Top Data in MSHT2020 PDFs

Robert Thorne

September 2nd 2021

University College London

With Shaun Bailey, Tom Cridge, Lucian Harland-Lang, and Alan Martin

I will discuss the inclusion of and effect of top quark data in what we call the MSHT2020 PDFs.

Mass Scheme Hessian Tolerance – now intended to be a permanent naming convention.

Includes many more cross sections, all at NNLO.

Problems with correlated uncertainties and tensions in data sets in some cases.

NLO clearly no longer sufficient for real precision.

New LHC data fit.

New data on $\sigma_{t\bar{t}}$ at 8 TeV, added to previous data mainly at 7 TeV and Tevatron combined data.

ATLAS single differential distributions in $p_{T,t}, M_{t\bar{t}}, y_t, y_{t\bar{t}}$

CMS double differential distributions in $p_{T,t}, y_t$ both at 8 TeV.

Also single differential ATLAS dilepton and CMS lepton $+$ jet both as a function of $y_{t\bar{t}}$ only.

For inclusive data allow top mass (pole) to vary. Central value $m_t = 172.5 \text{ GeV}$ with an error (with penalty) of 1 GeV. Gives $m_t = 172.9 \text{ GeV}$ at NNLO and $m_t = 169.9 \text{ GeV}$ at NLO.

For the differential distributions use grids for the cross sections with fixed $m_t = 172.5 \text{ GeV}$.

Include all our recent LHC data updates in the fit at NNLO (for default $\alpha_S(M_Z^2) = 0.118$).

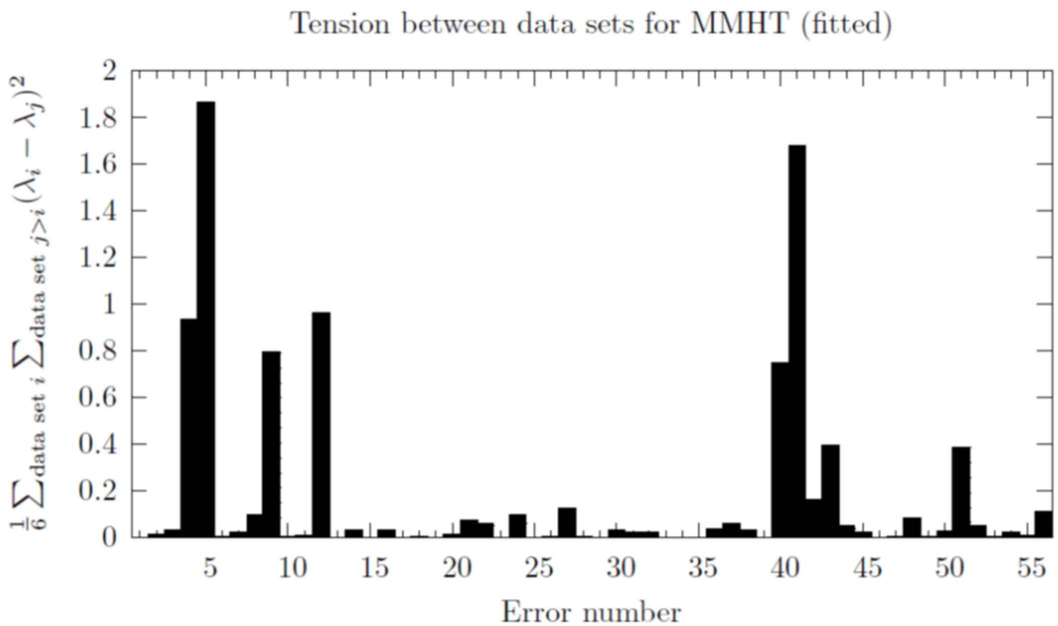
	no. points	NNLO $\chi^2/N_{\rm pts}$
D0 W asymmetry	14	0.86
$\sigma_{t\bar{t}}$ Tevatron +CMS+ATLAS $7, 8\rm TeV$	17	0.85
LHCb 7+8 TeV $W + Z$	67	1.48
LHCb 8 TeV e	17	1.54
CMS 8 TeV W	22	0.58
ATLAS 7 TeV jets $R = 0.6$	140	1.59
CMS 7 TeV $W + c$	10	0.86
ATLAS 7 TeV W, Z	61	1.91
CMS 7 TeV jets $R = 0.7$	158	1.11
ATLAS 8 TeV Zp_T	104	1.81
CMS 8 TeV jets	174	1.50
ATLAS 8 TeV $t\bar{t} \rightarrow l + {\rm j}$ single-diff	25	1.02
ATLAS 8 TeV $t\bar{t} \rightarrow l^+l^-$ single-diff	5	0.68
ATLAS 8 TeV high-mass Drell-Yan	48	1.18
ATLAS 8 TeV $W^{+,-} + {\rm jet}$	32	0.60
CMS 8 TeV $(d\sigma_{t\bar{t}}/dp_{T,t}dy_t)/\sigma_{t\bar{t}}$	15	1.50
ATLAS 8 TeV W^+, W^-	22	2.61
CMS 2.76 TeV jets	81	1.27
CMS 8 TeV $t\bar{t}$ y_t distribution	9	1.47
ATLAS 8 TeV double differential Z	59	1.45
Total, LHC data	1328	1.33
Total, all data	4363	1.17

Fit quality generally good. Relatively poor χ^2 values for some sets seemingly observed by other groups, rectified by treatment of correlated uncertainties.

ATLAS 8 TeV single differential data - systematic uncertainties

Generally the fit is good. However, most straightforward approach gives distinctly poor fit quality to this data set.

Clear tensions between different differential distributions. Clearly related to modelling-type systematic uncertainties, e.g. Parton Shower, ISR/FSR, hard scattering.



Discussed in detail in Eur.Phys.J.C 80 (2020) 60 - Bailey, Harland-Lang.

Decorrelated Parton shower systematic between distributions (maximal all three sources).

Baseline	No decor.	parton shower across	Max decor.
1.04	6.84	1.69	0.81

Table 4: $\chi^2/N_{\rm pts}$ for the ATLAS 8 TeV differential top data, with α_S free, in the lepton+jet channel. Results for different correlation scenarios, explained in the text, are given.

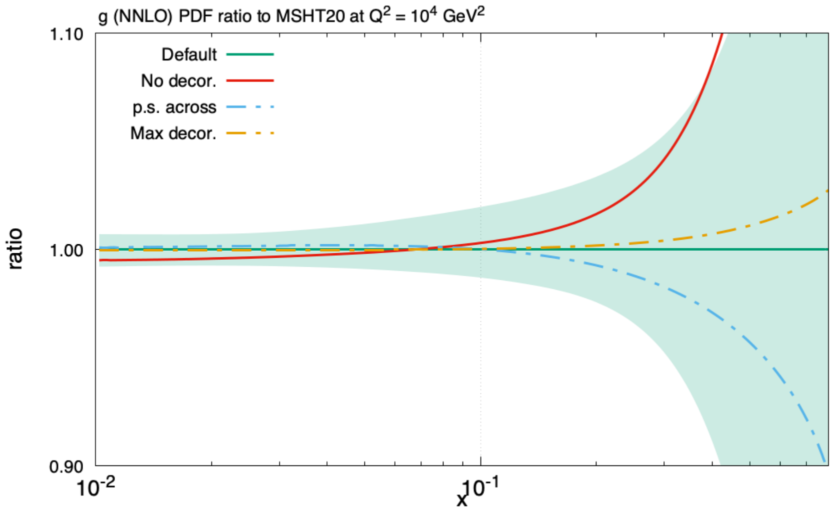
Baseline	NNLO QCD, LO EW	NLO QCD, LO EW
1.04	0.92	1.66

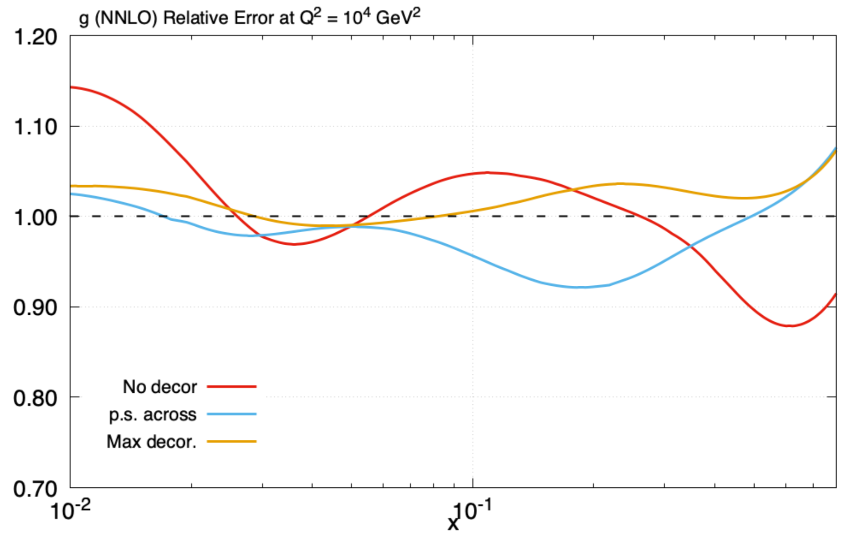
Table 5: $\chi^2/N_{\rm pts}$ for the ATLAS 8 TeV differential top data, with α_S free, in the lepton+jet channel. Results using different choices of theory for the matrix element calculation are shown.

Huge improvement in χ^2 when decorrelating the parton shower systematic across different distributions. Maximal not necessary, some redundancy between sources.

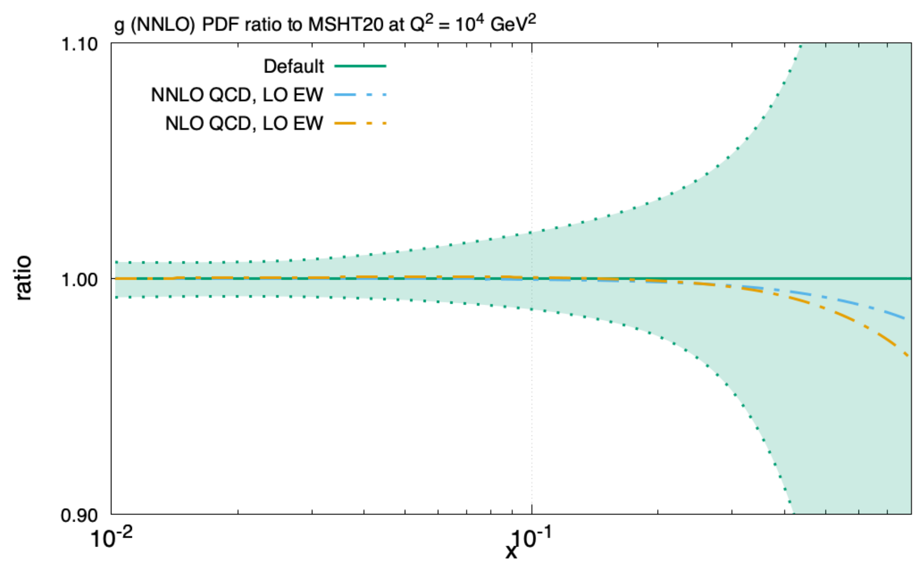
Even moreso when decorrelating the same systematic as a smooth function of rapidity, similar to smooth decorrelation advocated for jets in ATLAS – JHEP 09 020 (2017), i.e. split systematic uncertainty into $\cos(y_t/y_{range}), \sin(y_t/y_{range})$ components.

NNLO corrections clearly quite a large effect.





Effect on PDFs of the different systematic uncertainty treatments.



Effect of different orders on the gluon.

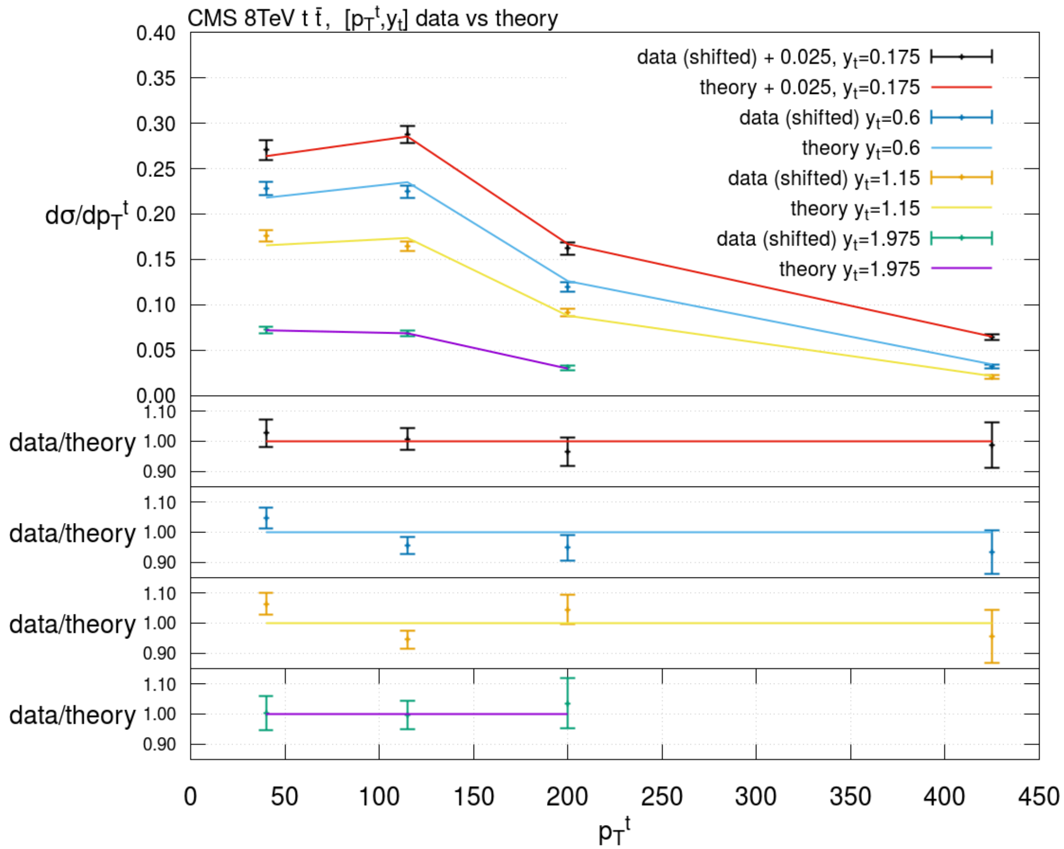
Other single Differential data.

Can only fit ATLAS dilepton and CMS lepton $+$ jet both as a function of $y_{t\bar{t}}$ only since correlations between different distributions not available.

ATLAS dilepton data lack statistical correlations between distributions. Fit to single distribution good.

CMS single differential data is normalized. We take systematic uncertainties as uncorrelated, as correlations destroy normalization of the data. Again, fit quality good.

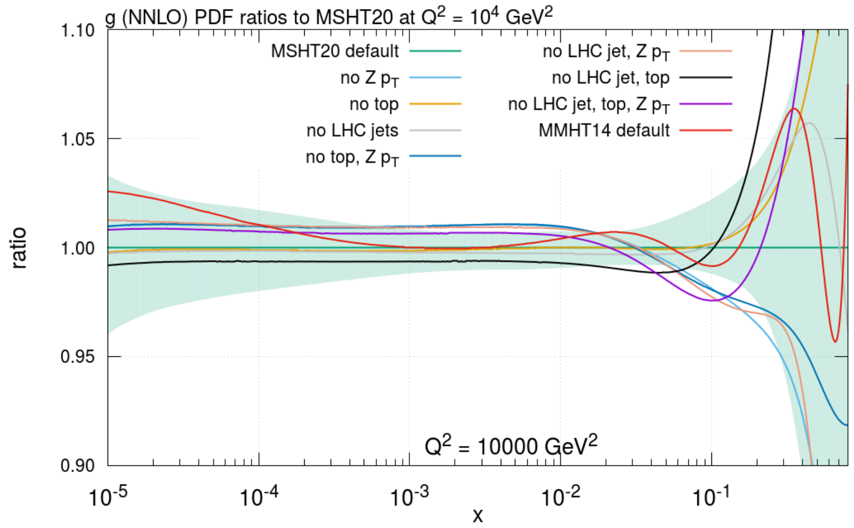
CMS double-differential top pair data.

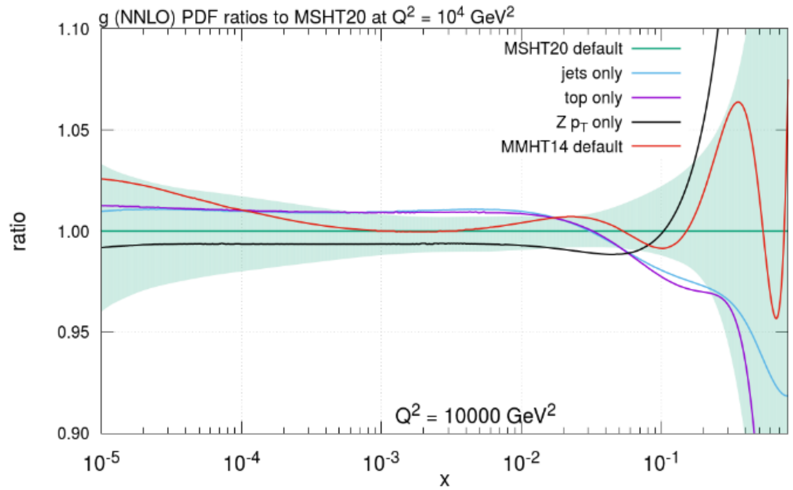


We choose to fit to the distribution differential in p_T^t, y_t, as the two variables are largely uncorrelated.

Fit good, with no systematic failings. Some systematics partially decorrelated (else not consistent with normalization of distribution).

Gluon tensions at high x.





Details in shape near and above $x = 0.1$ due to LHC jet, $Z\,p_T$ and differential $t\bar{t}$ data.

$Z\,p_T$ pulls gluon up, differential $t\bar{t}$ data pulls the gluon down. Affects lower x normalization via momentum sum rule.

Choice of jet data included in the fit and relative weight of data in the fit can improve the fit to single differential ATLAS data and improve the situation with correlated uncertainties, see Cridge yesterday.

Fits at NLO

We also produce PDFs at NLO (and also still at LO - fit very poor).

Start to notice significant deterioration in fit quality for some of the precision LHC data. NNLO now much preferred.

Data set	Points	NLO χ^2/N_{pts}	NNLO χ^2/N_{pts}
DØ W asymmetry	14	0.94	0.86
$\sigma_{t\bar{t}}$	17	1.34	0.85
LHCb 7+8 TeV $W + Z$	67	1.71	1.48
LHCb 8 TeV $Z \to ee$	17	2.29	1.54
CMS 8 TeV W	22	1.05	0.58
CMS 7 TeV $W + c$	10	0.82	0.86
ATLAS 7 TeV jets $R = 0.6$	140	1.62	1.59
ATLAS 7 TeV $W + Z$	61	5.00	1.91
CMS 7 TeV jets $R = 0.7$	158	1.27	1.11
ATLAS 8 TeV Z p_T	104	2.26	1.81
CMS 8 TeV jets $R = 0.7$	174	1.64	1.50
ATLAS 8 TeV $t\bar{t} \to l + j$ sd	25	1.56	1.02
ATLAS 8 TeV $t\bar{t} \to l^+l^-$ sd	5	0.94	0.68
ATLAS 8 TeV high-mass DY	48	1.79	1.18
ATLAS 8 TeV $W^+W^- +$ jets	30	1.13	0.60
CMS 8 TeV $(d\sigma_{\bar{t}t}/dp_{T,t}dy_t)/\sigma_{\bar{t}t}$	15	2.19	1.50
ATLAS 8 TeV W^+W^-	22	3.85	2.61
CMS 2.76 TeV jets	81	1.53	1.27
CMS 8 TeV $d\sigma_{\bar{t}t}/dy_t$	9	1.43	1.47
ATLAS 8 TeV double differential Z	59	2.67	1.45
Total, LHC data in MSHT20	1328	1.79	1.33
Total, non-LHC data in MSHT20	3035	1.13	1.10

Constraints on $\alpha_S(M_Z^2)$ and quark masses.

For MMHT2014 $\alpha_S(M_Z^2) = 0.1172 \pm 0.0013$ (NLO) and $\alpha_S(M_Z^2) = 0.1201 \pm 0.0015$ (NLO).

Final current value from global fit $\alpha_S(M_Z^2) = 0.1174 \pm 0.0013$ at NNLO and $\alpha_S(M_Z^2) = 0.1203 \pm 0.0015$ at NLO.

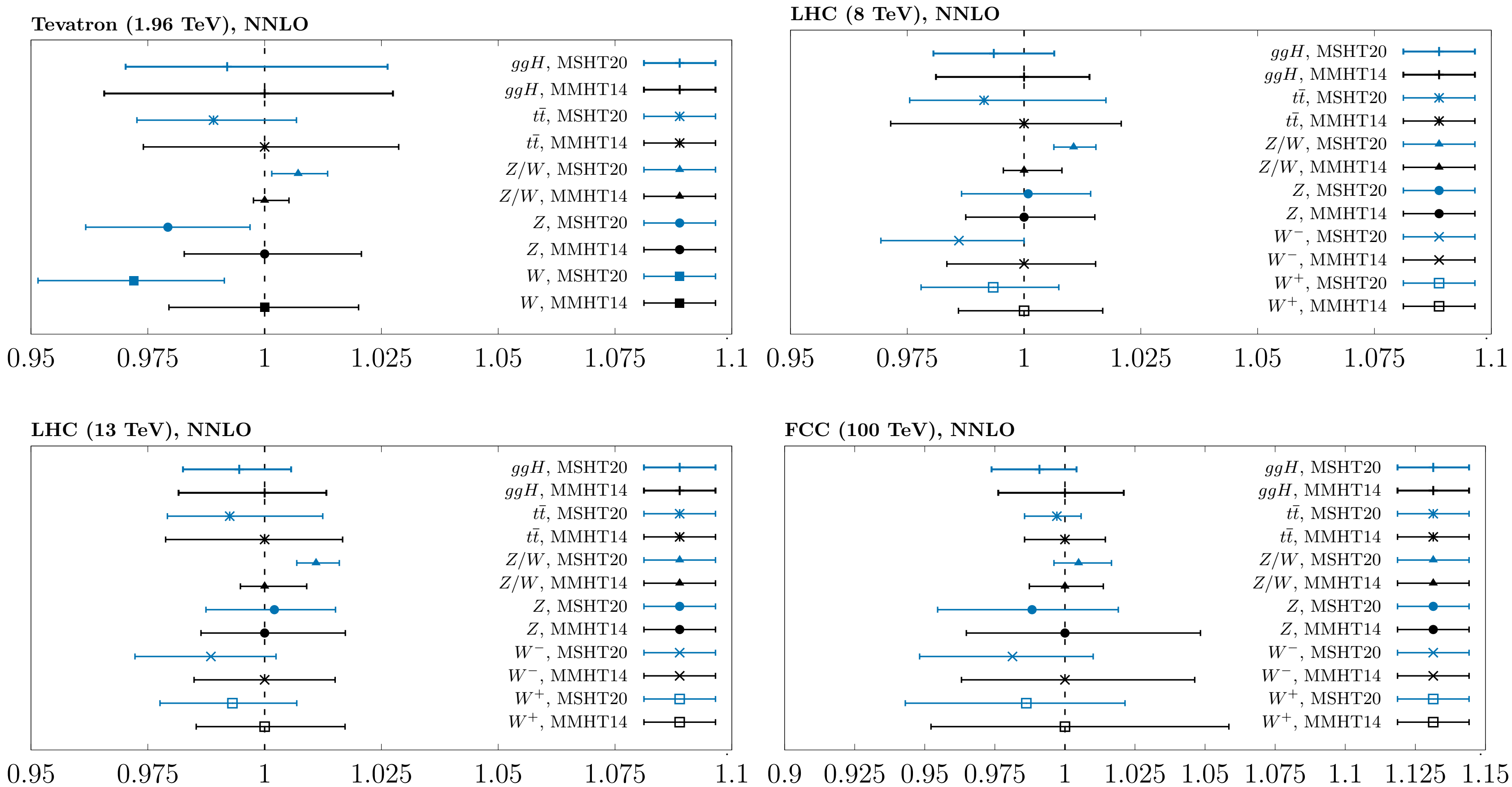
Data on top pair total cross sections constrains $\alpha_S(M_Z^2)$ quite well. Consistent with global fit at NNLO.

Constrains low values at NLO strongly, but anti-correlation with fit value of m_t due to opposite effects on cross sections and unrealistic values of m_t for low $\alpha_S(M_Z^2)$. Therefore constraint not taken as limiting.

Single differential data provide some constraint on upwards variation of $\alpha_S(M_Z^2)$ at NNLO and NLO, but for fixed top mass. Other data more constraining in both cases.

Predictions for Benchmark Processes.

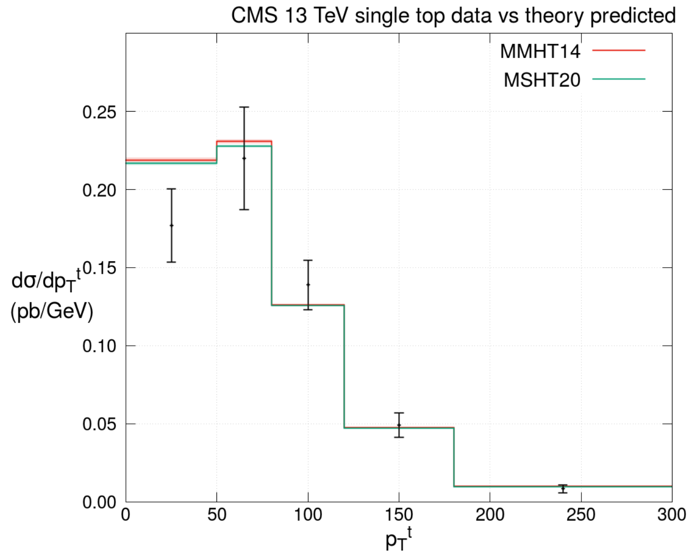
Some changes in σ_W, σ_Z and particularly their ratio largely due to changes in strange quarks.

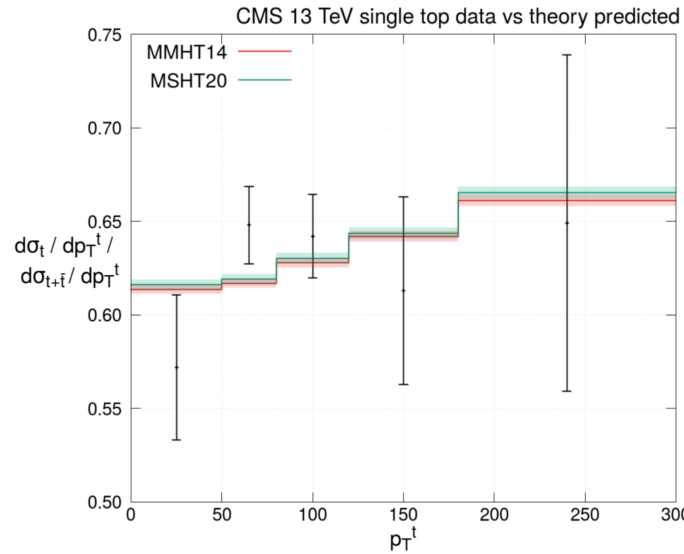


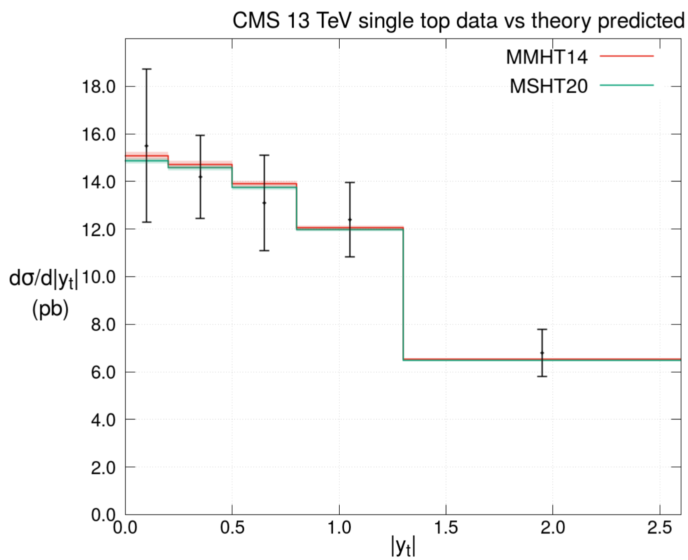
For gluon initiated top and Higgs cross sections an improvement in uncertainties but central values stable.

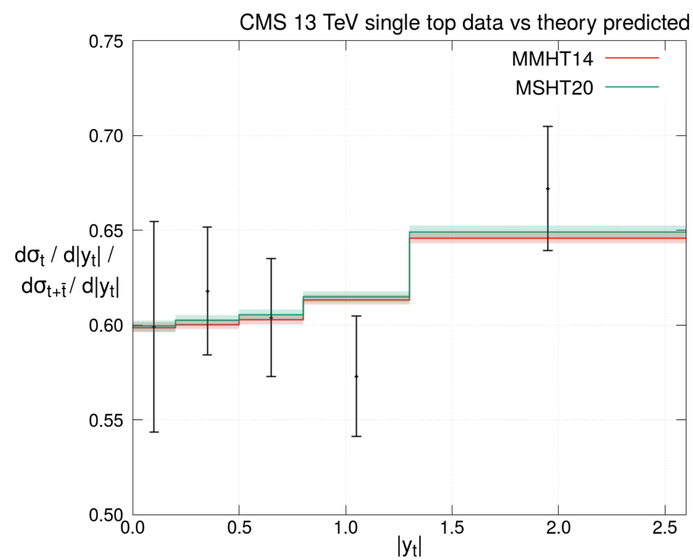
Predictions for other data

Single top data not fit (uncertainties much larger than PDF uncertainties), but good preditions (e.g. Eur.Phys.J.C 80 (2020), 370).

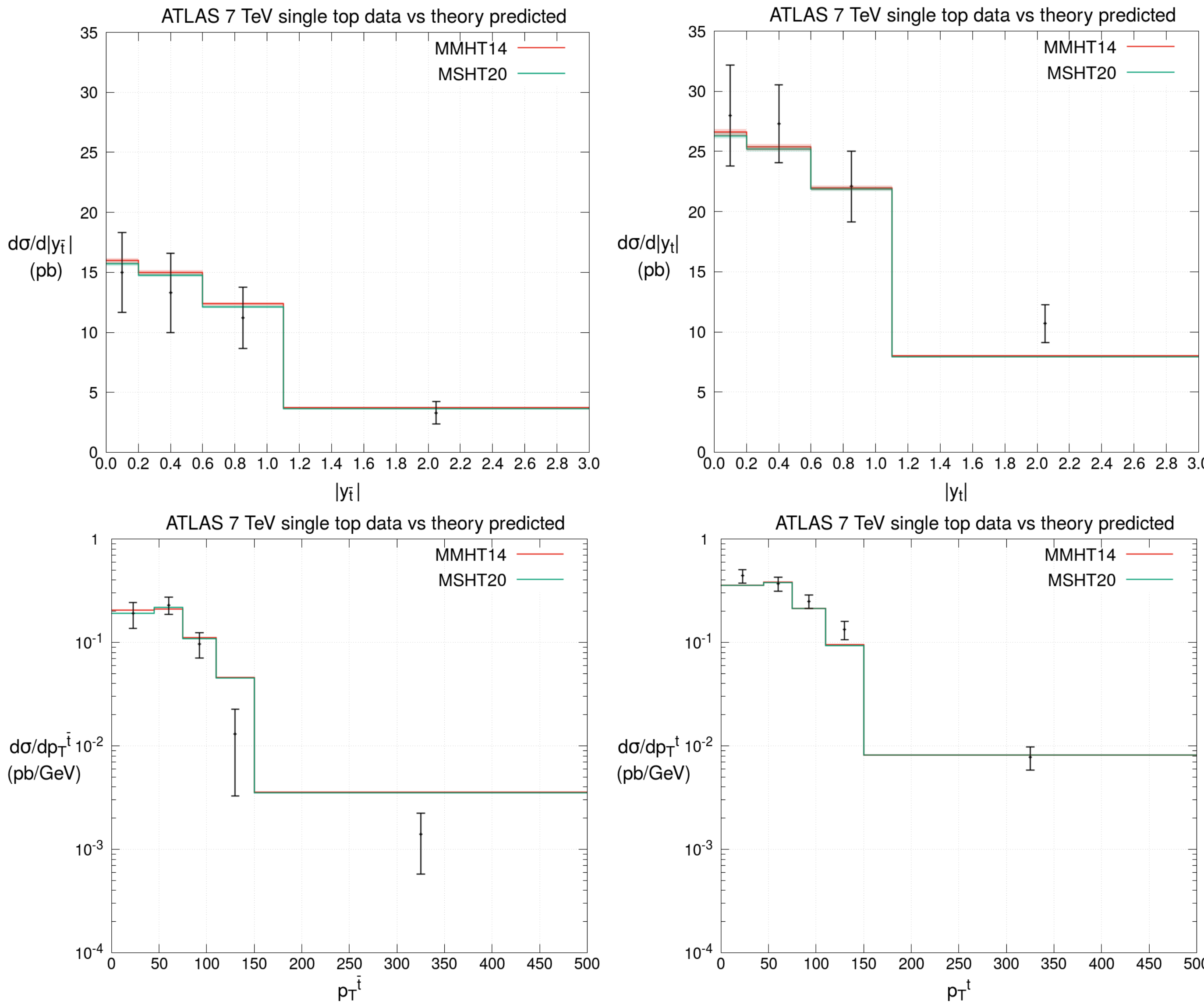








Also ATLAS, Phys. Rev. D 90, 112006 (2014)



In principle tests the u/d ratio.

Conclusions

LHC data starting to have a very significant impact on PDF extractions.

Theory catching up for precision data, e.g NNLO single and double-differential top.

Largely stability for gluon, but uncertainty reduction and definite pull from top data in tension with other data sets.

Precision data and theory causing problems in cases where correlated systematics (which increasingly dominate) are important. Top data an example. Improved interplay between theory/experiment on these seems a priority.