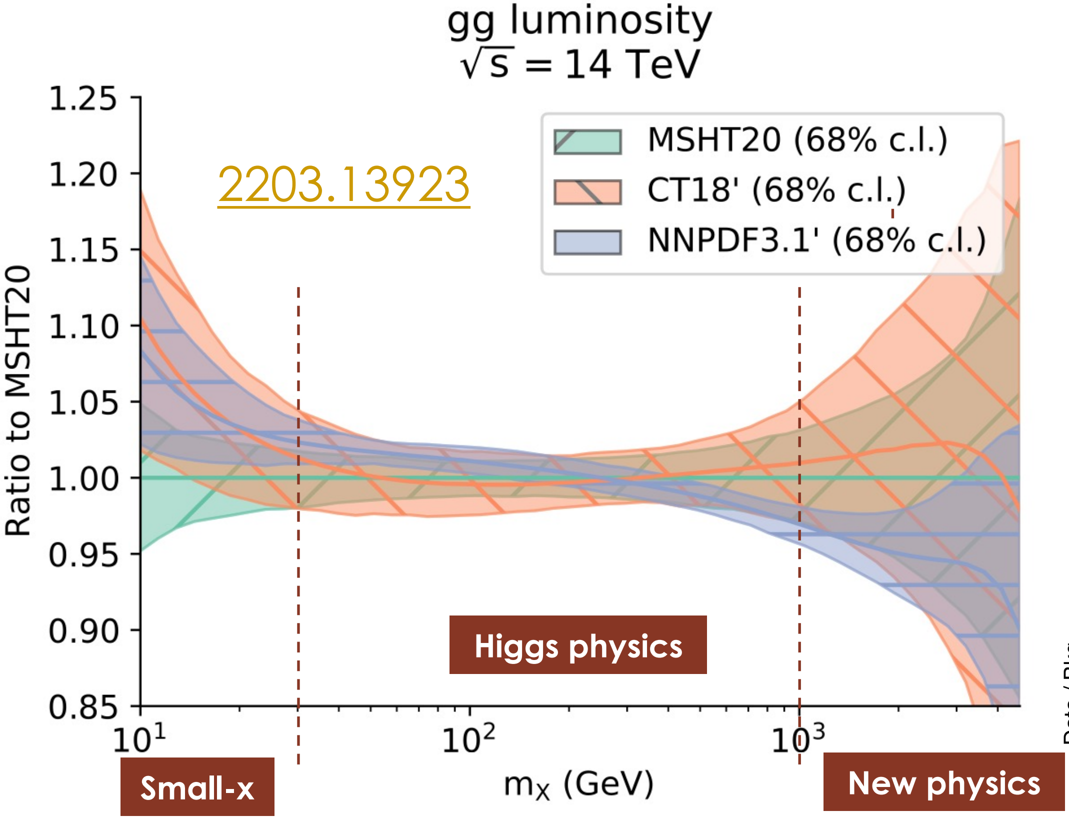
The impact of ATLAS $t\bar{t}$ data on PDFs

Francesco Giuli (on behalf of the ATLAS Collaboration)

CTEQ workshop
Michigan State University
09/11/2023

Why proton PDFs matter

- Precise knowledge of Parton Distribution Functions (PDFs) is essential
- PDFs have large uncertainties in the LHC kinematics regions
  - Significant source of uncertainty for Higgs and top production
  - Limits precision on fundamental parameters (m_W, α_S, etc.)
  - Limits searches for new massive particles

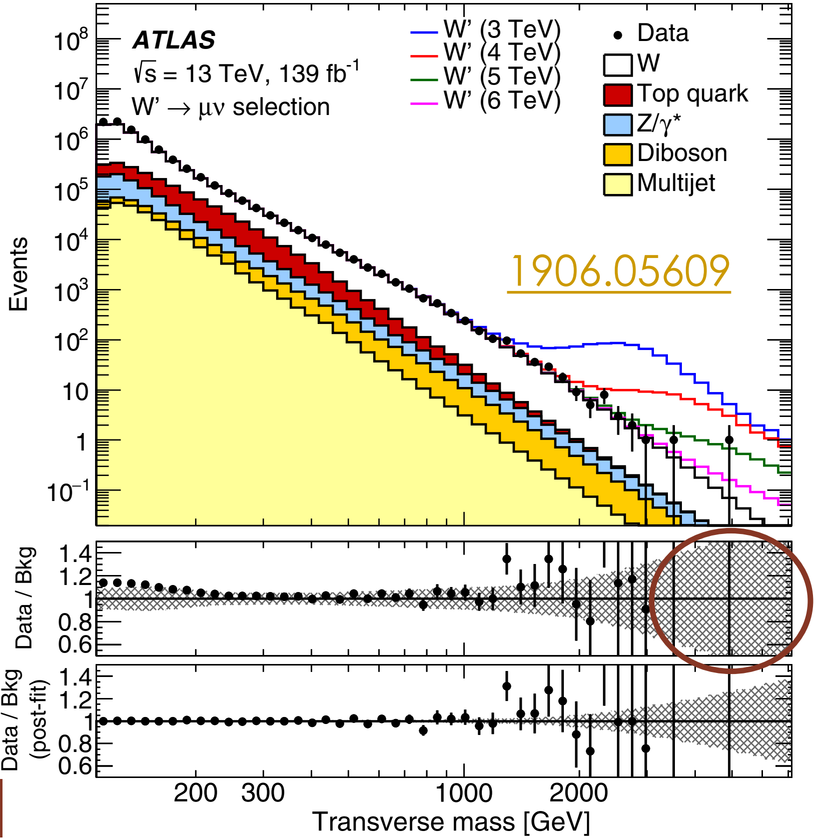


2203.13923

Higgs physics

Small-x

New physics



1906.05609

Proton PDF studies within ATLAS

- Several PDF analyses performed by ATLAS in the past

- QCD analysis of 2011 W,Z & strange sea density – ATLASepWZ12 fit
  - First evidence for unsuppressed strange at small-x

- Fit to inclusive W,Z data at 7 TeV – ATLASepWZ16 fit
  - Confirmation of unsuppressed strange at small-x with high precision data

- **Fit to inclusive W,Z (7 TeV) and top (8 TeV) data –** ATLASepWZtop18 fit
  - Gluon constraints at medium- and high-x (both shape and uncertainties)

- Fit to inclusive W,Z (7 TeV) and V+jets (8 TeV) data – ATLASepWZVjets20 fit
  - Strange suppression at high-x and resolved ambiguities in high-x shapes

- **Fit to several ATLAS data sets** – ATLASpdf21 fit
  - It includes $t\bar{t}$ data at both 8 and 13 TeV

$t\bar{t}$ data at 8 TeV in the ATLAS PDF fit

- To access the impact of including ATLAS $t\bar{t}$ production data in fits to extract the proton PDF

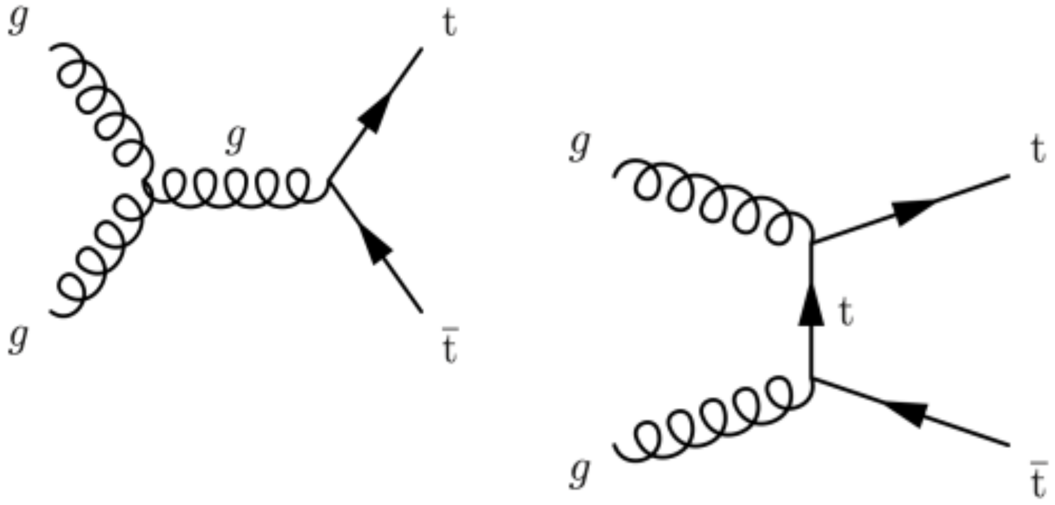
- QCD fit to DIS data from HERA and W,Z at 7 TeV and $t\bar{t}$ at 8 TeV

- $t\bar{t}$ data are complementary to the $W, Z/\gamma^*$ data

- Expected to be **sensitive to gluon distribution in the medium- and high-x regime** (x $\gtrsim 5 \cdot 10^{-2}$)

- Important to perform this fit now since the **NNLO predictions of pQCD** for $t\bar{t}$ production data are now available and usable in PDF fits (1704.08551)

- $t\bar{t}$ production input datasets:
  - lepton+jets channel at 8 TeV
  - dilepton channel at 8 TeV



$t\bar{t}$ data at 8 TeV in the ATLAS PDF fit

- All the results publicly available – epWZtop18 fit

- Available spectra for the fit:
  - Lepton+jets channel ($\boldsymbol{m_{t\bar{t}}}$, $\boldsymbol{p_T^t}$, $y_{t\bar{t}}$, y_t)
  - Dilepton channel ($m_{t\bar{t}}$, $\boldsymbol{y_{t\bar{t}}}$)

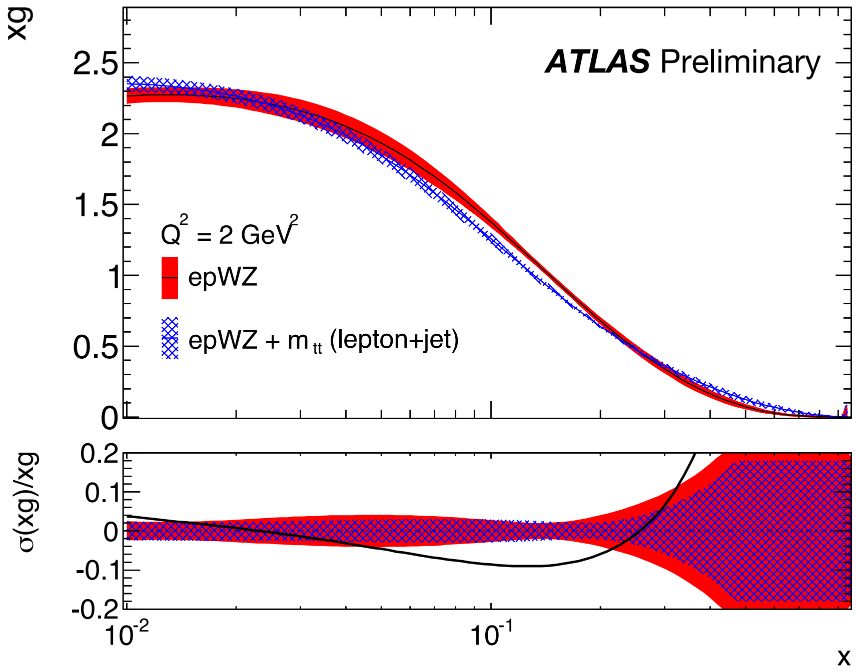
- Bin-to-bin statistical correlations **within each spectrum** and **between the spectra** (lepton+jets channel) available and included in the fit → effect is small but not negligible

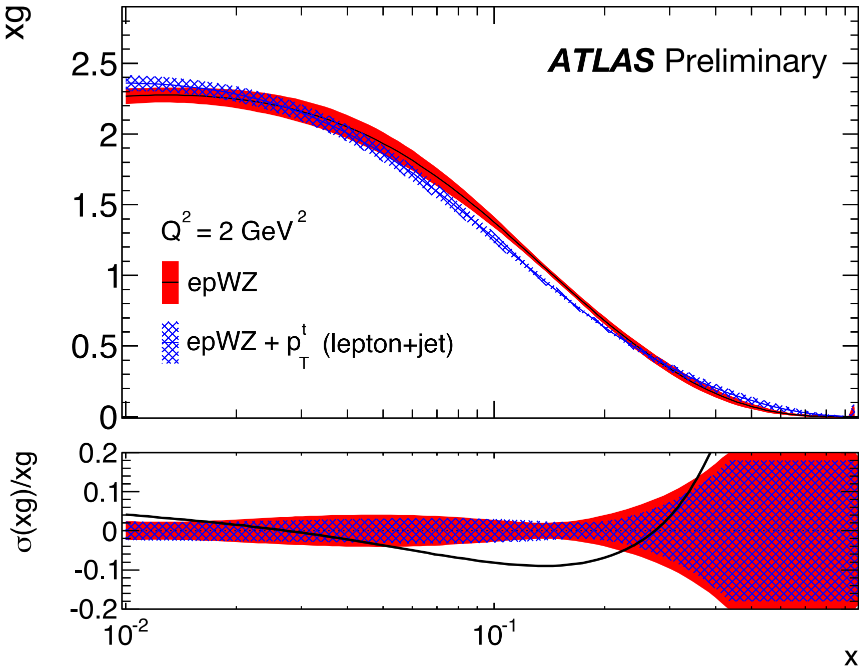
- We have the systematic correlations between the spectra - we would usually assume these to be 100% correlated

- The largest systematic uncertainties are due to:
  - Initial state/final state radiation (ISR/FSR) ~8%
  - Parton shower model (PS) ~5%
  - Hard-scattering model ~4%

- Effect of decorrelating this source of uncertainty investigated

Fits to individual lepton+jets $t\bar{t}$ spectra

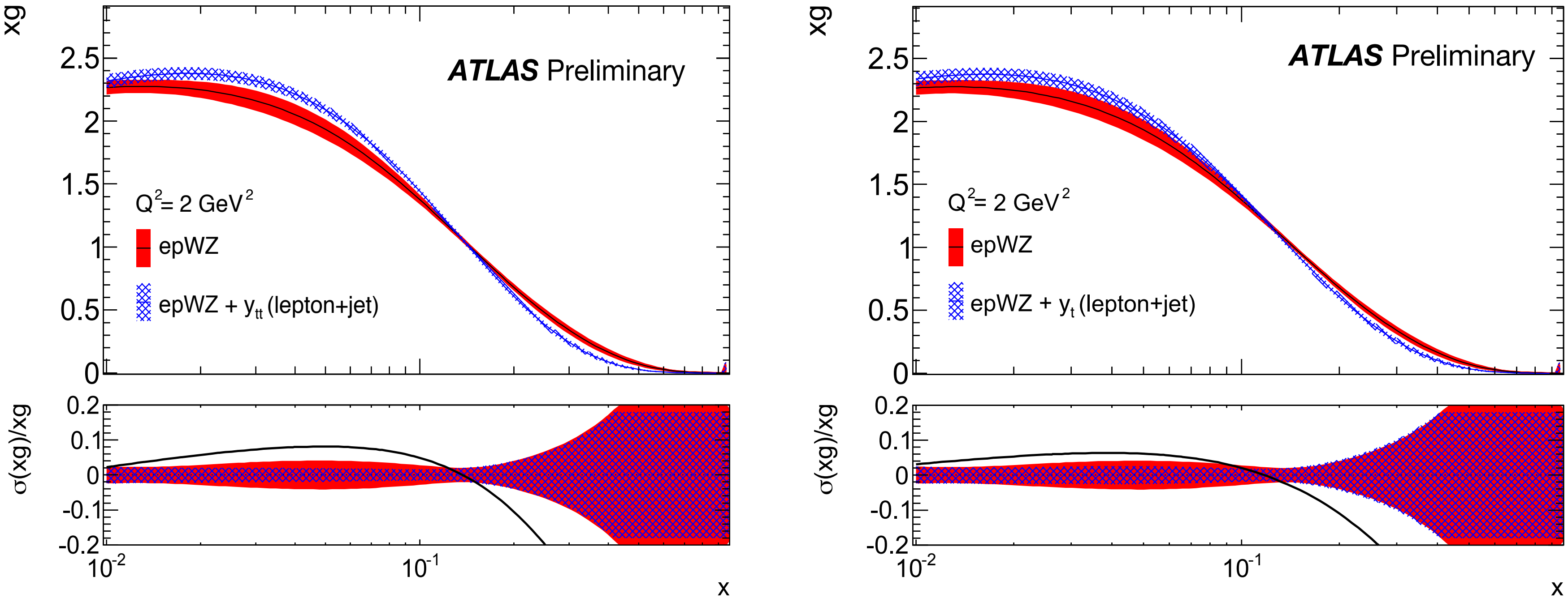




		lepton+jets spectrum			
		m_{tt}	p_T^t	y_{tt}	y_t
Total χ^2/NDF		1238.4 / 1062	1239.4 / 1063	1257.5 / 1060	1246.5 / 1060
Partial χ^2/NDP	HERA	1153 / 1016	1151 / 1016	1149 / 1016	1146 / 1016
Partial χ^2/NDP	ATLAS $W, Z/\gamma^*$	82.0 / 55	82.1 / 55	86.4 / 55	85.0 / 55
Partial χ^2/NDP	ATLAS $t\bar{t}$	3.4 / 7	7.9 / 8	19.7 / 5	18.3 / 5

- Partial χ^2 **good** for $m_{t\bar{t}}$ **and** p_T^t but fits to $y_{t\bar{t}}$ **and** y_t are **poor**

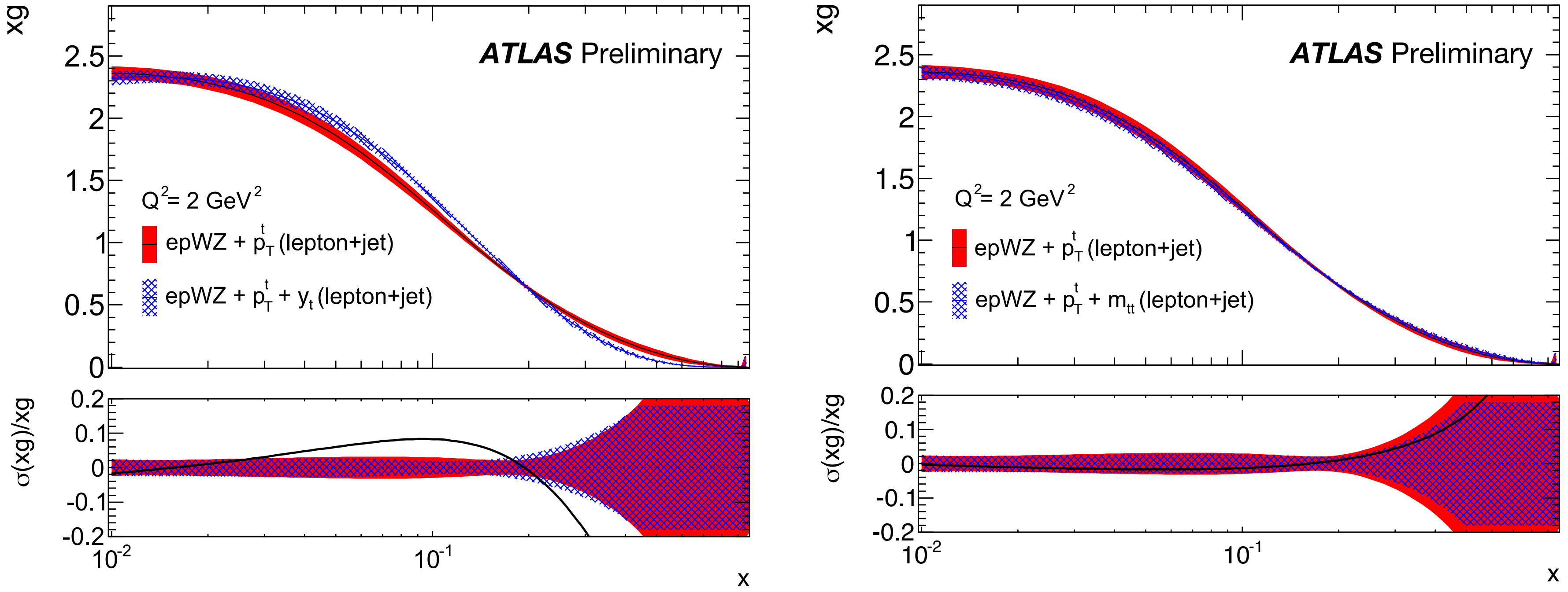
Fits to individual lepton+jets $t\bar{t}$ spectra



		lepton+jets spectrum			
		m_{tt}	p_T^t	y_{tt}	y_t
Total χ^2/NDF		1238.4 / 1062	1239.4 / 1063	1257.5 / 1060	1246.5 / 1060
Partial χ^2/NDP	HERA	1153 / 1016	1151 / 1016	1149 / 1016	1146 / 1016
Partial χ^2/NDP	ATLAS $W, Z/\gamma^*$	82.0 / 55	82.1 / 55	86.4 / 55	85.0 / 55
Partial χ^2/NDP	ATLAS $t\bar{t}$	3.4 / 7	7.9 / 8	19.7 / 5	18.3 / 5

- $y_{t\bar{t}}$ and y_t distributions pull the gluon PDF in a different direction wrt $m_{t\bar{t}}$ and p_T^t

Fits to various lepton+jets $t\bar{t}$ spectra

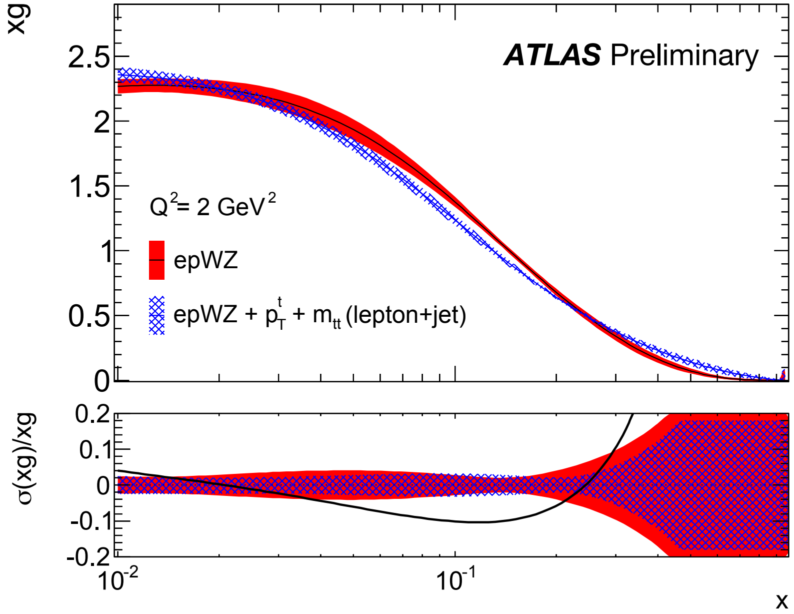


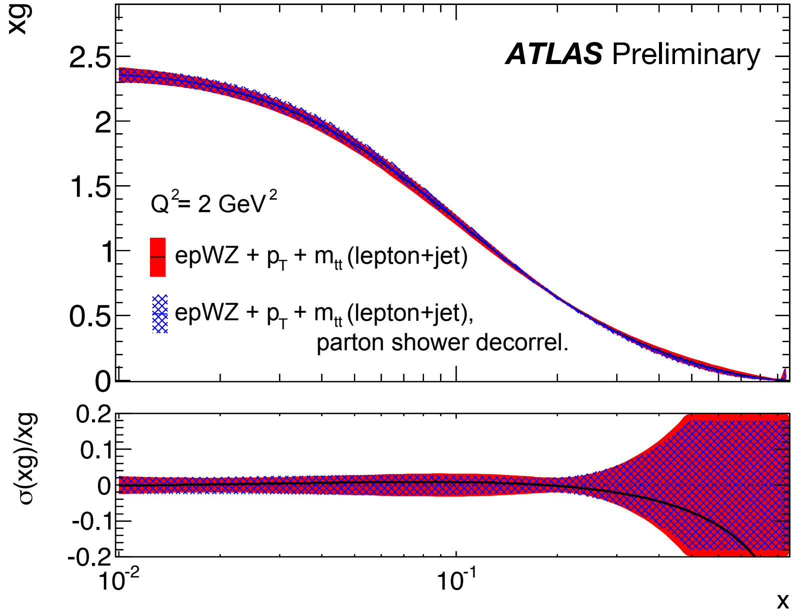
		lepton+jets spectra			
		p_T^t and y_t with statistical correlations	p_T^t and y_t without statistical correlations	p_T^t and m_{tt} with statistical correlations	p_T^t and m_{tt} without statistical correlations
Total χ^2/NDF		1264 / 1068	1260 / 1068	1290 / 1070	1287 / 1070
Partial χ^2/NDP	HERA	1148 / 1016	1147 / 1016	1162 / 1016	1162 / 1016
Partial χ^2/NDP	ATLAS $W, Z/\gamma^*$	82.7 / 55	83.5 / 55	83.2 / 55	83.1 / 55
Partial χ^2/NDP	ATLAS $t\bar{t}$	33 / 13	30 / 13	45 / 15	42 / 15

- Poor χ^2 of p_T^t, y_t consistent with separate χ^2, whereas the poor χ^2 of $p_T^t, m_{t\bar{t}}$ is not consistent with their separate χ^2

Impact of the decorrelation

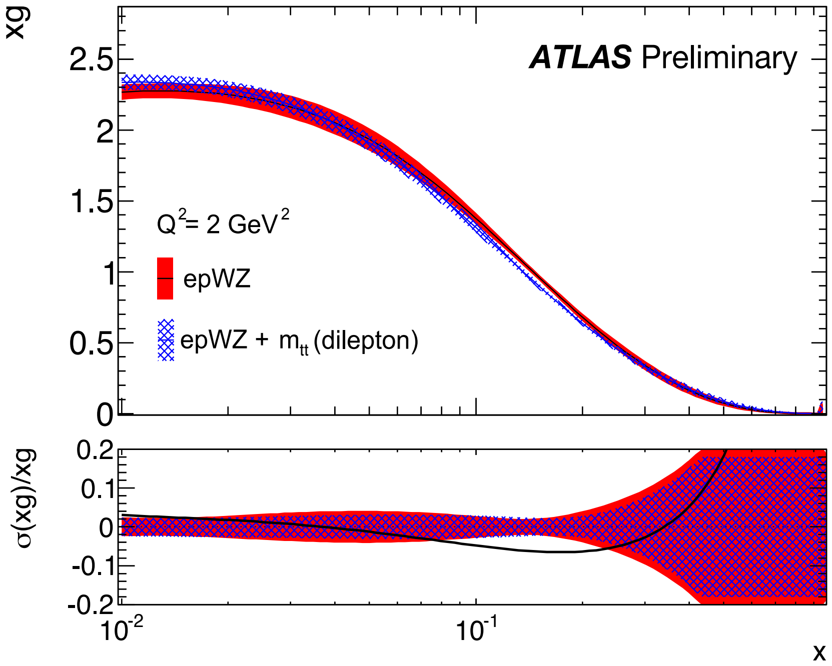
- Impact of decorrelating the systematic uncertainty due to PS
- Resulting gluon PDF fully compatible wrt PS uncertainty fully correlated
- The **main effect** of decorrelation is the **reduction of the χ^2 value**
- We rejected the $y_{t\bar{t}}$ and y_t spectra because of the poor fit

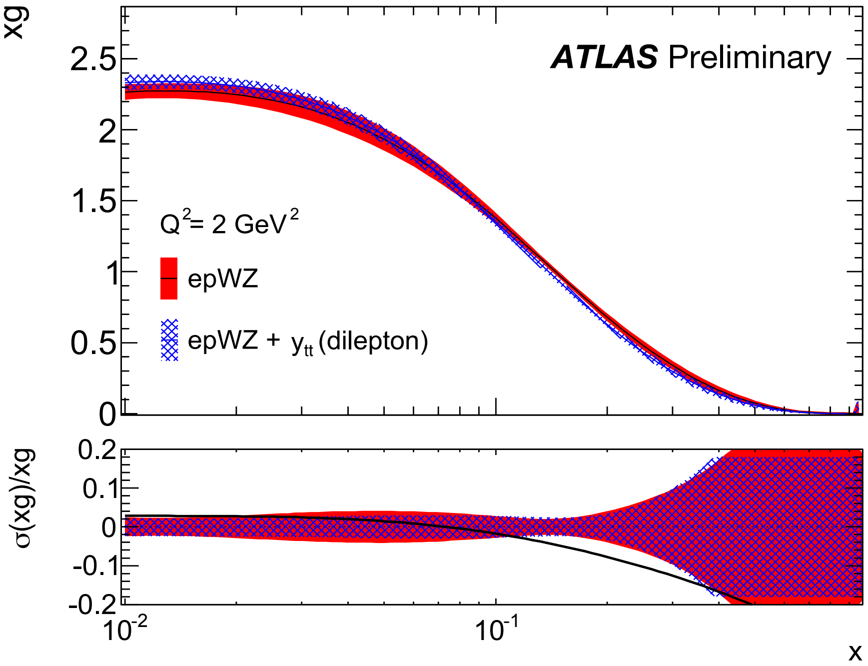




- Now also **global PDF fitters** e.g. CT18, MSHT20 **have done something similar**
  - MSHT decorrelated systematics across different bins of each $y_{t\bar{t}}, y_t$ spectrum to get a good fit - and that actually made these spectra have very little effect

Fits to various dilepton $t\bar{t}$ spectra





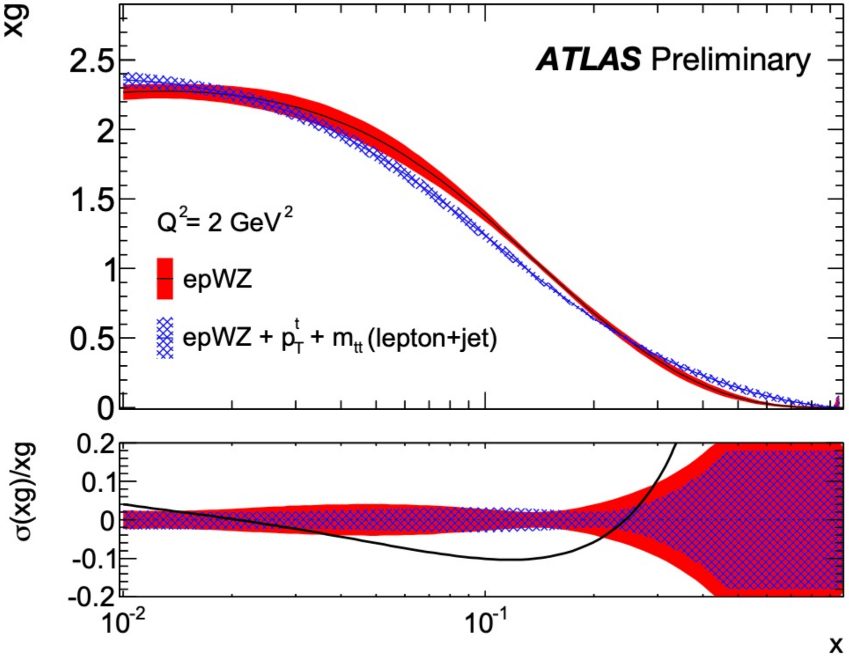
		dilepton spectrum	
		m_{tt}	y_{tt}
Total χ^2/NDF		1233.8 / 1061	1233.8 / 1060
Partial χ^2/NDP	HERA	1152 / 1016	1147 / 1016
Partial χ^2/NDP	ATLAS $W, Z/\gamma^*$	79.3 / 55	82.8 / 55
Partial χ^2/NDP	ATLAS $t\bar{t}$	2.6 / 6	4.5 / 5

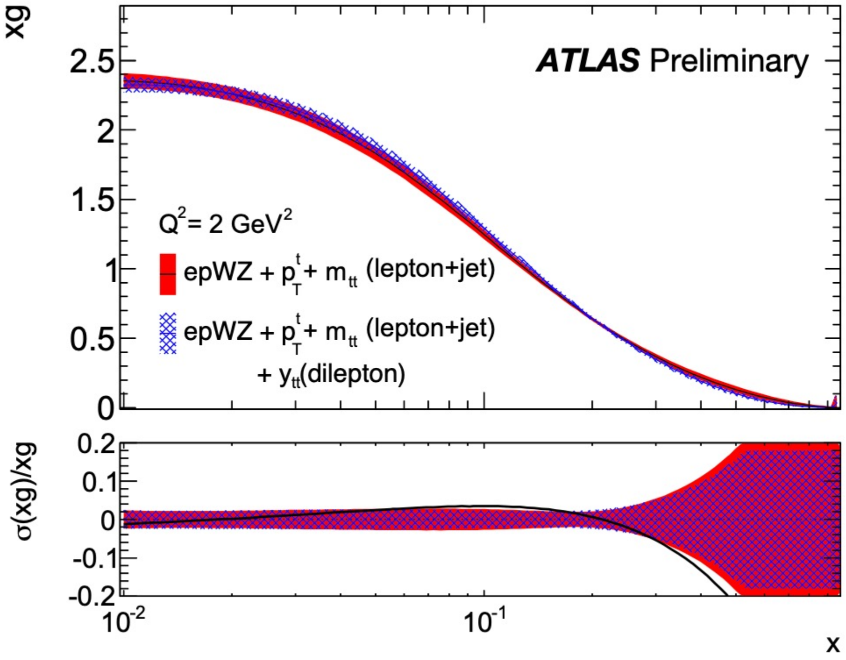
We do not have correlations for these spectra → we have to choose one!

- As for the lepton+jets spectra, the $m_{t\bar{t}}$ data support a harder gluon while the $y_{t\bar{t}}$ data prefer a softer gluon – **anyway both fits show good χ^2**

Fits to dilepton and lepton+jets spectra

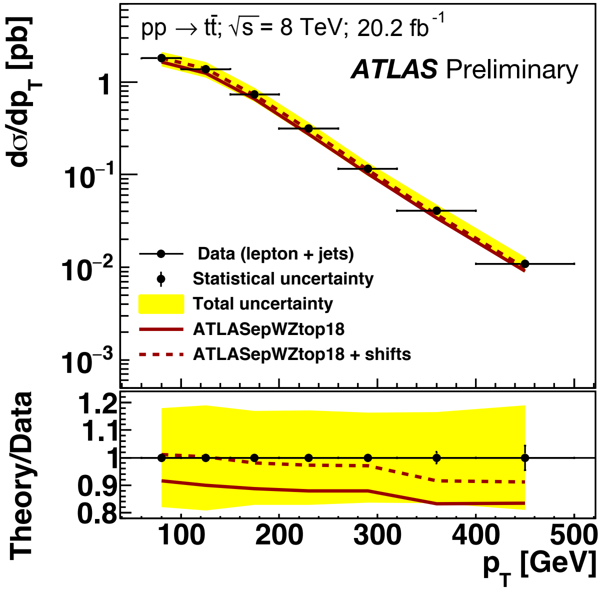
Total χ^2/NDF		**1253.8/1061**
Partial χ^2/NDF	HERA	1149/1061
Partial χ^2/NDF	ATLAS $W, Z/\gamma^*$	78.9/55
Partial χ^2/NDF	ATLAS $m_{t\bar{t}}$ and p_T^t	**16.0/15**
Partial χ^2/NDF	ATLAS $y_{t\bar{t}}$	5.4/5

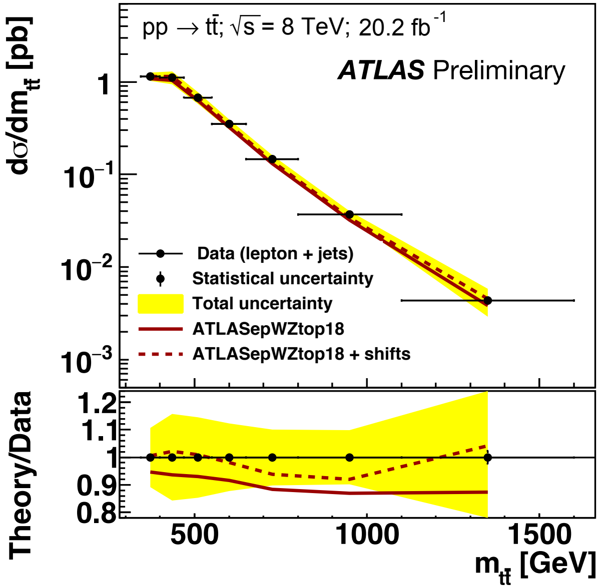


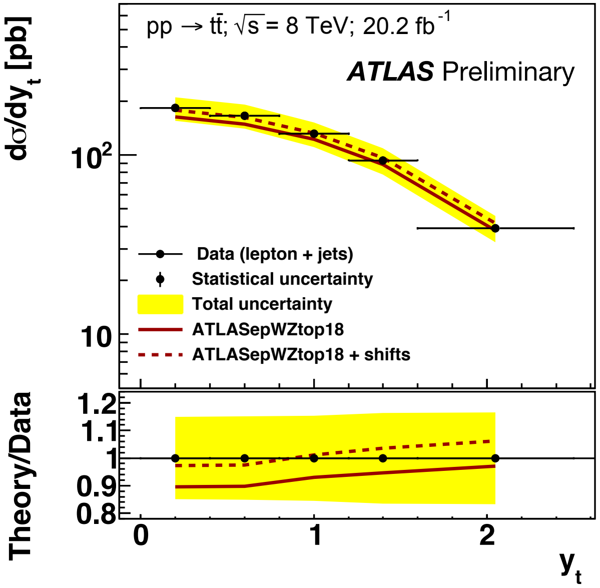


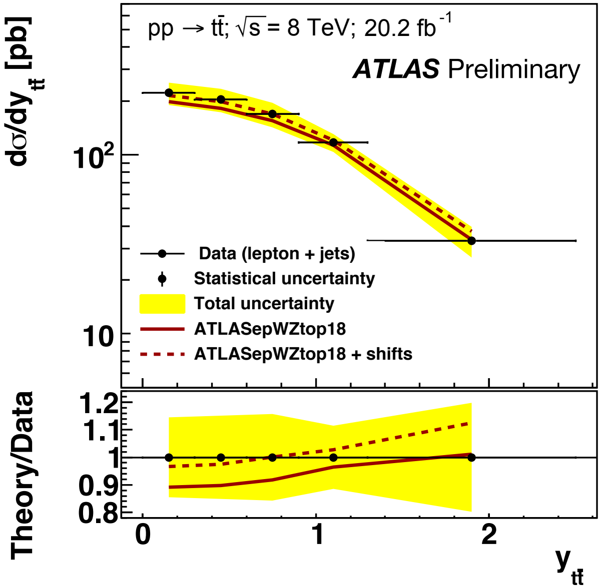
- **Harder gluon** and a **significantly reduced high-x uncertainty** on the gluon PDF
- Effect of the dilepton spectrum: change in the high-x shape but not further reduction in PDF uncertainties

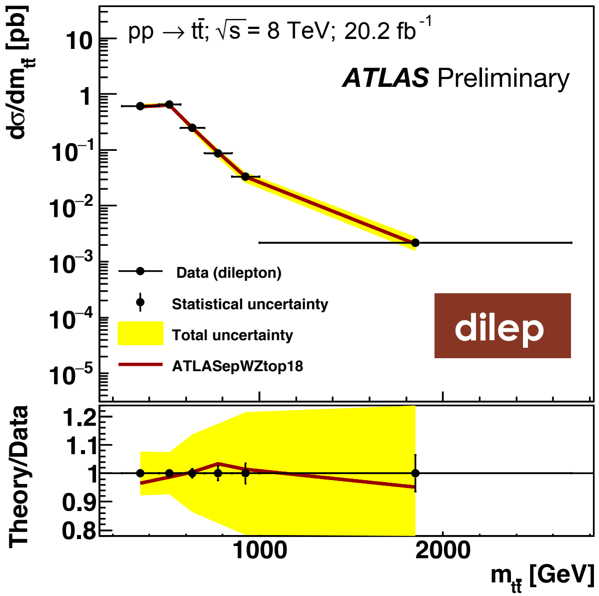
Data description

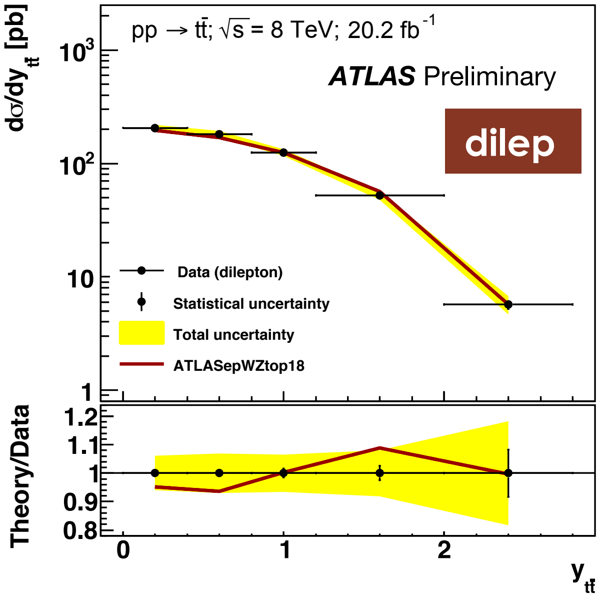












The ATLASpdf21 fit

- **ATLASpdf21** is a PDF fit to **multiple ATLAS data sets** - EPJC 82 (2022) 5, 438

- DIS HERA data are the backbone of ATLAS PDF fits – 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

- All the fits performed using xFitter

- Fit to **NNLO in QCD and NLO in EW** achieved either by direct NNLO grids or by k-factor corrections - current 'state of the art'

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 $\|y_{\text{jets}}\|$
$t\bar{t}$ [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 $\|y_{\text{jets}}\|$

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!

13 TeV $t\bar{t}$ data in the ATLASpdf21 fit

- The 13 TeV $t\bar{t}$ data (1908.07305) refer to partial Run 2 data set (2015 + 2016)

- The cross sections considered are:
  - $\frac{d\sigma}{dm_{t\bar{t}}}$
  - $\frac{d\sigma}{dp_T^{t,had}}$
  - $\frac{d\sigma}{dy^{t,had}}$
  - $\frac{d\sigma}{dy_{boost}^{t\bar{t}}}$
  - $\frac{d^2\sigma}{dm_{t\bar{t}}dp_T^{t,had}}$

36 fb^{-1} of data in the l+jets channel

1D cross sections – these 4 distributions are included in our fit

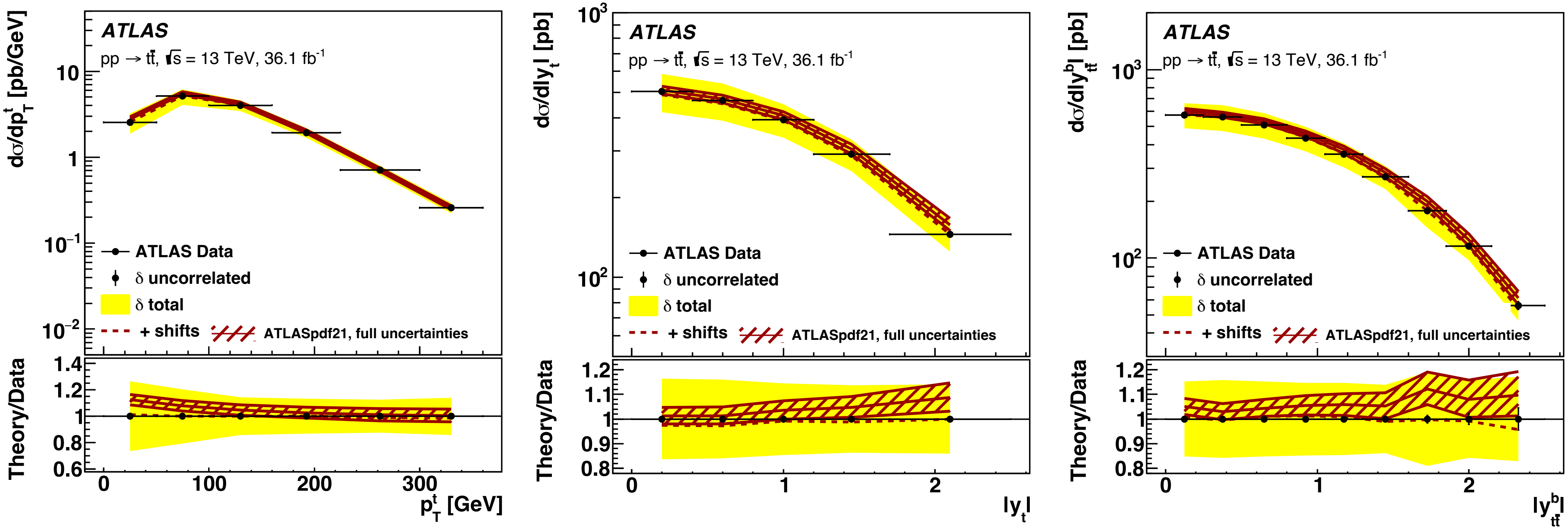
2D cross section

- Systematic uncertainties provided as full covariance matrices converted into nuisance parameters for all the spectra

- Systematic correlations available – **no need to do any decorrelation**

- Bin-to-bin statistical correlations among all the 1D spectra also available and included in the fit – negligible impact on PDFs and χ^2

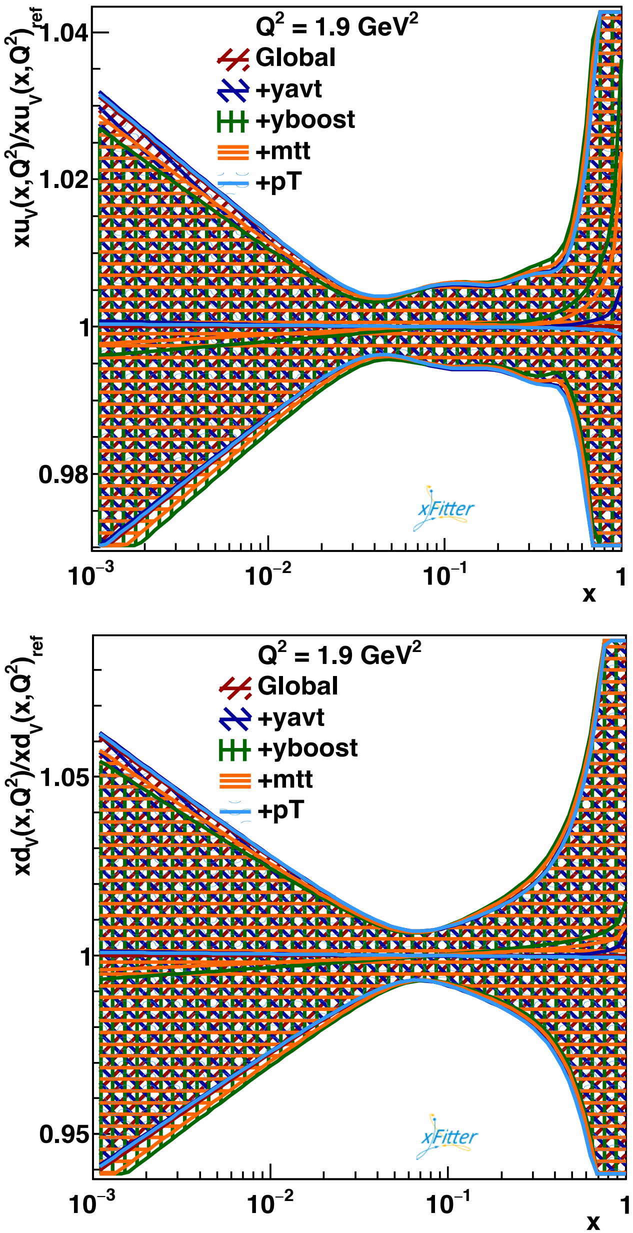
Data description

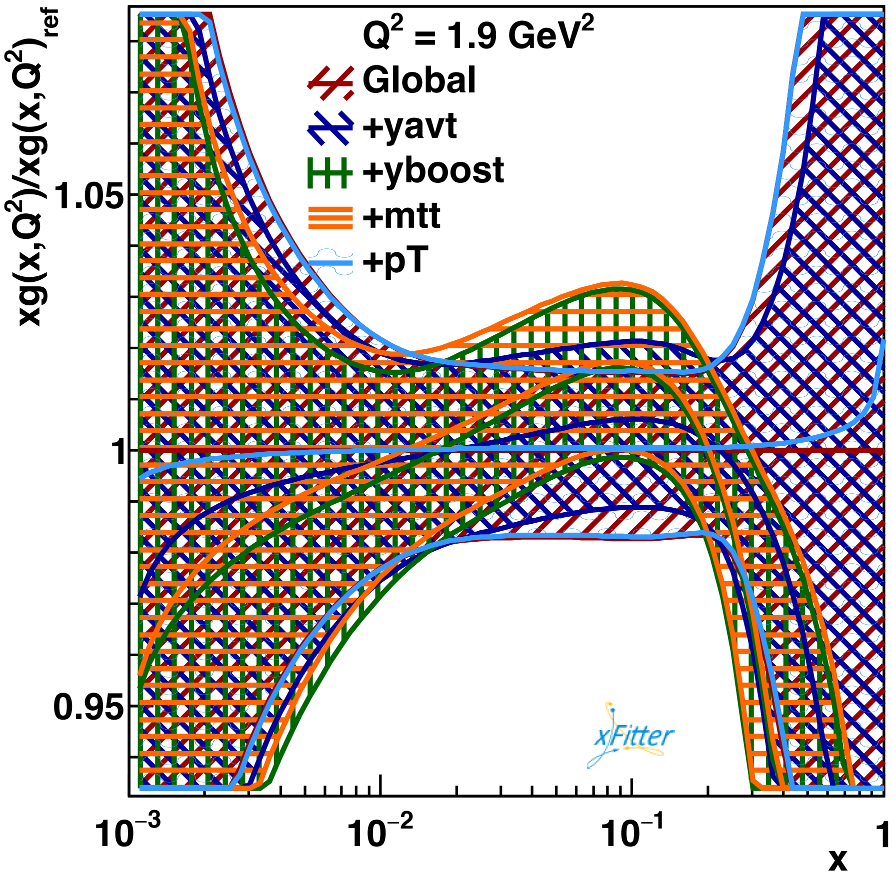


Dataset	Exp. χ^2/dof
$+\, m_{t\bar{t}}$	12.4/9
$+\, p_T^{t,had}$	4.8/6
$+\, y^{t,had}$	4.6/5
$+\, y_{boost}^{t\bar{t}}$	9.3/9

- Top data at 13 TeV nicely included in the ATLASpdf21 fit
- Inclusion of each 1D spectrum one at a time → similar results and no deterioration of the χ^2
- Inclusion of all the 1D spectra together → nice agreement with results from the previous step of the ATLASpdf21 fit

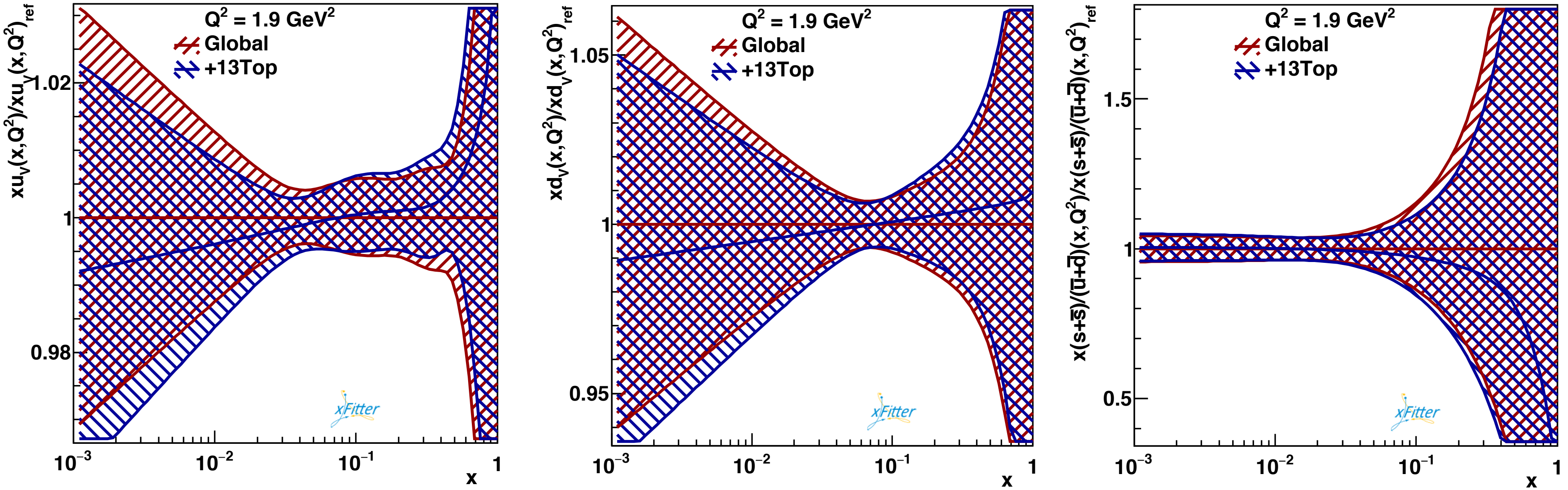
1D spectrum one at a time

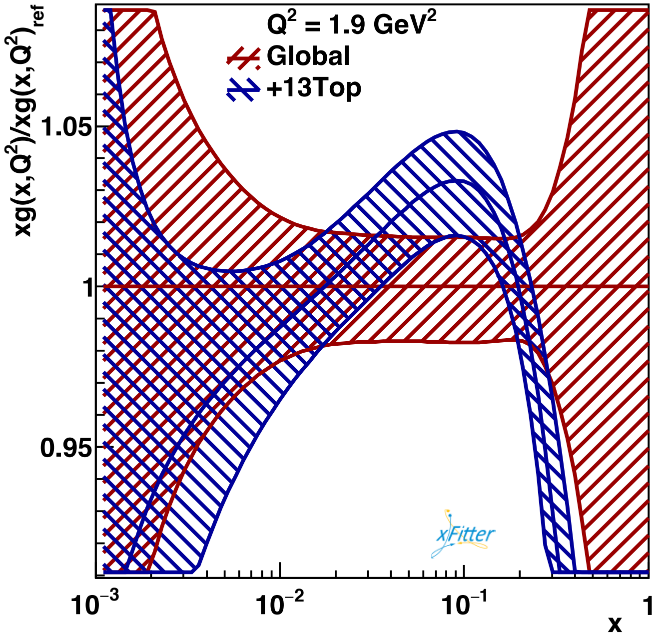




- Everything looks ok and consistent wrt the ATLAS global fit (‘Global’ fit does not include the 8 TeV inclusive jet data)
- As expected, **the most affected PDF is the gluon distribution**

1D spectra all together

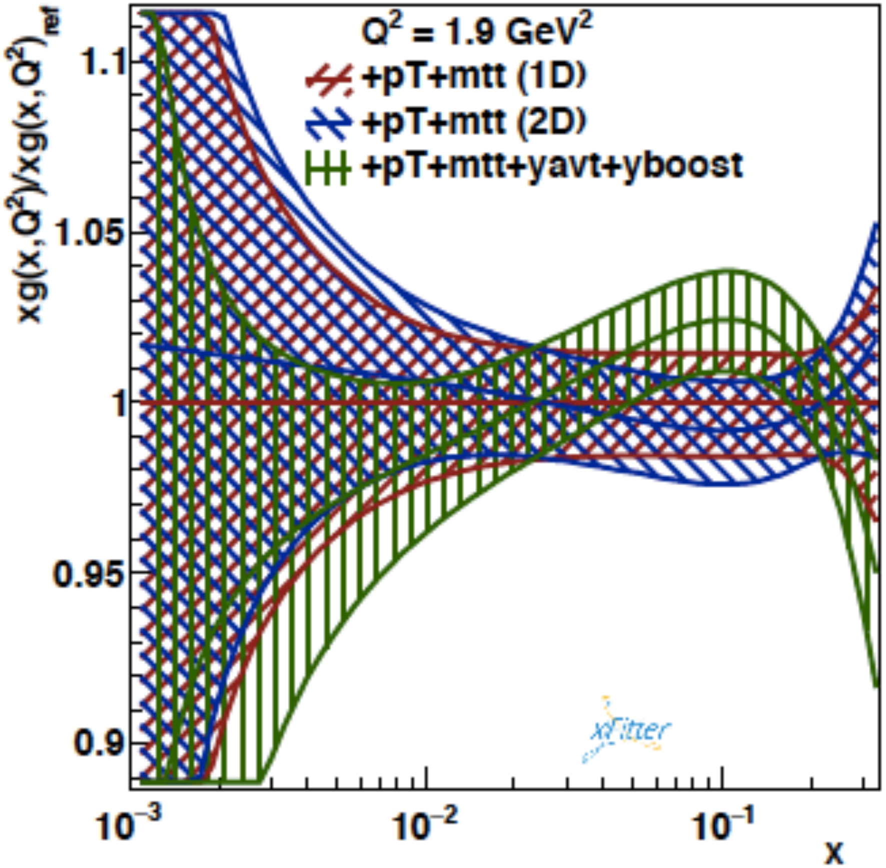




- <u>'Global' fit does not include the 8 TeV inclusive jet data</u>
- Gluon distribution slightly pulled up at x ~ 0.1 but compatible within uncertainties
- Once model and parameterisation uncertainties included, everything consistent (here T = 1, standard PDF set with T = 3)
- Mild effect on all the other PDFs is confirmed

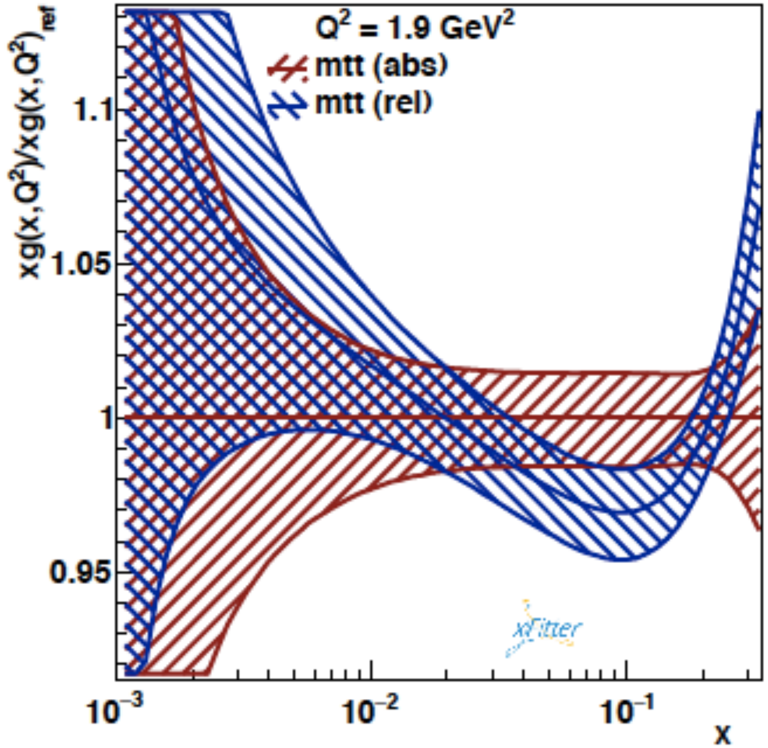
1D vs 2D

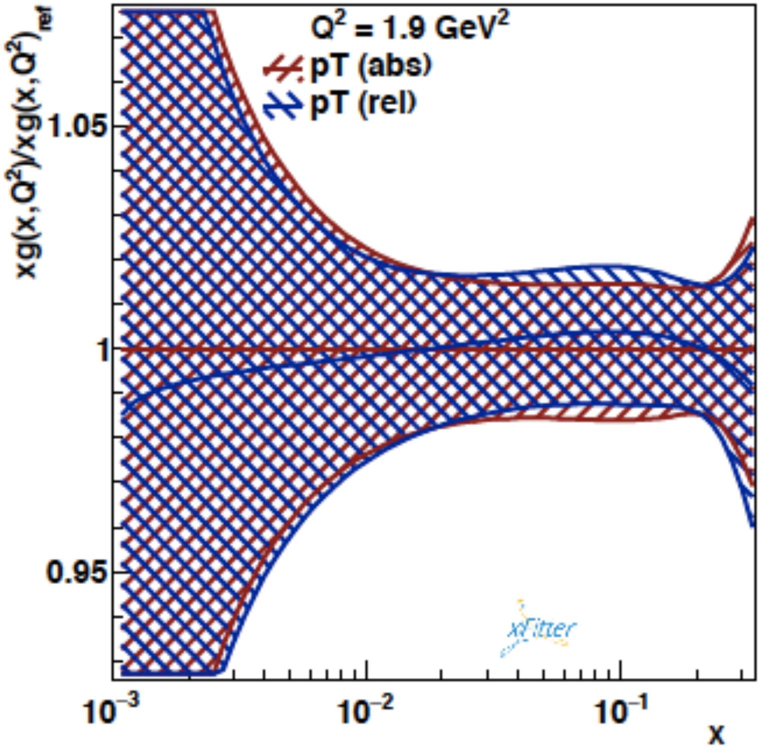
- We studied the inclusion of the $\frac{d^2\sigma}{dm_{t\bar{t}}dp_T^{t,had}}$ cross sections and I compared it with the results obtained when including the 1D distributions $\frac{d\sigma}{dm_{t\bar{t}}}$, $\frac{d\sigma}{dp_T^{t,had}}$
- Uncertainties not in form of full covariance matrix for the 2D distribution
- Identical results obtained when using 2D distribution or 1D spectra
- Size of uncertainties very similar
- Output parameters from the fits well consistent within uncertainties
- No deterioration in χ^2 when including these data in the fit



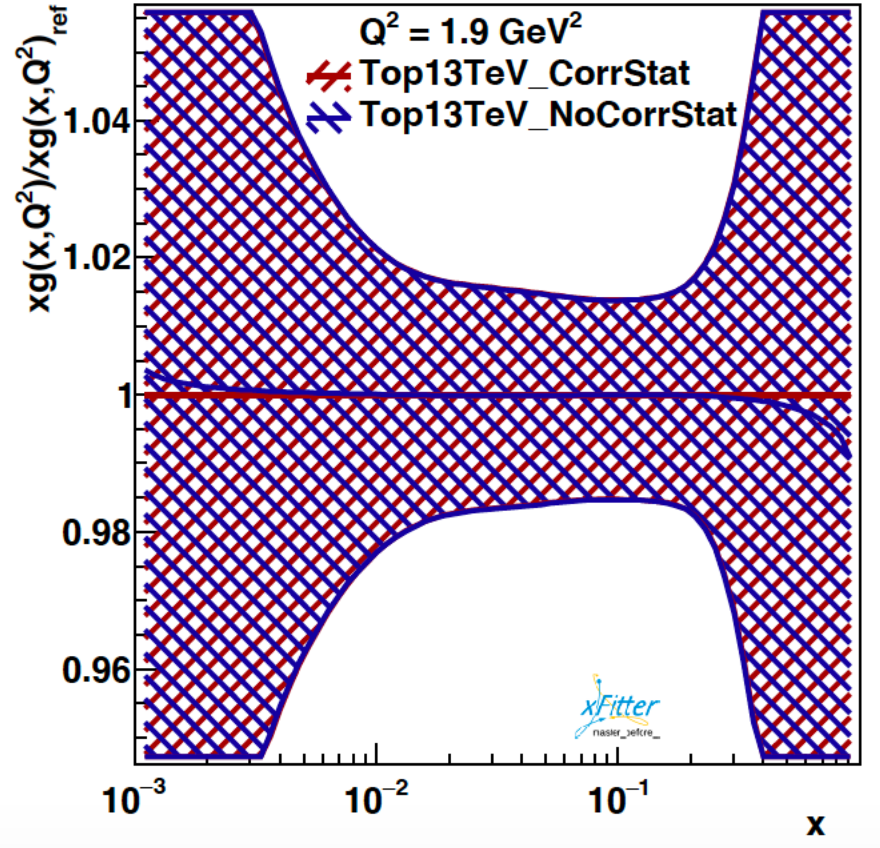
Normalised or absolute distributions?

- Predictions for a fit to some 1D spectra if using normalised or absolute distributions
- When including normalised cross sections, the last bin of the distribution is removed
- Same results for $y^{t,had}$ and $y^{t\bar{t}}_{boost}$ (backup)

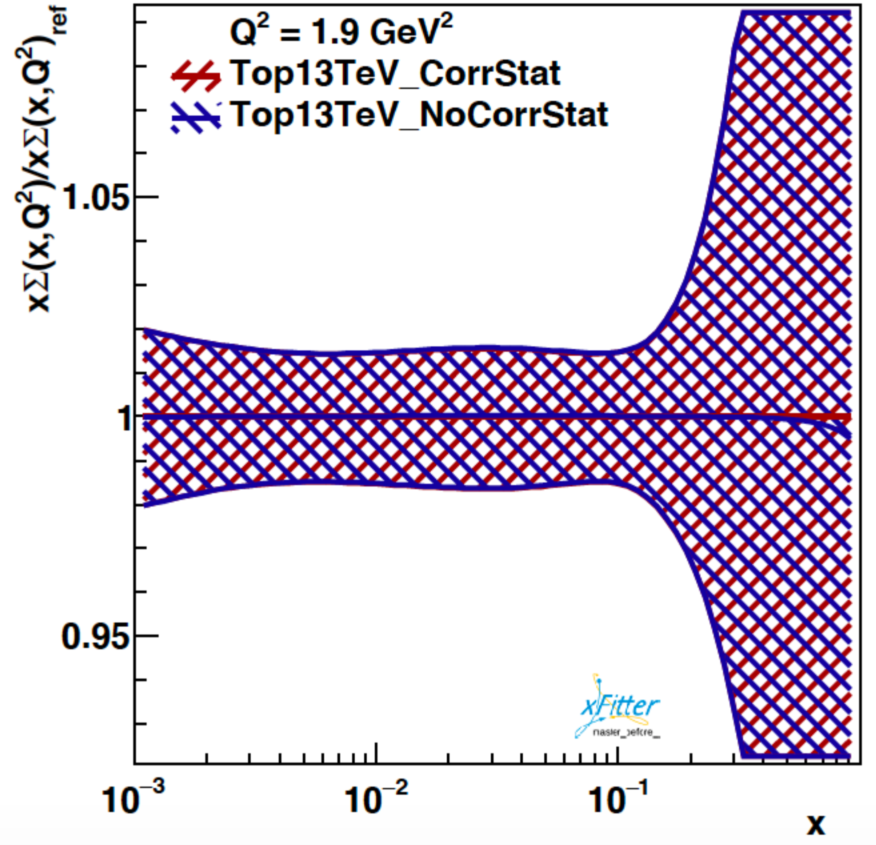




Impact of bin-to-bin statistical correlation



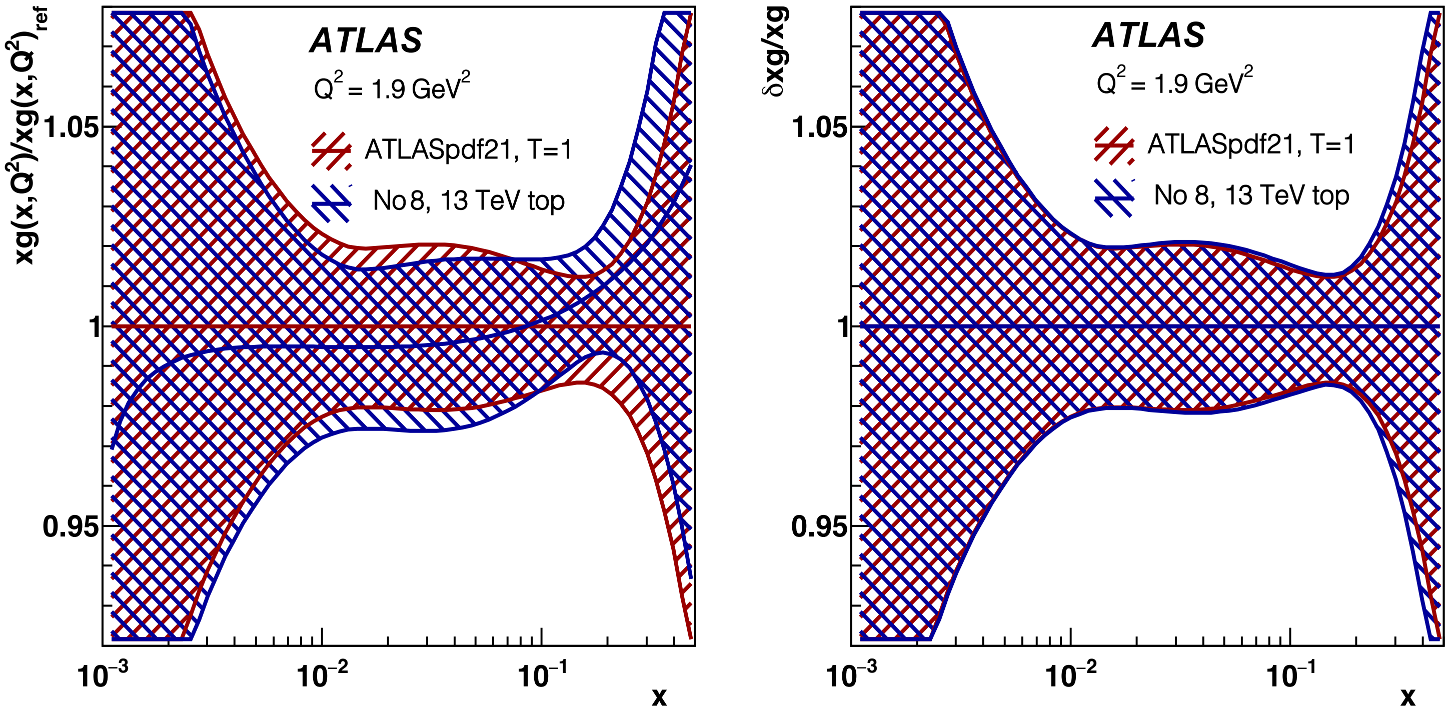
Even smaller impact wrt the one found in the epWZtop18 fit



No sensible difference between PDFs or χ^2

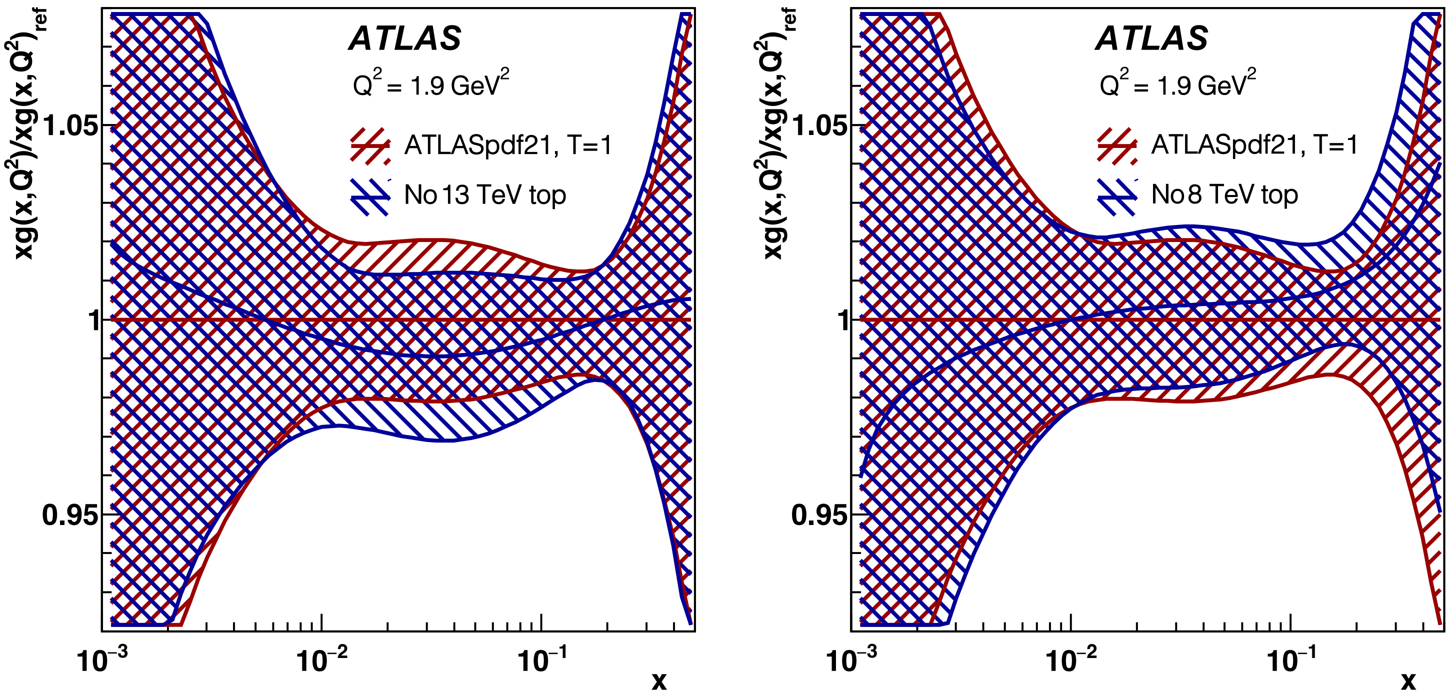
Impact of the top data on ATLASpdf21

- 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
- Milder impact wrt what was found for the ATLASepWZtop18 fit (but here we have many other data sets added)



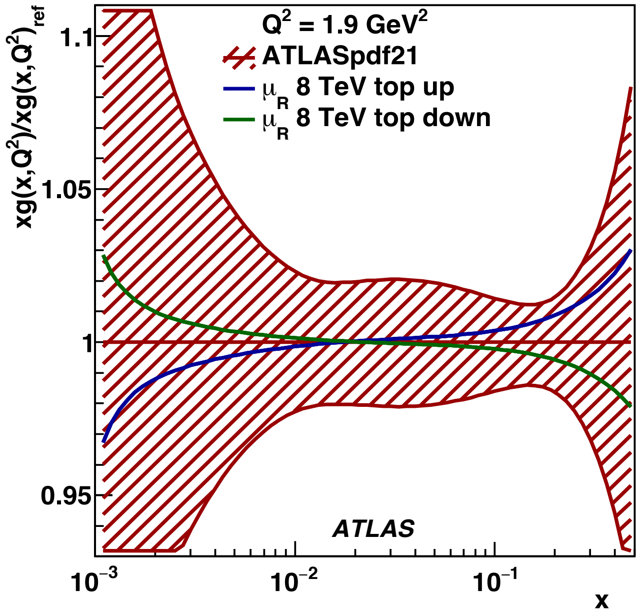
Impact of the top data sets on ATLASpdf21

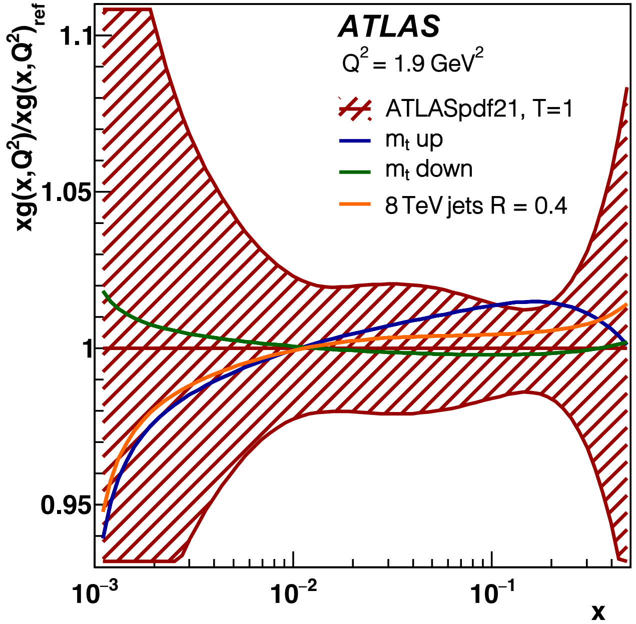
- We removed all the $t\bar{t}$ **data** from the fit
- Removing the 8 and 13 TeV $t\bar{t}$ data separately makes it clear that the **8 TeV has the bigger effect**
- 8 TeV looked to have more impact when combined with only a few data sets but not so much after many other data sets are added



Scale and m_{top} variations on top data

- μ_R variation by a factor 2 up/down
- Impact on the gluon PDF
- **Scale dependence of 8 TeV $t\bar{t}$ data is very modest**
- The effect of this scale change is well within the PDF uncertainties
- Study conducted on the 13 TeV $t\bar{t}$ data as well – even smaller impact (more in backup)
- We studied μ_R variations because it is the most impactful scale change (more in backup for μ_F)
- The variation of m_{top} considered as well
- Larger impact, still well within PDF uncertainties





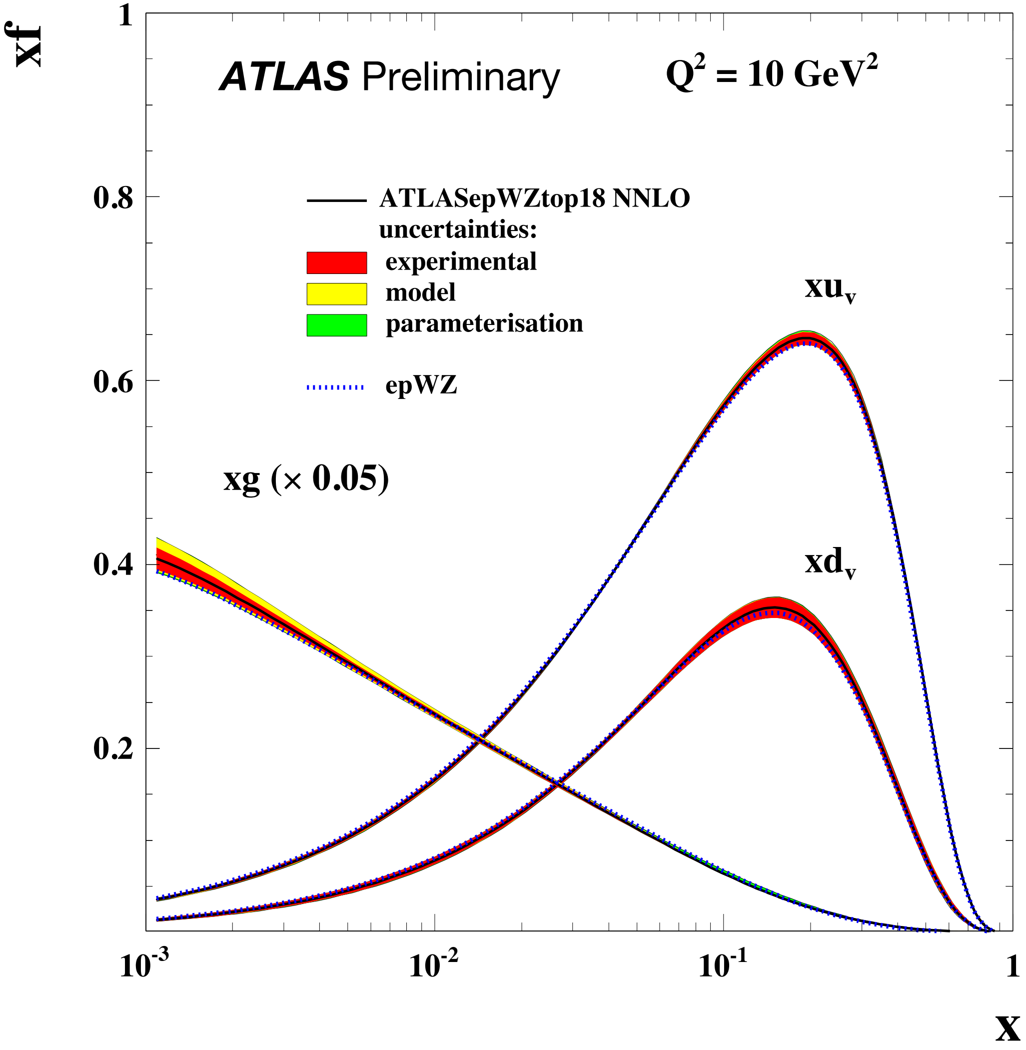
Conclusion and outlook

- Different $t\bar{t}$ data sets already included in recent ATLAS PDF studies
- Gluon constraints at medium- and high-x (both shape and uncertainties)
- 8 TeV data have larger impact on the the gluon PDF wrt 13 TeV data (being the former more precise than the latter)
- What next? Investigate impact of full Run 2 $t\bar{t}$ data (1D, 2D and **3D (!!!)** – lepton + jets channel)
  - NNLO predictions available through MATRIX + PineAPPL – very CPU-consuming
  - Possible extension to **simultaneous** extraction of **PDF + m_{top} (+ α_S)**
  - Check the complementarity with full Run 2 inclusive jet data
- Run 3 is halfway through → ~300 fb^{-1}expected by end of 2025
- **Even more impressive results with larger dataset**
- Interesting times ahead… Stay tuned! ☺

Backup Slides

epWZtop18: resulting PDFs

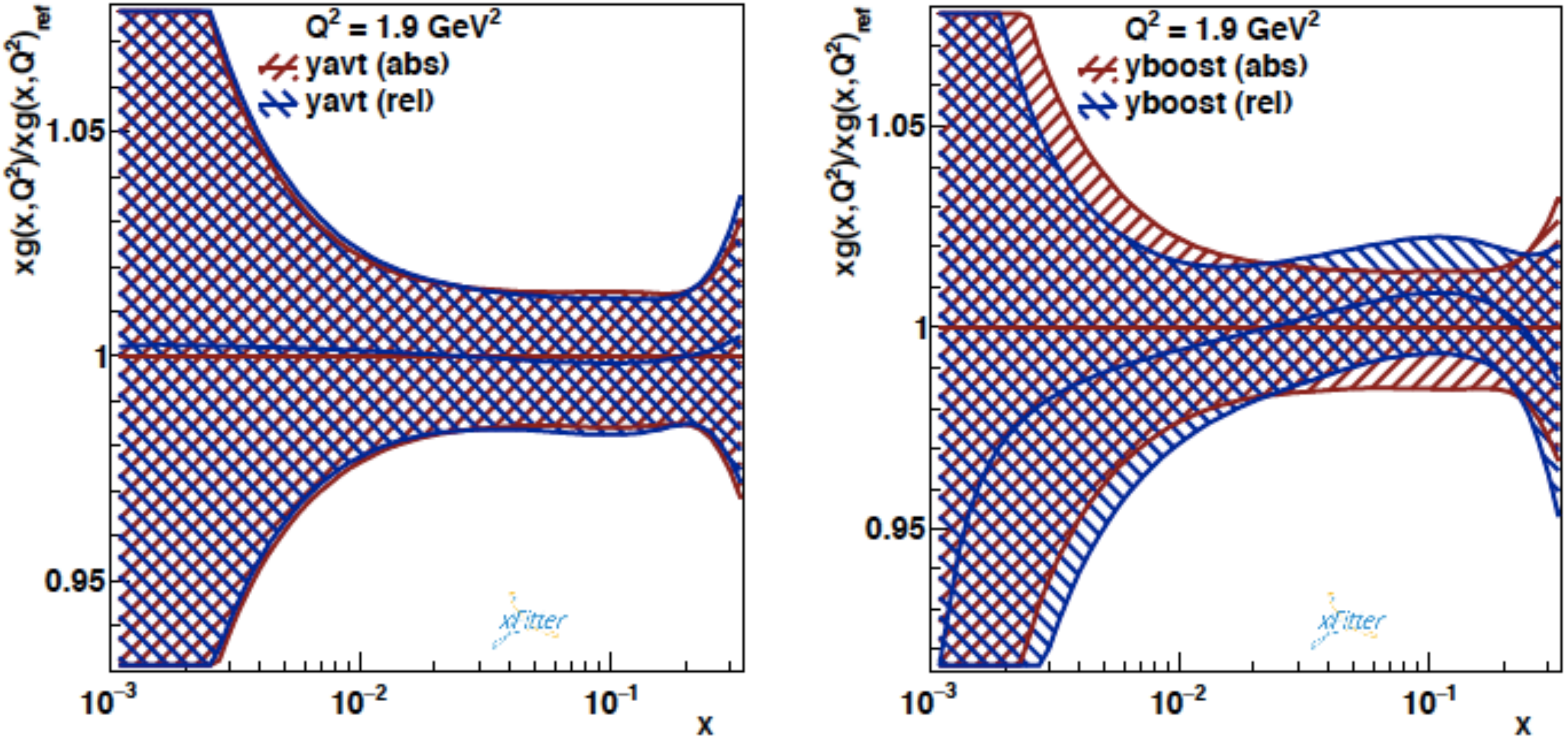
- The model uncertainties include variations of the charm/beauty masses, the minimum Q^2 cut value and the Q_0^2 starting scale
- The parametrisation uncertainties correspond to an envelope of results obtained with extra parameters
- The shapes of the extracted PDFs are not sensitive on the top quark mass, but the χ^2 of the fit is sensitive to it
- The strong coupling constant α_S was set to the PDG value and investigating its impact was beyond the scope of the study



Normalised or absolute distributions?

- Predictions for a fit to some 1D spectra if using normalised or absolute distributions

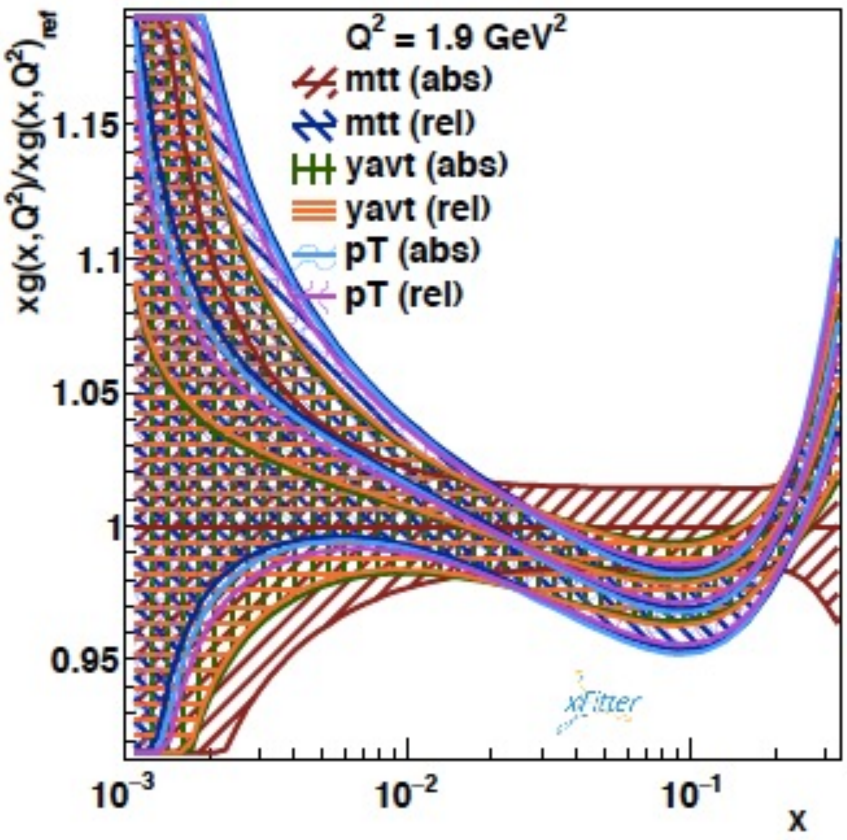
- When including normalised cross sections, the last bin of the distribution is removed



Normalised or absolute distributions?

- Predictions for a fit to some 1D spectra if using normalised or absolute distributions

- When including normalised cross sections, the last bin of the distribution is removed

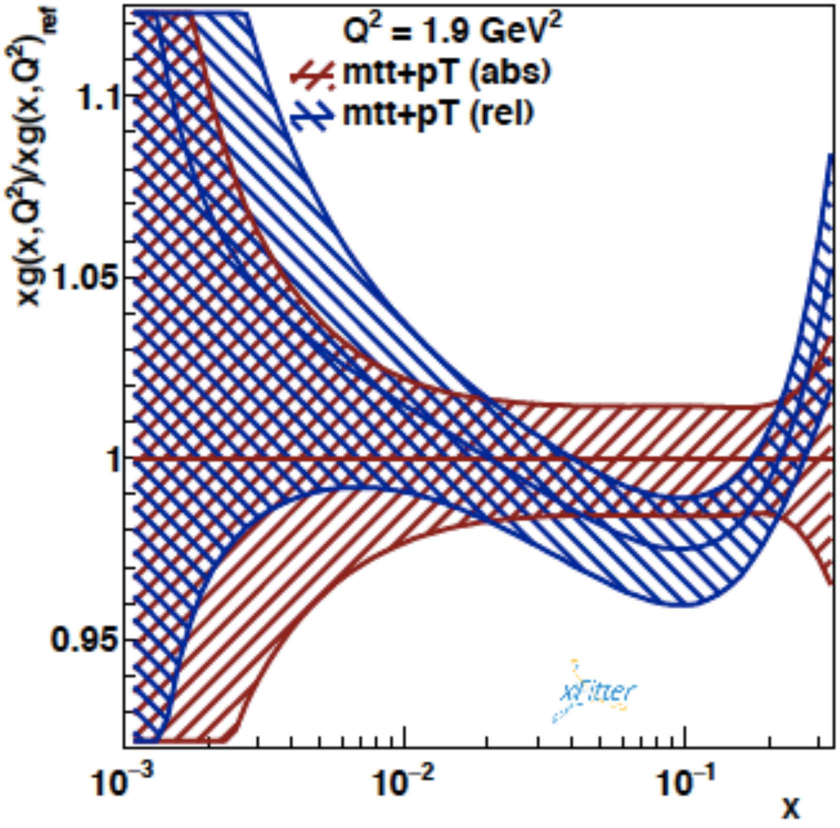


Dataset	mtt (abs)	mtt (rel)	yt (abs)	yt (rel)
Exp. χ^2	1873	1879	1865	1861
Corr. χ^2	213	208	208	207
Total χ^2	2086	2087	2074	2068

- Similar χ^2 (also the ones not reported)
- PDFs consistent within uncertainties
- No better description of the data if using normalised distributions instead of absolute ones
- Fit parameters compatible within uncertainties

Normalised or absolute distributions?

- Predictions for a fit to some 1D spectra if using normalised or absolute distributions
- When including normalised cross sections, the last bin of the distribution is removed

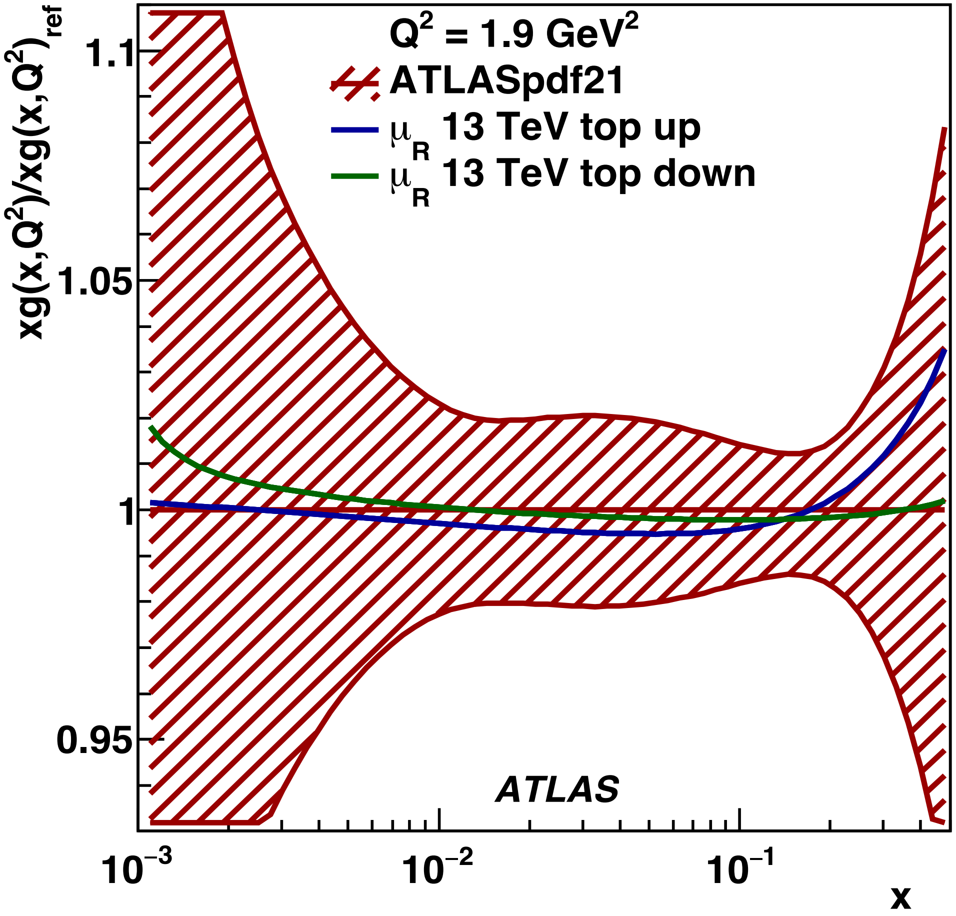


Dataset	mtt+pT (abs)	mtt+pT (rel)
Exp. χ^2	1877	1883
Corr. χ^2	212	208
Total χ^2	2089	2091

- We also tried to included two 1D spectra at a time…
- … and same conclusions found!

Impact of scale variations on top data

- Renormalisation scale uncertainty variation by a factor 2 up/down
- Impact on the gluon PDF
- **Scale dependence of 13 TeV $t\bar{t}$ data TeV is negligible**
- The effect of this scale change is well within the PDF uncertainties
- Smaller than the one observed for the 8 TeV $t\bar{t}$ data only



Impact of scale variations on top data

- Factorisation scale uncertainty variation by a factor 2 up/down
- Impact on the gluon PDF
- **Scale dependence of $t\bar{t}$ data (both 8 and 13 TeV) is very modest**
- The effect of this scale change is well within the PDF uncertainties and it is smaller than the one observed for μ_R