The ATLASpdf21 fit

Francesco Giuli (on behalf of the ATLAS Collaboration)

'alphaS-2022' workshop
ECT* (Trento, Italy)

01/02/2022

The ATLASpdf21 fit

2112.11266 (submitted to EPJC)

- **ATLASpdf21** is a PDF fit to **multiple ATLAS data sets**
- Attempt to simultaneously fit as many useful ATLAS data sets as possible
- DIS HERA data are the backbone of ATLAS PDF fits – we add ATLAS measurements on top of them
- HERA data provide constraints over a very wide range of x and Q^2
- LHC data provide additional constraints at medium and high-x and Q^2

Data set	$\sqrt{s}$ [TeV]	Luminosity [fb^{-1}]	Decay channel	Observables entering the fit
Inclusive $W, Z/\gamma^*$ [9]	7	4.6	e, μ combined	η_l (W), y_Z (Z)
Inclusive Z/γ^* [13]	8	20.2	e, μ combined	$\cos\theta$ in bins of $y_{\ell\ell}, M_{\ell\ell}$
Inclusive W [12]	8	20.2	μ	η_μ
$W^\pm$ + jets [23]	8	20.2	e	p_T^W
Z + jets [24]	8	20.2	e	p_T^{jets} in bins of $\lvert y_{\text{jets}}\rvert$
$t\bar{t}$ [25, 26]	8	20.2	lepton + jets, dilepton	$m_{t\bar{t}}$, p_T^t, $y_{t\bar{t}}$
$t\bar{t}$ [15]	13	36	lepton + jets	$m_{t\bar{t}}$, p_T^t, y_t, $y_{t\bar{t}}$
Inclusive isolated γ [14]	8, 13	20.2, 3.2	-	E_T^γ in bins of η^γ
Inclusive jets [16–18]	7, 8, 13	4.5, 20.2, 3.2	-	p_T in bins of $\lvert y_{\text{jets}}\rvert$

Advances wrt previous ATLAS PDF fits

- First ATLAS PDF fit which aim to include **many ATLAS data sets** (in contrast to previous papers where HERA + 1 or 2 other ATLAS data sets were fitted)
- First ATLAS PDF fit which includes **13 TeV data** (only NNPDF4.0 include 13 TeV data, other global fitters do not)
- **Inclusion of scale uncertainties** as additional correlated systematic uncertainties (where relevant e.g. W,Z data at 7 and 8 TeV)
- **Detailed study of the correlation** between the various ATLAS data sets (this is something only experimentalists can do) - gain experience on ATLAS systematic uncertainty treatment to make proposal for use in global fits
- **Extended PDF parametrisation** – 21 free parameters (previous ATLAS PDF fits with 15 or 16 free parameters)
- First ATLAS PDF fit with **enhanced tolerance** (determined following the MSHT dynamic tolerance procedure) – **T =** $\sqrt{\Delta\chi^2}$ **= 3**

Theory framework

- All the fits performed using xFitter
- Results checked with an independent fitting code
- LHC cross sections from fastNLO and APPLgrid

Data set	NLO QCD code	LO EW code	NNLO QCD code	NLO EW code
Inclusive $W, Z/\gamma^*$ [9]	MCFM	MCFM	DYNNLO 1.5, FEWZ 3.1.b2	DYNNLO 1.5, FEWZ 3.1.b2
Inclusive Z/γ^* [13]	MCFM	MCFM	NNLOJET	NNLOJET
Inclusive W [12]	MG5_AMC@NLO 2.6.4	MG5_AMC@NLO 2.6.4	DYNNLO 1.5	DYNNLO 1.5
$W^{\pm}$ + jets [24]	N_{jetti}	N_{jetti}	N_{jetti}	SHERPA
Z + jets [25]	Ref. [51]	Ref. [51]	Ref. [51]	SHERPA
$t\bar{t}$ (lepton + jets) [26]	-	Ref. [52]	Ref. [52]	Ref. [55]
$t\bar{t}$ (dilepton) [27]	MCFM	MCFM	Ref. [28]	Ref. [55]
$t\bar{t}$ [15]	-	Ref. [52]	Ref. [52]	Ref. [55]
Inclusive isolated γ [14]	MCFM	MCFM	Ref. [57]	Ref. [58]
Inclusive jets [16–18]	NLOJET++	NLOJET++	NNLOJET	Ref. [63]

- For 7,8 TeV W,Z data scale uncertainties are applied as additional correlated systematic uncertainties (comparable with experimental systematics)
- We evaluated the impact of scale uncertainties for the other data sets and it is found to be negligible

Fit model

- PDF parametrisation:

$$xf(x) = Ax^B(1-x)^C P(x) = Ax^B(1-x)^C(1 + Dx + Ex^2 + Fx^3)$$

(with an extra negative term for the gluon $-A'_g x^{B'g}(1-x)^{C'_g}$)

- Constraints for the central fit from sum rules
- At Q_0^2 = 1.9 GeV^2, we parametrise $\boldsymbol{xu_V, xd_V, x\bar{u}, x\bar{d}, xs}$ and $\boldsymbol{xg}$
- Central fits with **21 parameters** with $\boldsymbol{\alpha_S(m_Z)}$ = **0.118** (previous ATLAS fits with 15 or 16 free parameters)
- $P_g(x) = 1 + D_g(x),\ P_{u_V} = 1 + D_{u_V}x + E_{u_V}x^2$ and $P_{d_V} = 1 + D_{d_V}x$
- No constraints on the A and B parameters of the sea quarks, so no constraints on $x\bar{d} - x\bar{u}$ or on $x\bar{s}/(x\bar{d} + x\bar{u})$ as $x \to 0$ (either shape or normalisation)
- **Model assumptions and uncertainties:**
  - heavy quark masses – m_c= 1.41 GeV and m_b = 4.2 GeV
  - Q^2_{min} (10 GeV^2) cut off for inclusion of data in the fit
  - Starting scale Q_0^2
  - m_{top} (173.3 GeV)

Results

Total χ^2/NDF	2010/1620
HERA χ^2/NDP	1112/1016
HERA correlated term	50
ATLAS W, Z 7 TeV χ^2/NDP	68/55
ATLAS Z/γ^* 8 TeV χ^2/NDP	208/184
ATLAS W 8 TeV χ^2/NDP	31/22
ATLAS W and Z/γ^* 7 and 8 TeV correlated term	$71 = (38 + 33)$
ATLAS direct γ 13/8 TeV χ^2/NDP	27/47
ATLAS direct γ 13/8 TeV correlated term	6
ATLAS V+jets 8 TeV χ^2/NDP	105/93
ATLAS $t\bar{t}$ 8 TeV χ^2/NDP	13/20
ATLAS $t\bar{t}$ 13 TeV χ^2/NDP	25/29
ATLAS inclusive jets 8 TeV χ^2/NDF	207/171
ATLAS V+jets 8 TeV and $t\bar{t}$ + jets 8,13 TeV and $R = 0.6$ inclusive jets 8 TeV correlated term	$87 = (16 + 9 + 21 + 41)$

- This is a **better fit than that achieved by global fits to these data**

Correlation between various data sets

- Possible correlation between the ATLAS data sets carefully investigated

Systematic uncertainty	8 TeV $W+$jets	8 TeV $Z+$jets	8 TeV $t\bar{t}$ lepton + jets	13 TeV $t\bar{t}$ lepton + jets	8 TeV inclusive jets
Jet flavour response	JetScaleFlav2	Flavor Response	flavres-jes	JET29NP JET Flavour Response	syst JES Flavour Response*
Jet flavour composition	JetScaleFlav1Known	Flavor Comp	flavcomp-jes	JET29NP JET Flavour Composition	syst JES Flavour Comp
Jet punchthrough	JetScalepunchT	Punch Through	punch-jes	-	syst JES PunchThrough MC15
Jet scale	JetScalePileup2	PU OffsetMu	pileoffmu-jes	-	syst JES Pileup MuOffset
	-	PU Rho	pileoffrho-jes	JET29NP JET Pileup RhoTopology	syst JES Pileup Rho topology*
	JetScalePileup1	PU OffsetNPV	pileoffnpv-jes	JET29NP JET Pileup OffsetNPV	syst JES Pileup NPVOffset
	-	PU PtTerm	pileoffpt-jes	JET29NP JET Pileup PtTerm	syst JES Pileup Pt term
Jet JVF selection	JetJVFcut	JVF	jetvxfrac	-	syst JES Zjets JVF
B-tagged jet scale	-	btag-jes	JET29NP JET BJES Response	-	-
Jet resolution	-	jeten-res	JET JER SINGLE NP	-	-
Muon scale	-	-	mup-scale	MUON SCALE	-
Muon resolution	-	-	muonms-res	MUON MS	-
Muon identification	-	-	muid-res	MUON ID	-
Diboson cross section	-	-	dibos-xsec	Diboson xsec	-
$Z+$jets cross section	-	-	zjet-xsec	Zjets xsec	-
Single-t cross section	-	-	singletop-xsec	st xsec	-

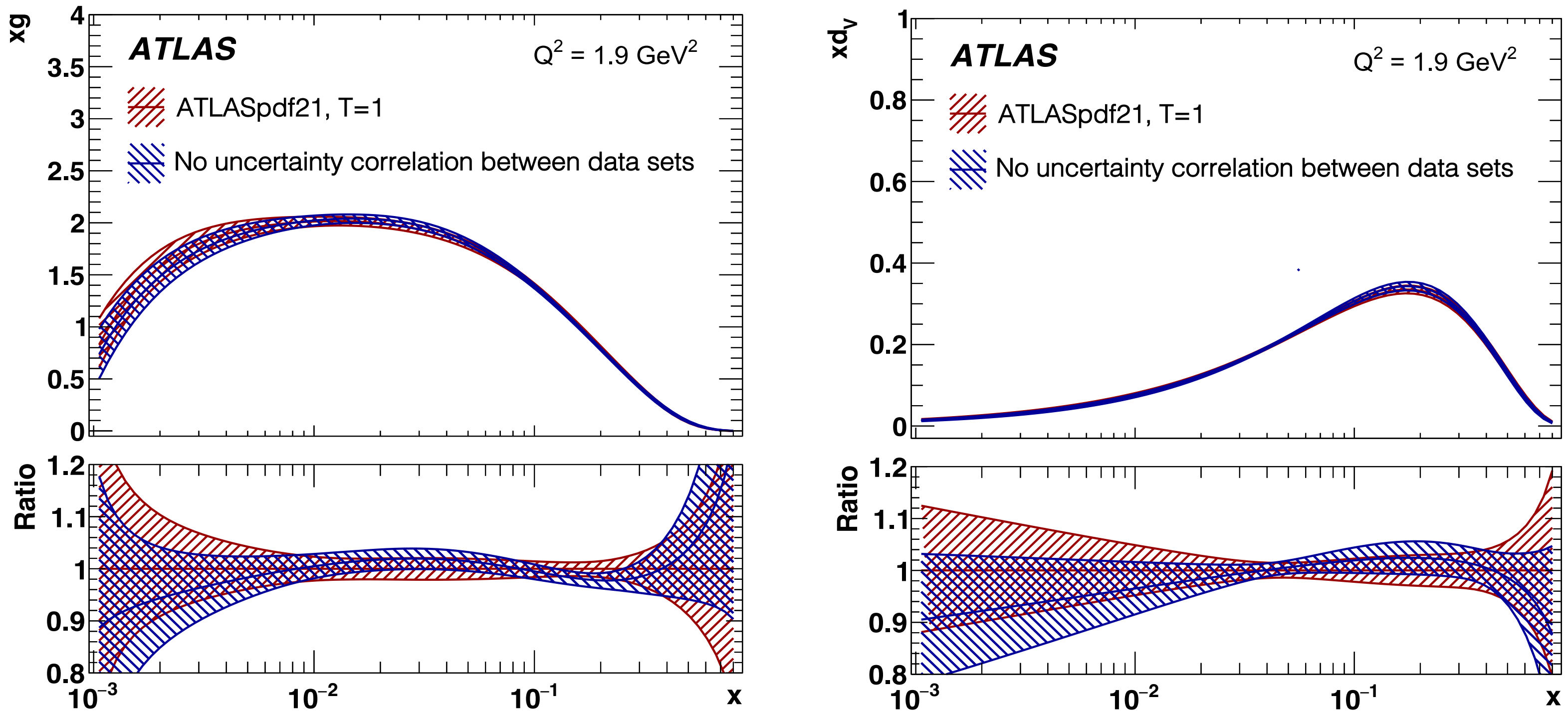
- Entries in the same raw taken 100%-correlated for V+jets and $t\bar{t}$+jets (R=0.4)

- Different degrees of correlation are considered of the inclusive jet data (R=0.6), because of the differing choice of the jet radius wrt V+jets and $t\bar{t}$+jets

- Exact degree of correlation to the inclusive jet data does not change the resulting PDFs

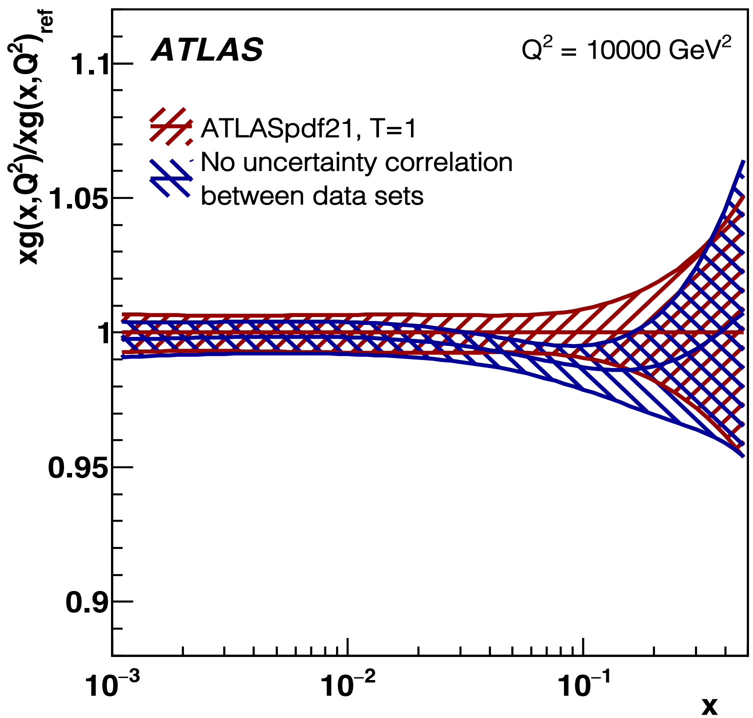
NOVELTY!

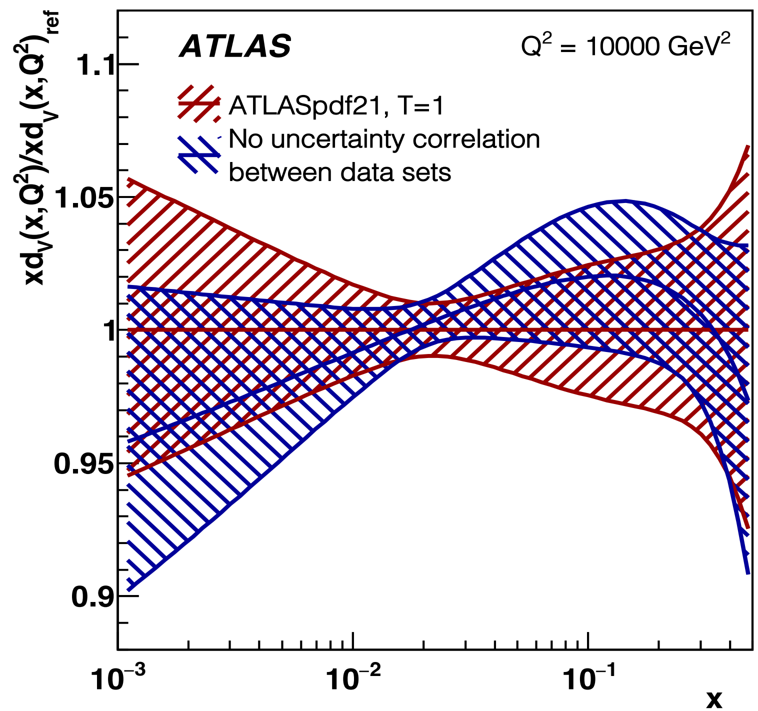
Correlation between various data sets



- The correlations do not give big effects on PDFs - always compatible within just experimental uncertainties evaluated with $\Delta\chi^2 = 1$ at 68% C.L.
- But visible differences in χ^2 e.g. the fit without correlation exhibits a smaller χ^2 by ~30 units
- So we have to **account for correlations properly**

Correlation between various data sets

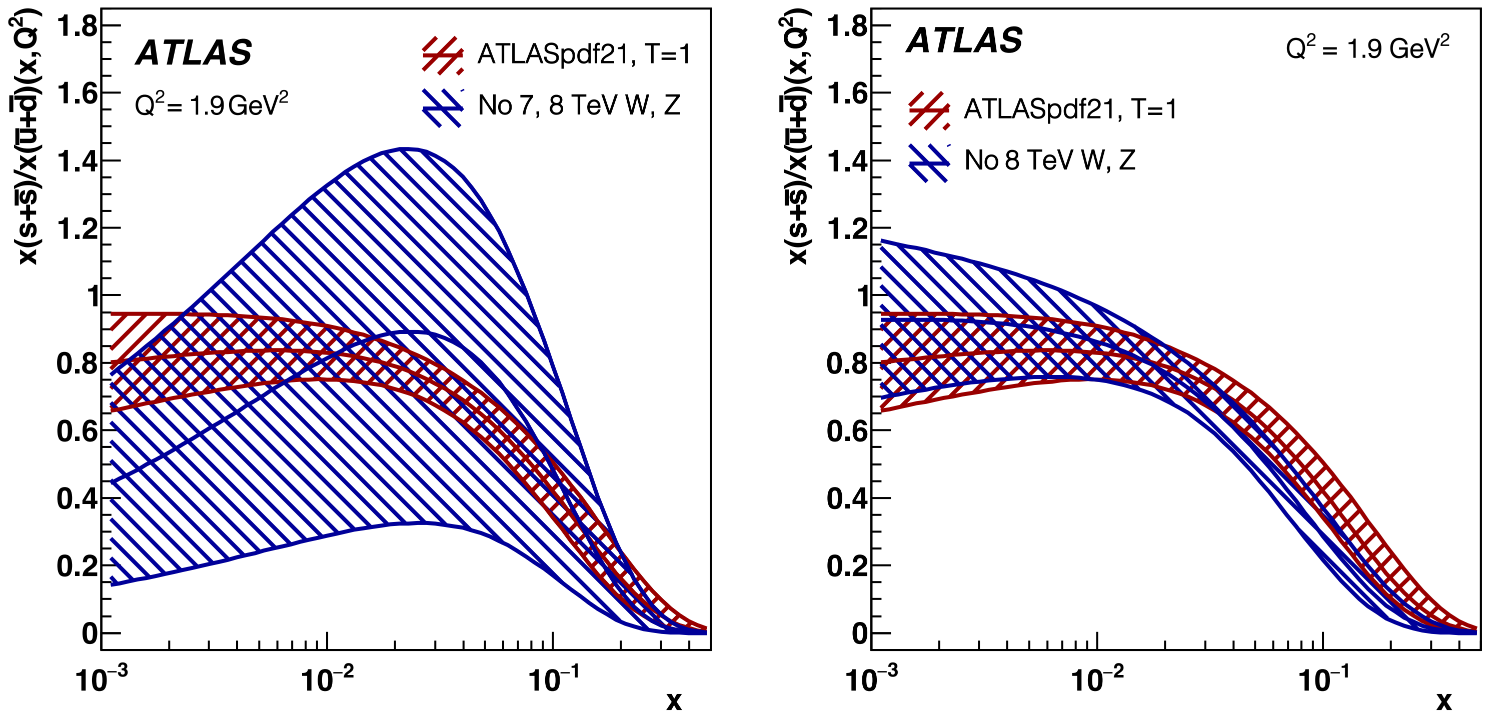




- The correlations do not give big effects on PDFs - always compatible within just experimental uncertainties evaluated with $\Delta\chi^2 = 1$ at 68% C.L.
- But visible differences in χ^2 e.g. the fit without correlation exhibits a smaller χ^2 by ~30 units
- At LHC scales, the central values shift by > 2% and middling x → **important for accuracy of m_W, $\sin^2\theta_W$ and α_S precision measurements**

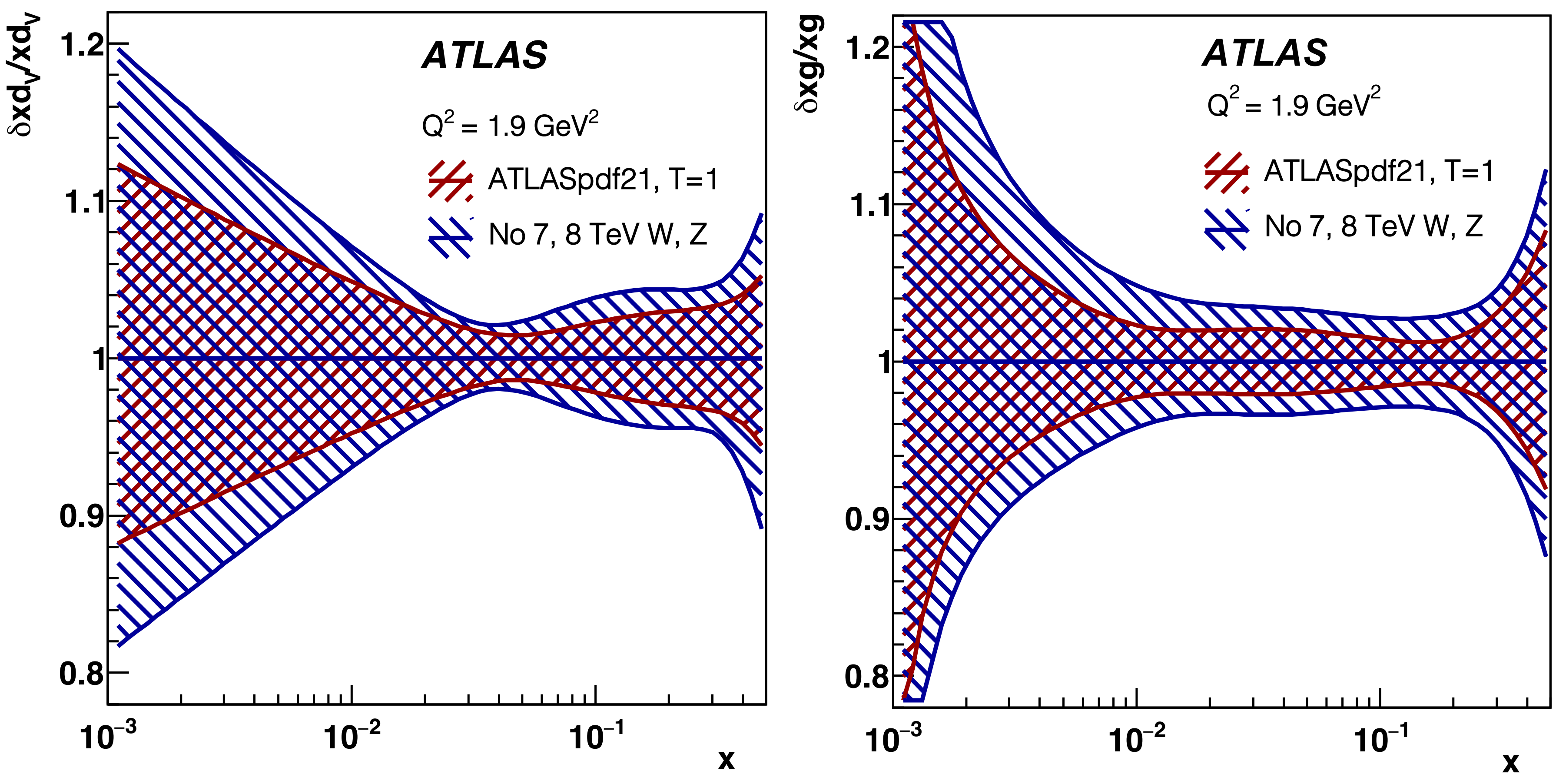
Impact of the various data sets on PDFs

- We first removed all the **inclusive W,Z data**
- Ratio of strange to light sea quarks very poorly determined (left plot)
- We retain the precise W,Z 7 TeV data and remove the W,Z 8 TeV data
- We see that whereas the W,Z 7 TeV serve to fix the low-x sea quarks, the 8 TeV data still has something to add (right plot)



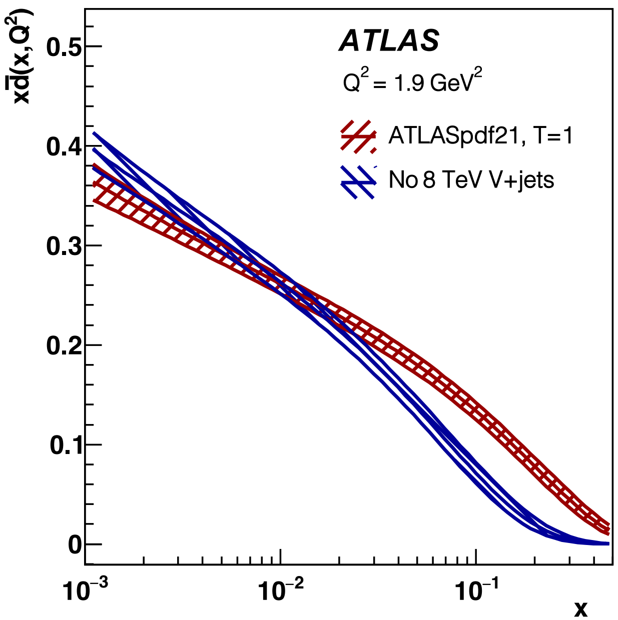
Impact of the various data sets on PDFs

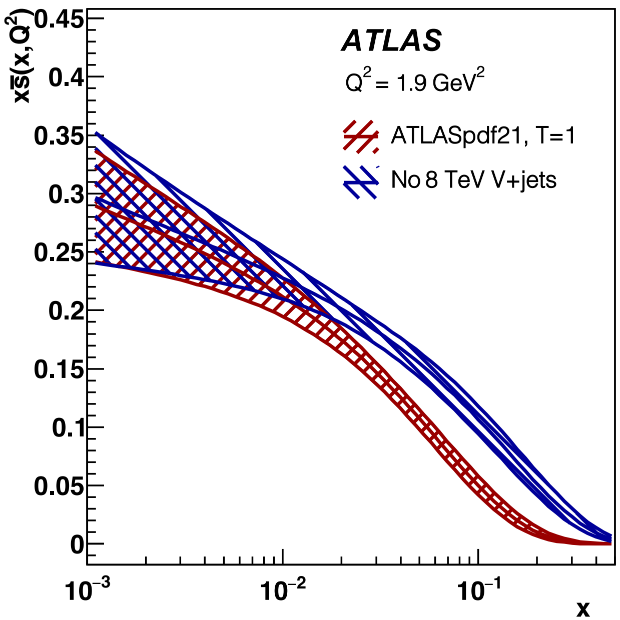
- We first removed all the **inclusive W,Z data**
- The fit without W,Z data can constraints valence and gluon PDF but input of these data results in substantial decrease of uncertainties

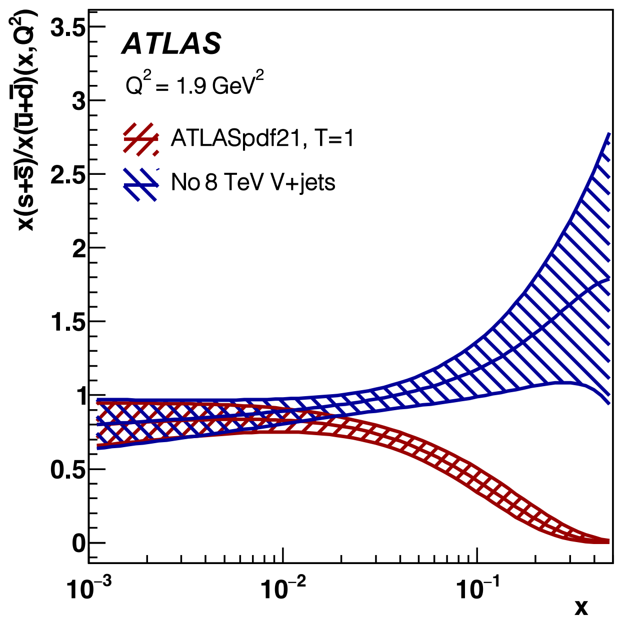


Impact of the various data sets on PDFs

- We removed all the **8 TeV V+jets data**
- As for the ATLASepWZVjets20 fit we see that the effect of these data (blue to red) is to harden the high-x $\bar{d}$ and soften the high-x $\bar{s}$
- Without them we cannot really determine the ratio of strange to light quarks at high-x

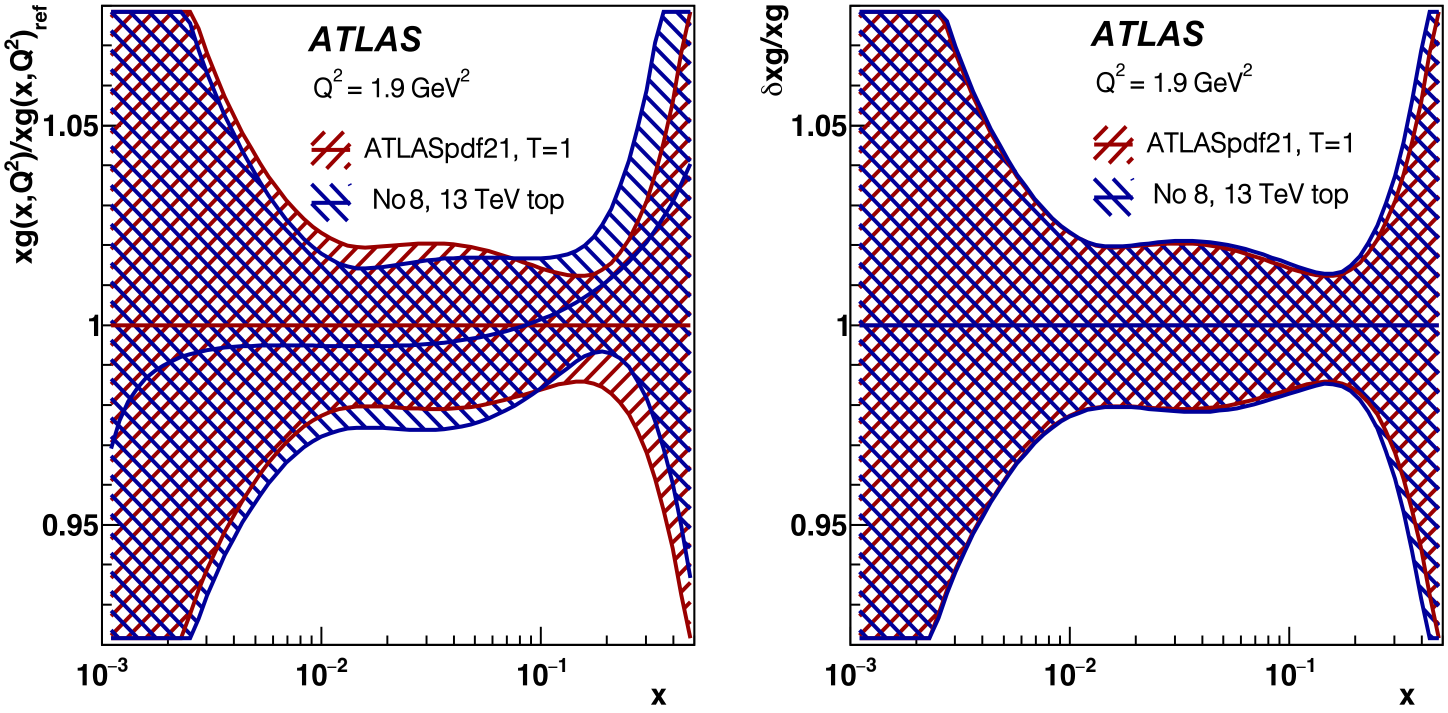






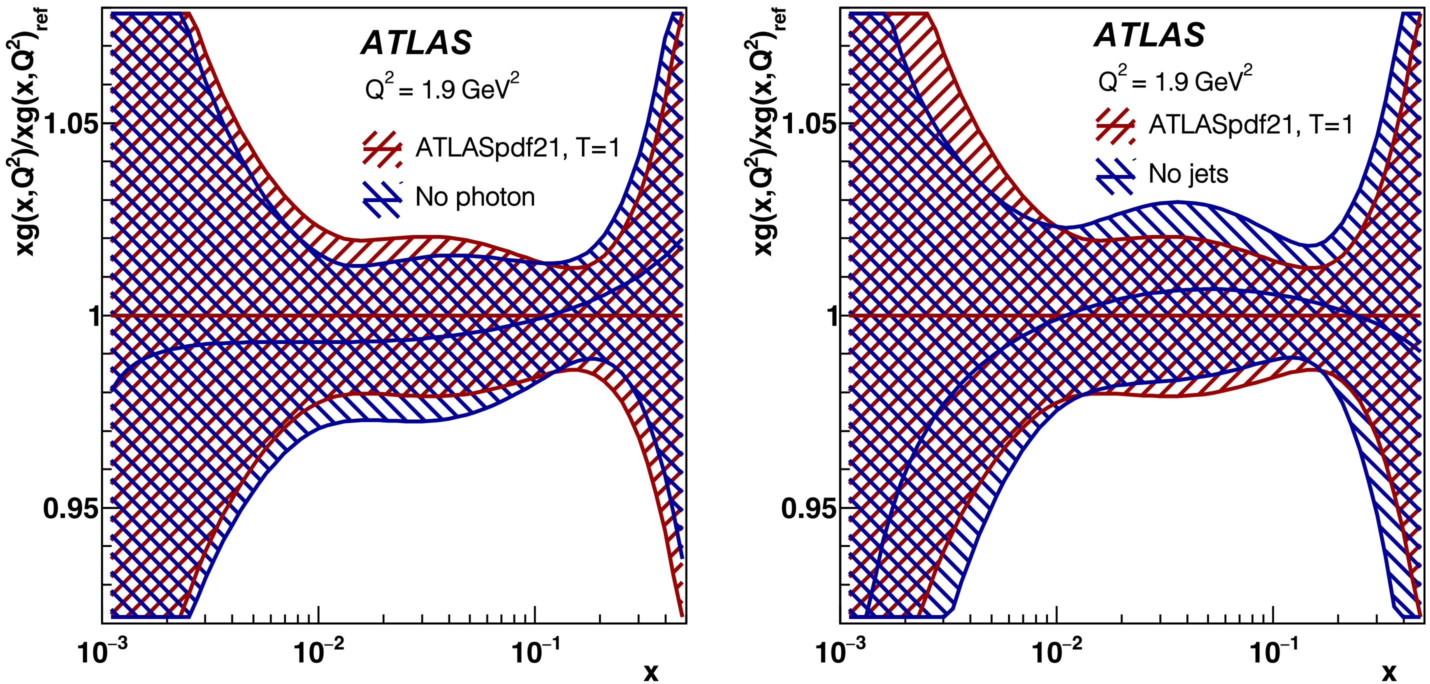
Impact of the various data sets on PDFs

- We removed all the ***$t\bar{t}$* data** from the fit
- These data marginally soften the high x gluon (blue to red) and **reduce its uncertainties at high-x** - 8 TeV has the bigger effect (backup)
- Milder impact wrt what was found for the ATLASepWZtop18 fit (but here we have many other data sets added)



Impact of the various data sets on PDFs

- We removed all the **13/8 TeV photon ratio data**
- This results in a marginal softening of the high-x gluon (blue to red) but no decrease in uncertainty
- We removed the **8 TeV inclusive jet data**
- This results in a marginal shape change of the gluon PDF (blue to red) and **very substantial decrease in its high-x uncertainty**



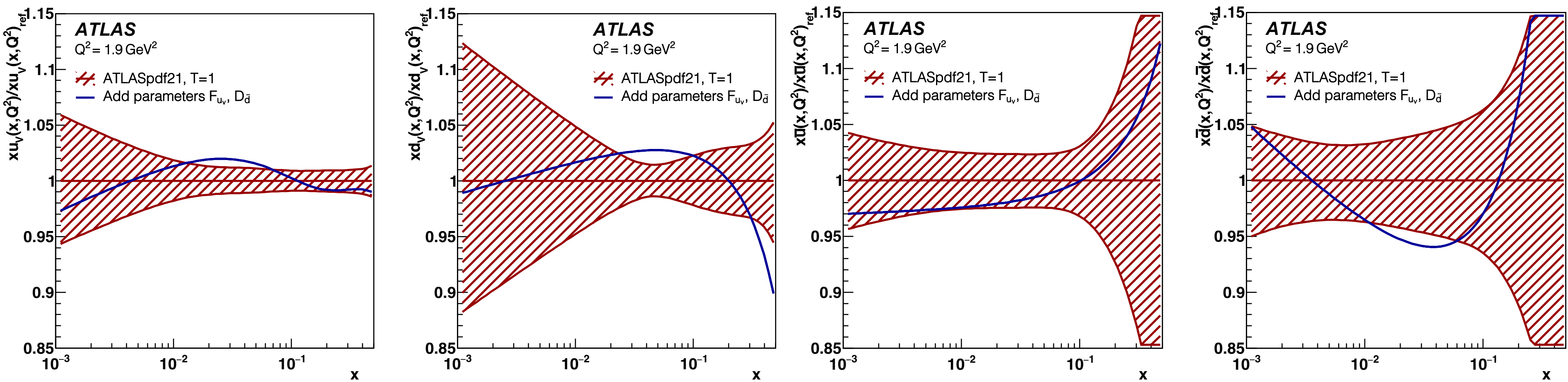
Parametrisation uncertainties

- PDF parametrisation:

$$xf(x) = Ax^B(1-x)^C P(x) = Ax^B(1-x)^C(1 + Dx + Ex^2 + Fx^3)$$

(with an extra negative term for the gluon $-A'_g x^{B'g}(1-x)^{C'_g}$)

- **Parametrisation uncertainties:** grey parameters in the equation above



- Adding extra parameters produces modest shape changes wrt uncertainties

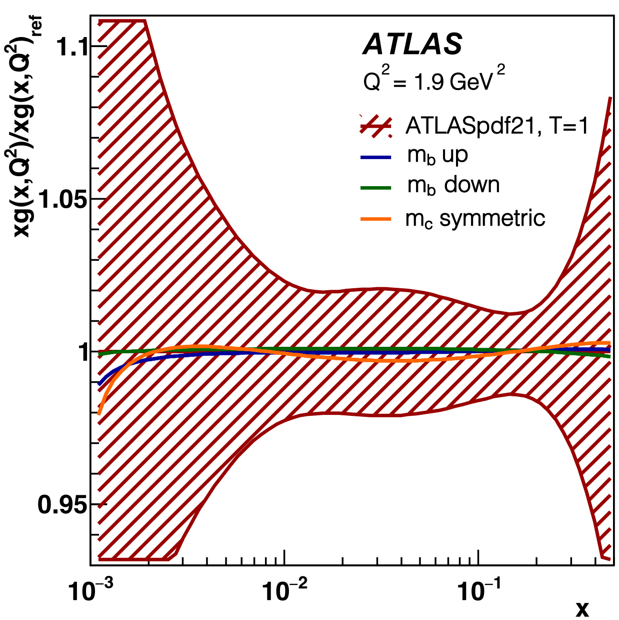
- These are included in the PDF uncertainty evaluation as an envelope of variations from the central fit

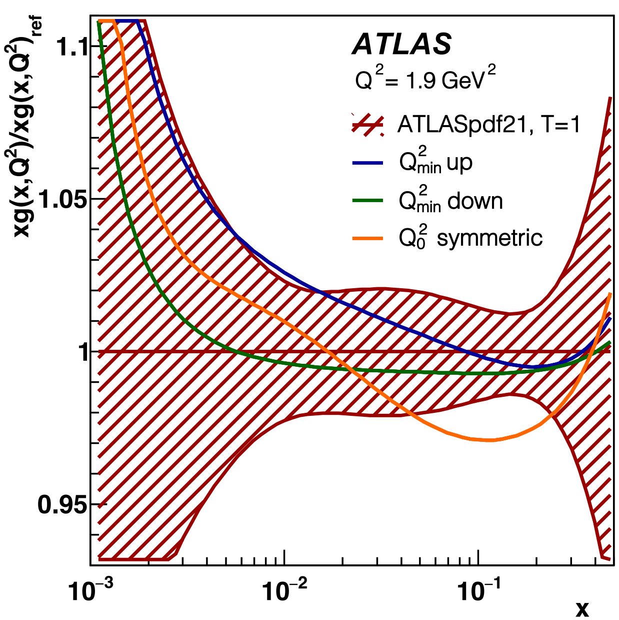
- **Check made using Chebyshev polynomials** – no improvement in χ^2 and PDFs compatible within uncertainties with the ordinary polynomial fit

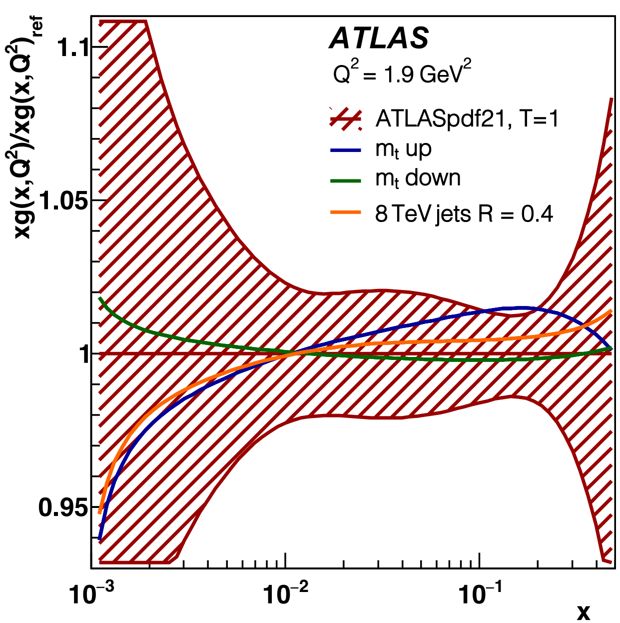
Model uncertainties

- We vary the input settings for the fit
- Impact on the gluon PDF in ratio below
- The variation of the starting scale Q_0^2 is the most significant
- We chose R = 0.6 since this is theoretically preferred – difference wrt R = 0.4 **ADDED as an additional uncertainty**

Central χ^2/NDF	2010/1620
Model variations	
$Q^2_{\rm min} = 12.5$ GeV^2	1947/1571
$Q^2_{\rm min} = 7.5$ GeV^2	2076/1660
$m_c = 1.45$ GeV (sym)	2025/1620
$Q^2_0 = 1.6$ GeV^2 (sym)	2018/1620
$m_b = 4.3$ GeV	2016/1620
$m_b = 4.1$ GeV	2014/1620
$m_t = 175.0$ GeV	2063/1620
$m_t = 172.5$ GeV	2018/1620
$R = 0.4$	2080/1620
Parameter variations	
F_{u_v}, $D_{\bar{d}}$	2007/1620

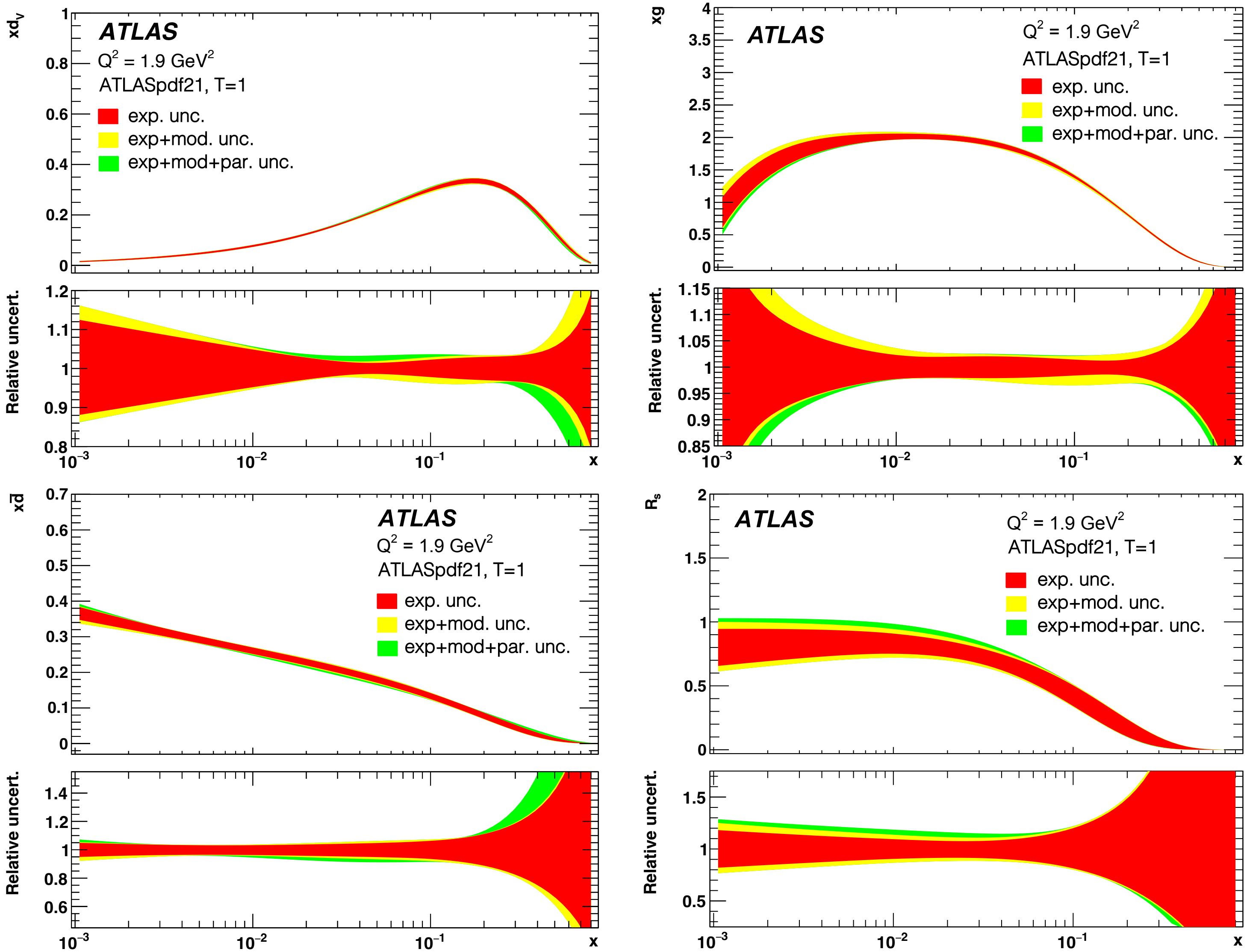






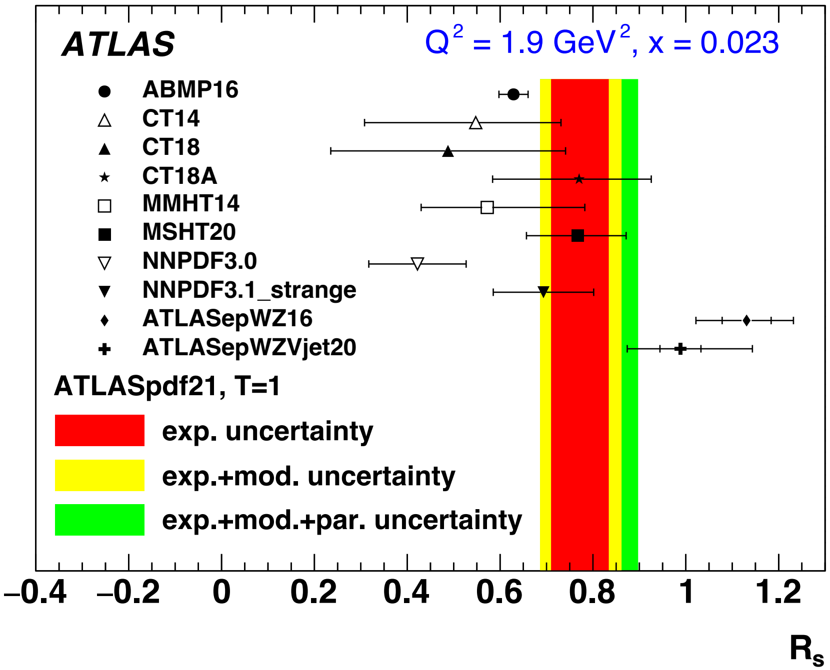
Combining uncertainties

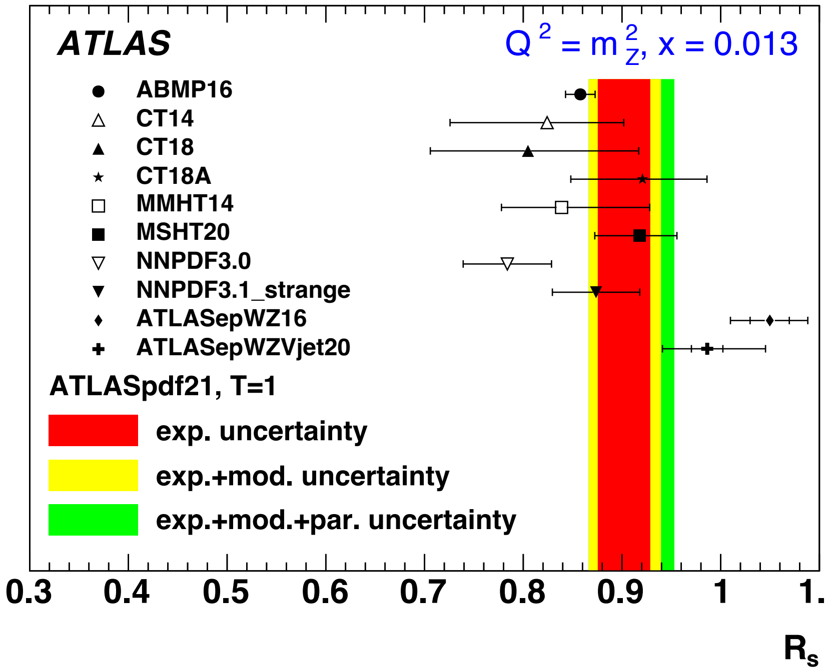
- Model and parametrisation added in quadrature and combined to the experimental uncertainties



A less strange result

- **ATLAS R_s** has come **DOWN** from ~1.0 to 0.8
- **MSHT, CT and NNPDF R_s** have come **UP** from ~0.5 to 0.8 when including W,Z 7 TeV ATLAS data
- Shift from epWZVjet20 to ATLASpdf21 due to a combination of adding W,Z 8 TeV data and our freer low-x parametrisation



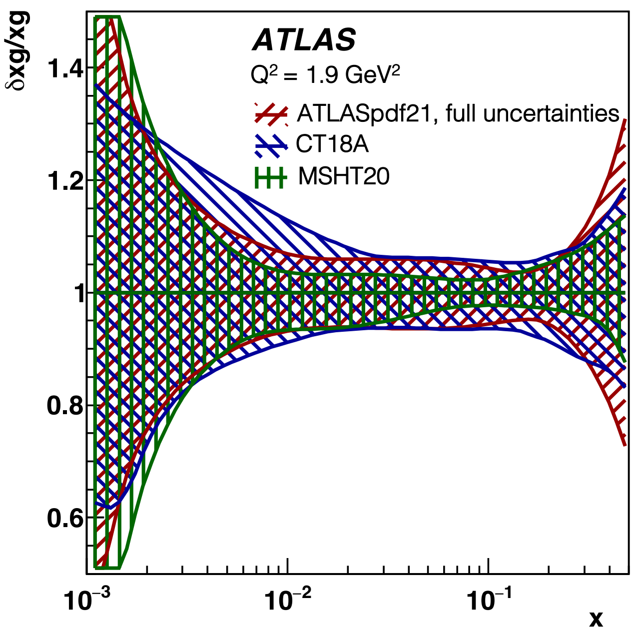


Considering an enhanced tolerance

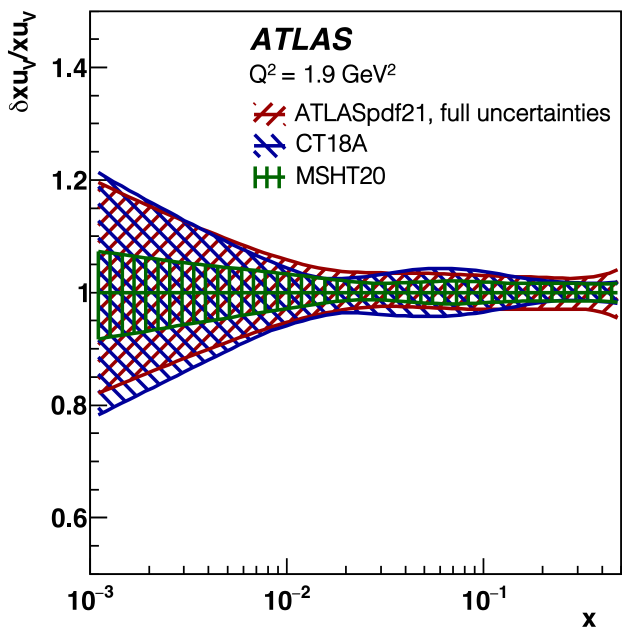
- Considering an enhanced tolerance following the **MSHT dynamic tolerance procedure**

- We decided to use **$T = \sqrt{\Delta\chi^2} = 3$**

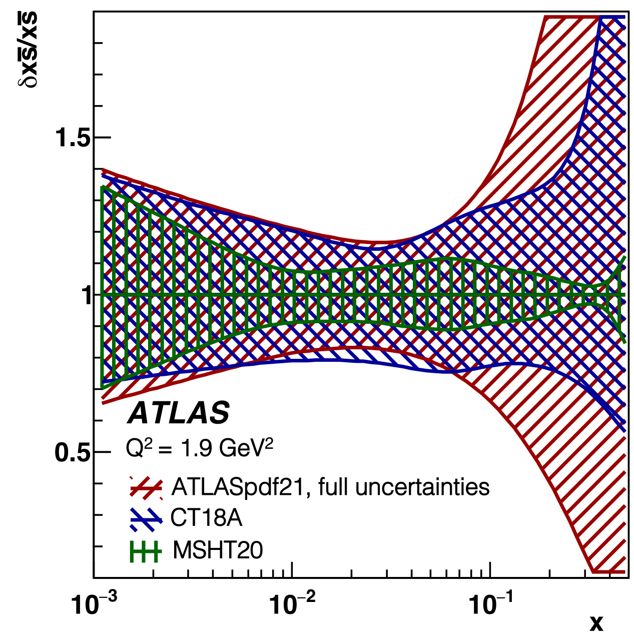
- ATLAS uncertainties expected to be larger at high-x because global PDF fits include fixed target and Tevatron data (both DIS and DY)



gluon PDF

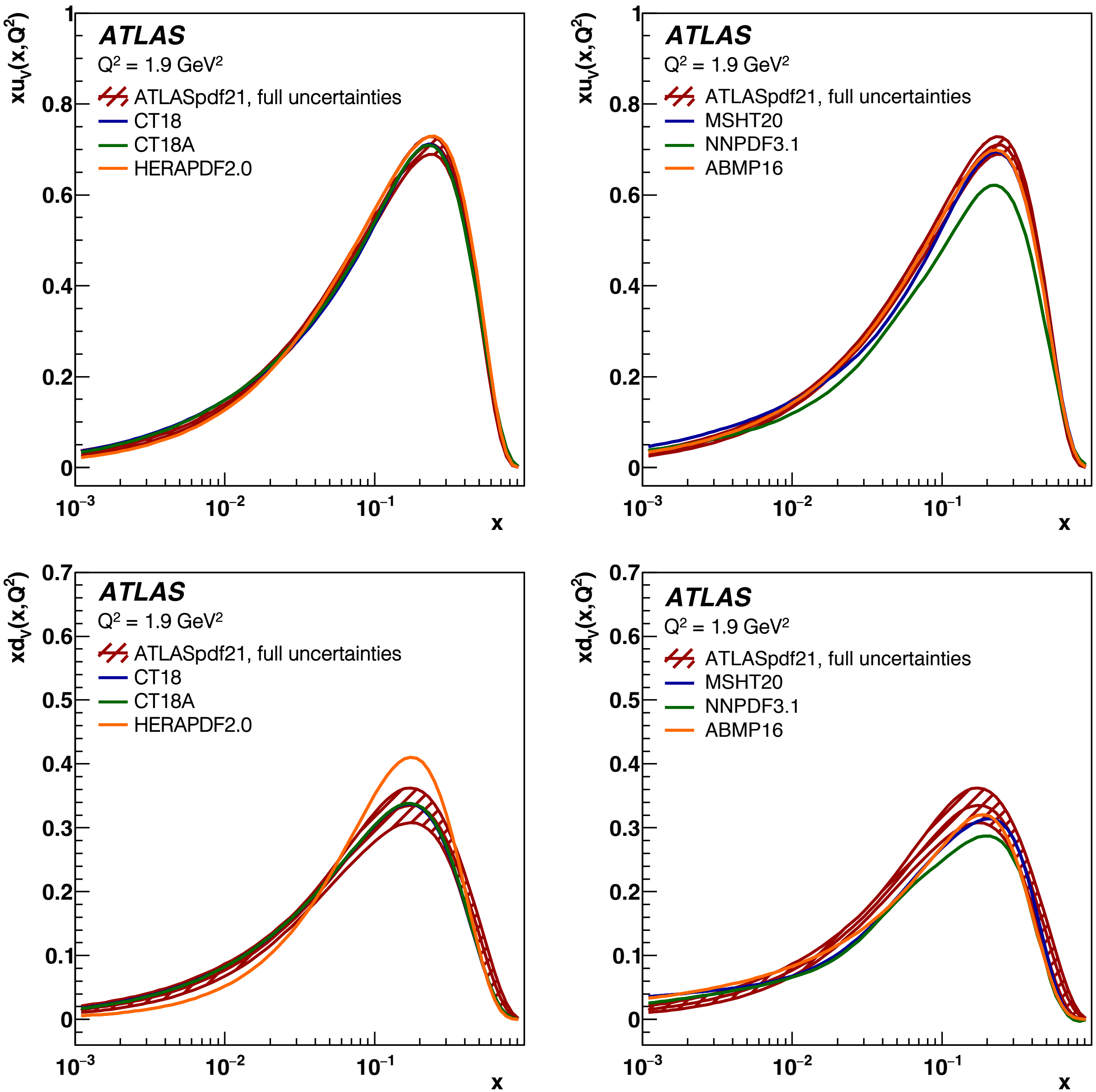


u_V PDF



strange PDF

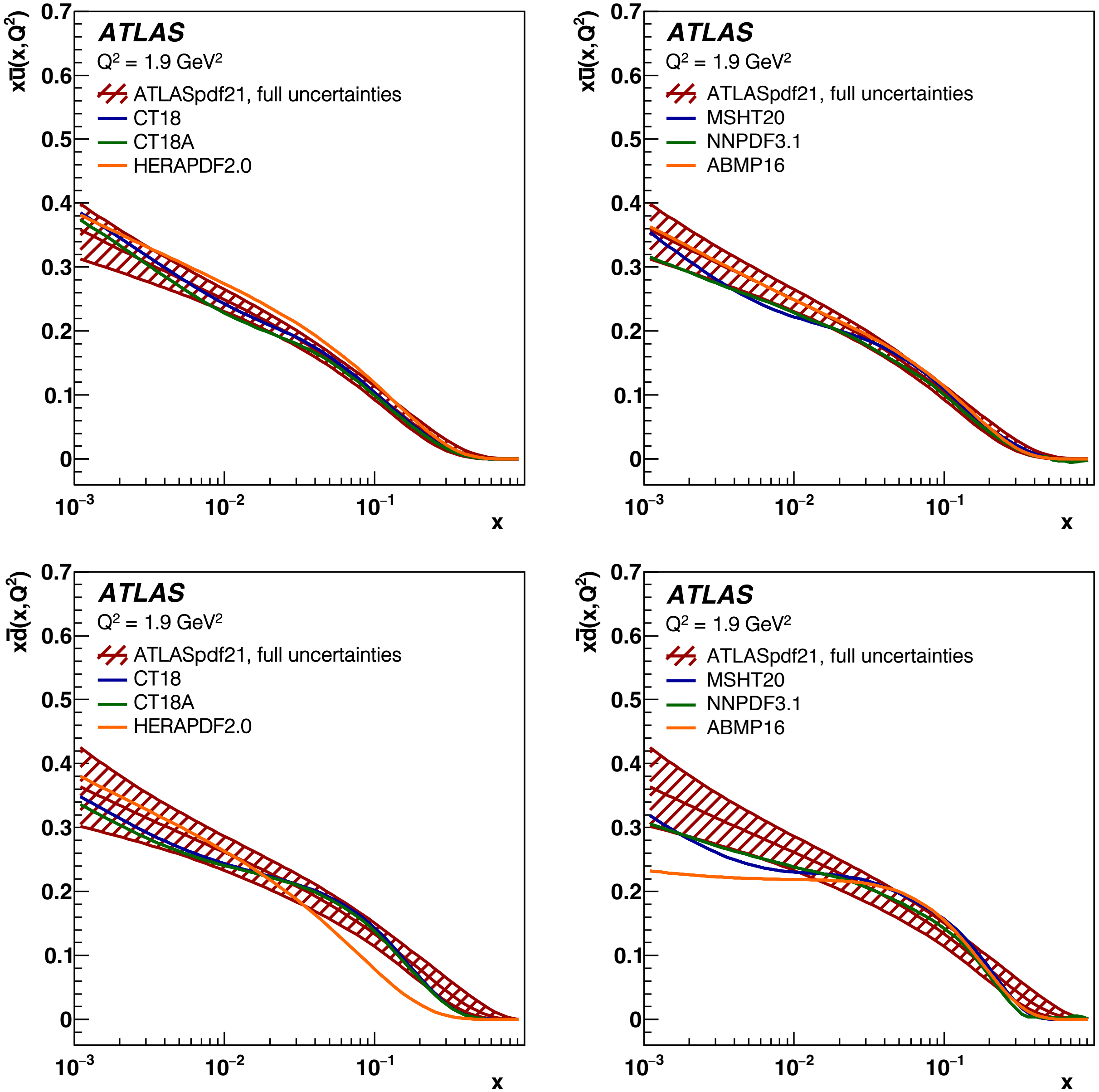
Comparison to global PDF sets



- The addition of ATLAS data has resulted in the d_V PDF moving away from the HERAPDF towards the global fits
- **Lower χ^2 for these data than the global fitters**

PDF (free pars)	χ^2/NDP
ATLASpdf21 (21)	**2010/1641**
CT18 (29)	2135/1641
CT18A (29)	2133/1641
MSHT20 (52)	2218/1641
HERAPDF2.0 (14)	2262/1641
NNPDF3.1	2109/1641

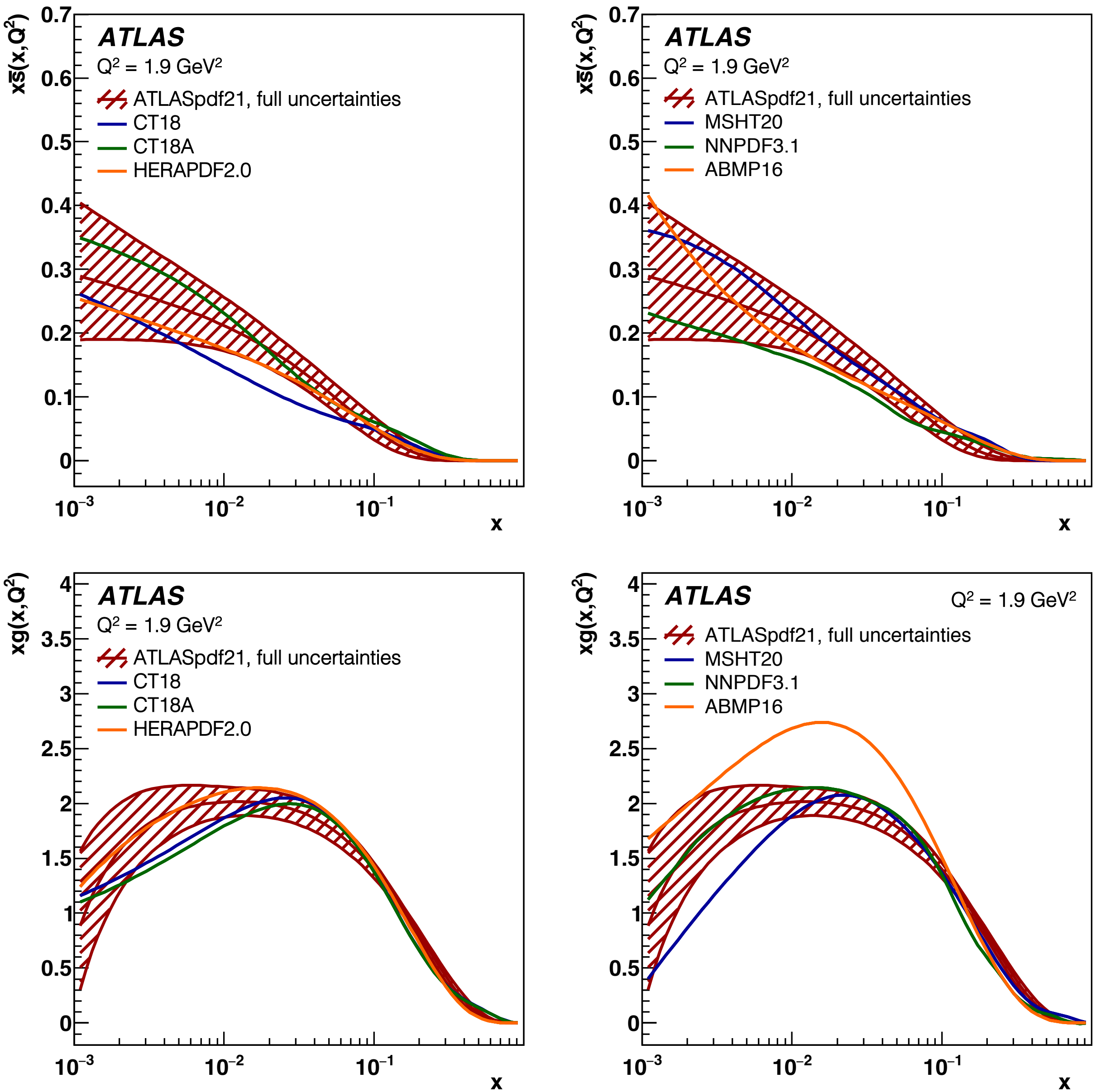
Comparison to global PDF sets



- Nice agreement between all the various PDF sets for $\bar{u}$
- The $\bar{d}$ also moves towards the global fits, being much harder than that of the HERAPDF at high-x

PDF (free pars)	χ^2/NDP
ATLASpdf21 (21)	**2010/1641**
CT18 (29)	2135/1641
CT18A (29)	2133/1641
MSHT20 (52)	2218/1641
HERAPDF2.0 (14)	2262/1641
NNPDF3.1	2109/1641

Comparison to global PDF sets



- xg is more similar to HERAPDF and NNPDF
- Note that ABMP uses a different value of $\alpha_S(m_Z)$
- xs agrees with MSHT20 and CT18A nicely

PDF (free pars)	χ^2/NDP
ATLASpdf21 (21)	**2010/1641**
CT18 (29)	2135/1641
CT18A (29)	2133/1641
MSHT20 (52)	2218/1641
HERAPDF2.0 (14)	2262/1641
NNPDF3.1	2109/1641

Conclusions & outlook

- This fit demonstrates that it is is possible to fit, simultaneously, and with small uncertainties, **a very wide portfolio or ATLAS data from different years**, and across the whole spectrum of QCD processes

- Making **proper use of the detailed correlated uncertainties** as recommended by ATLAS is **essential** in order to beat down these uncertainties and achieve the precision demanded by the experimental uncertainties

- This is **very important for any fits for instance for α_S**, since they can be very sensitive to small changes

- **Use of NNLO theory is essential**, as the ever shrinking experimental uncertainties continue to challenge those from the theoretical predictions

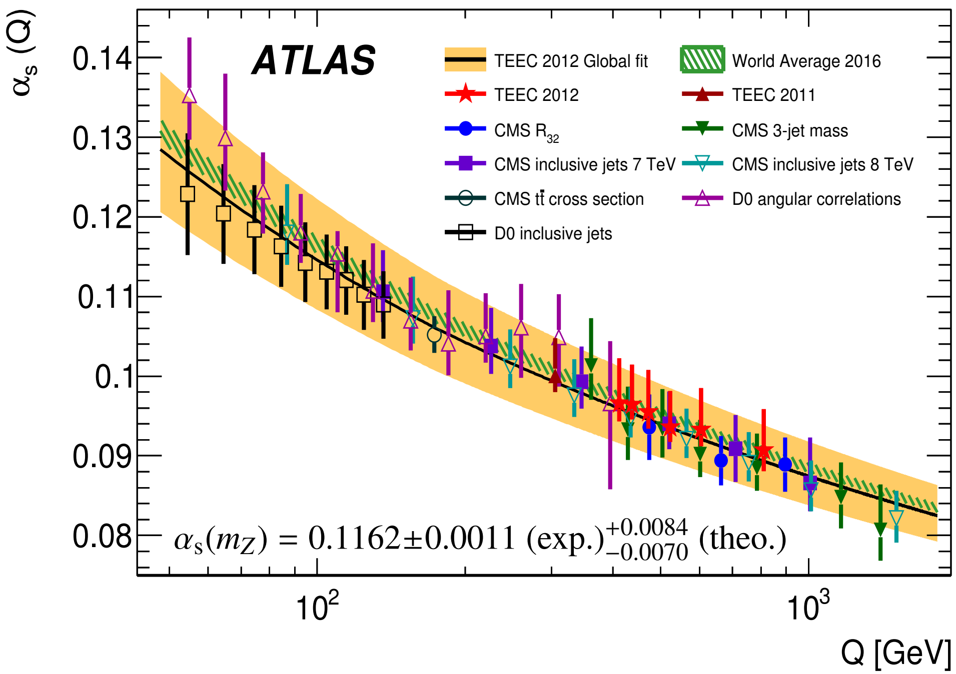
- We still have data available for analysis, and soon the start LHC Run 3 will allow us to push down the experimental uncertainties yet further, over even more of the LHC phase space

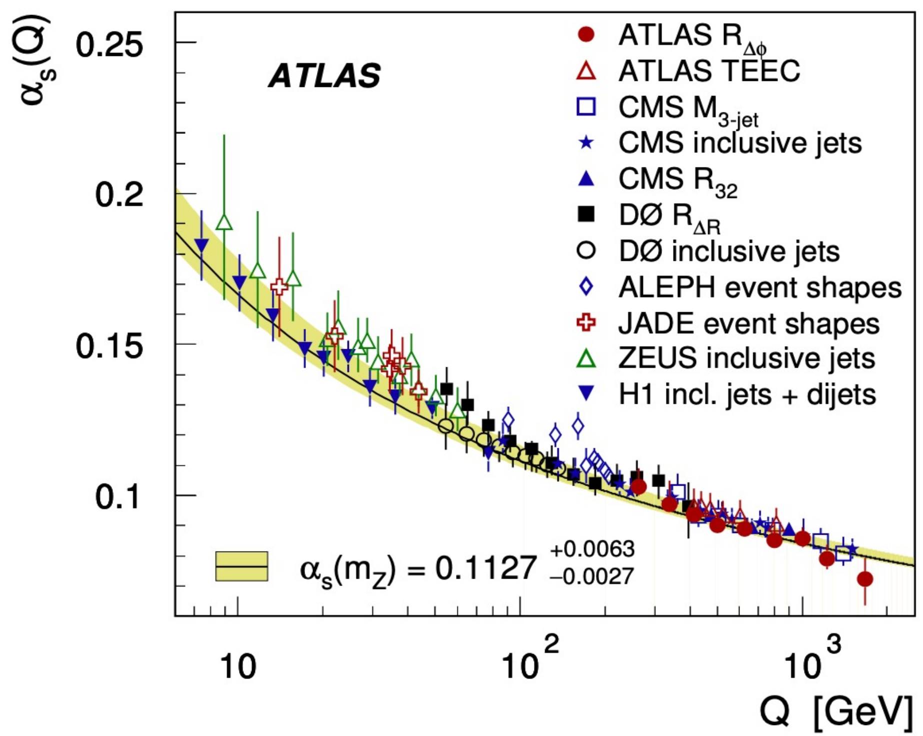
- Stay tuned!

Backup Slides

Some ATLAS α_S studies

- In this QCD analysis we do not attempt a simultaneous PDF and α_S fit, but ATLAS has produced α_S fits in the past
- Measurement of the transverse energy-energy correlations in multi-jet events - Phys. Lett. B 750 (2015) 427, Eur. Phys. J. 77 (2017) 872
- Measurement of the rapidity and transverse momentum dependence of dijet azimuthal decorrelations using $R_{\Delta\phi}$ - Phys. Rev. D 98 (2018) 092004





- More on Thursday afternoon!

The χ^2 formula

$$\chi^2 = \sum_{ik}\left(D_i - T_i(1-\sum_j \gamma_{ij}b_j)\right) C^{-1}_{\mathrm{stat},ik}(D_i, D_k)\left(D_k - T_k(1-\sum_j \gamma_{kj}b_j)\right) \quad \textbf{partial}$$

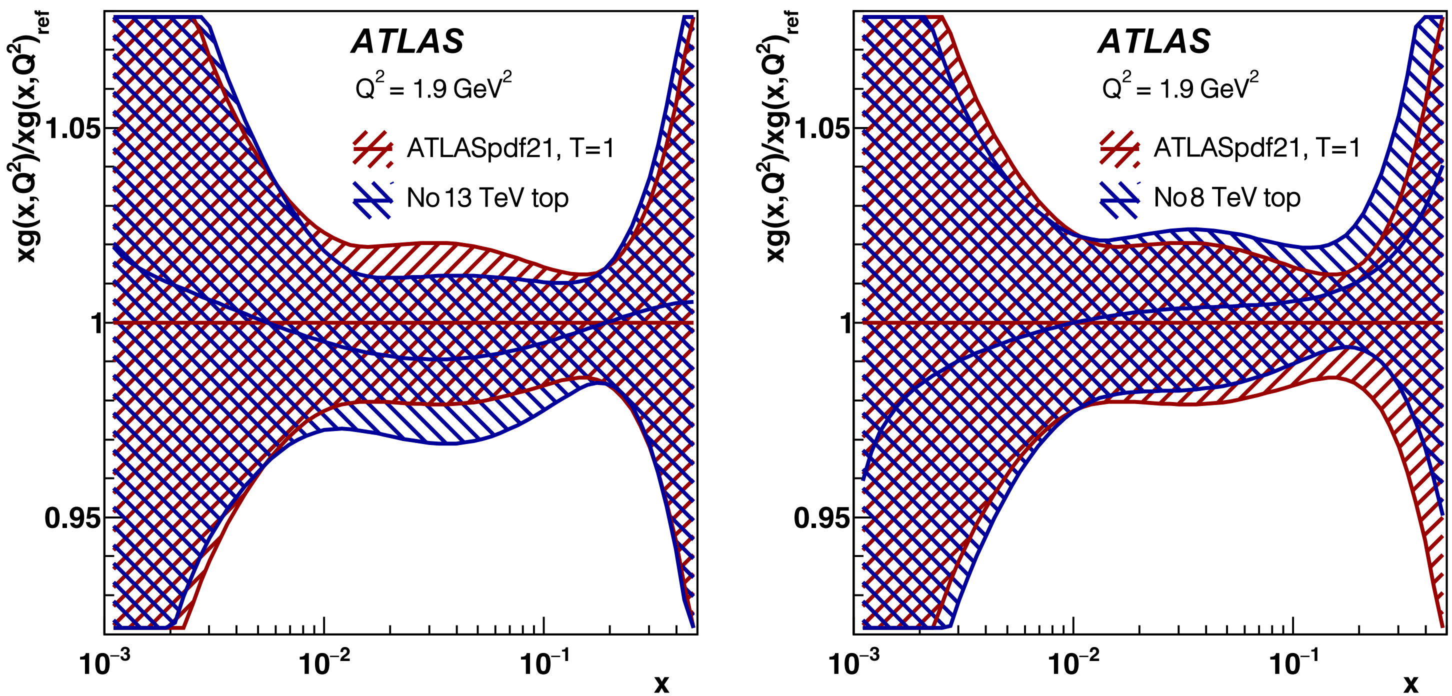
$$+ \sum_i \log \frac{\delta^2_{i,\mathrm{uncor}} T_i^2 + \delta^2_{i,\mathrm{stat}} D_i T_i}{\delta^2_{i,\mathrm{uncor}} D_i^2 + \delta^2_{i,\mathrm{stat}} D_i^2} \quad \textbf{log penalty term}$$

$$+ \sum_j b_j^2 \quad \textbf{correlated term}$$

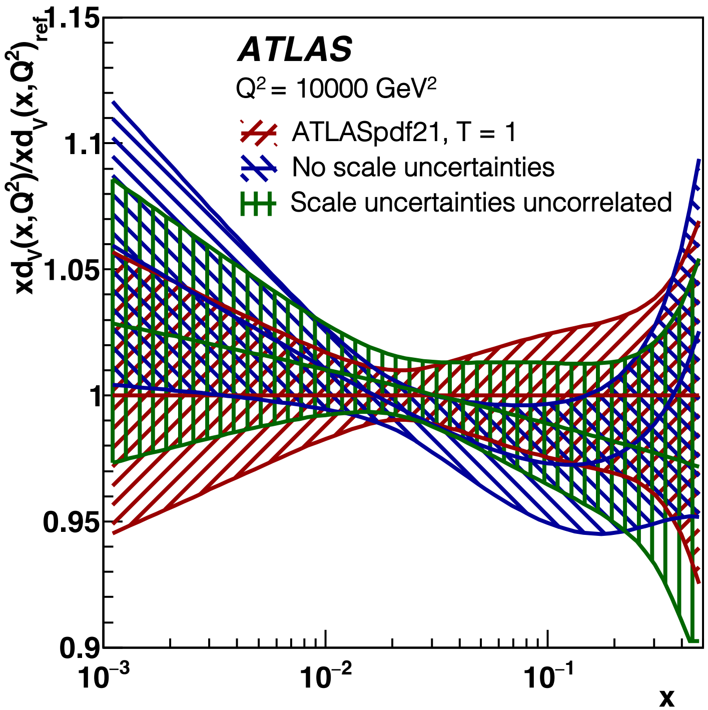
- D_i represents the measured data and T_i represents the corresponding data predictions
- $\delta_{i,uncor}$ and $\delta_{i,stat}$ are the uncorrelated systematics and statistical uncertainties on D_i
- Correlated systematics, described by γ_{ij}, are accounted for using the nuisance parameters b_j
- $C_{stat,ik}$ is the statistical (plus uncorrelated) covariance matrix
- The log penalty term is a small bias correction term

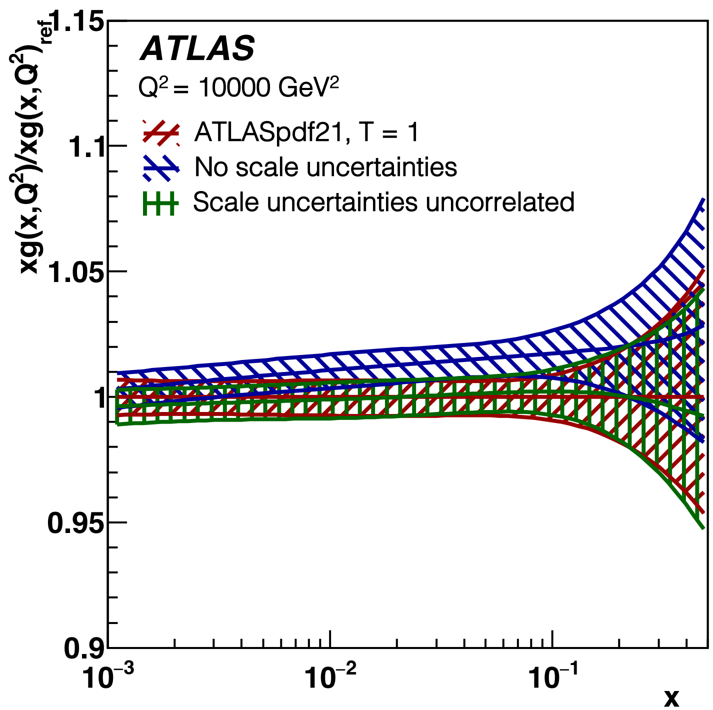
Impact of the various data sets on PDFs

- We removed all the ***$t\bar{t}$* data** from the fit
- Removing the 8 and 13 TeV $t\bar{t}$ data separately makes it clear that the **8 TeV has the bigger effect**
- 8 TeV looked to have more impact when combined with only a few data sets but not so much after many other data sets are added



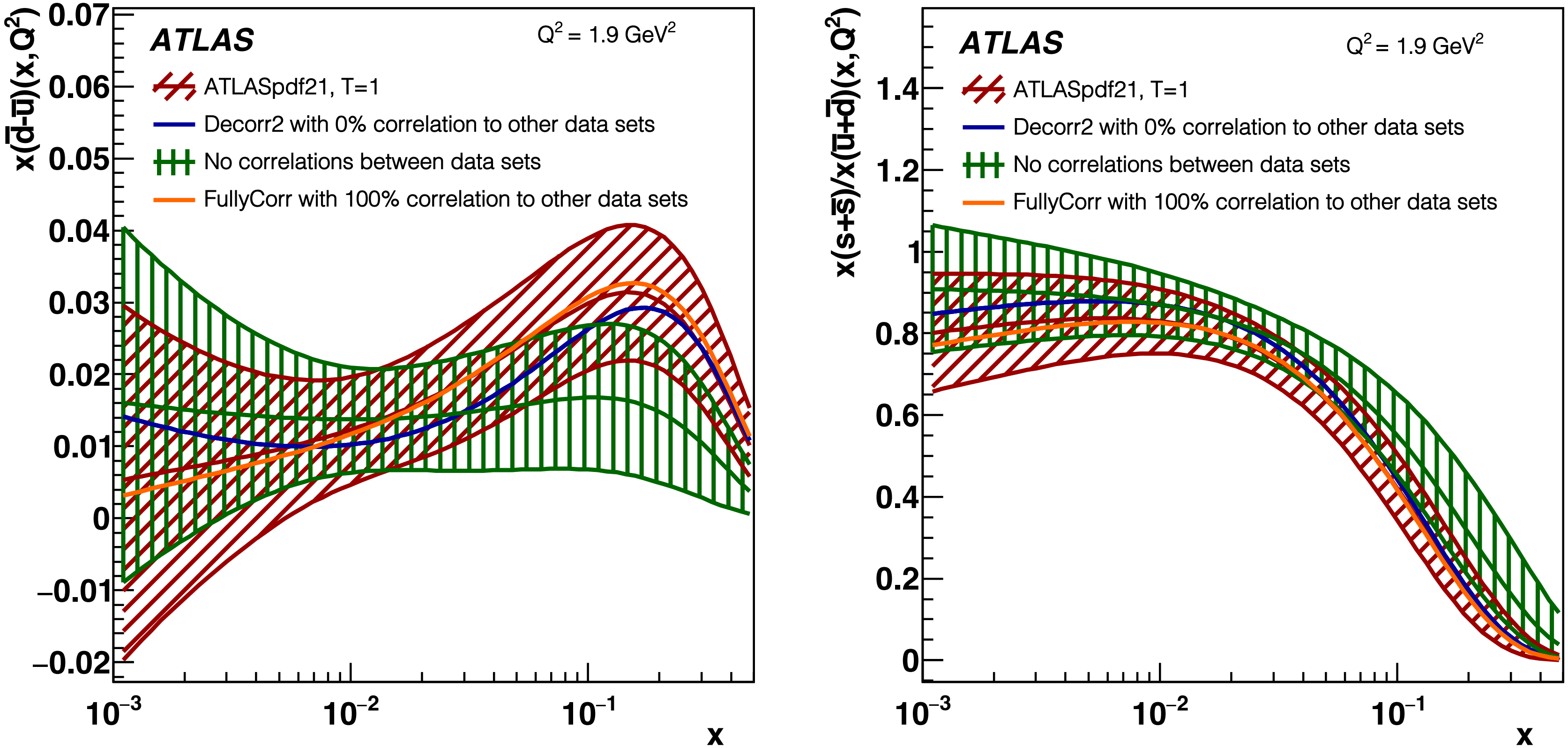
Scale uncertainties





- The differences in magnitude of the PDF uncertainties are very small
- The differences between the PDF shapes are not large, but they can be important if O(1%) is sought on PDFs
- The case where scale uncertainties are included but not correlated between 7 and 8 TeV W,Z data sets is also shown – smaller effect

Correlation between various data sets

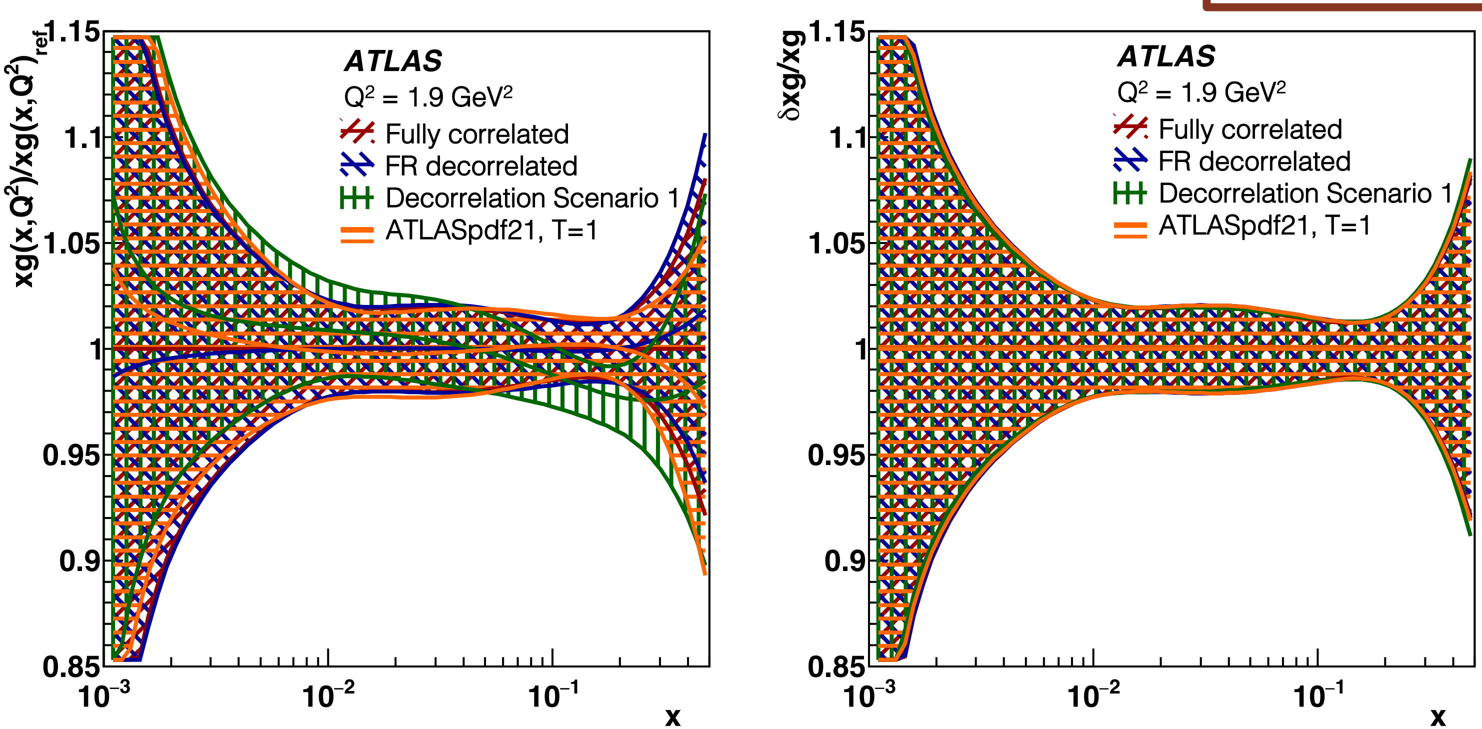


- The major effect of the correlations come from the correlations between V+jets and $t\bar{t}$+jets
- The exact degree of correlation to the inclusive data does not change the resulting PDFs significantly
- The choice of correlating all the inclusive jet systematics is also not important

Talking about inclusive jets

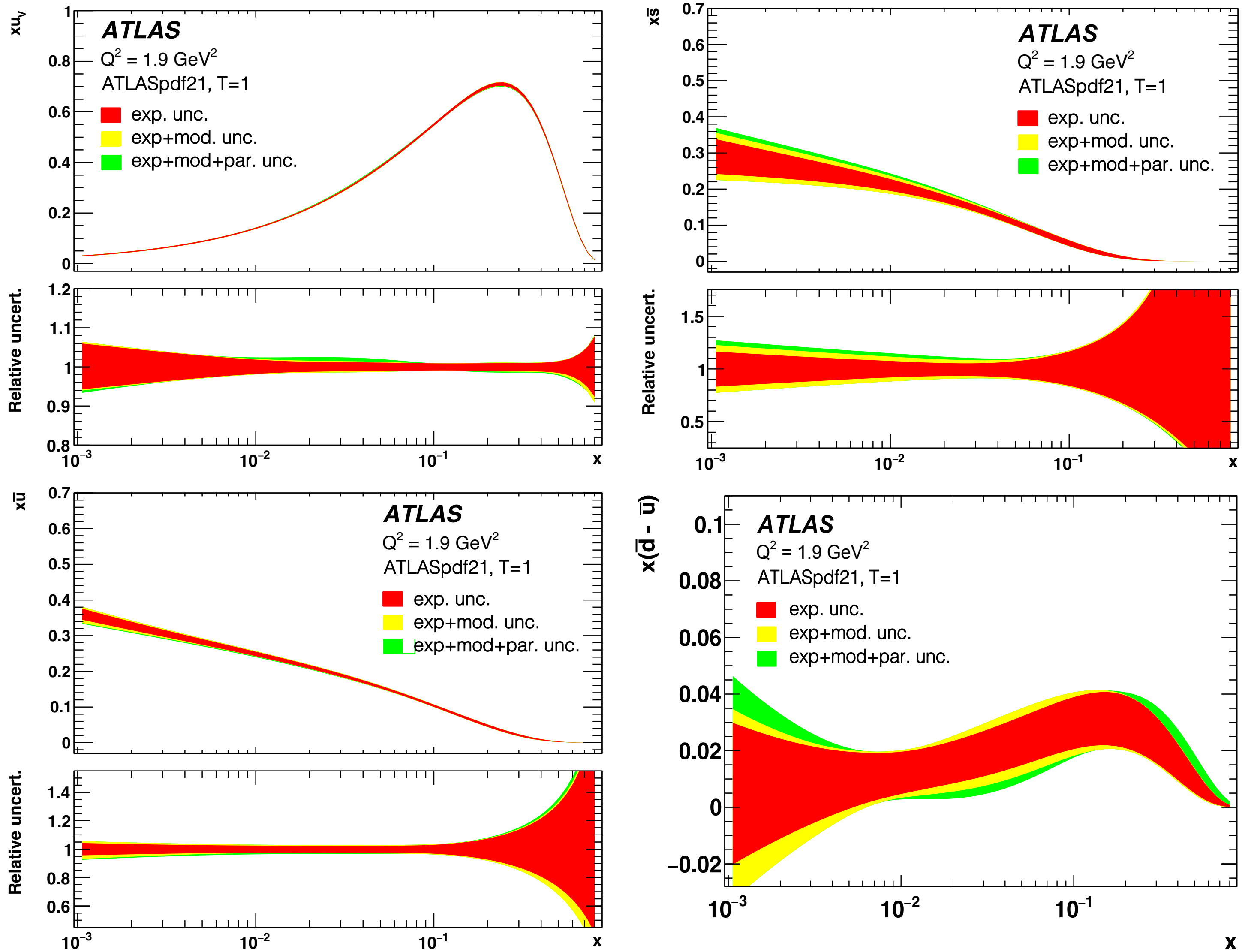
- Different model to treat correlated systematics:
  - keeping them fully correlated
  - decorrelating the Jets Flavour Response (FR) between rapidity bins
  - Two decorrelation scenarios as recommended in the 8 TeV jet paper
- This affects the χ^2 but has little effect on the PDFs

jets 8 TeV R=0.6	Fully Correlated	FR Decorrelated	Decorrelation Scenario 1	Decorrelation Scenario 2
χ^2/NDP	289/171	227/171	250/171	248/171

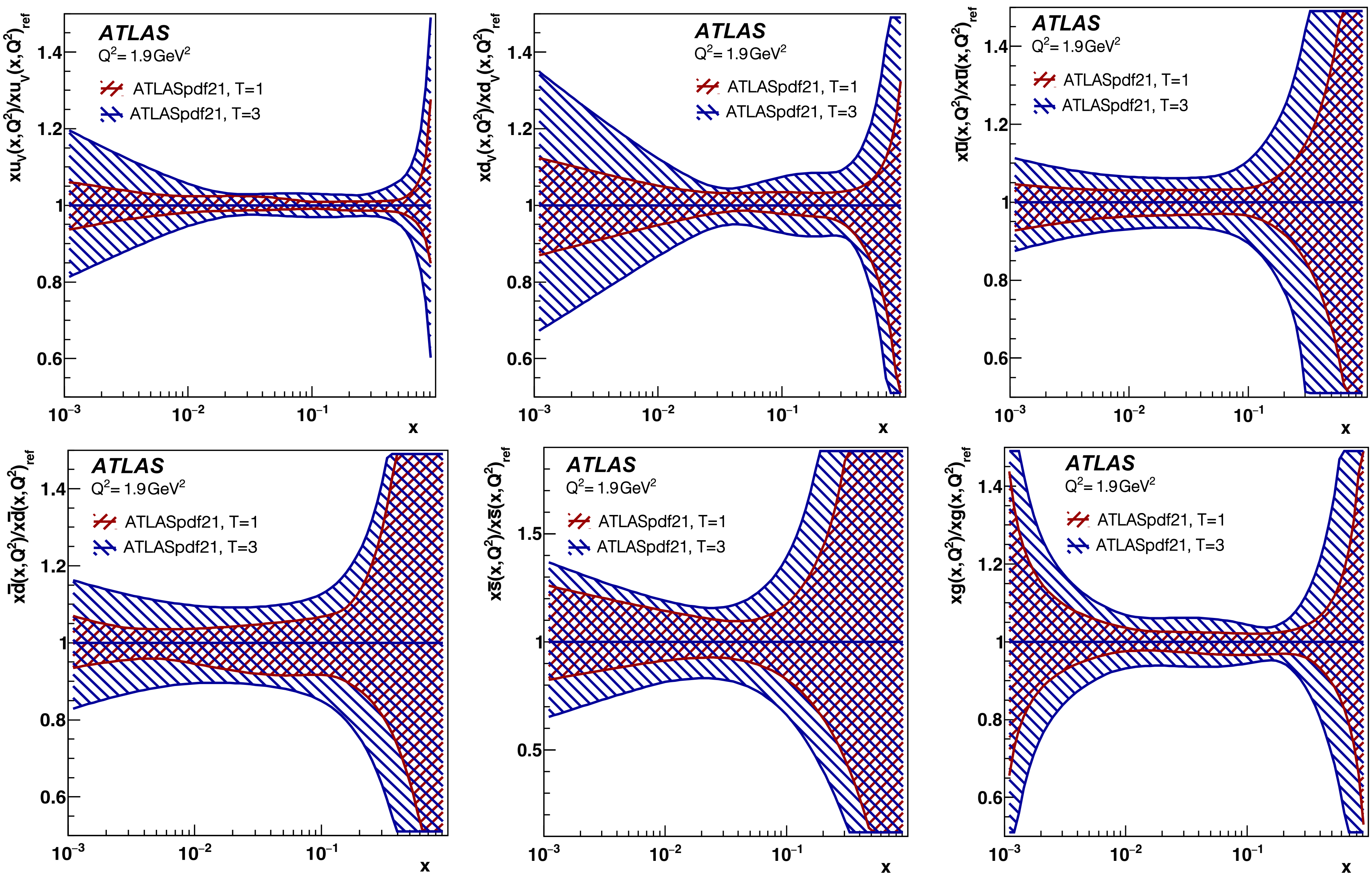


Combining uncertainties

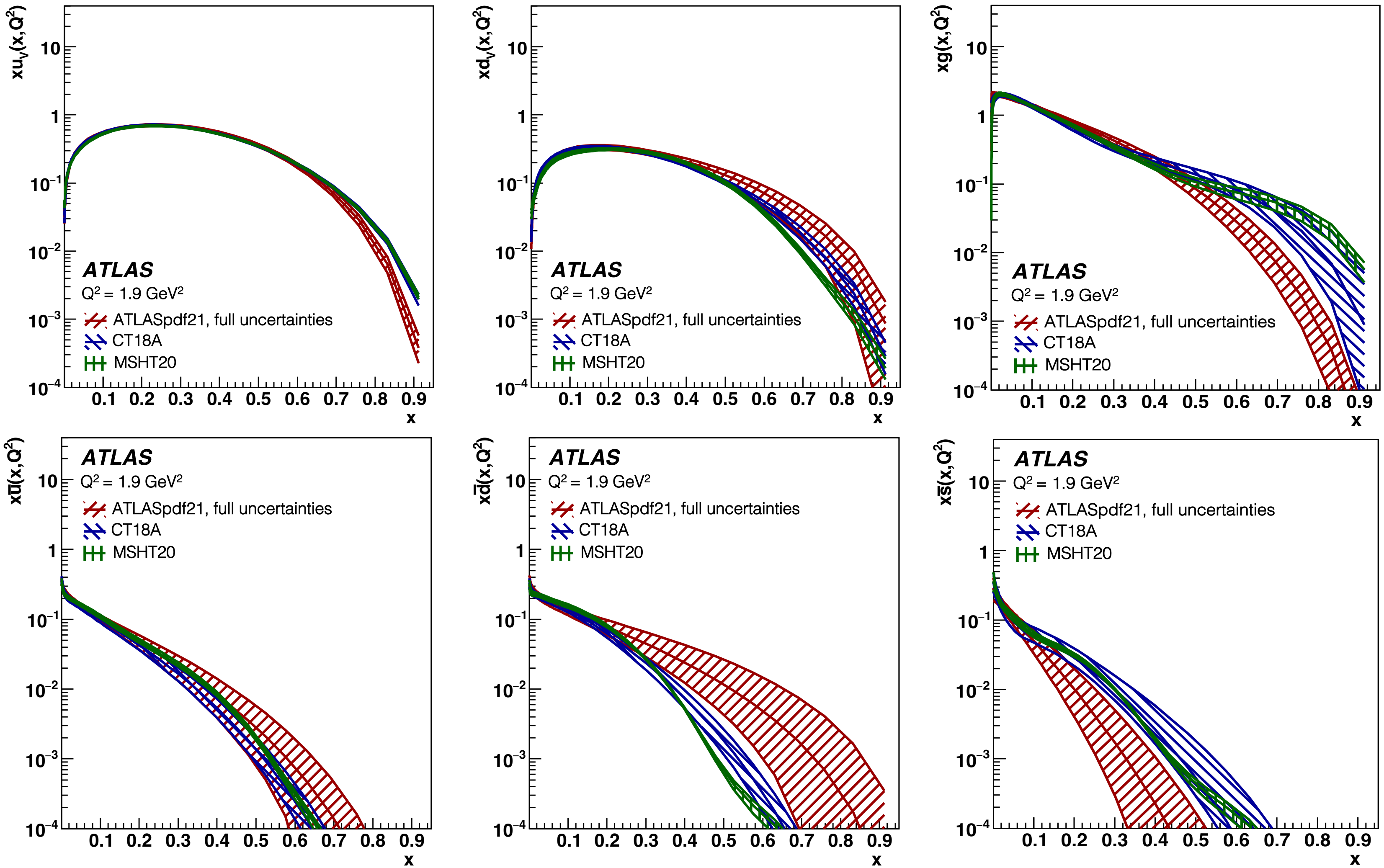
- Model and parametrisation added in quadrature and combined to the experimental uncertainties



T = 1 vs T = 3


ATLAS
Q^2= 1.9 GeV^2
ATLASpdf21, T=1
ATLASpdf21, T=3
$xu_V(x,Q^2)/xu_V(x,Q^2)_{ref}$
$xd_V(x,Q^2)/xd_V(x,Q^2)_{ref}$
$x\bar{u}(x,Q^2)/x\bar{u}(x,Q^2)_{ref}$
$x\bar{d}(x,Q^2)/x\bar{d}(x,Q^2)_{ref}$
$x\bar{s}(x,Q^2)/x\bar{s}(x,Q^2)_{ref}$
$xg(x,Q^2)/xg(x,Q^2)_{ref}$
x
10^{-3}
10^{-2}
10^{-1}

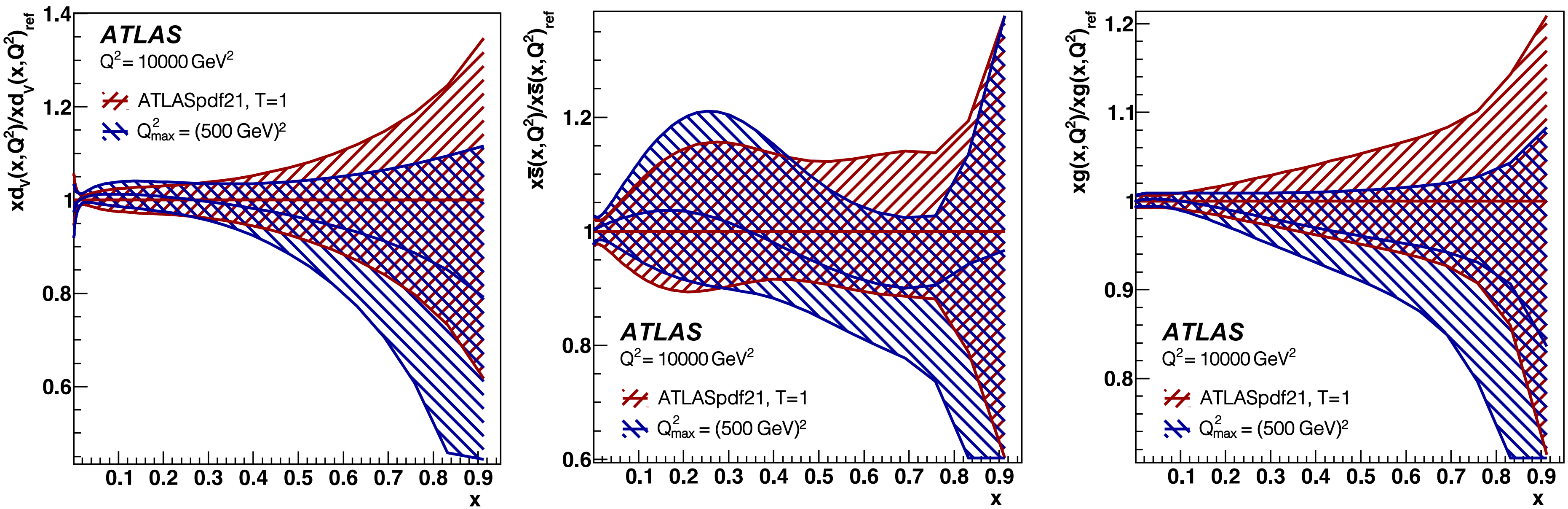
Focus on the very high-x region



- Discrepancies appear, but they are not so severe for T=3
- Discrepancies most evident in the sea sector (even between CT18A and MSHT)

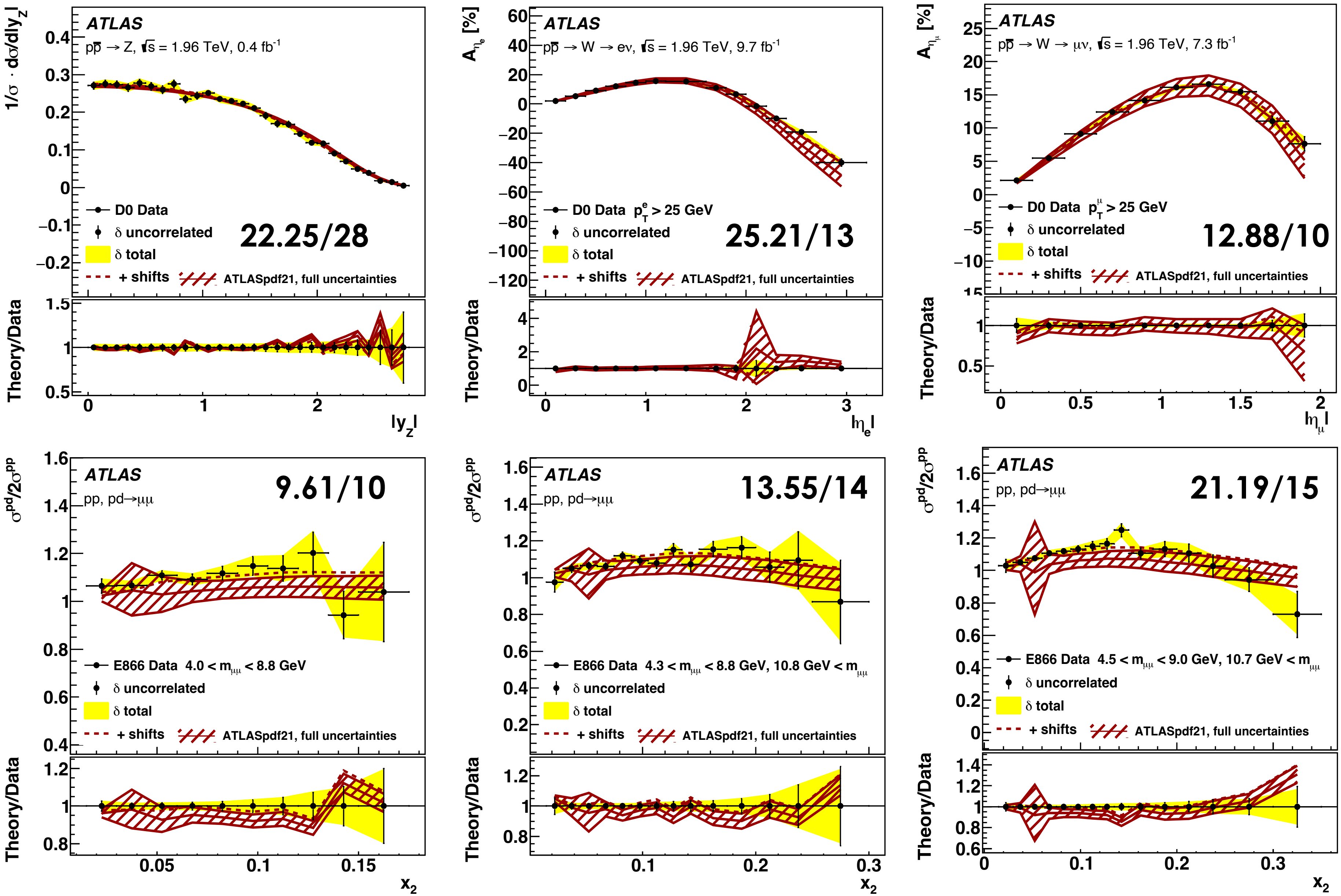
High-Q^2 data

- Subtle effects of BSM physics may be present in the high-scale data → if these data are included in a PDF fit, the estimate of backgrounds in searches could be distorted
- Check if the PDFs differ when we cut out possible hidden new physics
- Data with $Q^2 > (500\ \text{GeV})^2$ removed (mainly jet data)

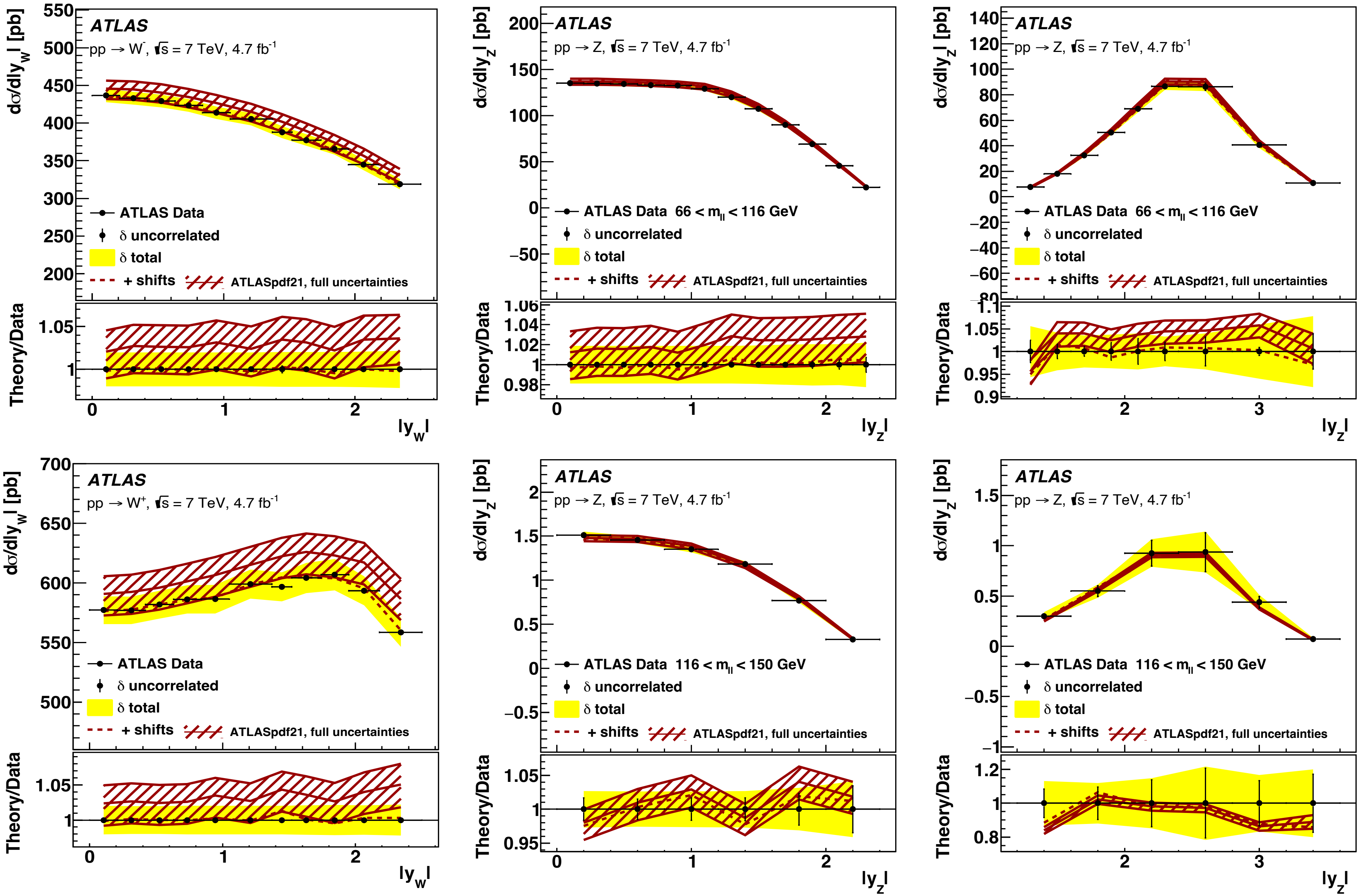


- PDF shapes and uncertainties not strongly affected

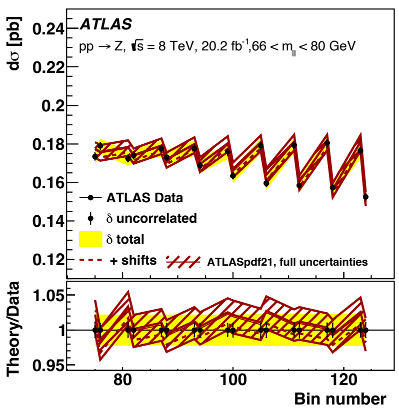
Description of Tevatron/fixed-target data

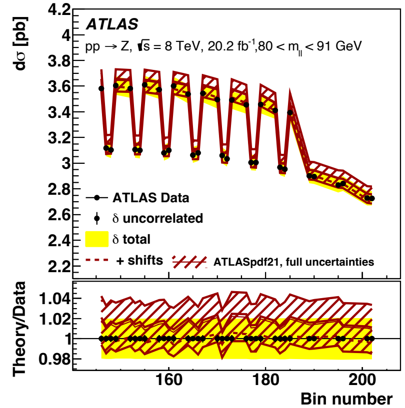


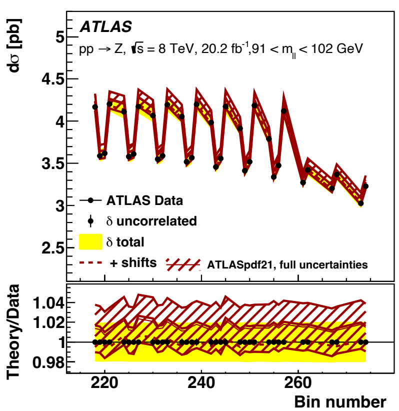
Description of W,Z 7 TeV data

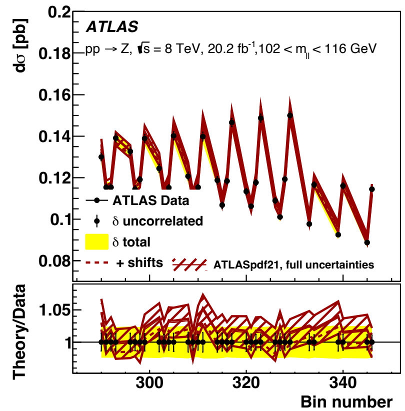


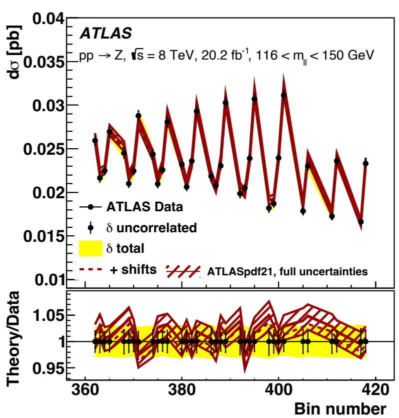
Description of W,Z 8 TeV data

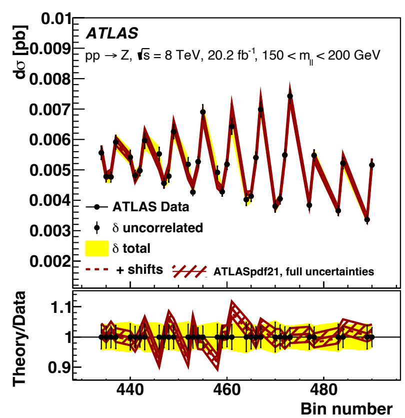


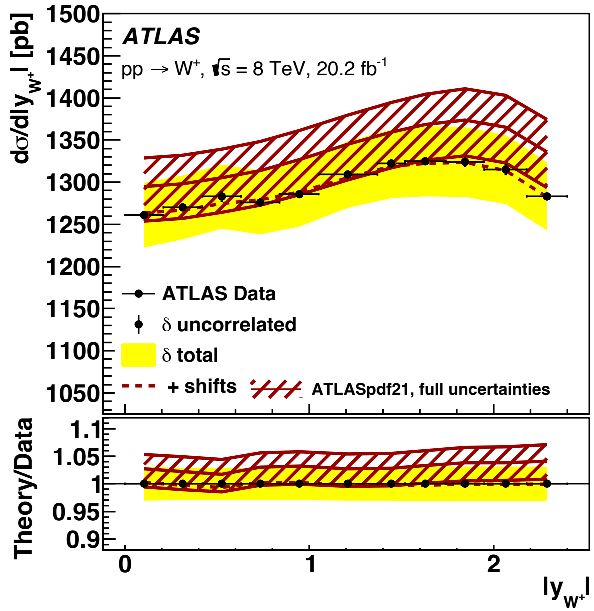


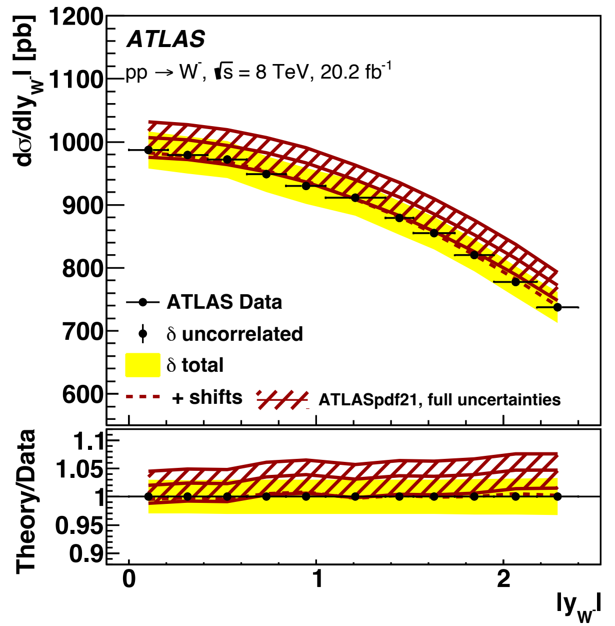




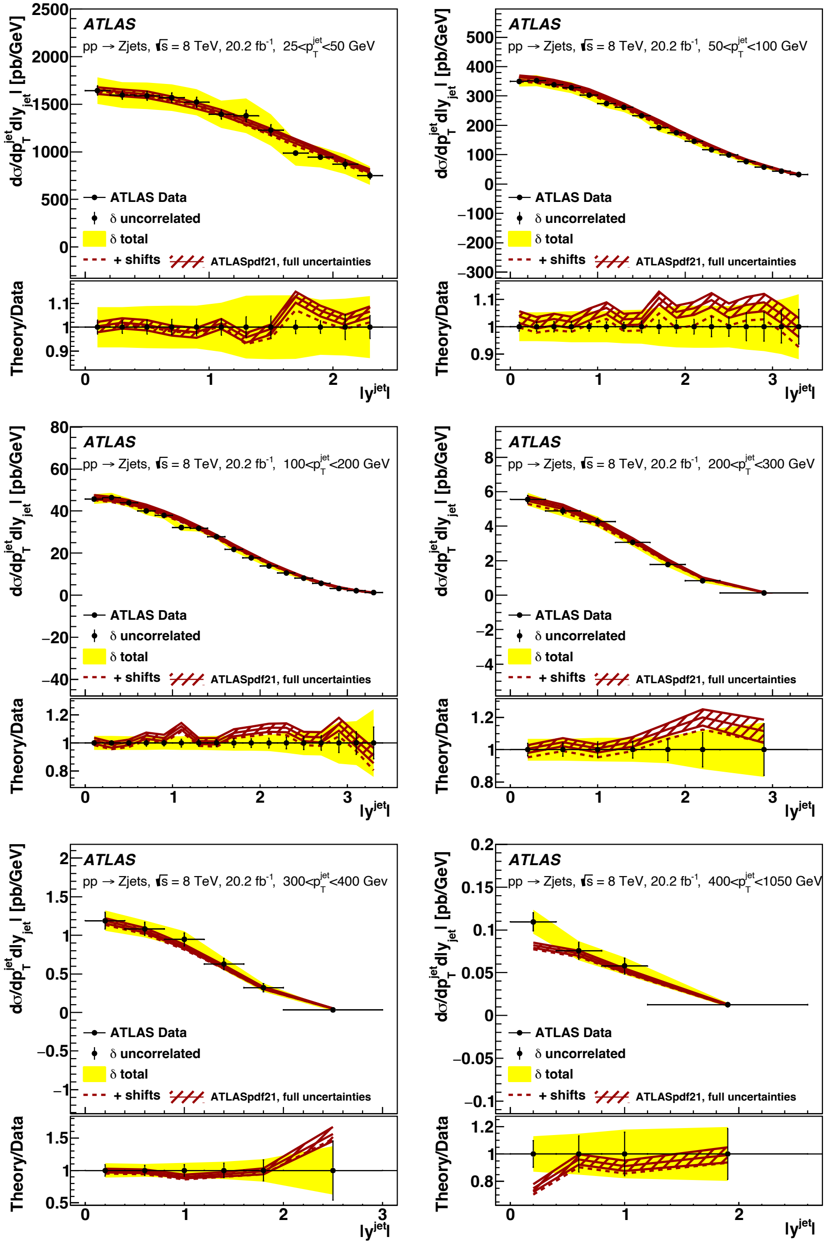


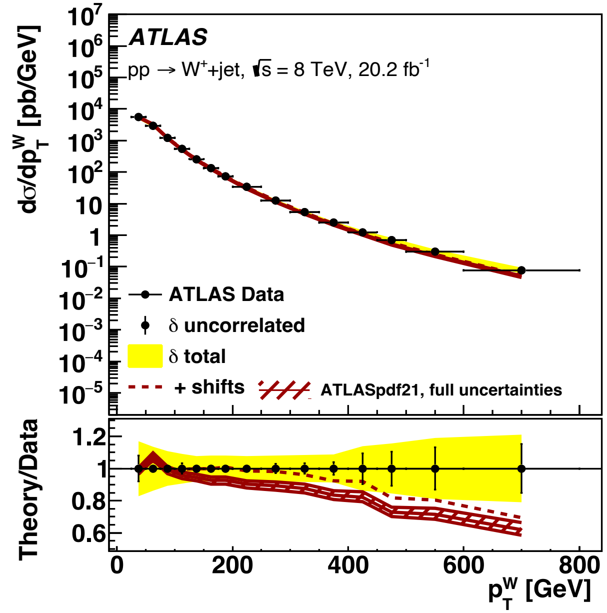


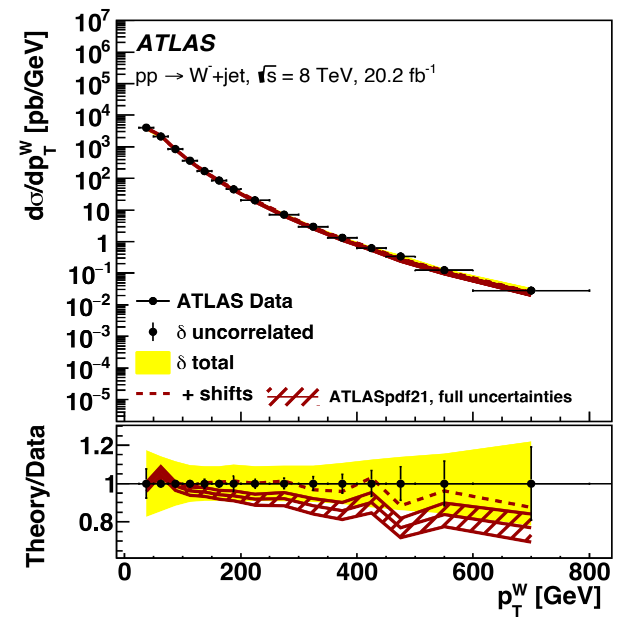




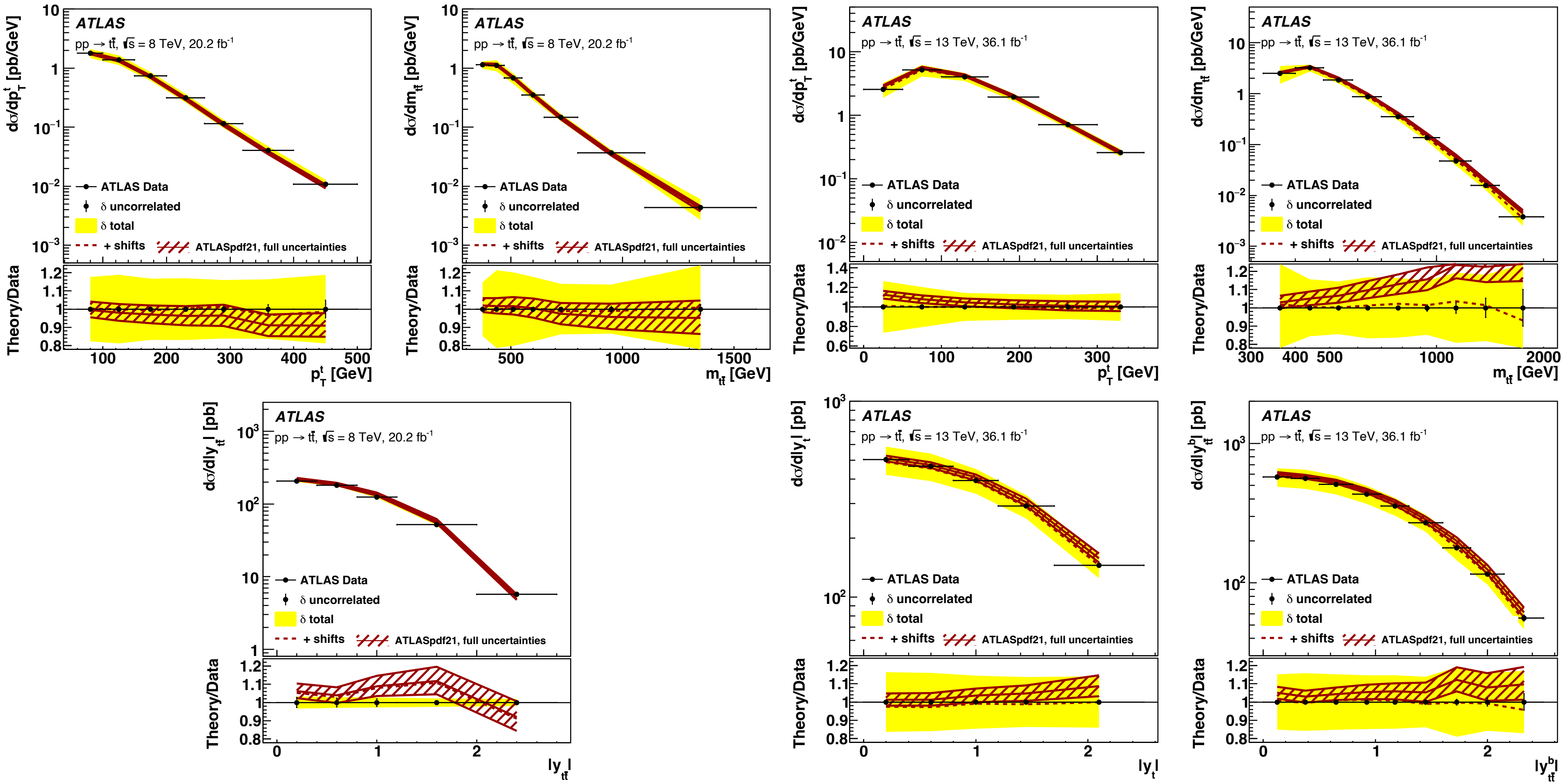
Description of V+jets data



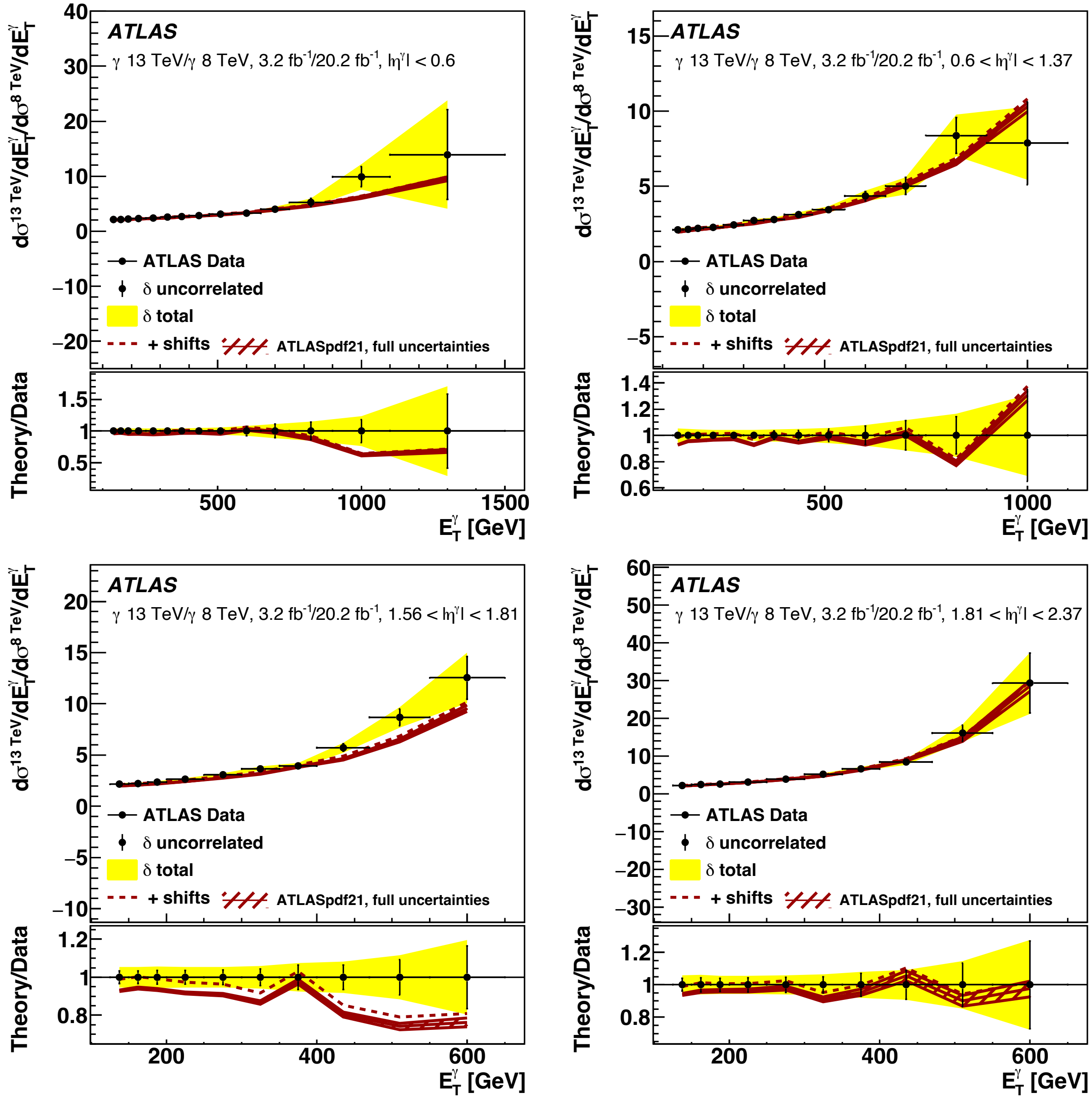




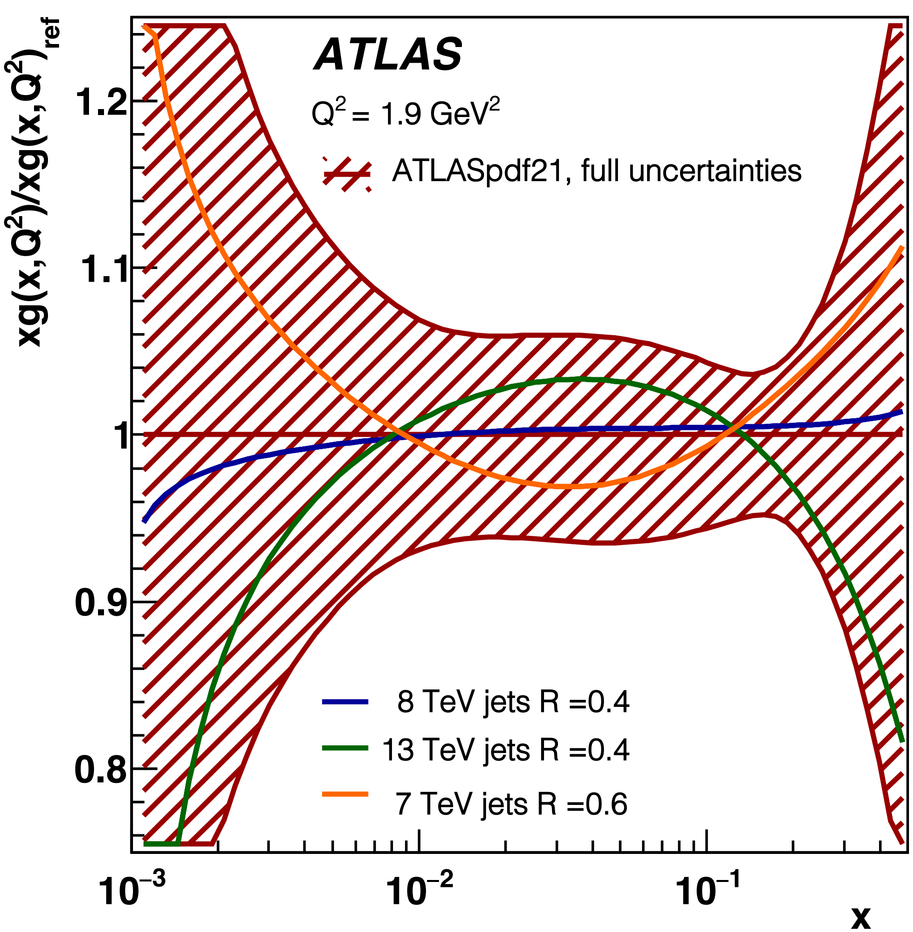
Description of top data

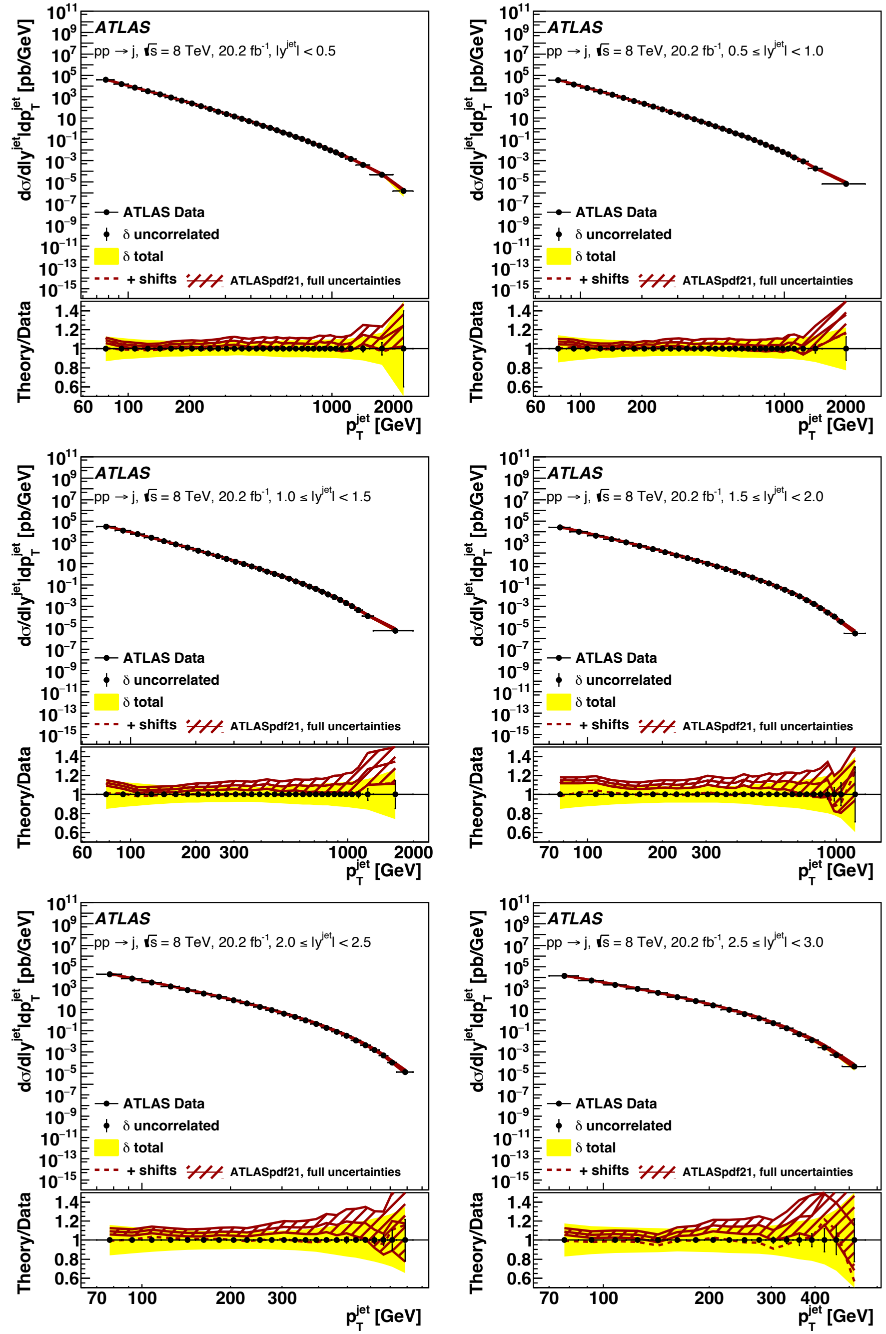


Description of photon data

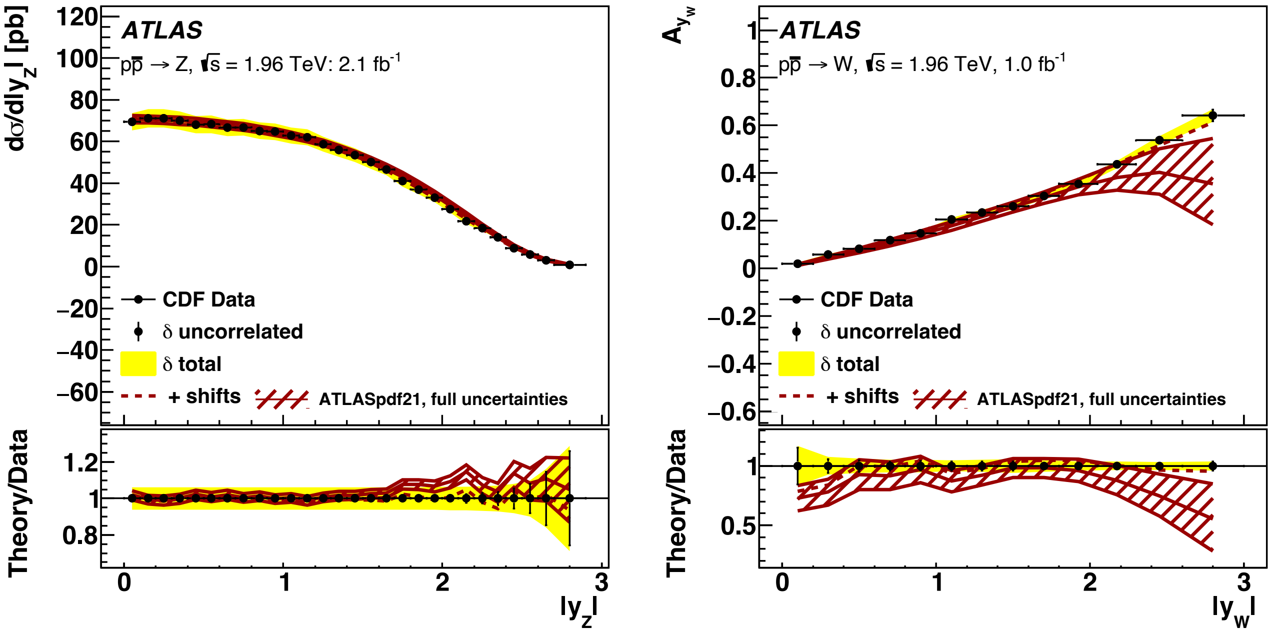


Description of inclusive jet data





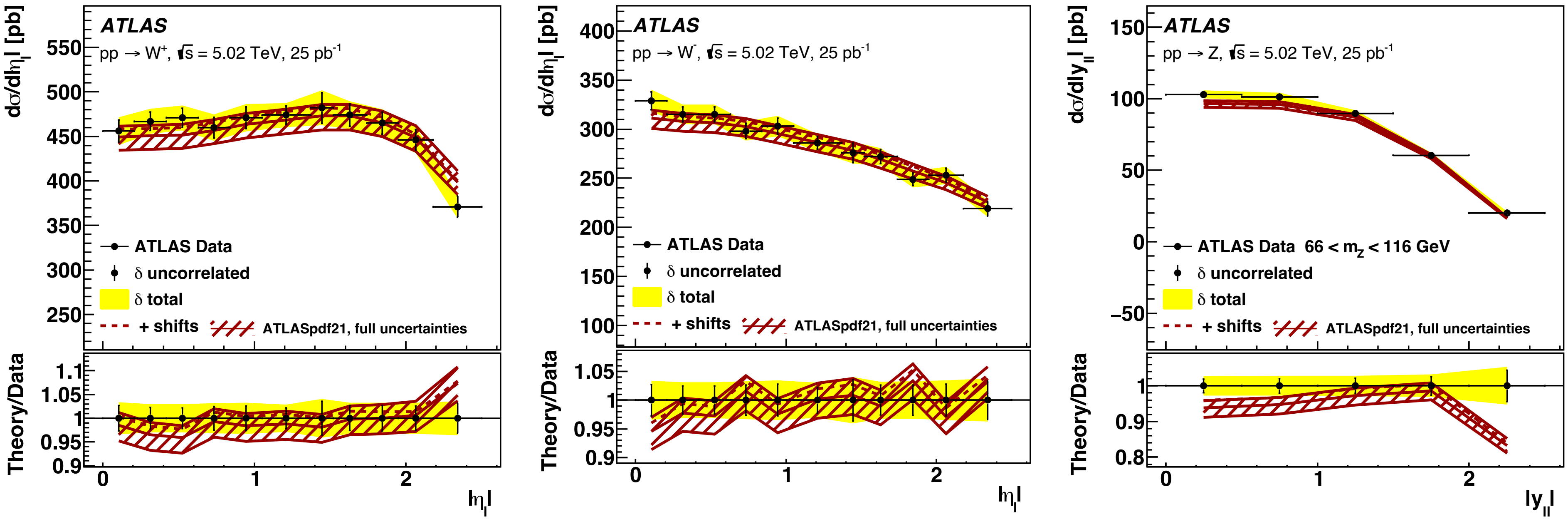
Description of CDF data



Data set	χ^2/ndf
Z	30.46/28
W	35.24/13

- Other PDFs e.g. MSHT20 and NNPDF3.1 also have a less good description of the W-electron asymmetry (e.g. 34/12)
- The ATLAS fit provides a fair description, **χ^2/NDF = 126/92**, of Tevatron data, which mostly influence the high-x valence quarks

Description of W,Z 5.02 TeV data

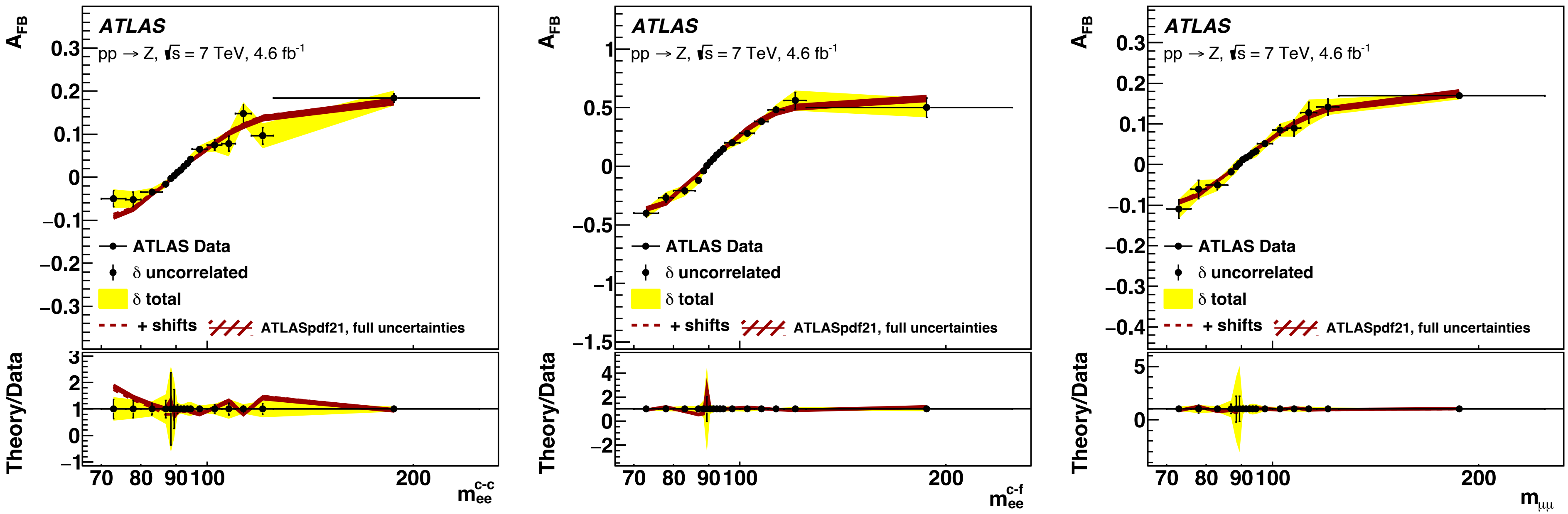


- ➢ The data have not been included because there is no information on full correlated systematics

- ➢ They are well described, apart from a single data point

Data set	χ^2/ndf
Z	22.11/5
W^+	11.18/11
W^-	11.65/11

Description of A_{FB} 7 TeV data



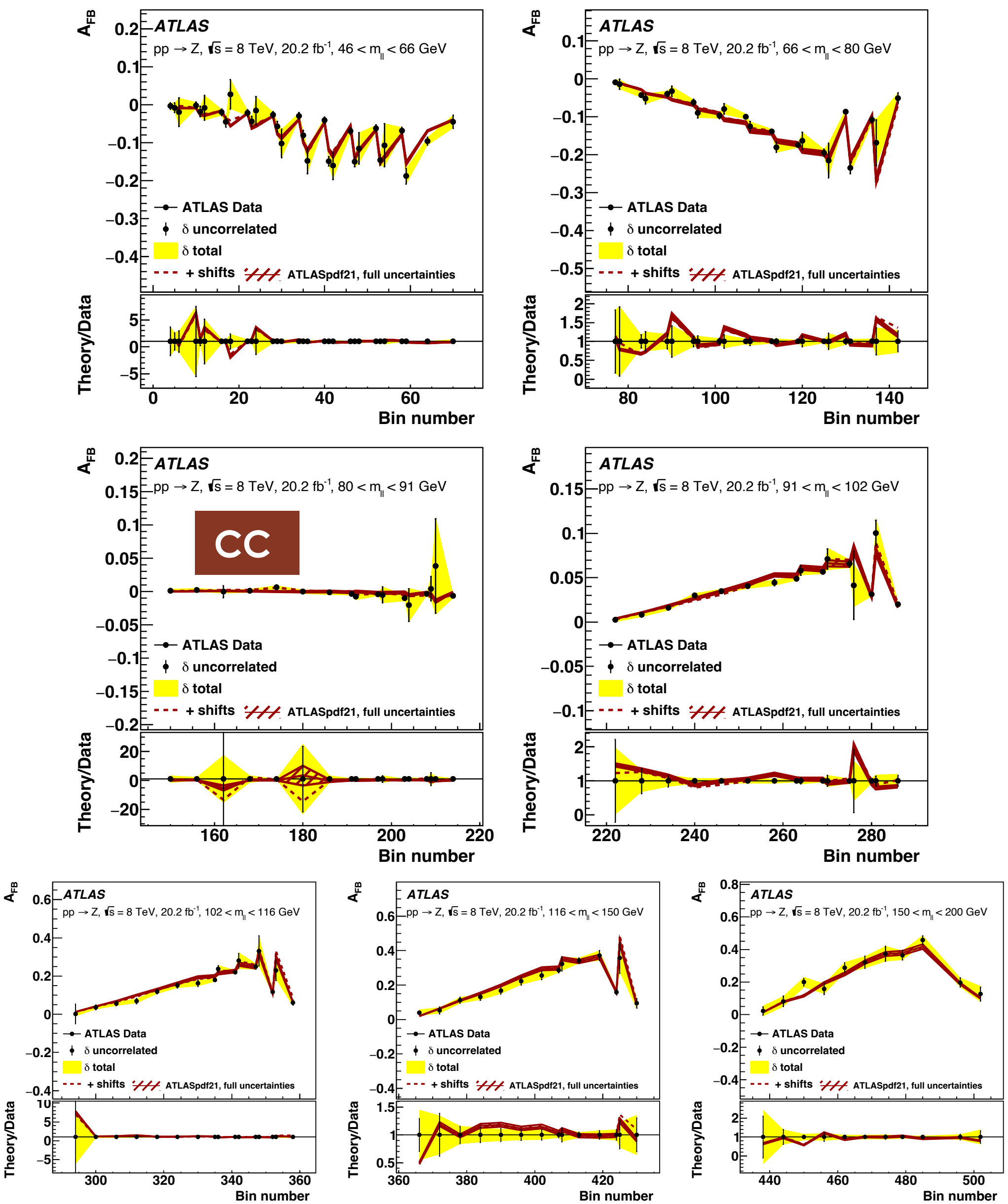
- AFB data is sensitive to $\sin^2\theta_W$ - a lot more data from the 8 TeV Z3D analysis

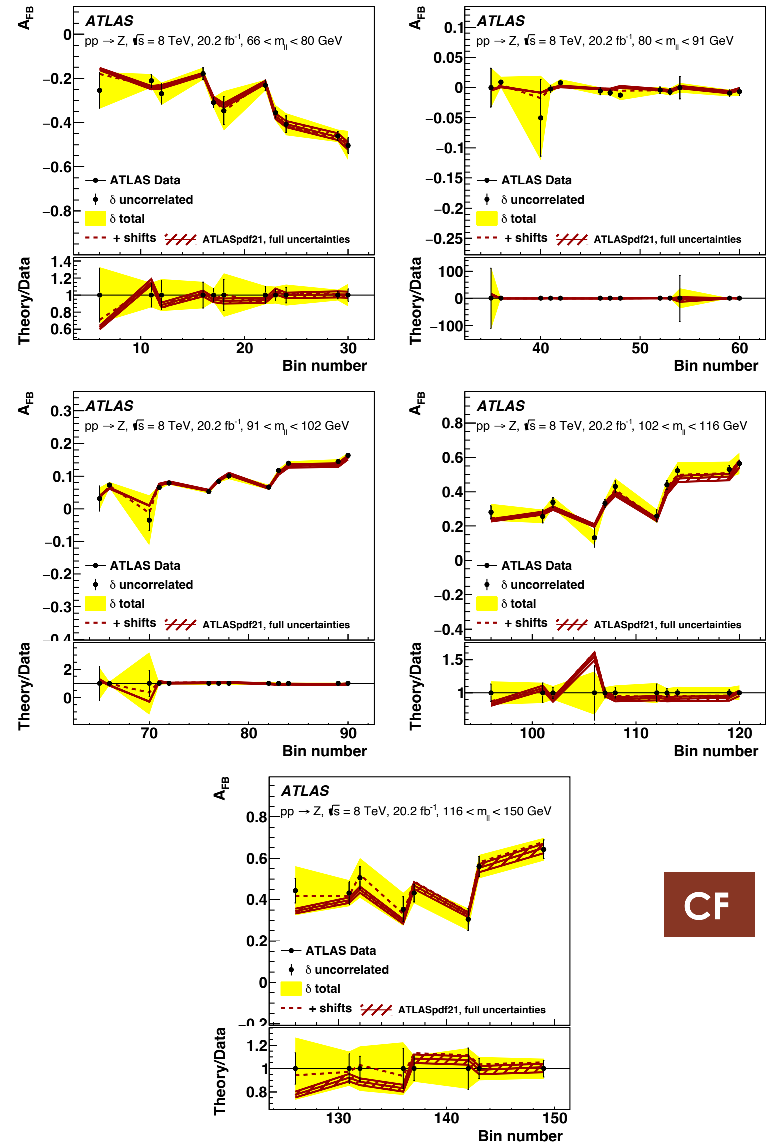
- These were not included in the fit both because of this sensitivity and because these asymmetry data cover kinematic regions which are not fully predicted at NNLO

- They are well described as far as one can tell within the current theoretical limitations

Data set	χ^2/ndf
ee CC	18.95/17
ee CF	21.10/17
$\mu\mu$	3.81/17

Description of Z3D asymmetry 8 TeV data

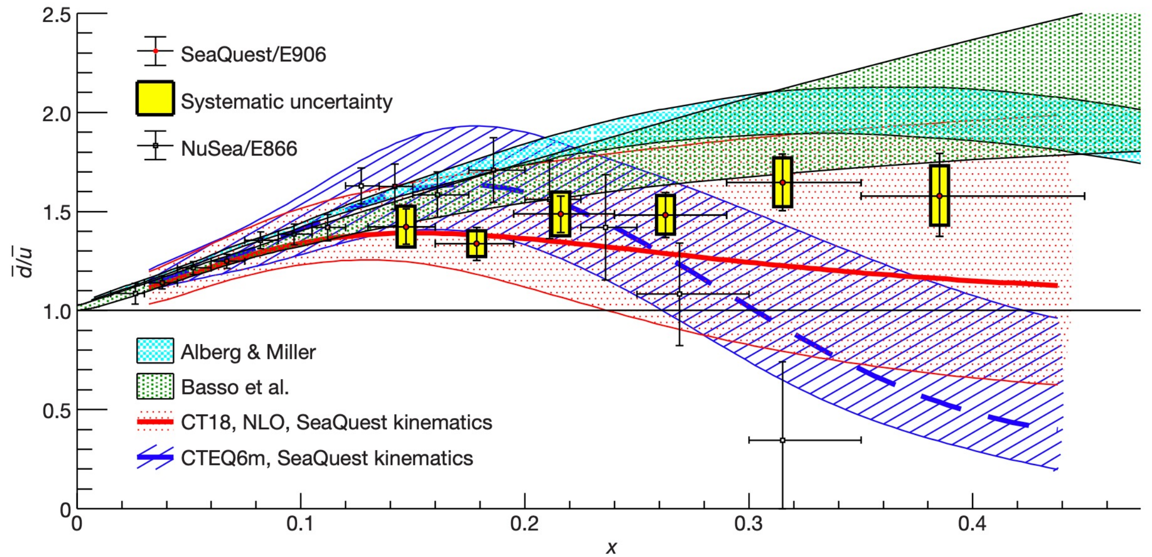


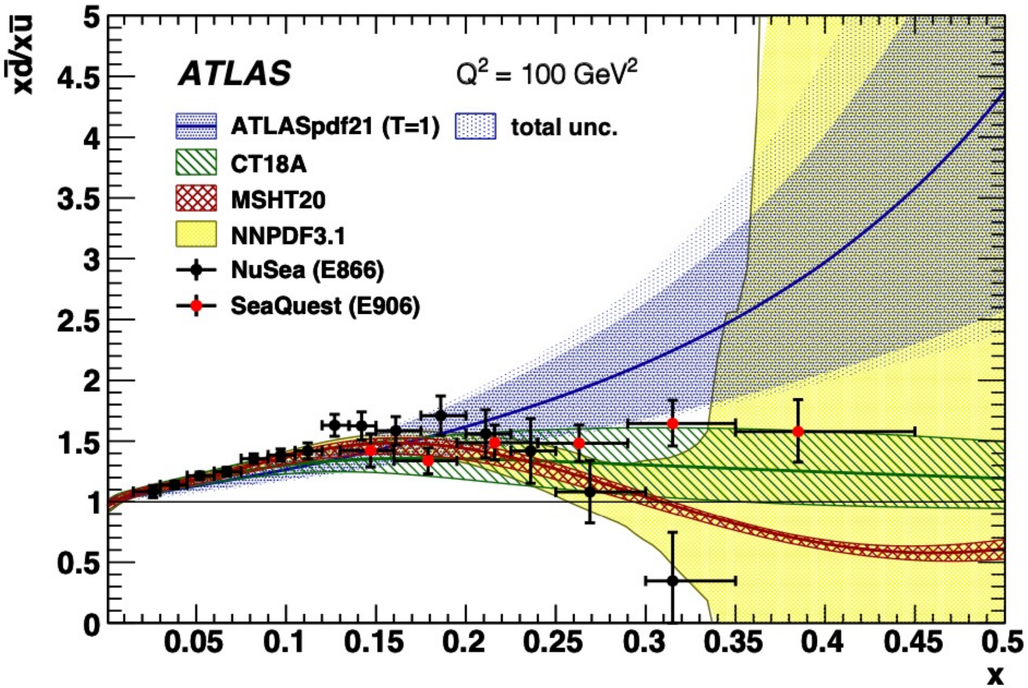


χ^2/ndf = 294/188

The asymmetry of antimatter in the proton

- Paper published on Nature
- It shows $\bar{d}(x)/\bar{u}(x)$ compared to various predictions
- **CT18** describes E906 for $x >$ 0.2 within its error band but older variants follow E866 more closely





- MHST20 closer to CTEQ6m
- **ATLASpdf21** closer to SeaQuest
- It seems that because we did not try to fit high-x data perfectly, our PDFs are free enough to accommodate SeaQuest