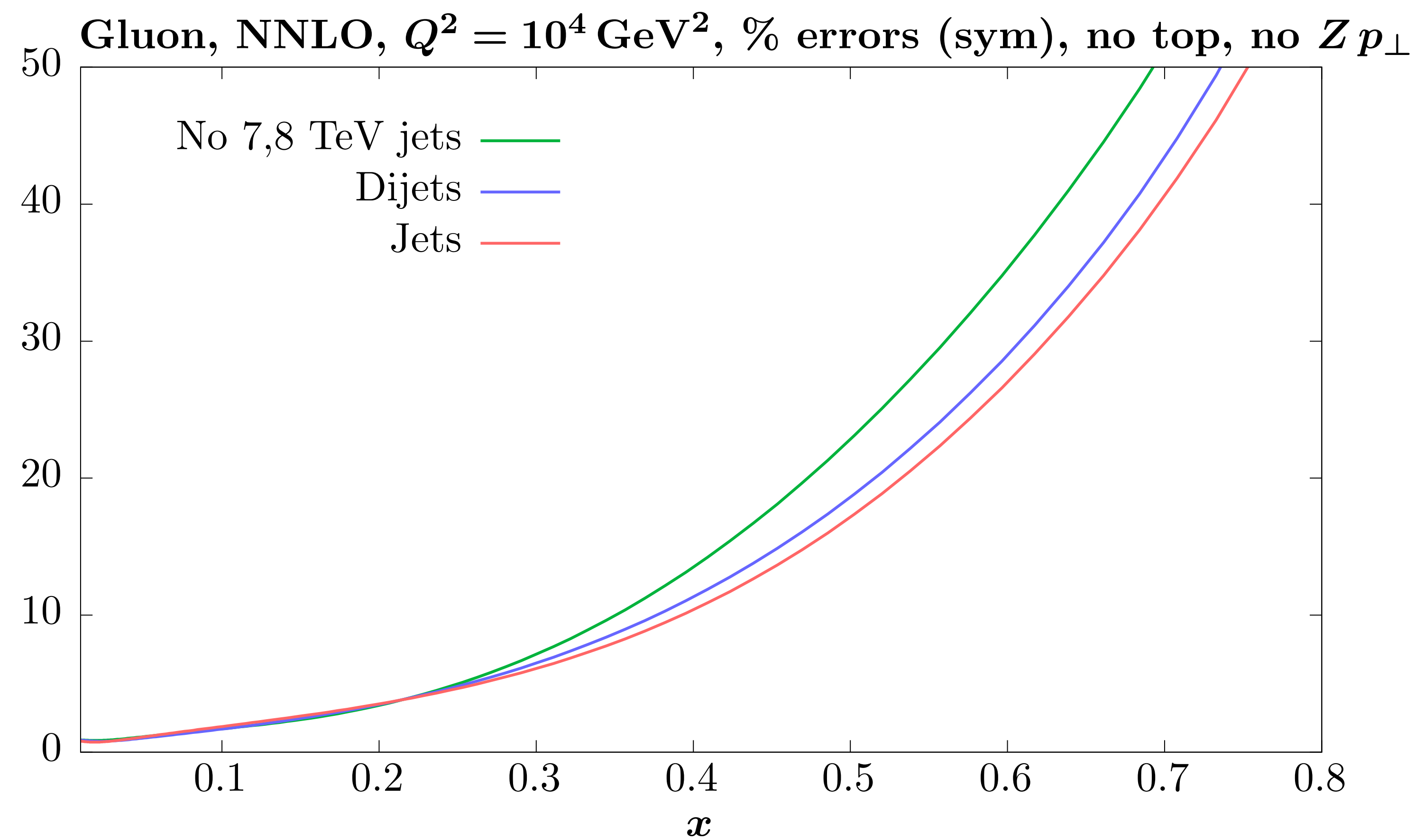
- Add dijets or jets to MSHT20 (no jets) baseline. Focus on gluon here.
- Overall consistency between two cases.
- But pull **rather different**.



- Impact of jets data on gluon uncertainty very mild. Larger for dijets.



- Now consider fits but added to MSHT20 baseline with no diff. top or $Z\,p_\perp$.
- Basic pulls as before.



- However relative impact on uncertainty different. Now jets more significant at high x .

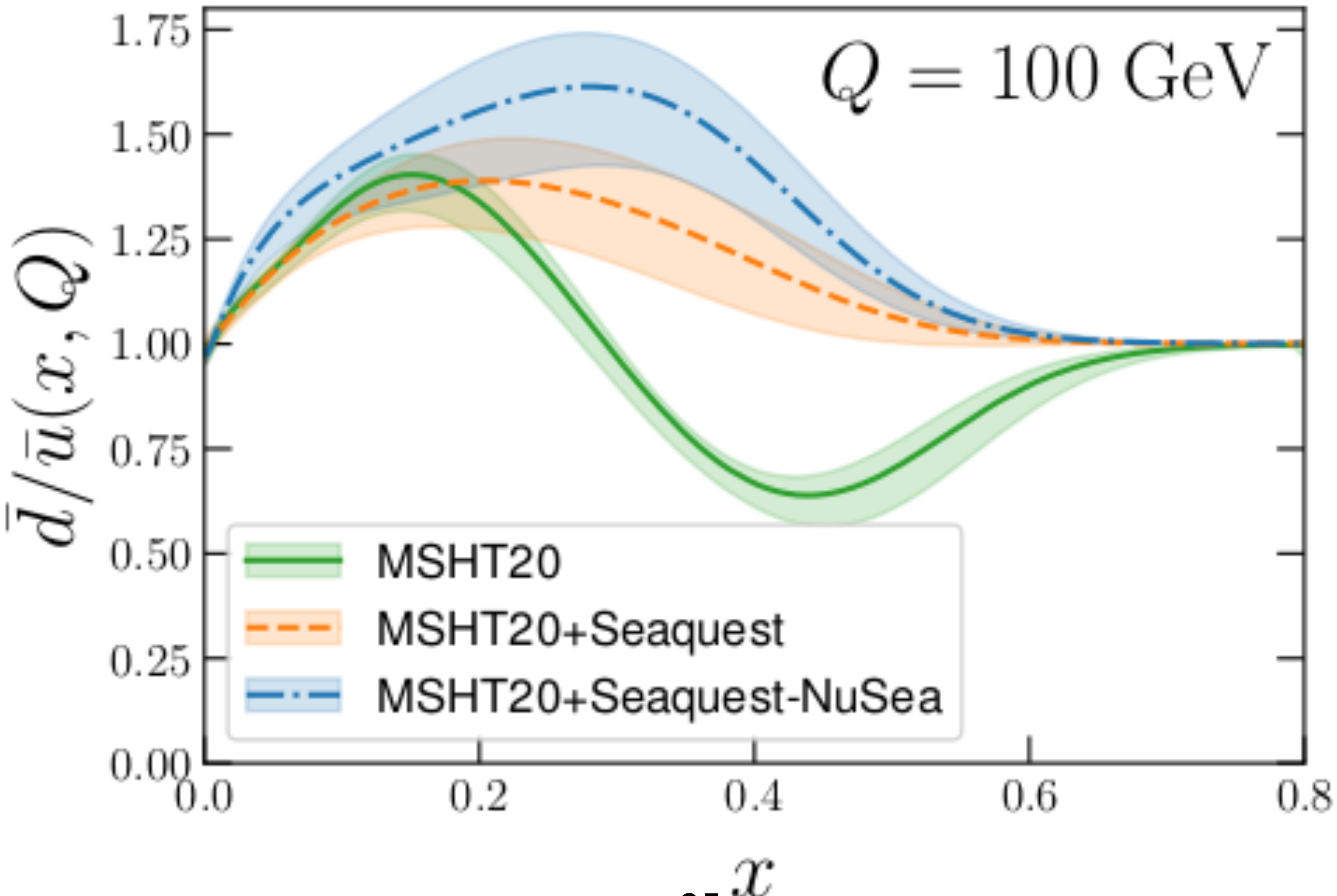
$\longrightarrow$ Clear that interplay with other data in fit drives impact. In particular greater consistency between dijets and $Z\,p_\perp$.

Backup

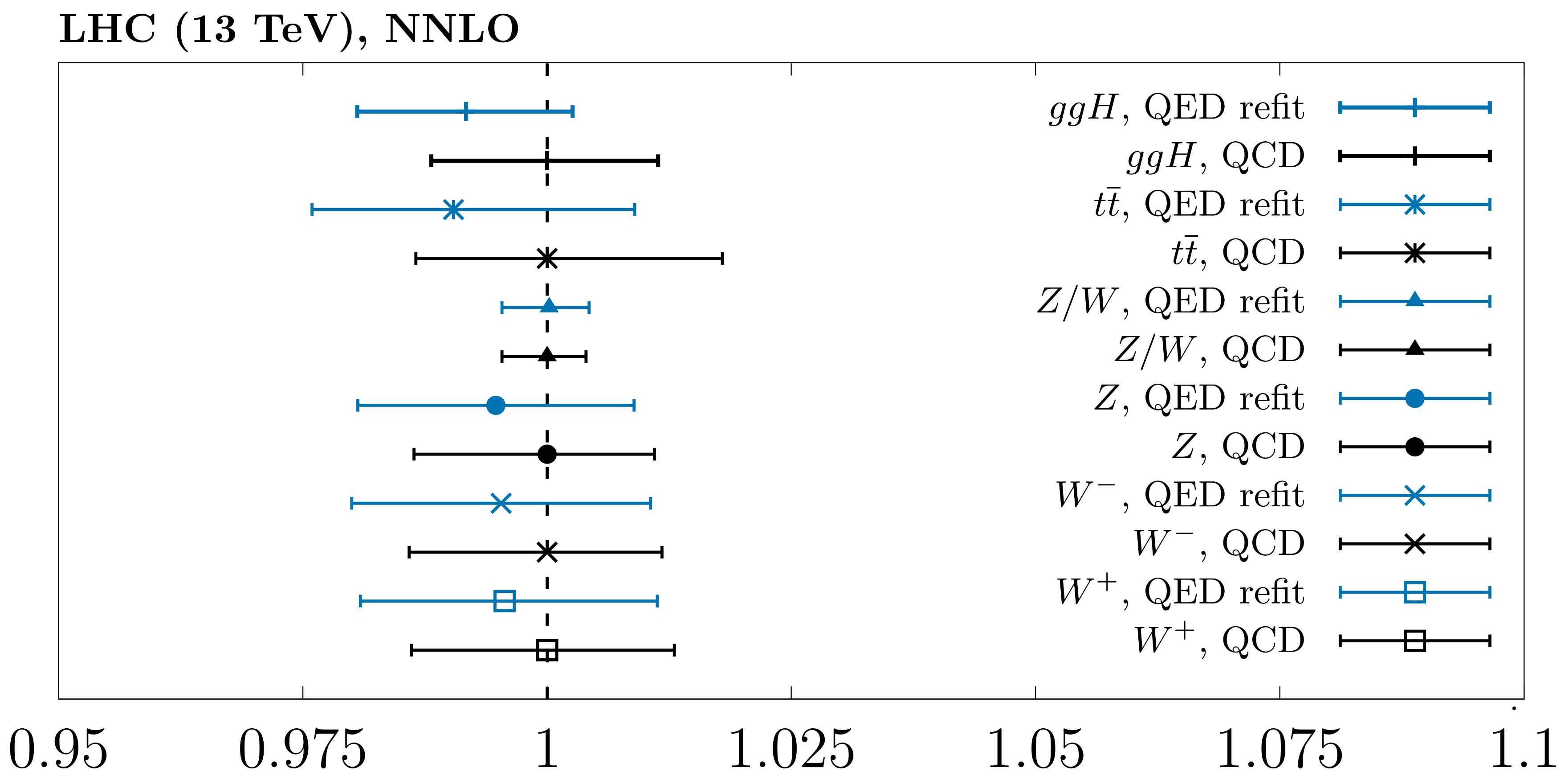
Other Studies

Fitting SeaQuest Data

- **Seaquest** fixed target DY data: sensitivity to high x quarks and sea quark asymmetry in particular.
- We have now included this on top of baseline MSHT20 fit.
- Impact on $\overline{d}/\overline{u}$ ratio at high x clear.
- Removing **NuSea** from fit increases effect, i.e. indicating difference in pulls.



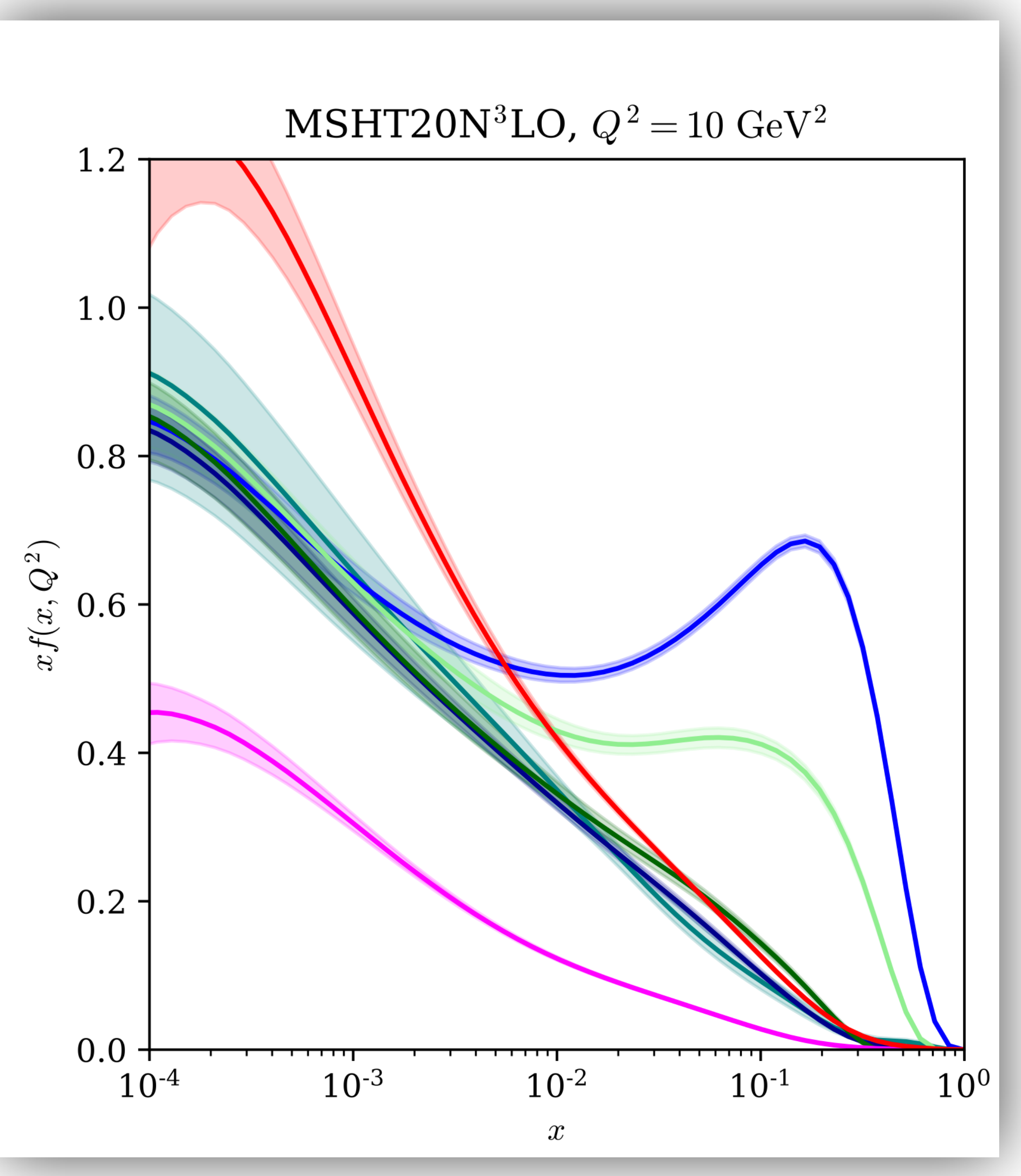
MSHT20qed



- QED corrections + photon PDF included in baseline MSHT20qed fit.
- Impact moderate put not negligible in light of high precision LHC requirements!

For more details see talk by LHL tomorrow - WG3.

MSHT20: approximate N3LO



For more details see talk by Jamie McGowan in next session!

- **New study**: inclusion of approximate N3LO theory in MSHT20 - full where available and approximate + uncertainty where not.

Summary

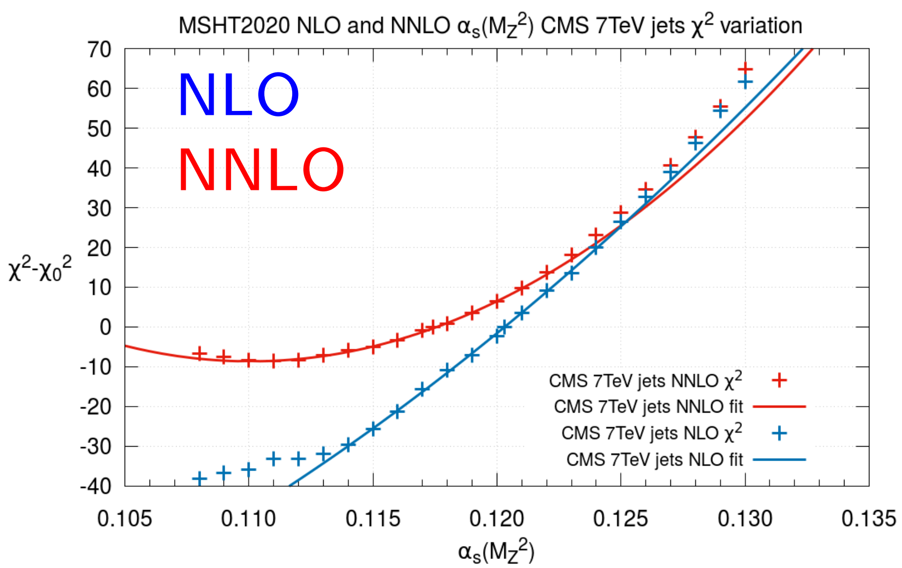
- I have covered a range of topics related to the MSHT project, both ongoing + completed:
  - ★ **MSHT20** - recap.
  - ★ Dedicated study on **strong coupling/heavy quark** masses.
  - ★ Including **dijet** production in MSHT20.
  - ★ Fitting **SeaQuest** data: first look.
  - ★ **MSHT20qed** - including QED corrections.
  - ★ Theoretical uncertainties: approximate higher order corrections.
- In the latter cases stay tuned for details talks today + tomorrow!

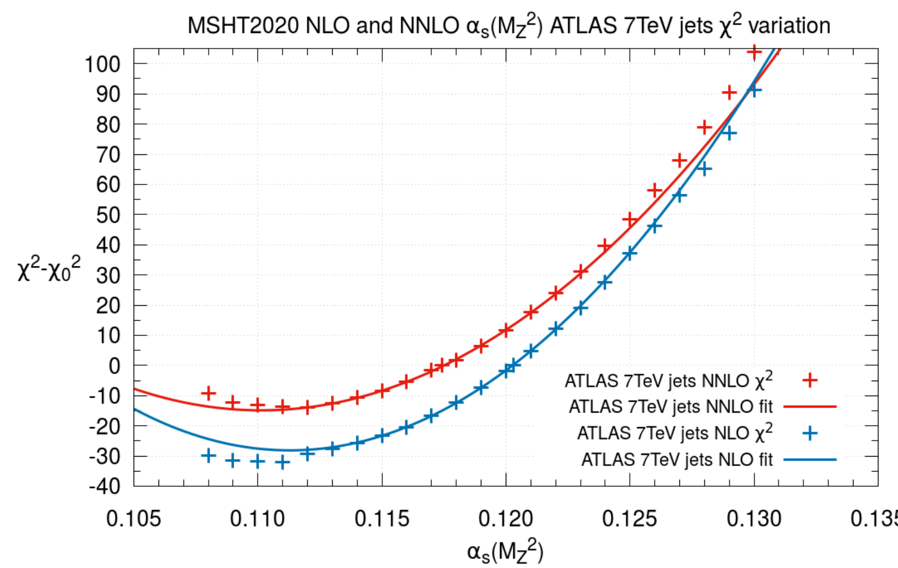
Thank you for listening!

Backup

- In more detail, can determine χ^2 profiles for individual datasets.
- For example, LHC jets + $Z\,p_\perp$:

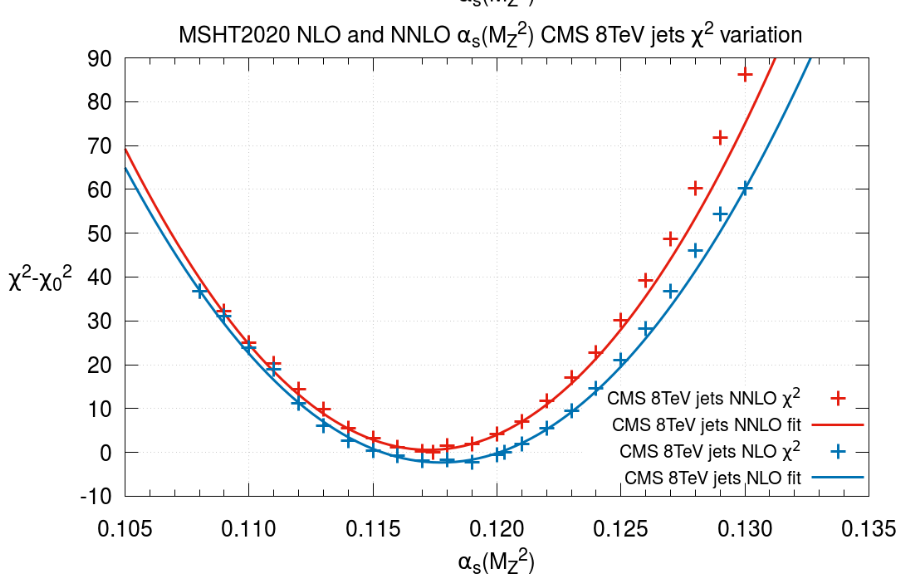
CMS 7 TeV jets prefers lower α_S, better quadratic profile at NNLO.

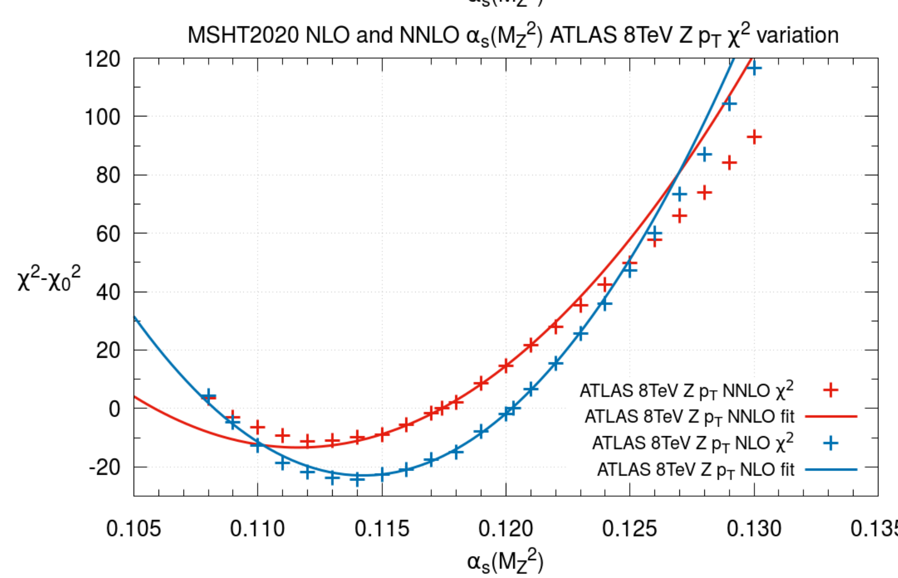




ATLAS 7 TeV jets prefers lower α_S.

CMS 8 TeV jets prefers α_S near best fit. Weak dependence around min, perhaps gluon moderates.





ATLAS 8 TeV Zp_T prefers lower α_S, allows increased high x gluon.

- Jets, Zp_T datasets have direct sensitivity to α_S, prefer lower α_S.

Impact of HO corrections

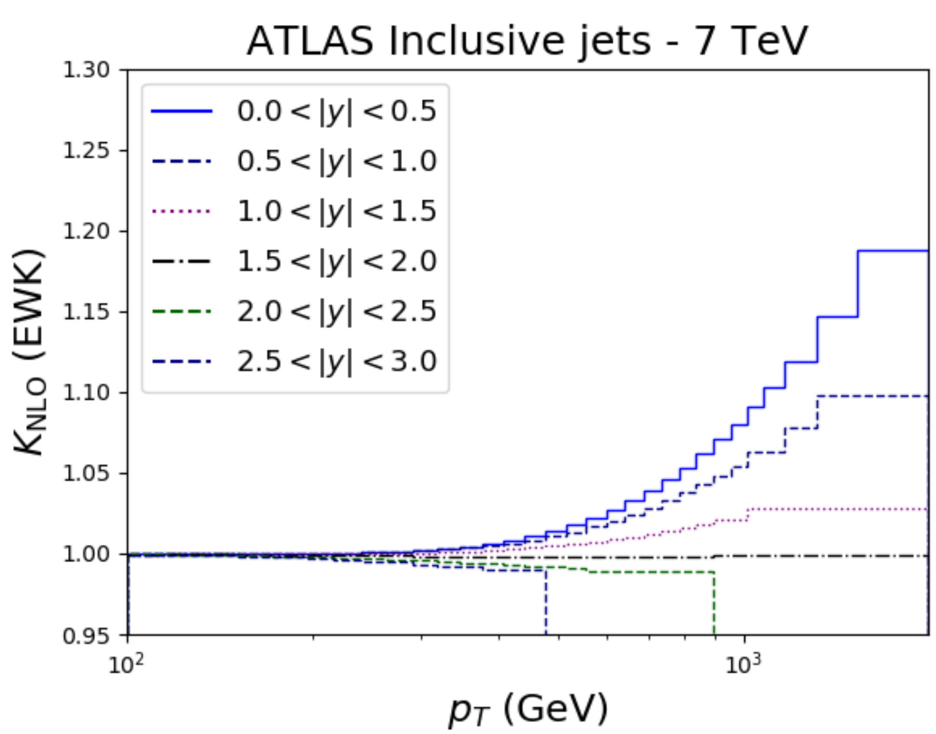
★ **EW** corrections:

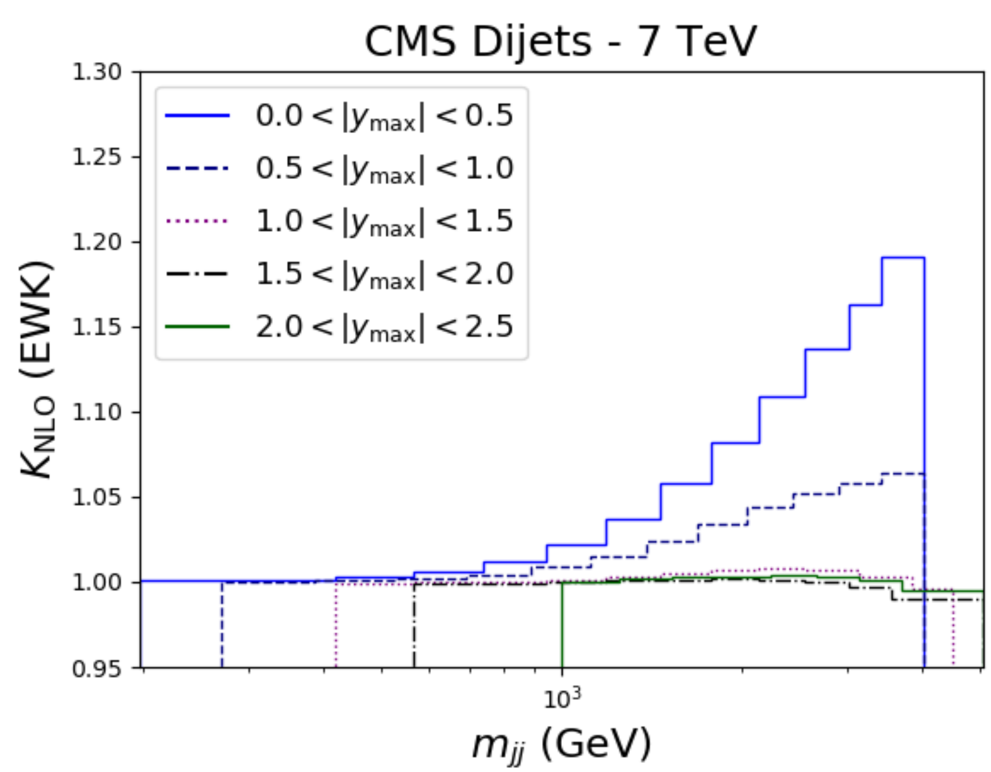
Dijet fit: $\chi^2(\text{no EW}) \to \chi^2(\text{EW}) : 1.34 \to 1.12$

Jet fit: $\chi^2(\text{no EW}) \to \chi^2(\text{EW}) : 1.39 \to 1.50$

(NNLO QCD)

i.e. we find dijet fit quality improved (driven by CMS 8 TeV), but inclusive (uniformly) deteriorates! Unclear why, but clearly impacts on discussion of relative fit quality.

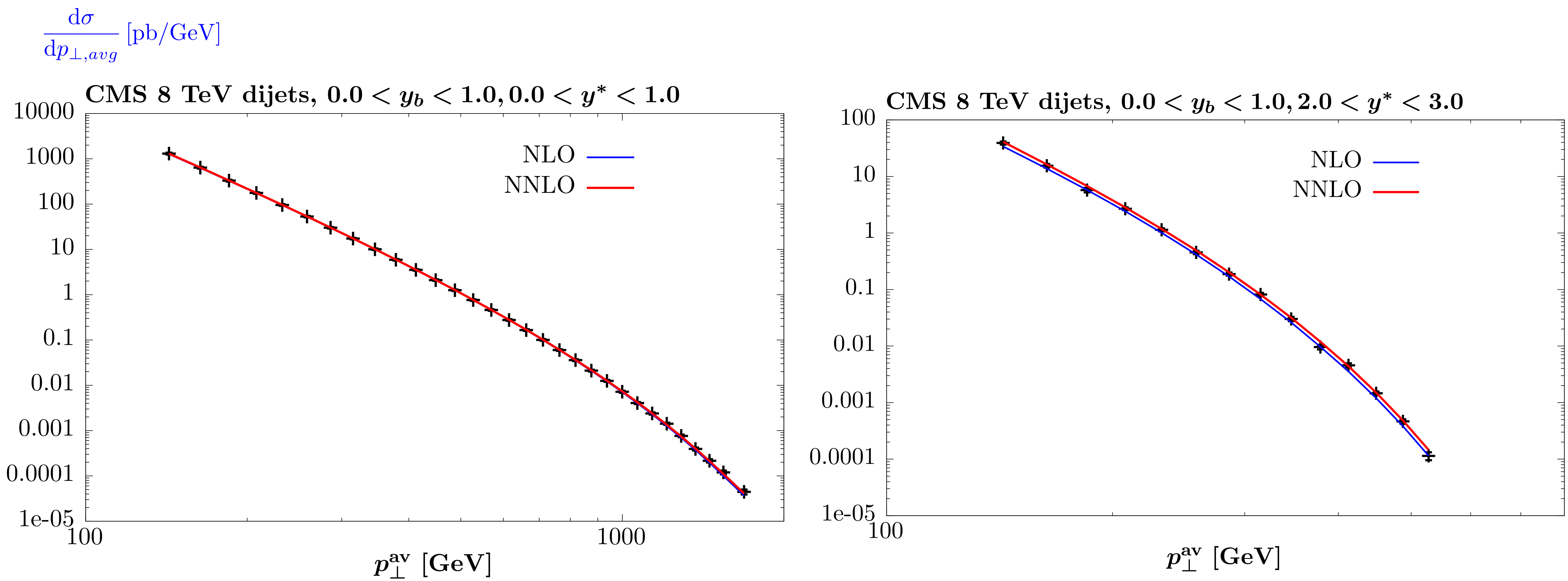




R. Abdul Khalek et al., *Eur.Phys.J.C* 80 (2020) 8, 797

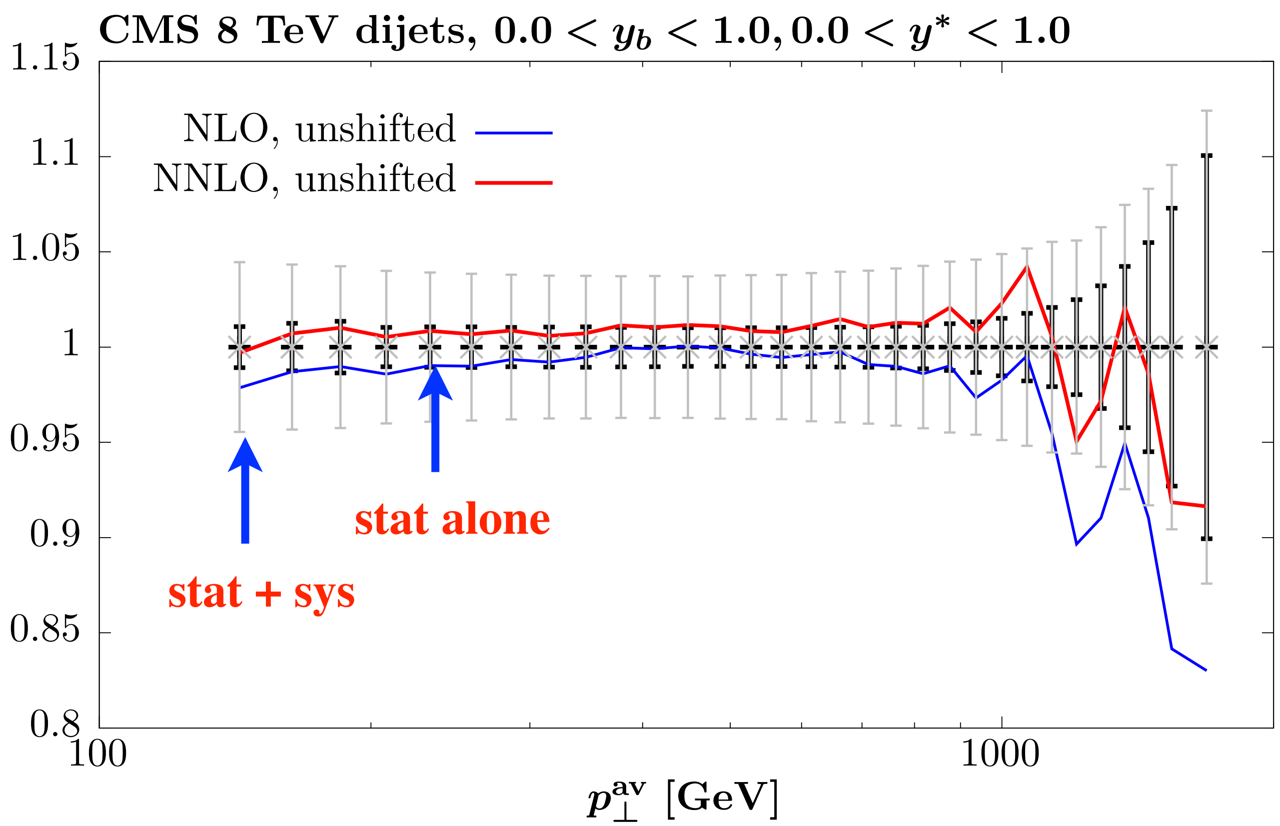
CMS 8 TeV Dijets

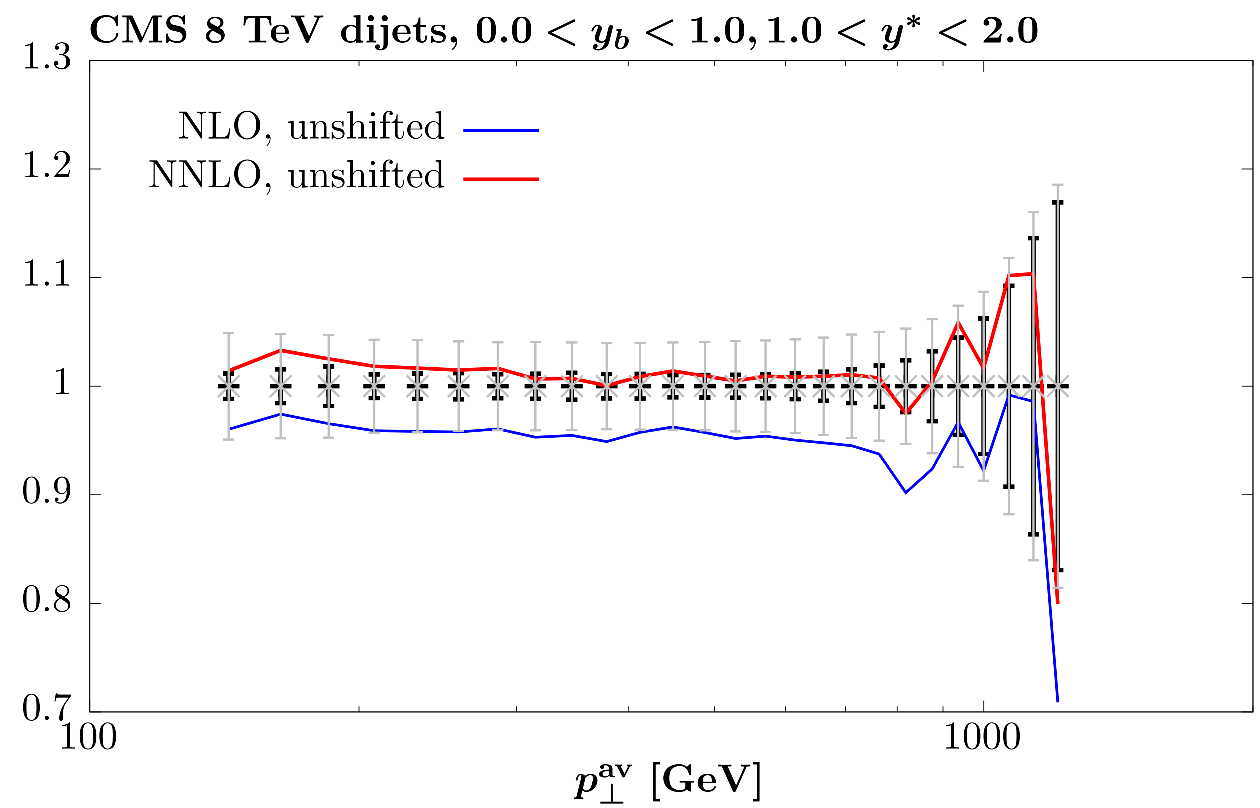
- What is driving this improvement?

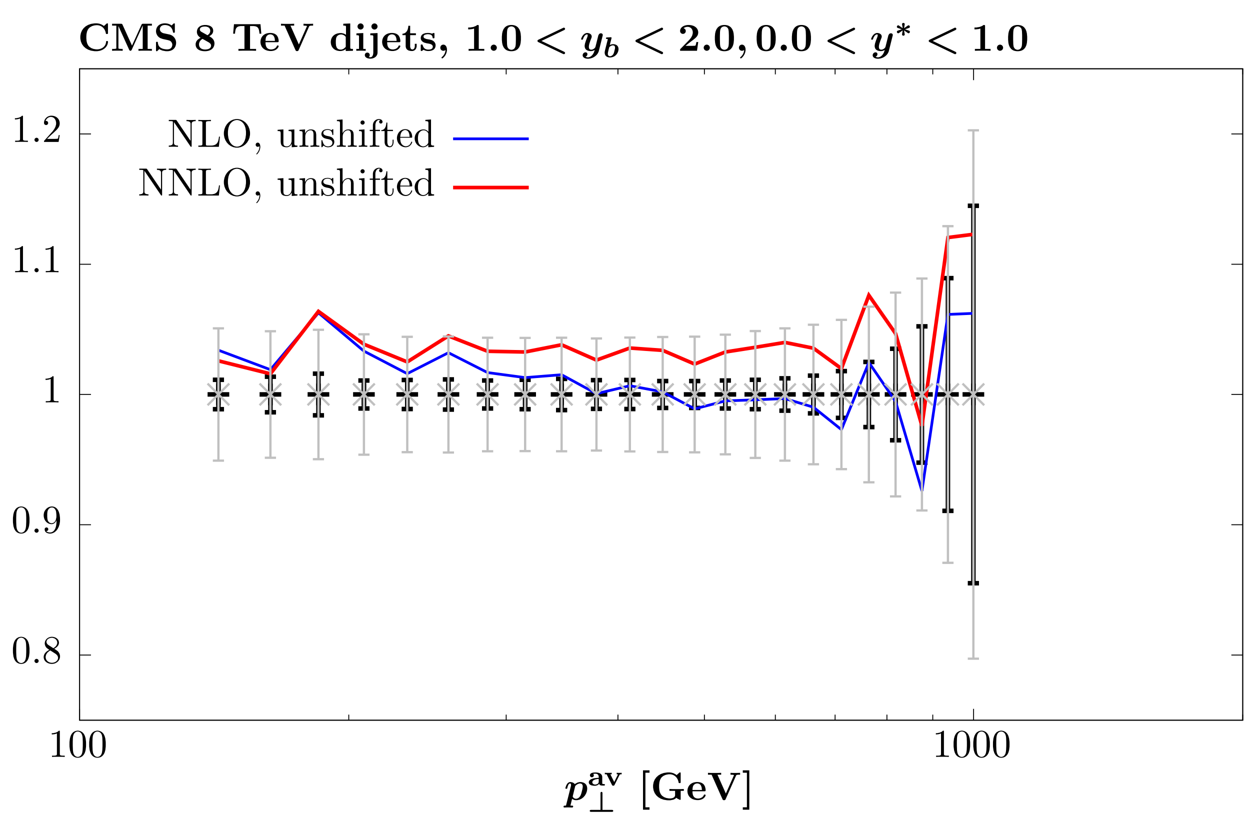


- Focus in on data/theory comparison…

• Overlaying data/theory no clear, by eye, trend for better description at NNLO.

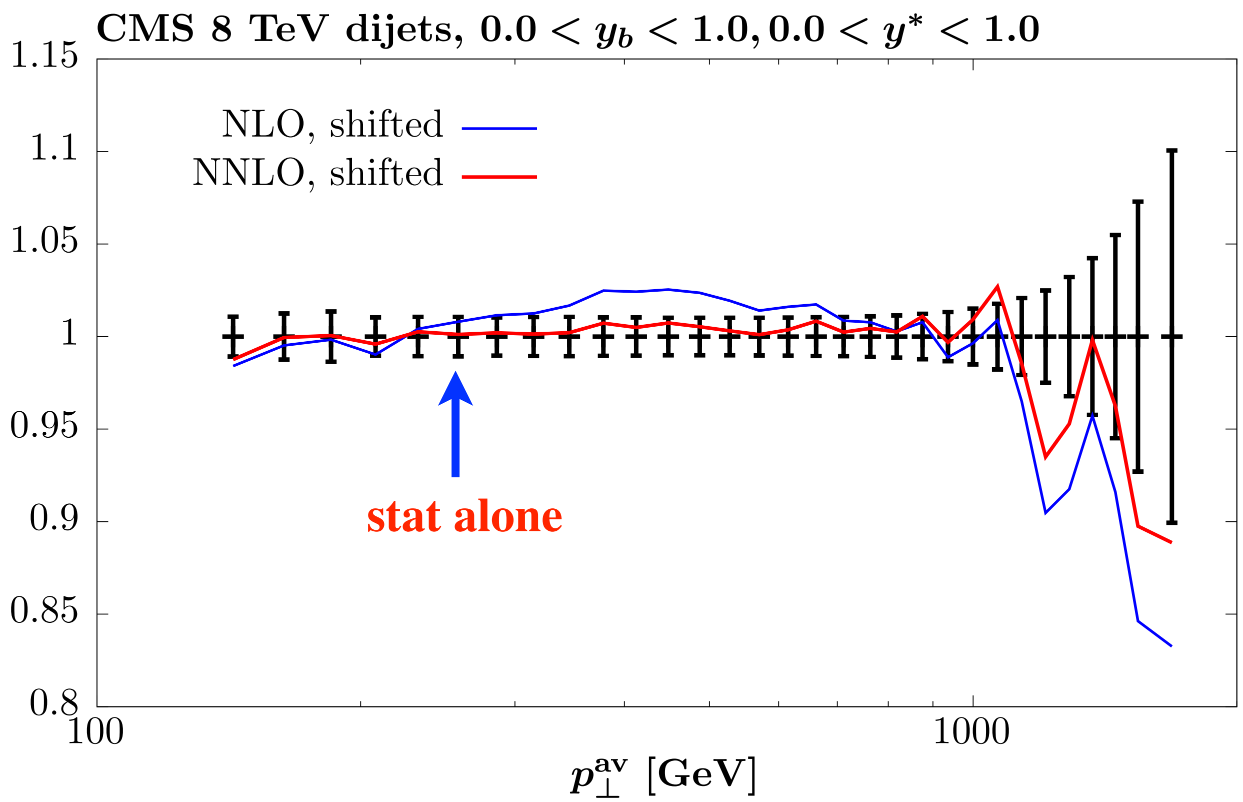


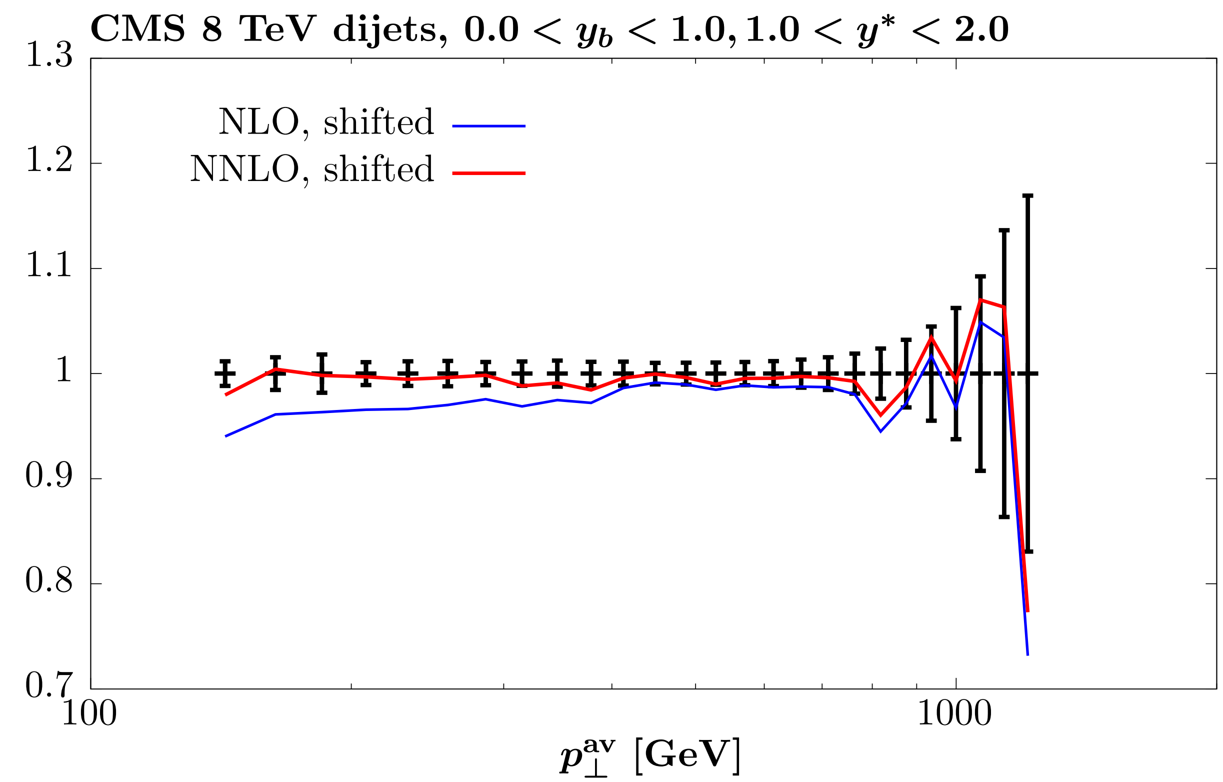


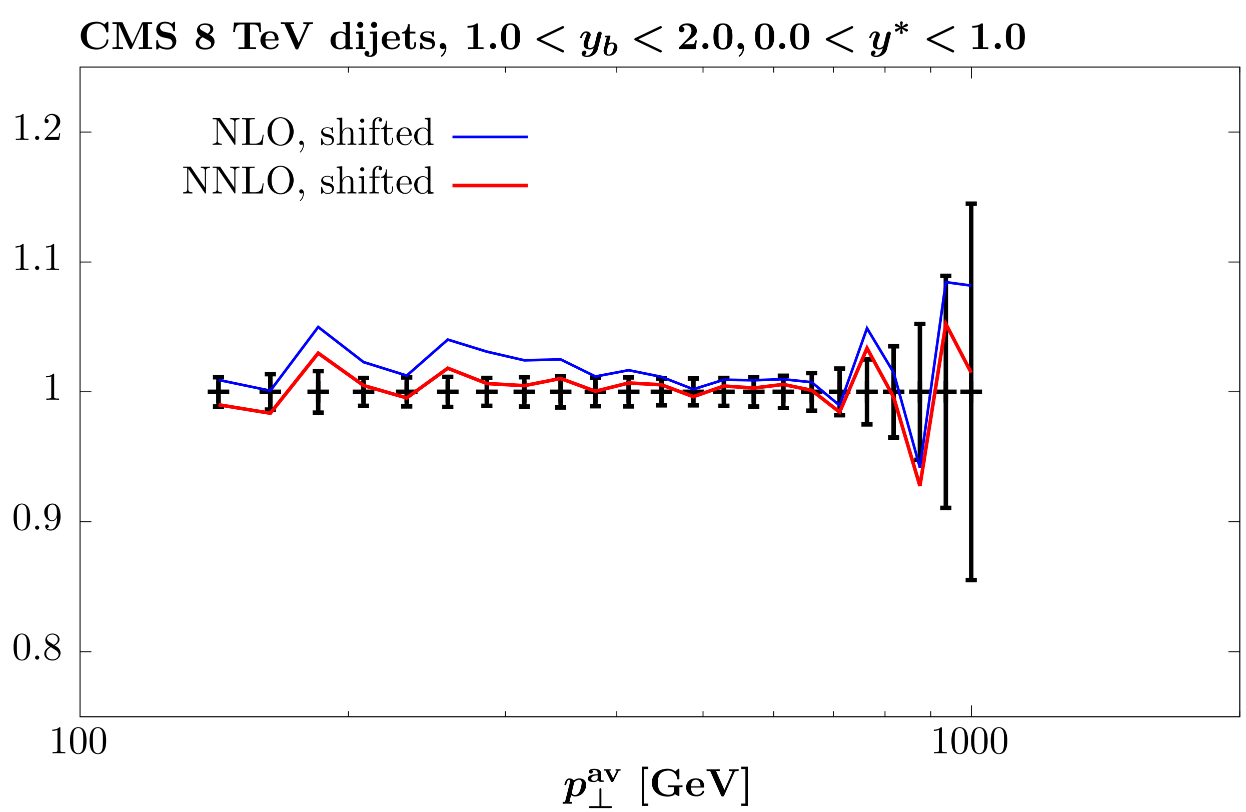


• However this is **before** shifting by correlated systematics.

• After including shifts from correlated systematics:

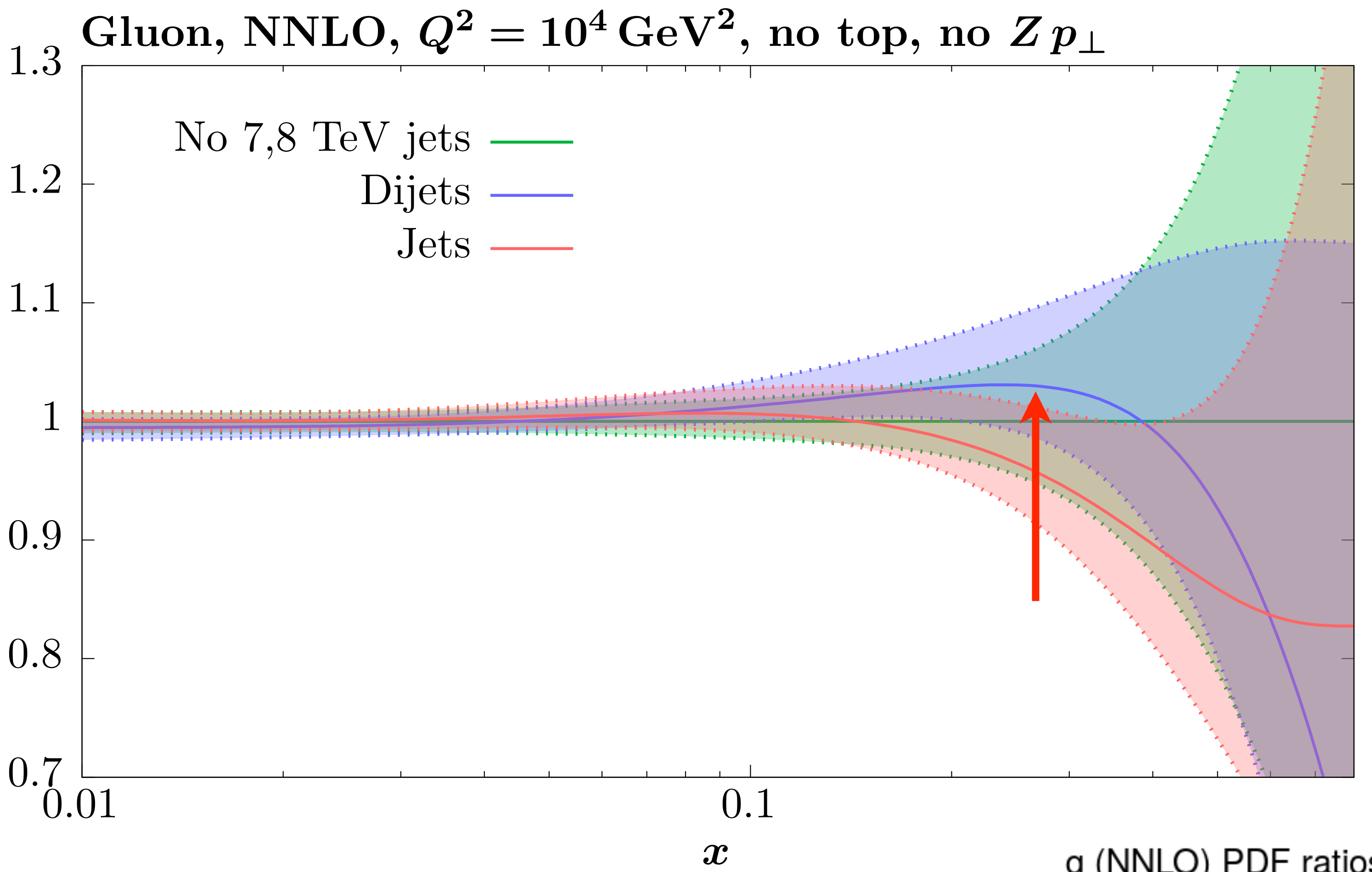




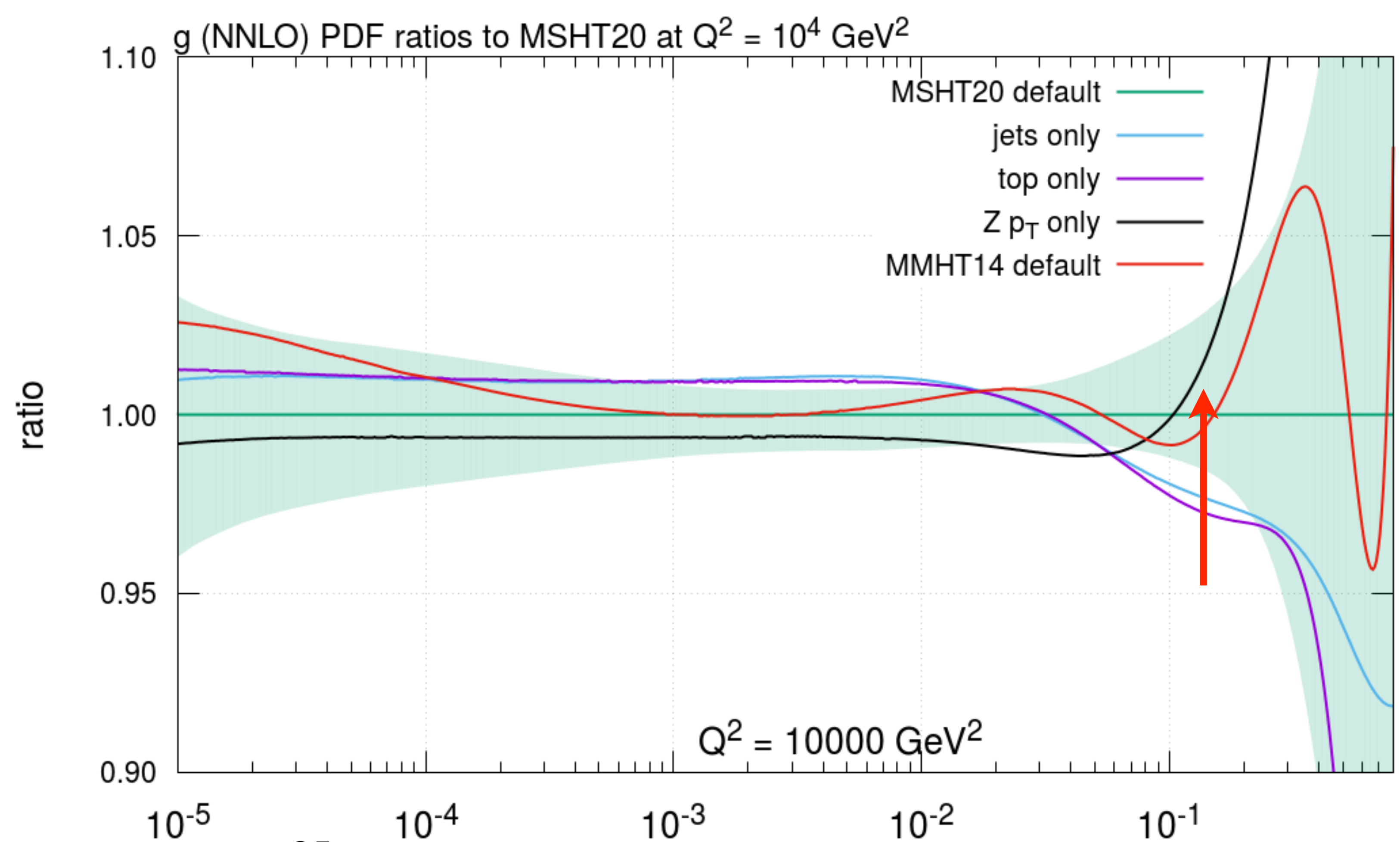


• Now improvement of NNLO is clear.

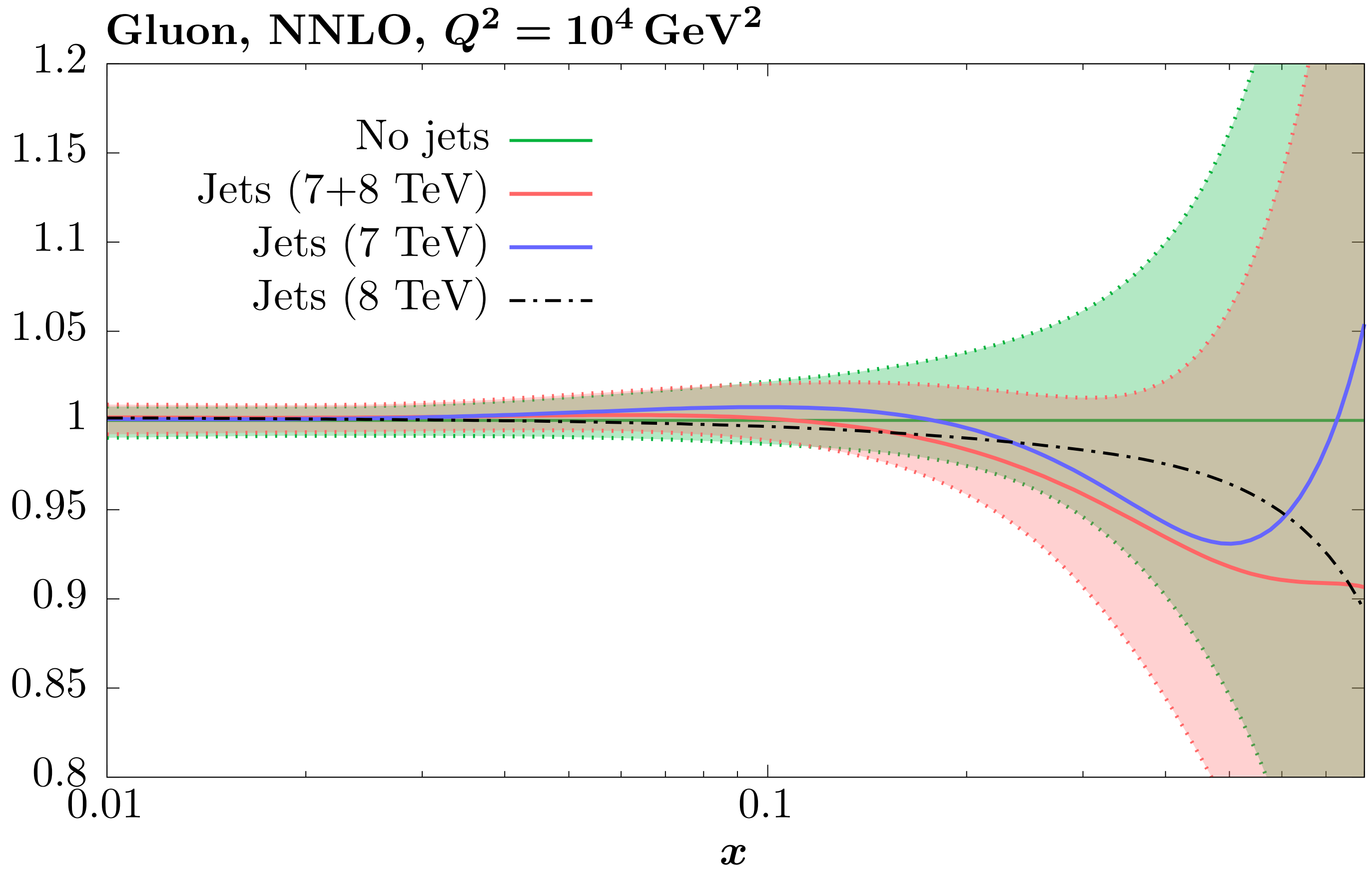
• Impact on shape of distributions in 3D kinematic space and interplay with correlated systematics drives this.



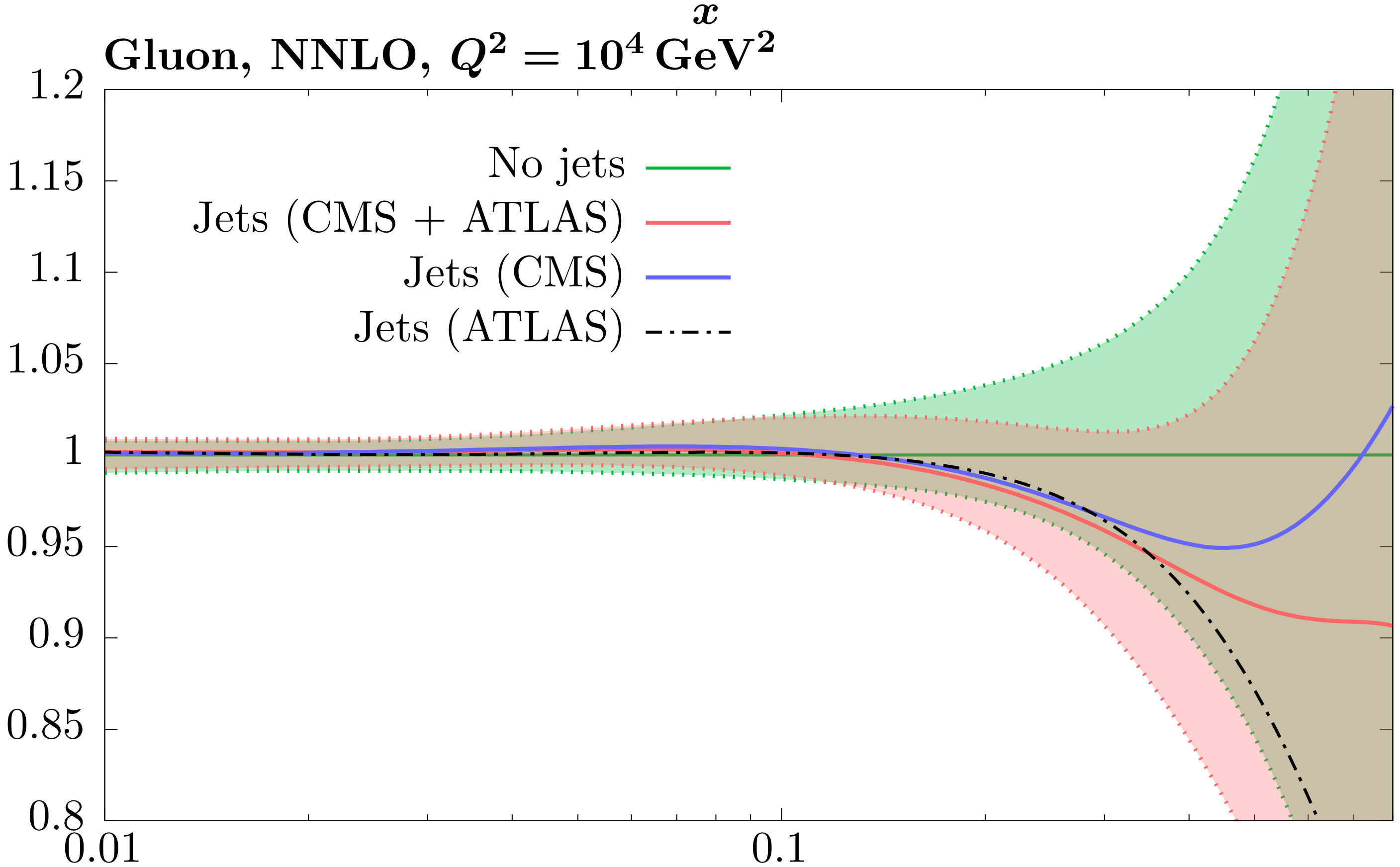
- In more detail, Dijets and both prefer rather higher $Z\,p_\perp$ gluon in $0.1 \lesssim x \lesssim 0.4$ region ($Z\,p_\perp$ out to higher x).



Consistency within datasets



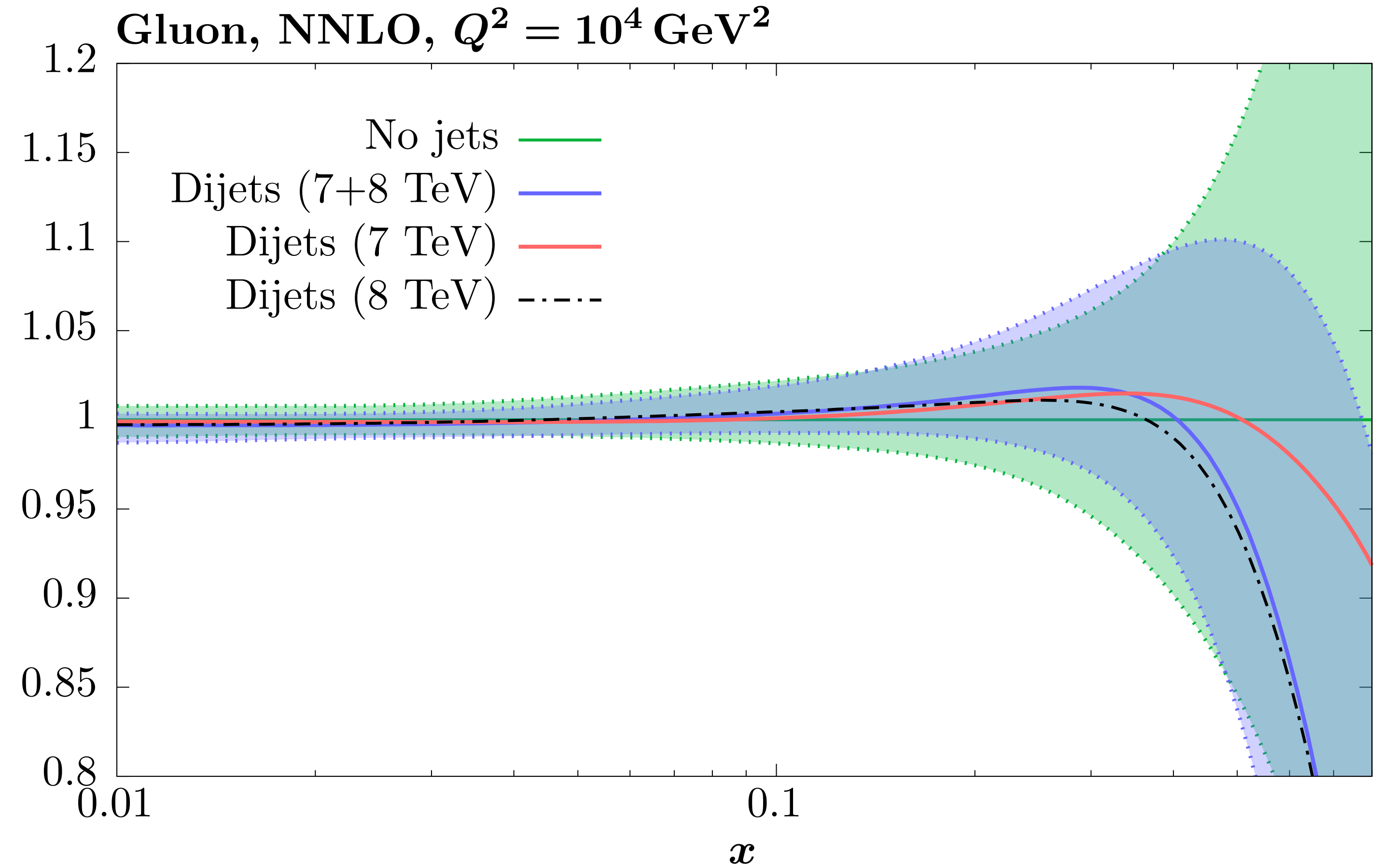
- *7* & 8 TeV data ~ consistent for inclusive jets.



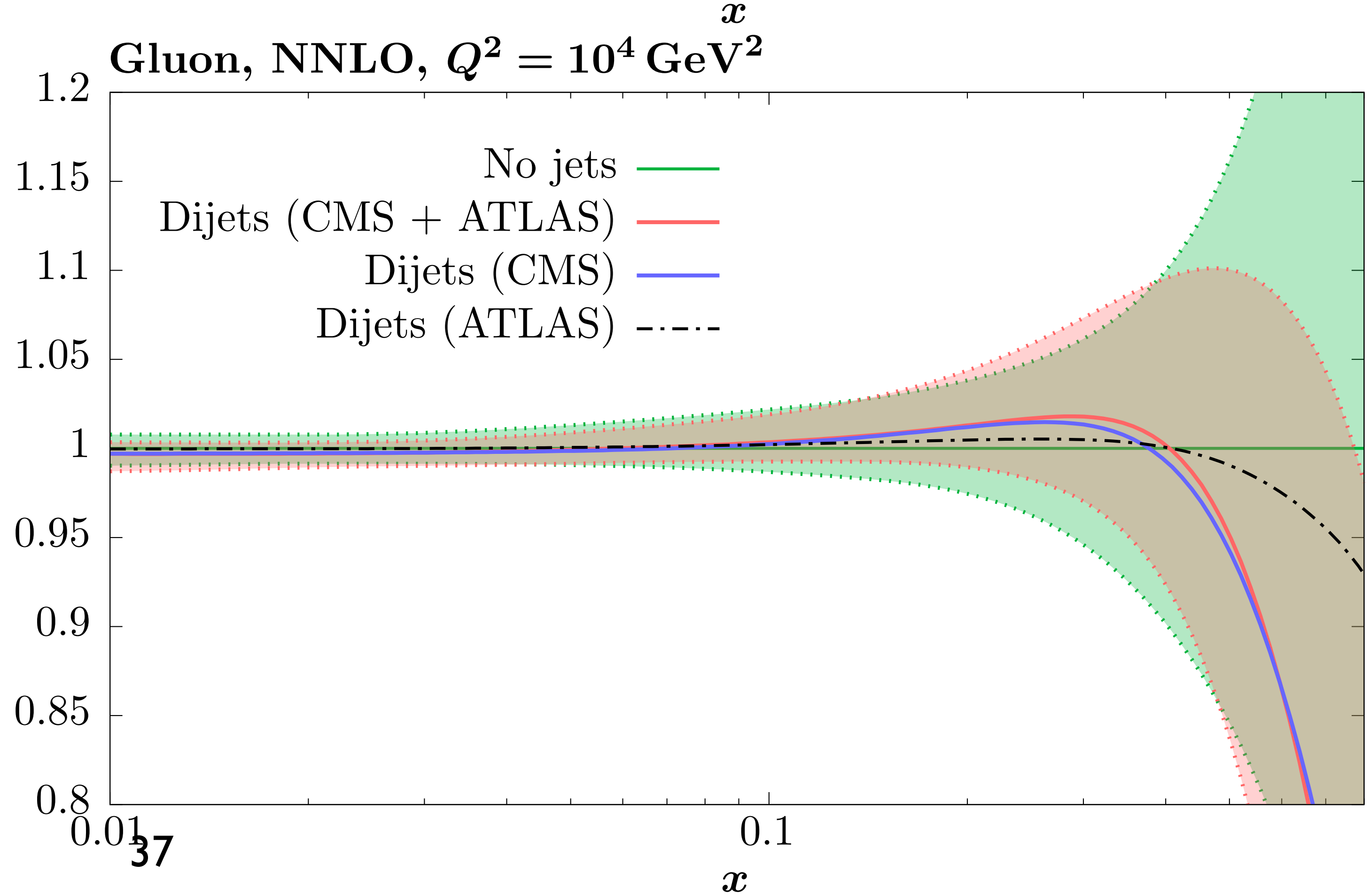
- At higher x clear difference between pulls of ATLAS and CMS (also seen in MSHT20).
- Final result compromise between these.

Consistency within datasets

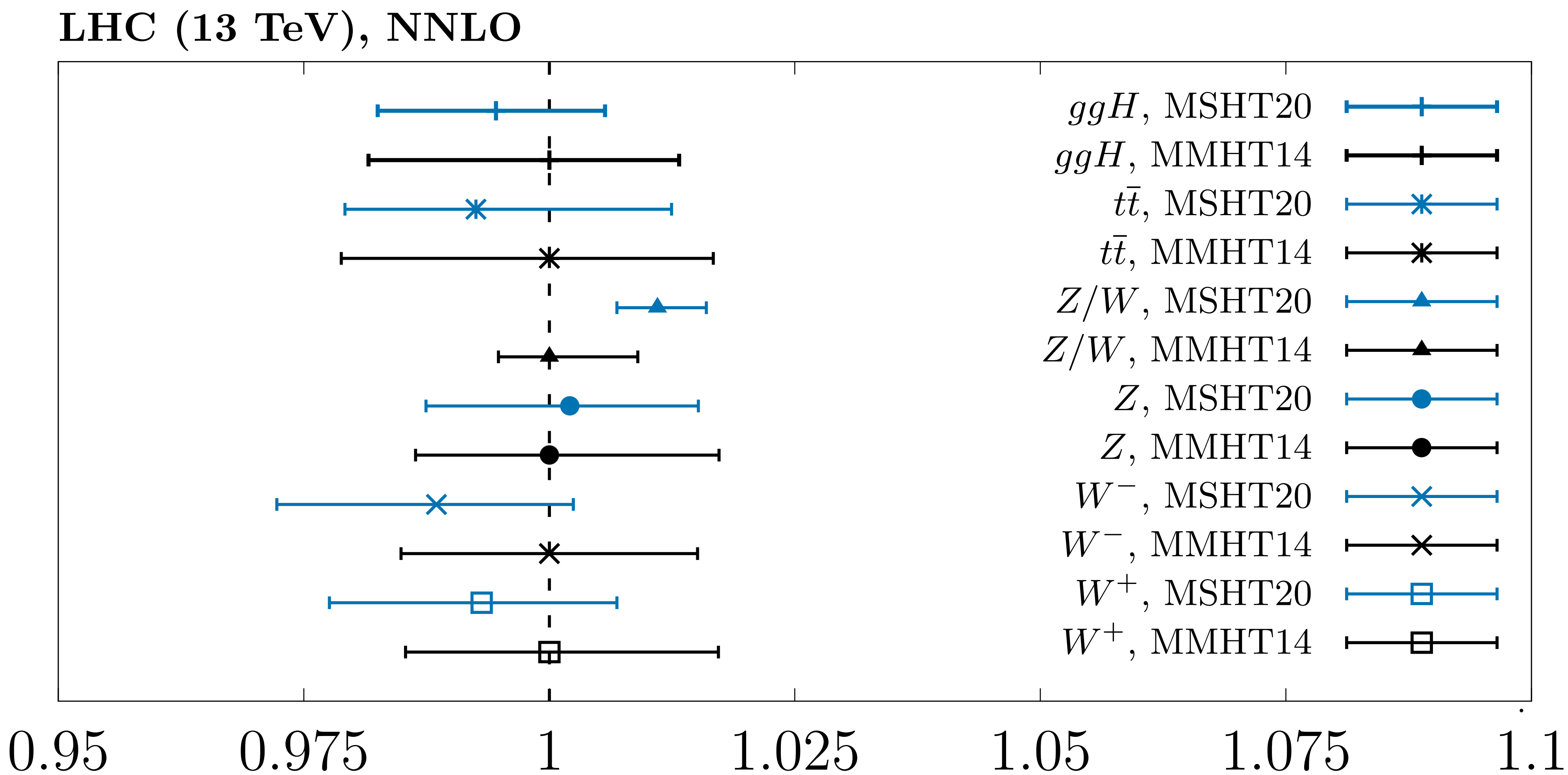
- *7* & 8 TeV data consistent for dijets, but this is due to broader result.
- That is: all dijet fits completely driven by CMS 8 TeV data.



- Similar picture, for same reason, between ATLAS/ CMS.



Results: Benchmarks



- For benchmark total cross sections at e.g. 13 TeV **encouraging picture**:
  - ★ General moderate reduction in PDF uncertainties.
  - ★ Central values relatively stable.
- Impact of LHC gluon sensitive data clear on ggH PDF uncertainty and high precision W,Z on in particular the Z to W ratio.