# Large-$x$ PDFs

S. Alekhin[1], M.V. Garzelli[2,1], J. Mazzitelli[3], S.-O. Moch[1], O. Zenaiev[1]

[1] II Institut für Theoretische Physik, Universität Hamburg
[2] CERN, Theory Department
[3] PSI, Villigen

Mainly on the basis of [arXiv:2311.05509[hep-ph]],
[arXiv:2407.00545[hep-ph]]
and work in progress.

MITP Program “Next Generation pQCD for Hadron Structure: preparing for the EIC”, October 21-31, 2024

# PDFs as ingredient in collinear factorization

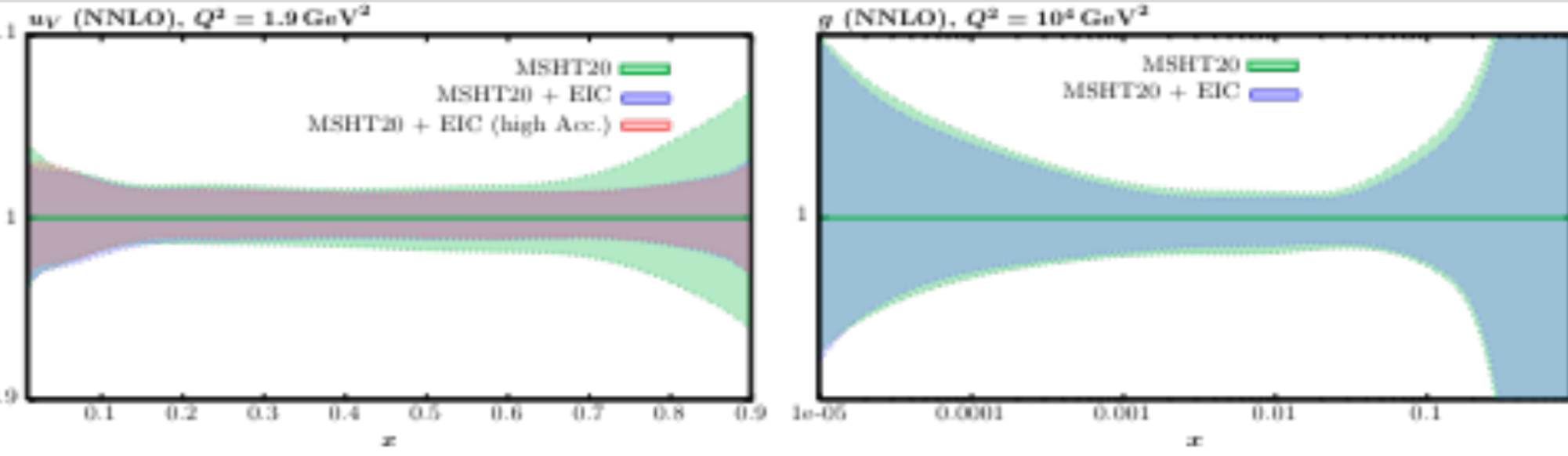
$*$ PDFs are necessary ingredients for the calculation of hadron-hadron interaction cross-sections in collinear factorization.
For $t\bar{t} + X$ production, one has e.g.

$$d\sigma_{pp\to X} = \sum_{i,j} PDF(x_i, \mu_F) \otimes PDF(x_j, \mu_F) \otimes d\hat{\sigma}_{ij\to X}(x_i P_i, x_j P_j, \mu_F, \mu_R, m_X)$$

$*$ Their dependence on $Q^2$ is regulated by DGLAP evolution equations, while their dependence on the longitudinal momentum fraction $x = p^+/P^+$ is fitted.

$*$ Lattice QCD calculations could also provide useful input for PDF fits, but this is not yet used in the standard so-called “global PDF fits”.

# ....while waiting for the Electron-Ion Collider (EIC)




*from T. Cridge et al., MSHT20, arXiv:2309.11269*

*ep* DIS at EIC will impact valence quarks distributions, more than gluon one. Larger impact on the gluon will come from HL-LHC *pp* data.

$\Rightarrow$ synergy EIC/HL-LHC is needed

# Constraints on PDFs at large $x$

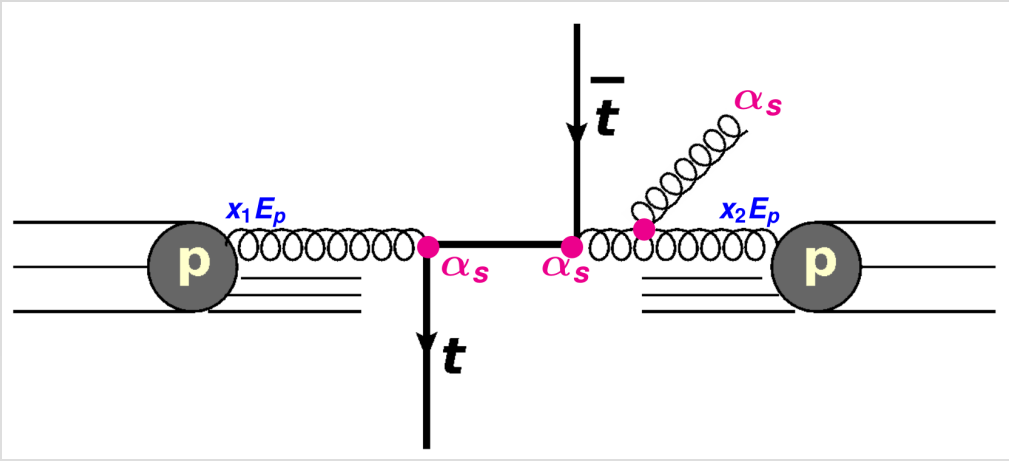
* PDF uncertainties blow up at large $x$
  - 5 - 10% relative uncertainty for $x \sim 0.7$ for $q_V$ PDFs, much larger for $g$ PDFs.
  - lack of data
  - big errorbars on the existing data, even due to the fact that
  - PDFs $\to 0$ for $x \to 1 \Rightarrow$ low parton luminosities for large $x \Rightarrow$ low $\sigma$'s.
  - theory issue: threshold logs $\alpha^k ln^{2k-1}(1-x)$ become large due to incomplete cancellation between real and virtual diagrams

* Loose constraints from HERA in the $0.1 < x < 1$ region

* While waiting for EIC,
  - fixed-target inclusive DIS experiments (SLAC, BCDMS, NMC, etc.)
  - semi-inclusive DIS experiments with $\nu$ beams (CCFR, NuTeV, CHORUS, NOMAD, etc.)
  - fixed-target DY experiments (CERN-NA51, FNAL-E605, FNAL-E866, etc.)
  - collider DY (+ jets) experiments (Tevatron, LHC)
  - collider single-top, $t\bar{t}$, inclusive jet and dijet production

* For gluon distributions: top-quark data $\leftarrow$ focus of this talk

* For sea-quark distributions: DY data $\leftarrow$ focus of this talk




- $m_t$, $\alpha_s(M_Z)$, $g$, $q$ and $\bar{q}$ PDFs are inputs for the computation of $pp \to t\bar{t} + X$ cross sections already at LO.
- $m_t$, $q$ and $\bar{q}$ PDFs also appear in the computation of cross sections for single-top production at LO, whereas in the $s$- and $t$-channels the dependence on $\alpha_s(M_Z)$ and $g$ PDFs appear only at higher orders.

$\Rightarrow$ If we want to use the cross-section data to extract PDFs, we have to take into account the correlations with $m_t$ and $\alpha_s(M_Z)$ (unless one supposes to know already the values of $m_t$ and $\alpha_s(M_Z)$, e.g. from independent measurements).

$\Rightarrow$ Simultaneous fits of PDFs, $m_t(m_t)$ and $\alpha_s(M_Z)$ have been performed:
- ABMP16, using total inclusive top data [S. Alekhin et al., PRD 96 (2017) 014011],
- ABMPtt, using multidifferential top data $\to$ this talk.

## Main messages from the ABMPtt fit, including top-quark data

$*$ The precision of current top-quark LHC experimental data (full Run 2 + start of Run 3) has allowed to reduce our uncertainties on the $g$ PDF at large $x$ and on $m_t(m_t)$ w.r.t the baseline ABMP16 by a factor $\sim 2$, remaining consisting.

$*$ The use of $t\bar{t} + X$ multi-differential data is a key aspect in this respect.
It also allows to reduce the large correlations between $\alpha_s(M_Z)$ and $m_t(m_t)$.

$*$ To facilitate the fitting work, it is very important that the experimental collaborations provide normalized cross-section data and, as much as info as possible on correlations of the uncertainties within an analysis and between different analyses. We also encourage cross-calibrations and combinations of ATLAS and CMS results.

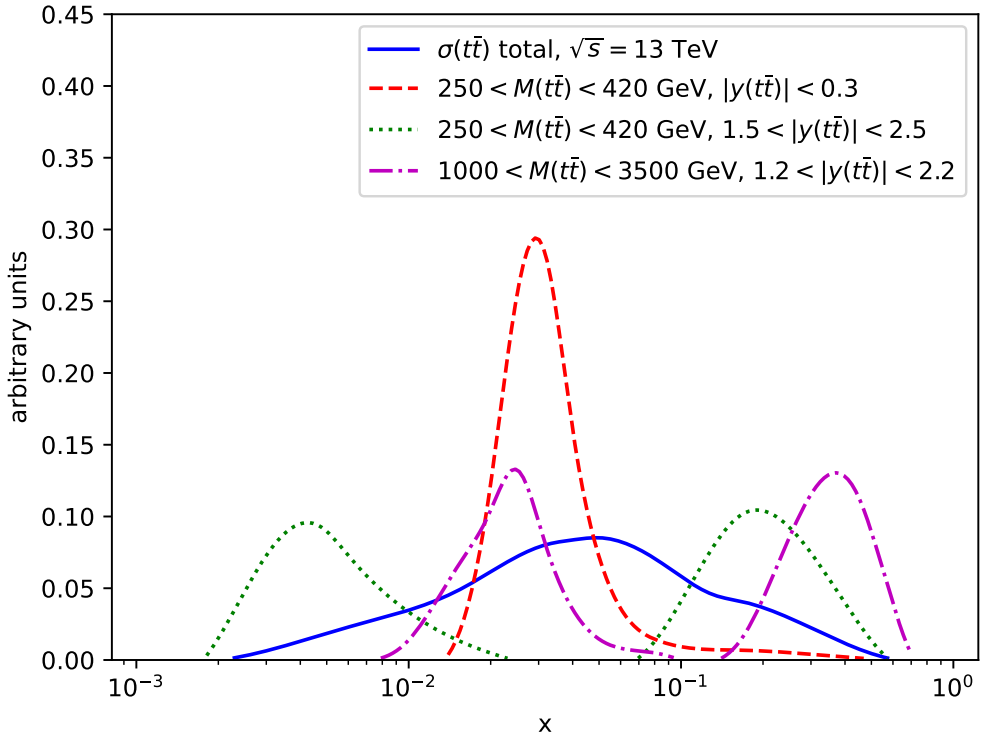
$*$ More studies, both at the experimental level and at the theory level, have to be done on single-top hadroproduction, to reduce the systematic uncertainties. This allows us to study flavour dependence of $q$ PDFs and might be crucial, once the experimental uncertainties will be reduced, to further reduce the correlations between extracted values of $\alpha_s(M_Z)$ and $m_t(m_t)$.

$*$ PRELIMINARY: accomodating a photon in our PDFs (new fit in preparation: ABMPttgam) produces results consistent with those shown in the following.

# $x$ intervals probed by $t\bar{t} + X$ hadroproduction

- $pp \to t\bar{t} + X$ @ 13 TeV probes $0.002 \lesssim x \lesssim 0.7$
  - $gg$ contributes $\approx 90\%$
- (double)-differential data probe different $x$ subintervals
- in particular we consider distributions double-differential in $M(t\bar{t})$ and $y(t\bar{t})$.
- Scales $m_H$, $M_W$, $M_Z$ and $m_t$ are similar among each other
- Higgs production at the LHC probes $x \sim m_H/\sqrt{s} \sim 0.01$ which is well covered by differential $t\bar{t} + X$ data
- DY production at the LHC probes a similar region $x \sim m_{W,Z}/\sqrt{s}$
  - mostly sensitive to quark PDFs
  - helps with light flavor separation

LO: $x_{1,2} = (M(t\bar{t})/\sqrt{s}) \exp[\pm y(t\bar{t})]$



# Theory framework for $t\bar{t} + X$ hadroproduction

- NNLO computations for total inclusive $pp \to t\bar{t} + X$ cross sections can be obtained with theory tools already publicly available since long (HATHOR, Fasttop, Top++).

- NNLO computations for total and multi-differential $pp \to t\bar{t} + X$ cross sections can now be performed thanks to the publicly available MATRIX framework [Catani, Devoto, Grazzini, Kallweit, Mazzitelli Phys.Rev.D 99 (2019) 5, 051501; JHEP 07 (2019) 100]
  - fully differential NNLO calculations were also published in JHEP 04 (2017) 071 [Czakon, Heymes, Mitov] , but no public code available. However, the HighTEA database [Czakon et al., arXiv:2304.05993] has recently appeared.

- Master formula for $t\bar{t} + X$ hadroproduction in MATRIX:

$$d\sigma^{t\bar{t}}_{(N)NLO} = \mathcal{H}^{t\bar{t}}_{(N)NLO} \otimes d\sigma^{t\bar{t}}_{LO} + \left[ d\sigma^{t\bar{t}+jet}_{(N)LO} - d\sigma^{t\bar{t},CT}_{(N)NLO} \right]$$

  $*$ based on $q_T$-subtraction for cancelling IR divergences, where $\vec{q}_T = \vec{p}_{t,T} + \vec{p}_{\bar{t},T}$, $\vec{q}_T = 0$ at LO.

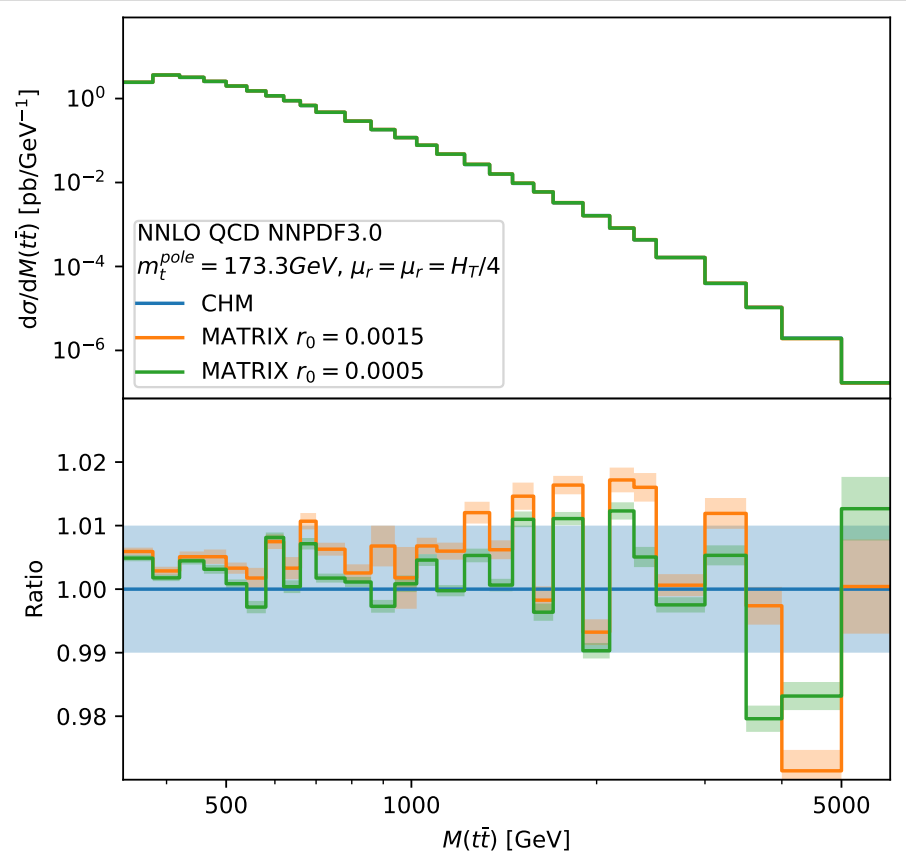
  $*$ $d\sigma^{t\bar{t}+jet}_{(N)LO}$ is IR divergent for $q_T \to 0$ The counterterm $d\sigma^{t\bar{t},CT}_{(N)NLO}$ compensating for the divergence is known from the fixed-order expansion of the resummation formula of the logarithmic contributions of the form $\alpha_s^{n+2}(1/q_T^2) ln^k(M_{t\bar{t}}^2/q_T^2)$ affecting the $q_T$ distribution, which are large in the limit $q_T \to 0$. $\Rightarrow$ The square bracket is finite for $q_T \to 0$.

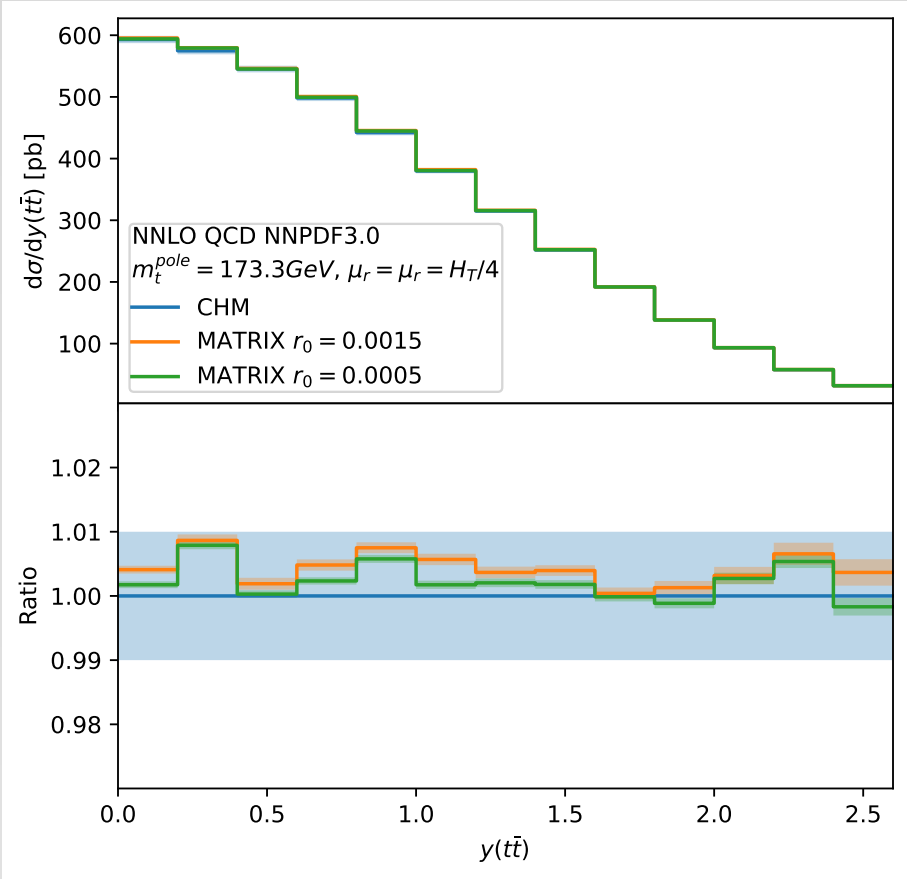
  $*$ in practice the calculation is performed by introducing cuts in $r = q_T/M$, with $r_{cut} \in [0.01\%, r_{max}]$ with $r_{max}$ varying between 0.5% and 1%.

# Our theory calculations with MATRIX + PineAPPL framework

- We use private version of MATRIX [Grazzini, Kallweit, Wiesemann, EPJC 78 (2018) 537]
- Interfaced to PineAPPL [Carrazza at al., JHEP 12 (2020) 108] to produce interpolation grids which are further used in xFitter https://gitlab.com/fitters/xfitter
  - reproduce NNLO calculations using any PDF + $\alpha_s(M_Z)$ set and/or varied $\mu_r$, $\mu_f$ in $\sim$ seconds
  - interface implemented privately and only for the $pp \to t\bar{t} + X$ process
- Further modifications to MATRIX to make possible runs with $\Delta\sigma_{t\bar{t}} < 0.1\%$
  - adapted to DESY Bird Condor cluster and local multicore machines
  - technical fixes related to memory and disk space usage, etc.
- We did runs with different $m_t$ values with step of 2.5 GeV and $\Delta\sigma_{t\bar{t}} = 0.02\%$
  - $\approx$ 350000 CPU hours/run ($\sim$30 years on a single CPU)
  - for differential distributions, statistical uncertainties in bins are $\lesssim 0.5\%$
- Differential distributions obtained with fixed $r_{cut} = 0.0015$ ($q_T$ subtraction)
  - checked that extrapolation to $r_{cut} = 0$ for total $\sigma(t\bar{t} + X)$ produces differences $< 1\%$, see also S. Catani et al., JHEP 07 (2019) 100
- $\mu_r = \mu_f = H_T/4, H_T = \sqrt{m_t^2 + p_T^2(t)} + \sqrt{m_{\bar{t}}^2 + p_T^2(\bar{t})}$, varied up and down by factor 2 with $0.5 \leq \mu_r/\mu_f \leq 2$ (7-point variation)

# Predictions for differential distributions with different $r_{cut}$ values



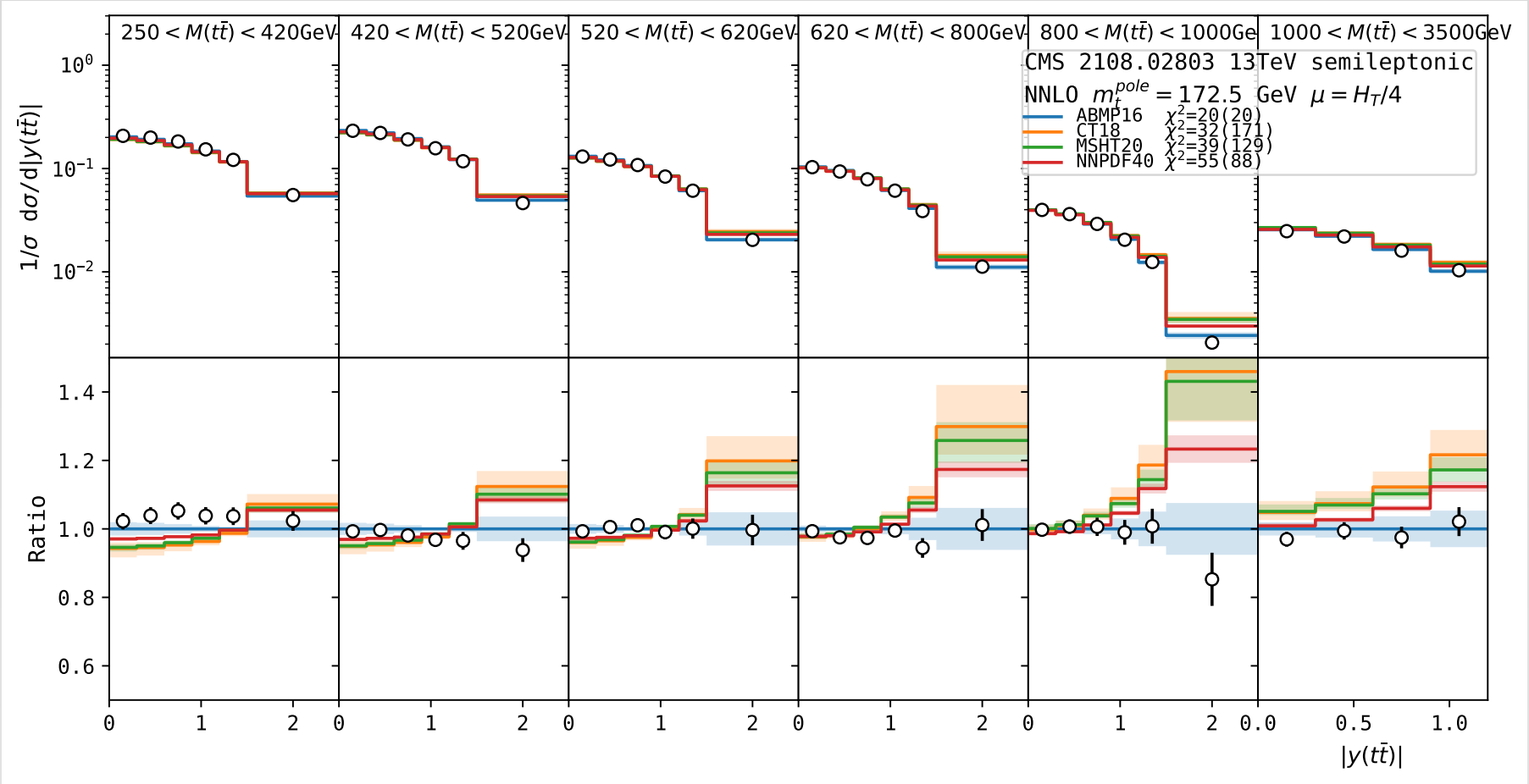

* In principle, the $q_T$-subtraction-based computation of (differential) cross-sections for finite $r_{cut}$ introduces power corrections, which vanish in the limit $r_{cut} \to 0$.

* In practice, good agreement with the exact calculation (local) by Czakon, Heymes, Mitov (CHM) (at least considering their quoted 1% uncertainty).

# ATLAS and CMS data used in this work

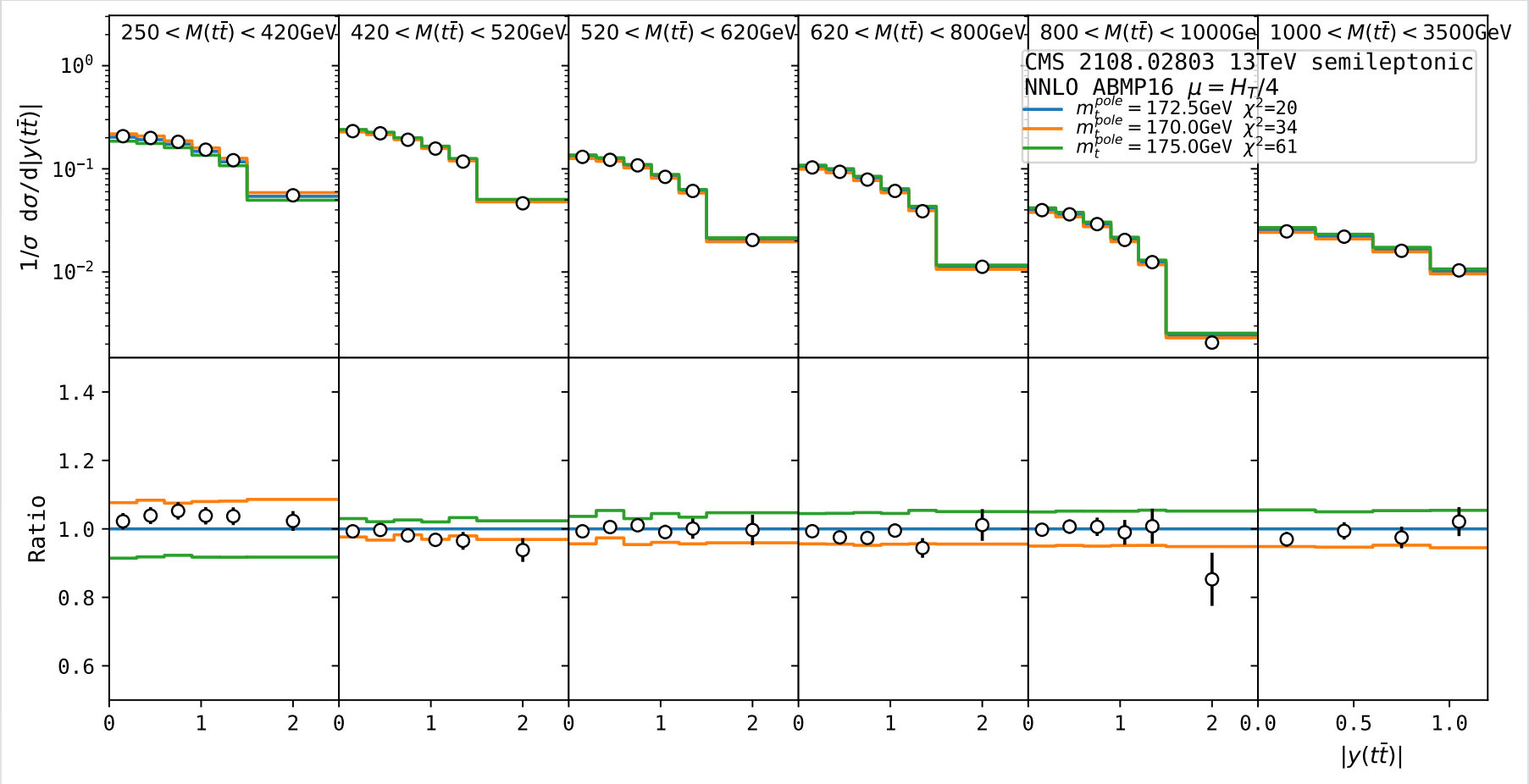
- We focus especially on measurements at 13 TeV where double-differential $M(t\bar{t})$, $y(t\bar{t})$ cross sections at parton level are available
  - **(1)** CMS EPJ C80 (2020) 658 [1904.05237, TOP-18-004]:
    2D cross sections in dileptonic channel, $L = 35.9\ \text{pb}^{-1}$
    - for 3D $M(t\bar{t})$, $y(t\bar{t})$, $N_{\text{jet}}$ cross sections, NNLO is not available for $t\bar{t}$ + jets + $X$
  - **(2)** CMS Phys.Rev.D104 (2021) 9, 092013 [2108.02803, TOP-20-001]:
    2D cross sections in l+jets channel, $L = 137\ \text{pb}^{-1}$
  - **(3)** ATLAS EPJ C79 (2019) 1028 [1908.07305]:
    2D cross sections in l+jets channel, $L = 36\ \text{pb}^{-1}$
  - **(4)** ATLAS JHEP 01 (2021) 033 [2006.09274]:
    2D cross sections in all-hadronic channel, $L = 36.1\ \text{pb}^{-1}$
- For all measurements, we use normalised cross sections unfolded to the final-state parton level
- We use information on correlations of experimental uncertainties as provided in the paper (1) or in the HEPDATA database (2,3,4)
  - assumed no correlation between different measurements
    (reasonable assumption for normalised cross sections)
- it would be interesting to also add LHCb data (sensitivity to larger $x$ and to $m_t$), but they are only available in the fiducial phase-space (cuts on leptons)
- Additionally, we use total inclusive $t\bar{t}$ + $X$ and single-top cross-section data at all energies, according to summary plots by the LHC Top Working Group + Tevatron.

# CMS TOP-20-001 vs NNLO predictions using different PDFs



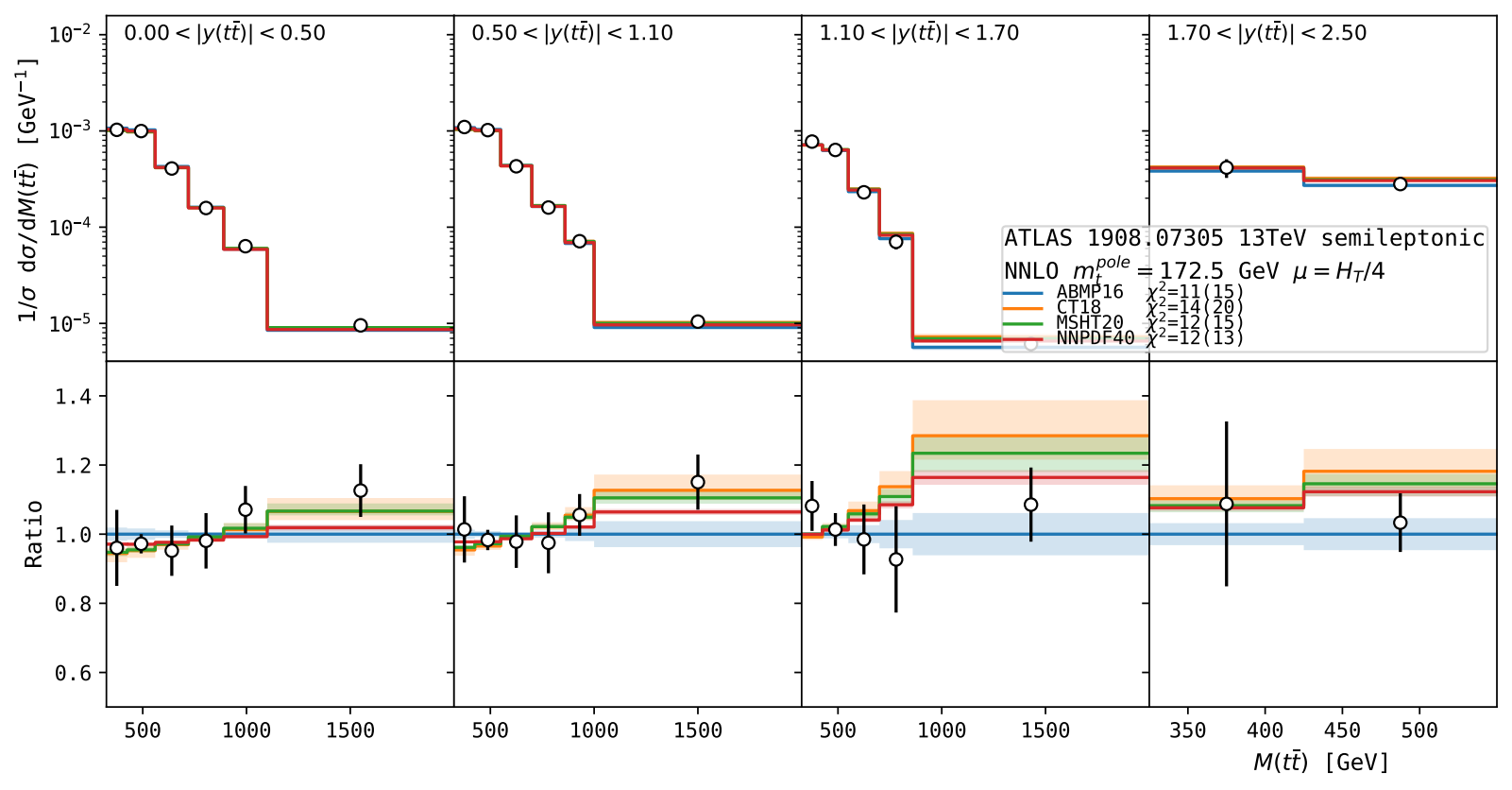

- Fixed $m_t^{\text{pole}} = 172.5$ GeV, $\mu_r = \mu_f = H_T/4$
- Reported $\chi^2$ values with (and without) PDF uncertainties
- All PDF sets describe data reasonably well, with best description by ABMP16
  - CT18, MSHT20 and NNPDF40 show clear trend w.r.t data at high $y(t\bar{t})$ (large $x$)
- This is most precise currently available dataset with finest bins

# CMS TOP-20-001 vs NNLO predictions with ABMP16 and different $m_t^{\text{pole}}$



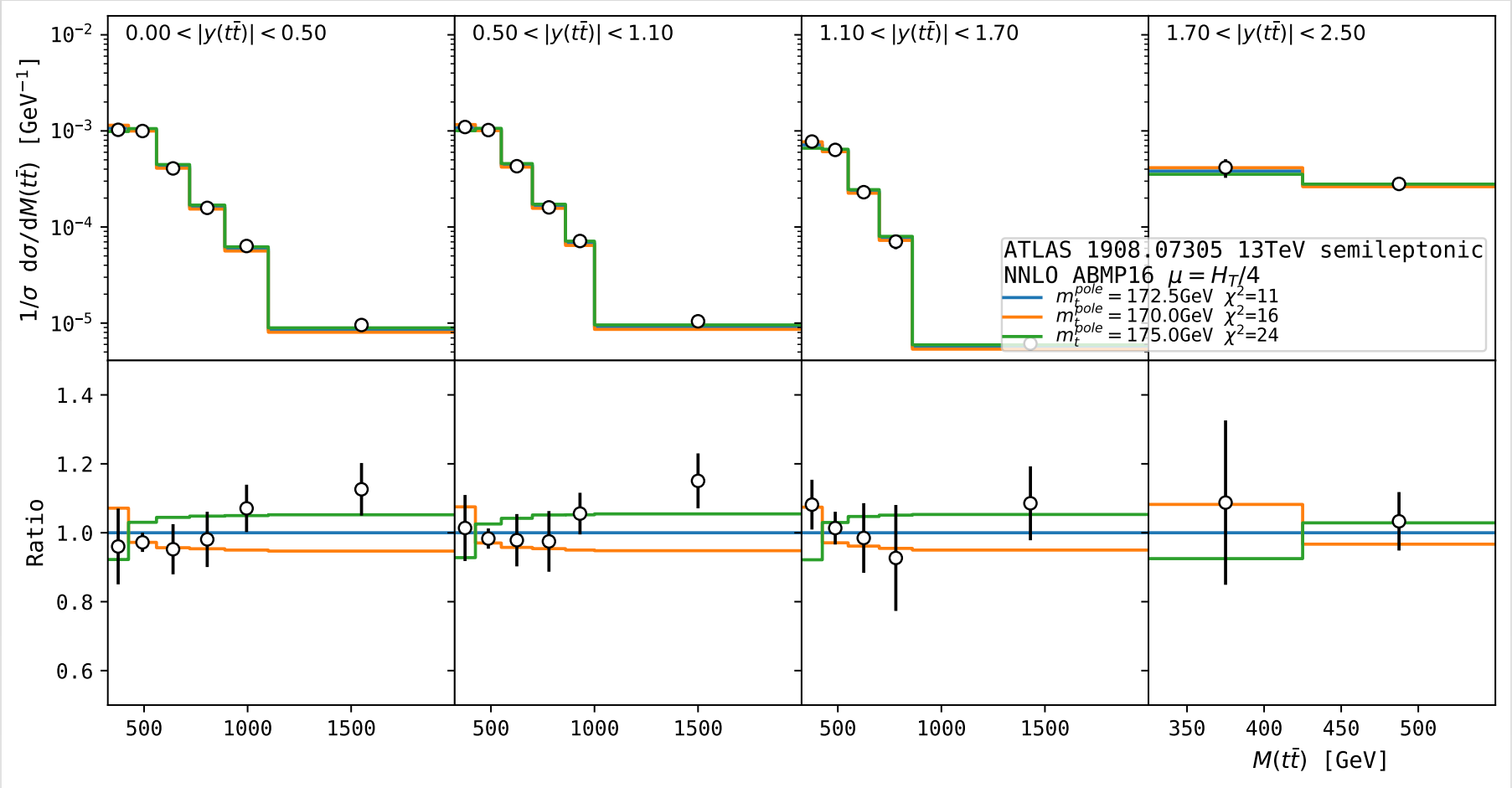

- Using ABMP16, $\mu_r = \mu_f = H_T/4$
- Reported $\chi^2$ values with PDF uncertainties
- Large sensitivity to $m_t^{\text{pole}}$ in the first $M(t\bar{t})$ bin (and even in other $M(t\bar{t})$ bins, thanks to cross-section normalisation)
- Fluctuations of theory predictions are $\lesssim 1\%$ and covered by the assigned uncertainty of 1%

# ATLAS 1908.07305 vs NNLO predictions using different PDFs




- Fixed $m_t^{\text{pole}} = 172.5$ GeV, $\mu_r = \mu_f = H_T/4$
- Reported $\chi^2$ values with (and without) PDF uncertainties
- All PDF sets describe data equally well
- $\chi^2/\text{dof} < 1$ indicating possible overestimate of experimental uncertainties (additionally, the data covariance matrix is not singular, i.e. $\det(\text{cov}) \neq 0$: we suspect this is related to numerical inaccuracy of data stored in Hepdata. This affects estimates of correlated uncertainties. Same issue in the $\sqrt{s} = 8$TeV ATLAS analysis [arXiv:1607.07281].

# ATLAS 1908.07305 vs NNLO predictions with ABMP16 and different $m_t^{\text{pole}}$




- Using ABMP16, $\mu_r = \mu_f = H_T/4$
- Reported $\chi^2$ values with PDF uncertainties
- Large sensitivity to $m_t^{\text{pole}}$ in the first $M(t\bar{t})$ bin (and even in other $M(t\bar{t})$ bins, thanks to cross section normalisation). The sensitivity does not increase with rapidity due to cross-section normalization.

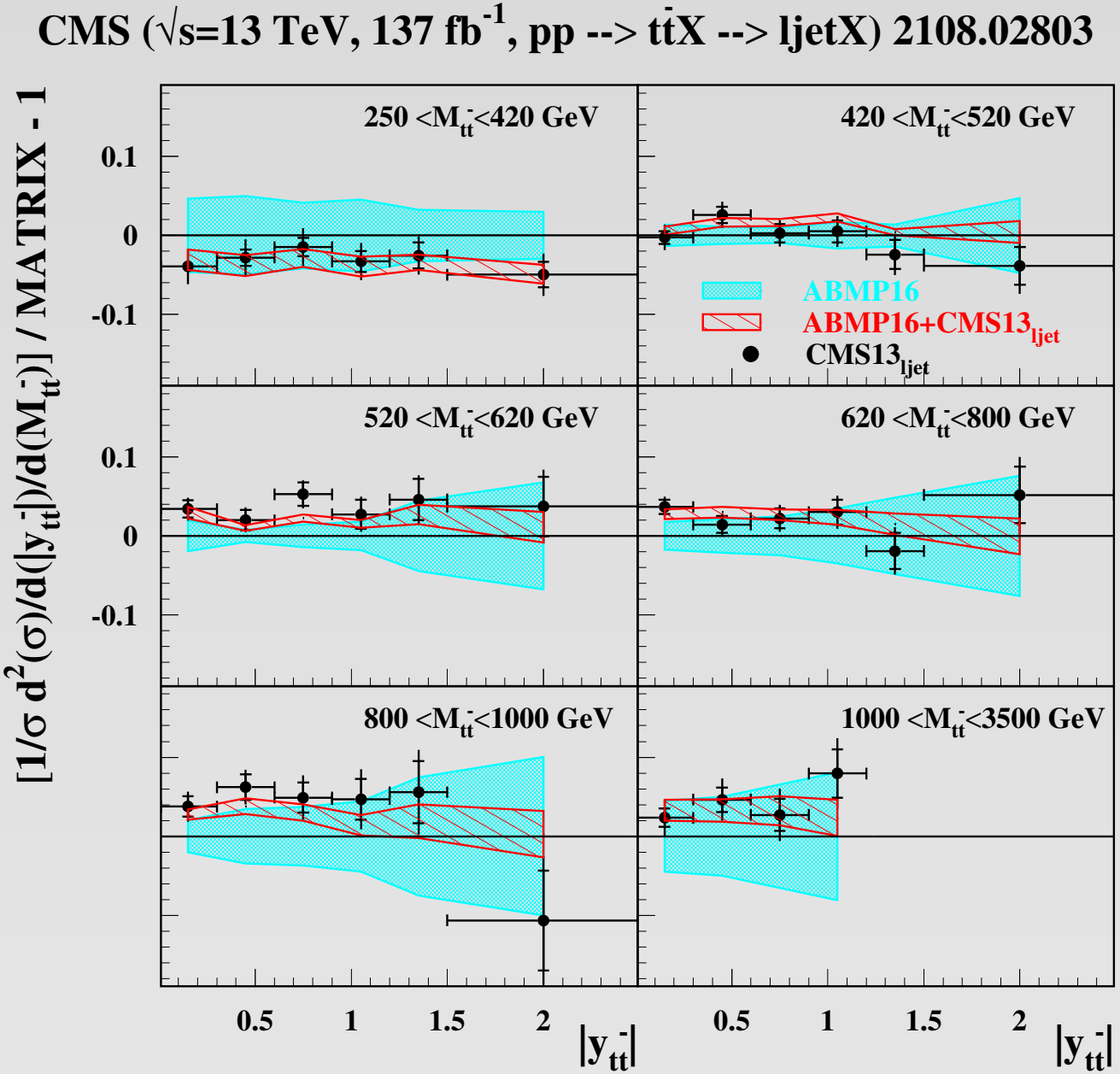
# Data vs NNLO predictions using different PDFs at fixed $m_t$: summary

| PDF | $t\bar{t}$ data in PDF fit | $\chi^2$/NDP (all data) | |
|---|---|---|---|
| | | w/ PDF unc. | w/o PDF unc. |
| ABMP16 | only total $\sigma(t\bar{t}+X)$ | 56/78 | 61/78 |
| CT18 | total and diff. $\sigma(t\bar{t}+X)$ | 80/78 | 252/78 |
| MSHT20 | total and diff. $\sigma(t\bar{t}+X)$ | 92/78 | 196/78 |
| NNPDF4.0 | total and diff. $\sigma(t\bar{t}+X)$ | 104/78 | 139/78 |

- None of these PDF fits includes the datasets (1)-(4) that we considered
  - NNPDF4.0 include single-differential data from CMS studies [1803.08856, 1811.06625], using 2016 events, with partial overlap with the events used in the independent CMS Run 2 analyses that we considered. Additionally they include the double-differential Run 1 CMS dataset [arXiv:1703.01630].
- To test the robustness of their fit and refine/improve the shape of their central PDFs, it will be important that the global PDF fitters start including these double-differential data.

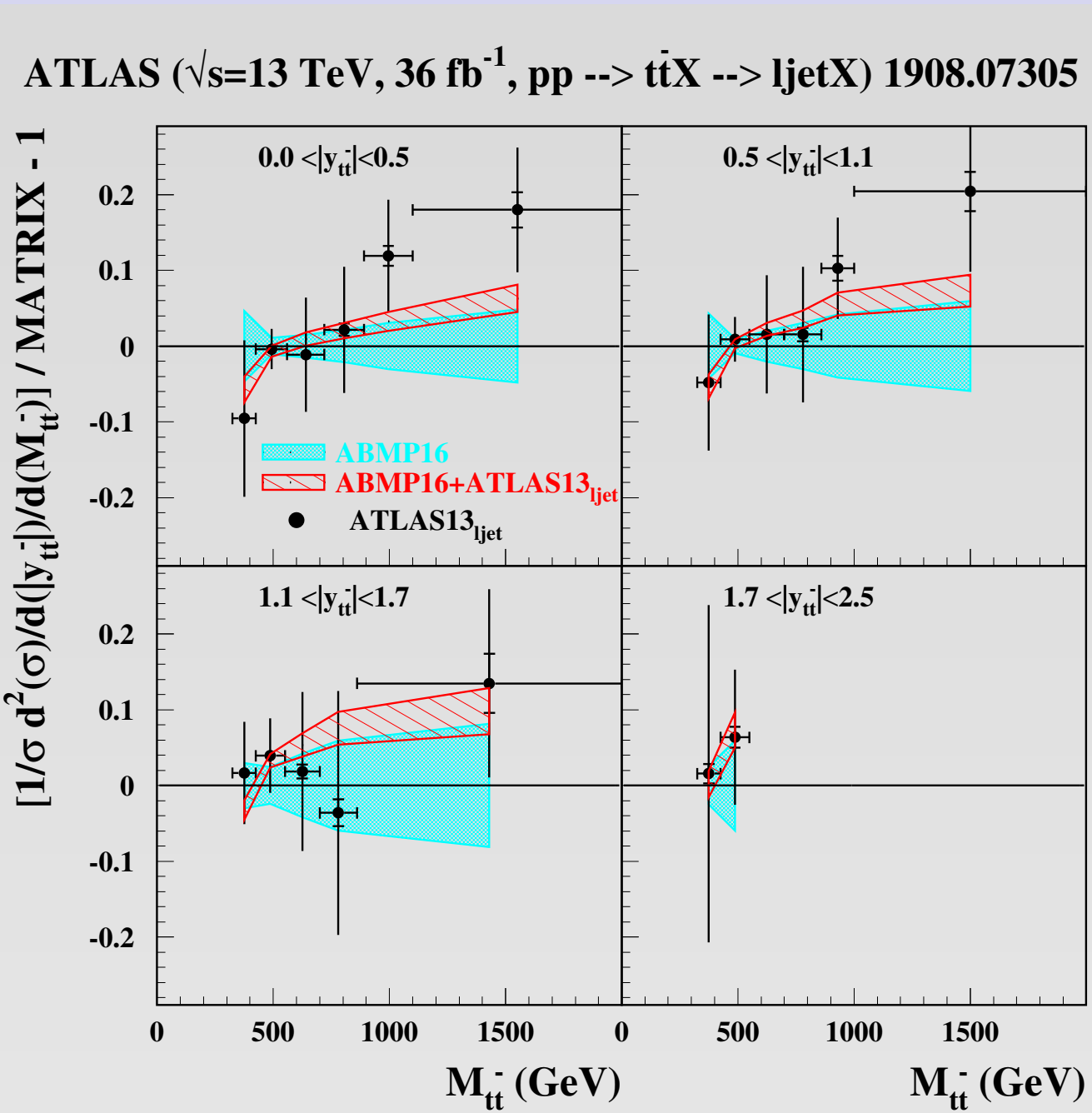
  $\rightarrow$ We have done this in the ABMPtt fit, as detailed step-by-step in the following.

# Pulls of CMS TOP-20-001 data with respect to ABMP predictions



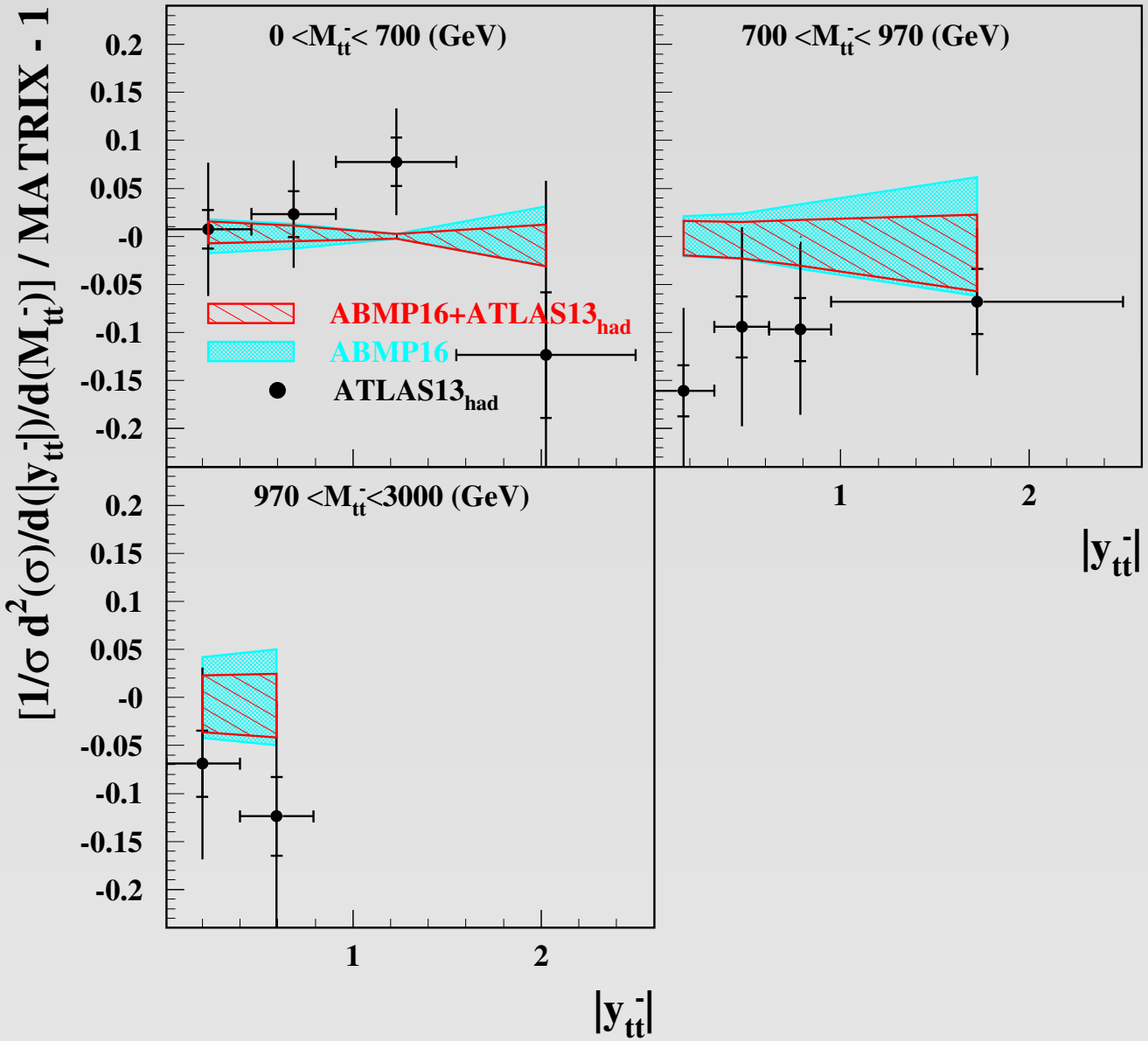

- ABMP PDF fit variant incorporating this specific dataset, w.r.t. already available ABMP16 PDF fit without it

# Pulls of ATLAS 1908.07305 data with respect to ABMP predictions




- ABMP PDF fit variant incorporating this specific dataset, w.r.t. already available ABMP16 PDF fit without it.
- ATLAS $\ell + j$ data tend to be larger than central theory predictions at large $M(t\bar{t}) \sim 1500$ GeV. But the data uncertainties are still large.
- ATLAS $\ell + j$ analysis with better statistics wanted.

# Pulls of ATLAS 2006.09274 data with respect to ABMP predictions




- ABMP PDF fit variant incorporating this specific dataset, w.r.t. already available ABMP16 PDF fit without it
- ATLAS hadronic data smaller than central theory predictions at large $M(t\bar{t})$.
- ATLAS $(\ell + j)$ data larger than central theory predictions at large $M(t\bar{t})$.

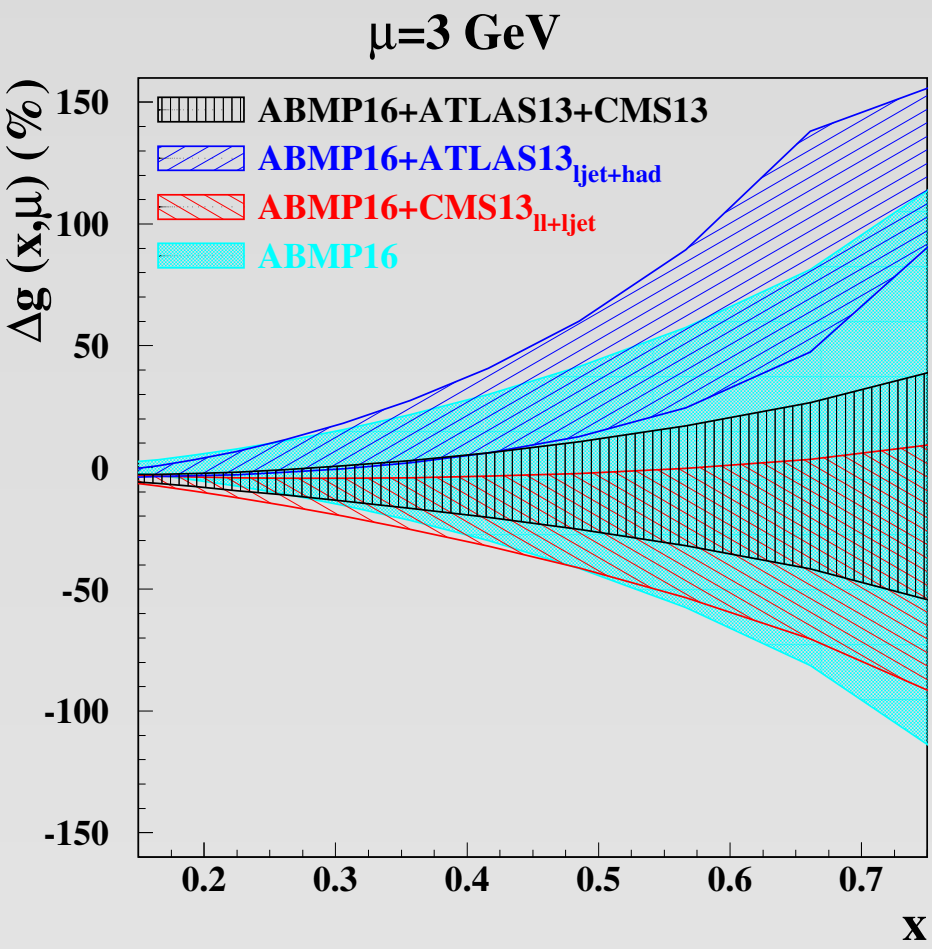
# Partial $\chi^2$ for variants of the new ABMP analysis including double-differential $t\bar{t} + X$ data at 13 TeV

| Experiment | Dataset | $\sqrt{s}$ (TeV) | *NDP* | $\chi^2$ | | |
|---|---|---|---|---|---|---|
| | | | | I | II | III |
| ATLAS | ATLAS13$_{ljet}$ | 13 | 19 | 34.0 | 28.2 | – |
| | ATLAS13$_{had}$ | 13 | 10 | 11.9 | 11.6 | – |
| CMS | CMS13$_{ll}$ | 13 | 15 | 20.7 | – | 19.6 |
| | CMS13$_{ljet}$ | 13 | 34 | 44.3 | – | 42.4 |

**Table:** The values of $\chi^2$ obtained for various $t\bar{t} + X$ datasets included in the present analysis (column I: both ATLAS and CMS datasets; column II: only ATLAS ones; column III: only CMS ones).

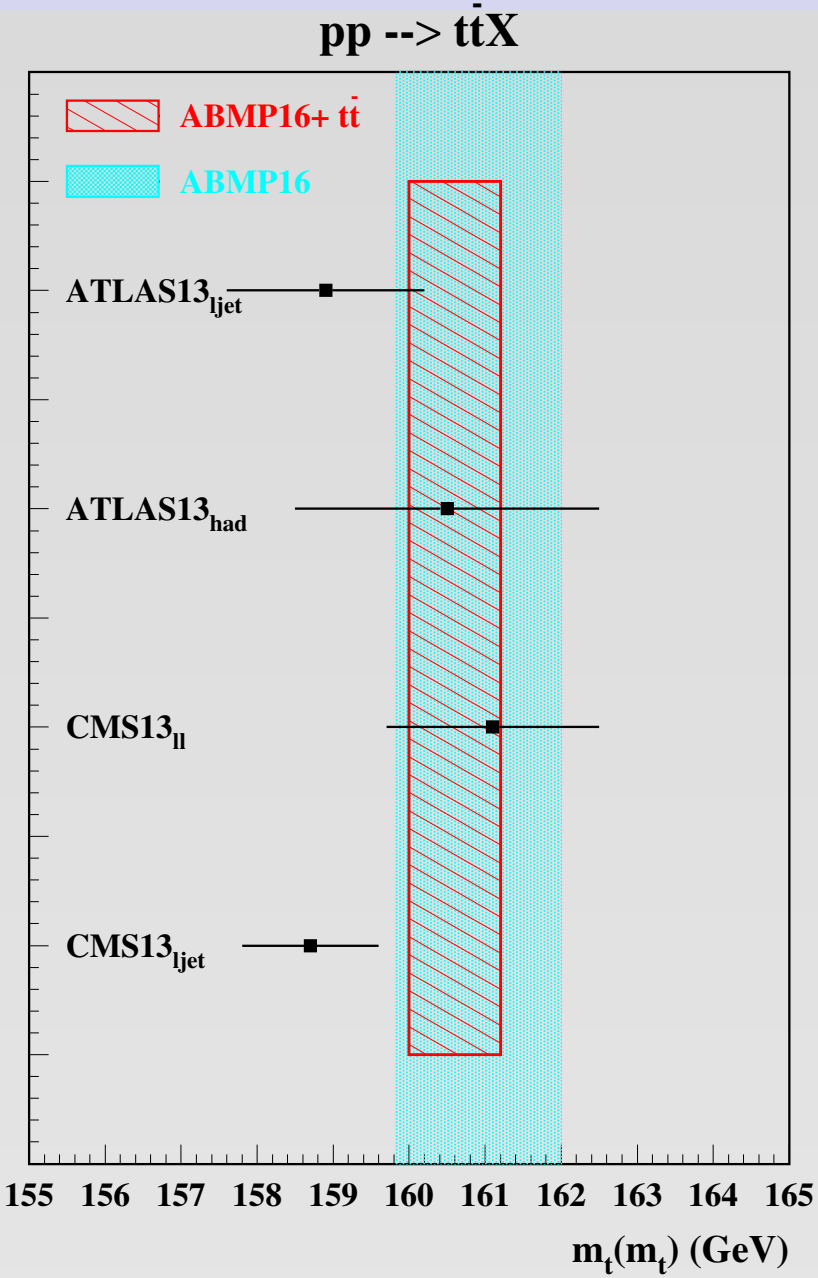
- In comparison to the fit including both CMS datasets (III), the $\chi^2$ slightly deteriorates when including also the datasets of the ATLAS analyses (I), but is still compatible within statistical uncertainties.
- In comparison to the fit including both ATLAS datasets (II), the $\chi^2$ for the all-hadronic dataset remains compatible within statistical uncertainties when including also the datasets of the CMS analysis (I). Viceversa the $\chi^2$ for the ATLAS $\ell + j$ dataset worsens. $\Rightarrow$ Tension of the ATLAS $\ell + j$ dataset with all other datasets

# Extracted $g(x)$ in variants of the ABMP fit




- $g(x)$ at the starting scale $\mu = 3$ GeV.
- $g(x)$ in the new ABMP fit variants compatible with ABMP16 previous fit.
- uncertainties on $g(x)$ decreased by a factor $\sim 2$ w.r.t. ABMP16 previous fit.
- ATLAS and CMS data points towards opposite trends of $g(x)$ at large $x$. ATLAS prefers a larger $g(x)$, related to the fact that ATLAS $(\ell + j)$ data tend to be larger than theory predictions at large $M(t\bar{t}) \sim 1500$ GeV. Note that this trend is not visible for ATLAS hadronic data.
- fit including both ATLAS and CMS data dominated by the CMS $\ell + j$ differential data.
- Observe that new $m_t(m_t)$ and $\alpha_s(M_Z)$ values are extracted simultaneously. In particular, the smaller $g(x)$ of the "global" fit is accompanied by a smaller $m_t(m_t)$ value (see next slides).

# Extracted values of $m_t(m_t)$ in variants of the ABMP fit

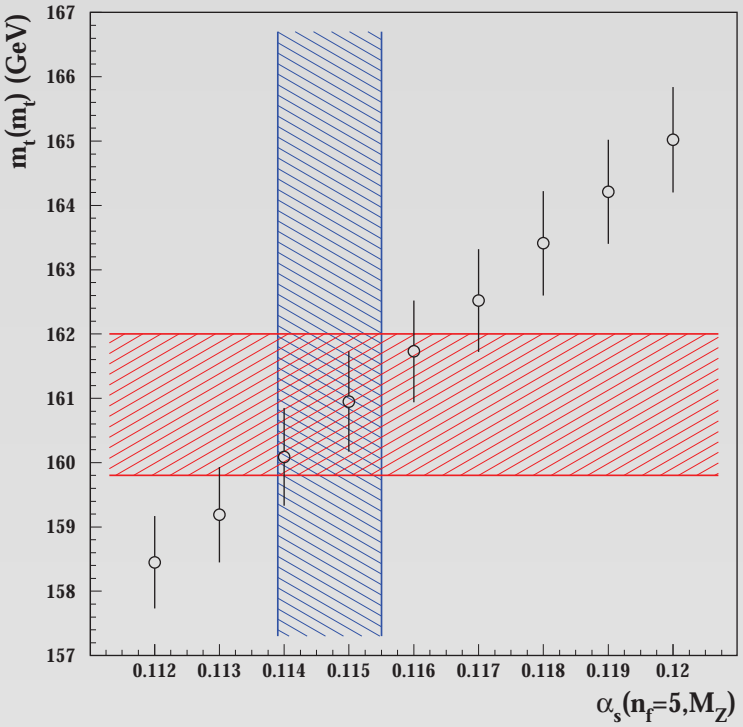



- Legenda:
  Black: ABMP PDF fit variant incorporating a single specific dataset,
  light-blue: previous ABMP16 PDF fit,
  red: new ABMP PDF fit, incorporating all $t\bar{t} + X$ double-differential data at 13 TeV.

- Good compatibility of $m_t(m_t)$ extracted in the different variants of the fit.
- ATLAS hadronic data are too uncertain to play a constraining role on $m_t(m_t)$.
- New central value of $m_t(m_t) = 160.6$ GeV slightly smaller than 160.9 GeV obtained in the previous ABMP16 fit, due to effect of the ATLAS and CMS $\ell + j$ differential data.
- Including all 13 TeV $t\bar{t} + X$ double-differential data allow to decrease by a factor 2 the uncertainty band on $m_t(m_t)$, varying from 1.1 GeV to 0.6 GeV.
- Observe that new PDFs and $\alpha_s(M_Z)$ values are extracted simultaneously.

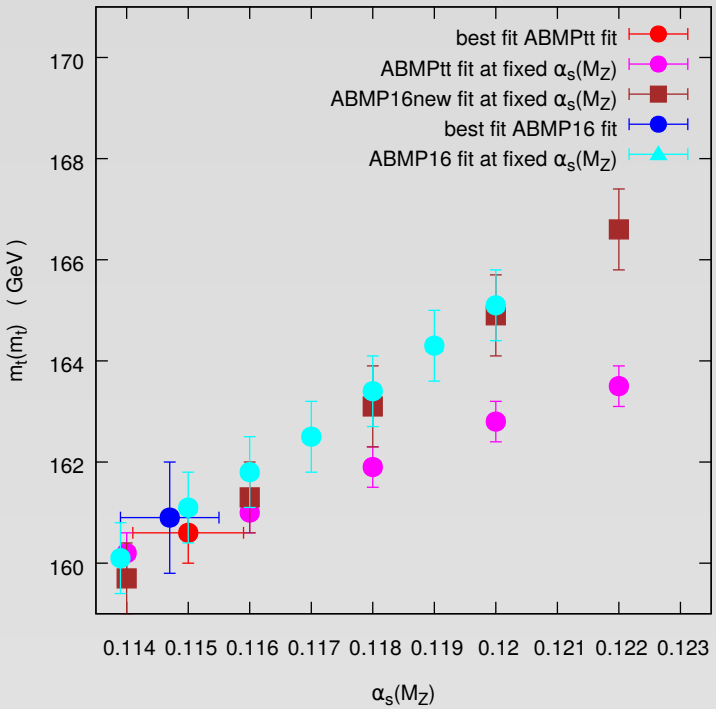
# Correlation between $m_t(m_t)$ and $\alpha_s(M_Z)$ in the old ABMP16 fit

from ABMP16 fit




- Correlations between PDF $g(x)$, $\alpha_s(M_Z)$ and $m_t(m_t)$ follows from the factorization theorem.
- Fit of $m_t(m_t)$ at fixed $\alpha_s(M_Z)$ shows positive correlation between $\alpha_s(M_Z)$ value and $m_t(m_t)$.

# Correlation between $m_t(m_t)$ and $\alpha_s(M_Z)$ in the new ABMP fit (vs. old ABMP16)

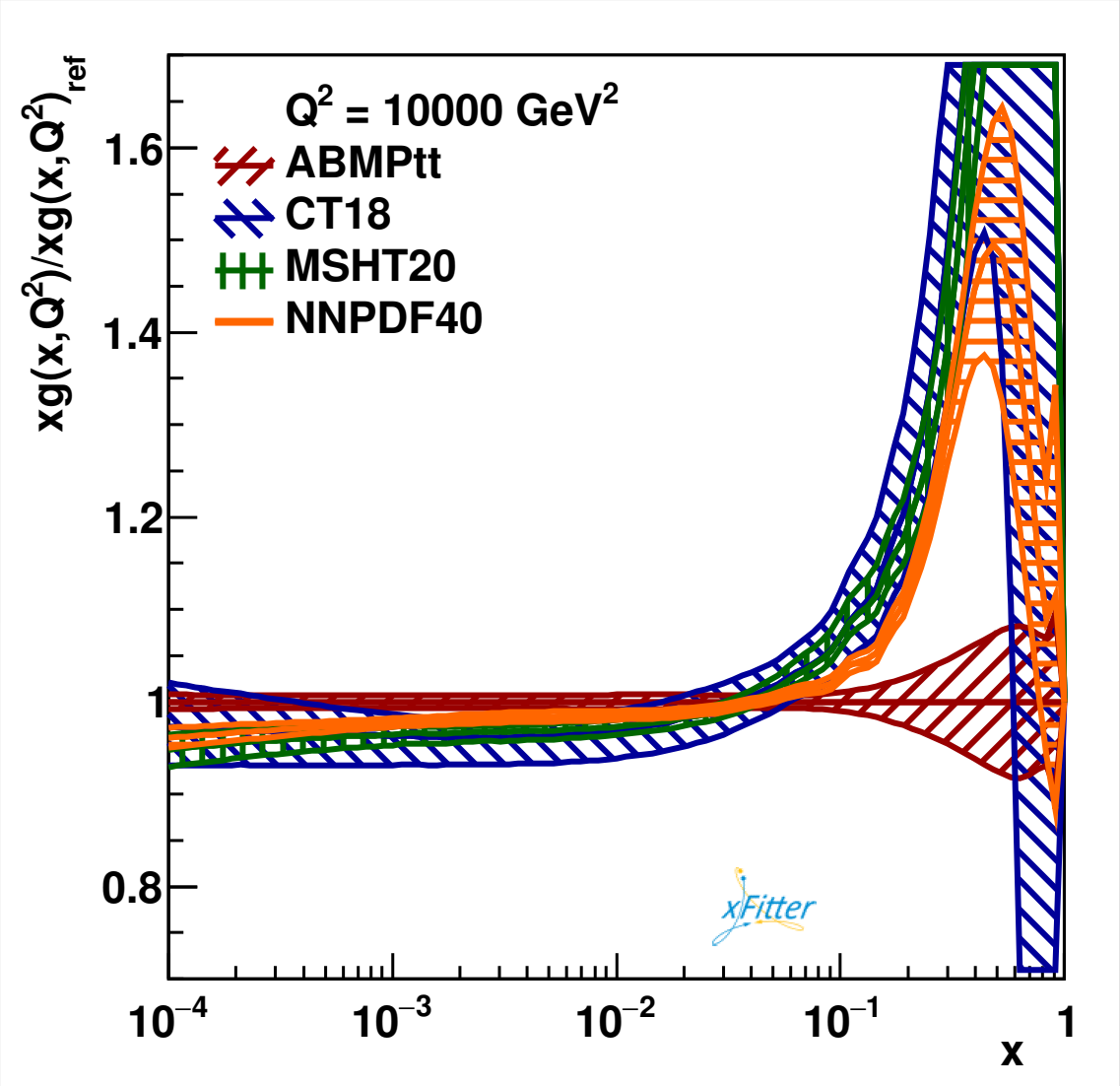



| $\alpha_s(M_Z, N_f = 5)$ | | $m_t(m_t)$ (GeV) |
|---|---|---|
| Fitted | 0.1150(9) | 160.6(6) |
| $\alpha_s(M_z)$ fixed | 0.114 | 160.2(4) |
| | 0.116 | 161.0(4) |
| | 0.118 | 161.9(4) |
| | 0.120 | 162.8(4) |
| | 0.122 | 163.5(4) |

**Table:** The values of $m_t(m_t)$ obtained with different values of $\alpha_s$ in the new ABMP fit.

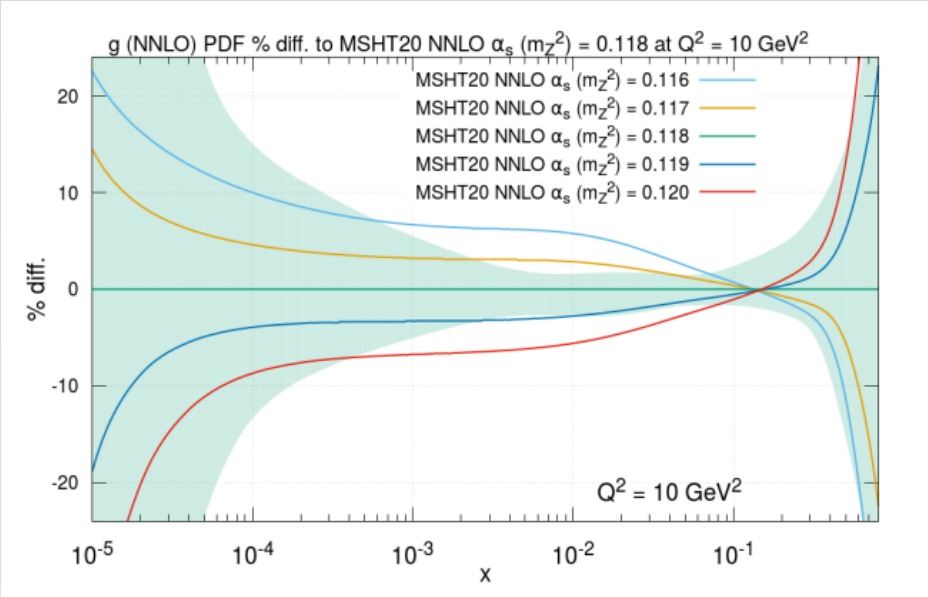
- Correlations between PDF $g(x)$, $\alpha_s(M_Z)$ and $m_t(m_t)$ follows from the factorization theorem.
- Fit of $m_t(m_t)$ at fixed $\alpha_s(M_Z)$ shows positive correlation between $\alpha_s(M_Z)$ value and $m_t(m_t)$.
- When including the $t\bar{t} + X$ differential data, the correlation coefficient decreases w.r.t. to the ABMP16 analysis, whereas the best-fit $\alpha_s(M_z)$ value remains approximately the same.
- With improved precision of data on single-top production in the $t$-channel, the impact of $\alpha_s(M_Z)$ on the $m_t$ determination could be further leveled.

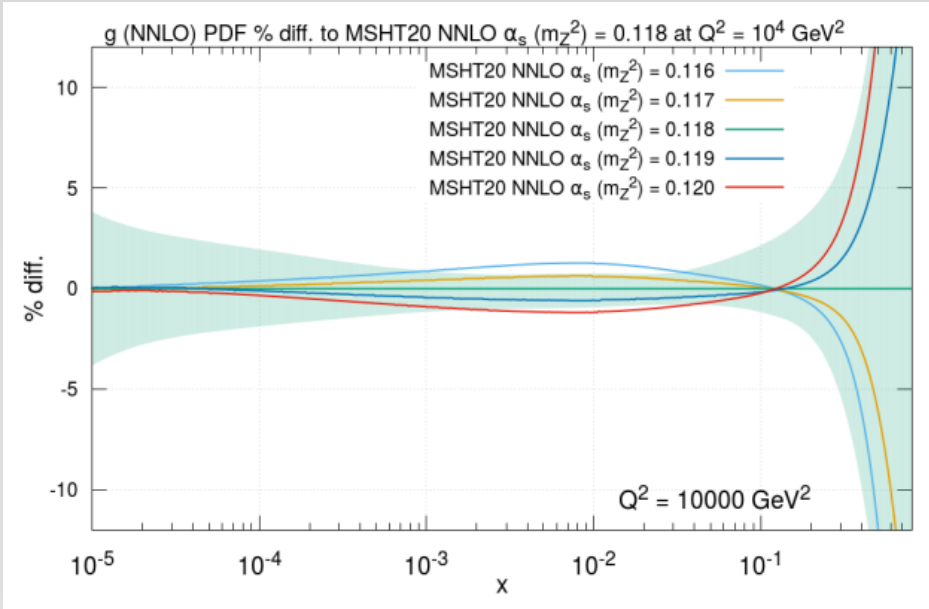
# Extracted $g(x)$ in comparison with global PDF fits




$*$ Large differences at large $x$: Besides the effect of the $t\bar{t} + X$ data, these are due to different $\alpha_s(M_Z)$ treatment, heavy-flavour DIS scheme, etc.

# PDF fits using as input different $\alpha_s(M_Z)$ values

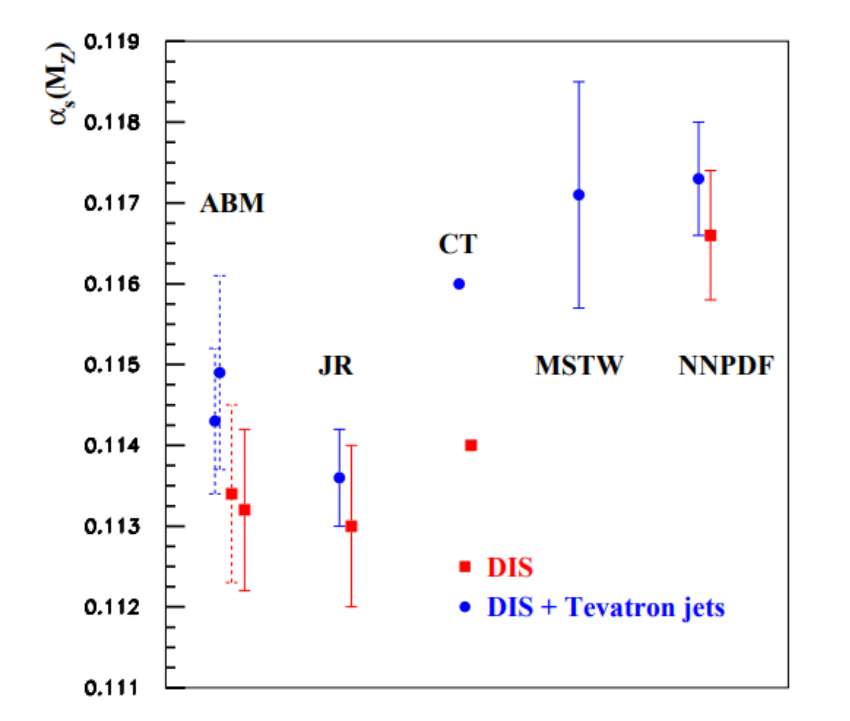



*from T. Cridge et al., MSHT20, arXiv:2106.10289*

$*$ Different $\alpha_s(M_Z)$ values as input play a large impact on the gluon at all $x$ values, especially at small $Q^2$

$\Rightarrow$ If $\alpha_s(M_Z)$ in MSHT20 would be similar to the one in ABMP16, the $g(x)$ would also look more similar to the latter (at least in the region covered by $t\bar{t}$ data).



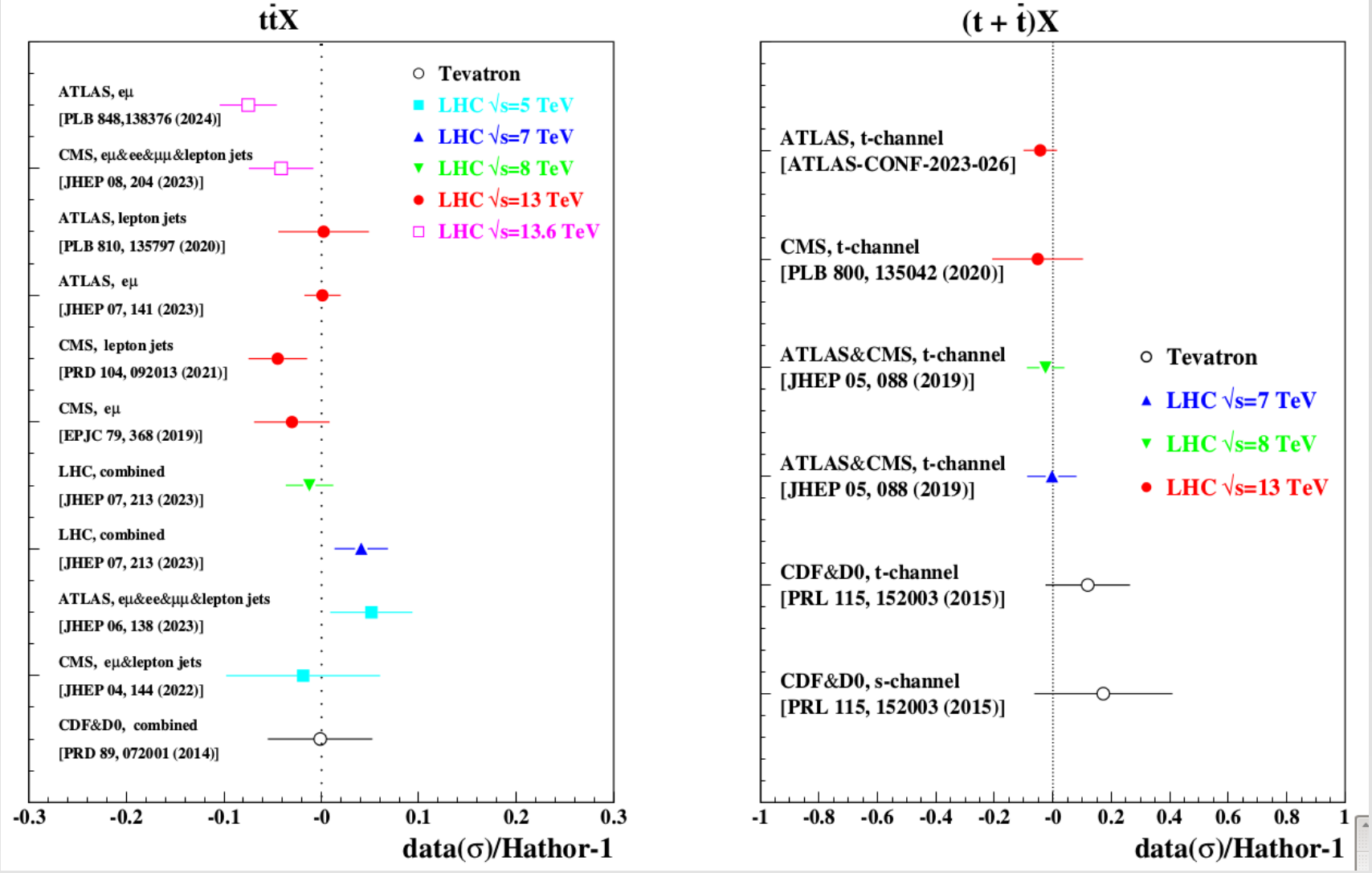

*from S. Alekhin et al., PRD 89 (2014) 054028*

∗ Differences in $\alpha_s(M_Z)$ between ABM and other PDF+$\alpha_s(M_Z)$ sets date back to 15 years...., in relation to:

- $F_L$ treatment
- Effects of including/not including jet data from hadronic collisions (Tevatron and LHC)
- Effects of including/not including higher-twist corrections: an analysis without the latter brings back $\alpha_s(M_Z)$ at large values an analysis without the latter but with cuts on $Q^2 > 10$ GeV$^2$, $W^2 > 12.5$ GeV$^2$ lead to low $\alpha_s(M_Z)$ values.
- Other power corections to DIS: target mass corrections, due to finite nucleon mass (see talk by O. Zenaiev)

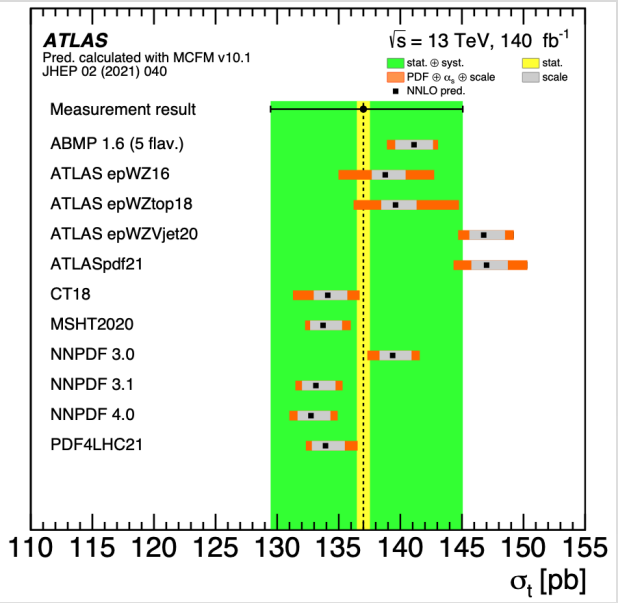
∗ Almost no impact of $t\bar{t} + X$ data on $\alpha_s(M_Z)$: we would need to analyze $t\bar{t}j$ data.

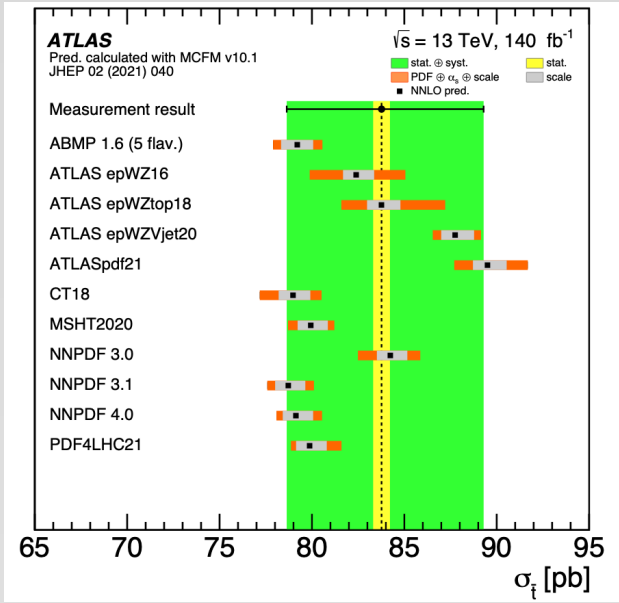
# ABMPtt fit: agreement with total inclusive cross-section data

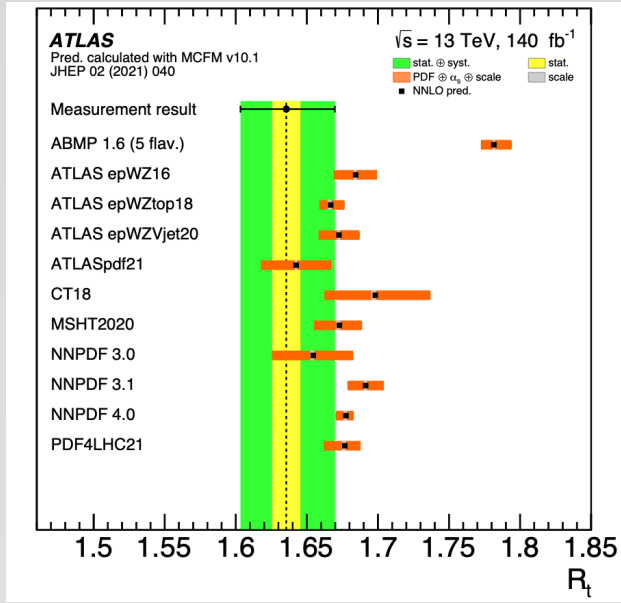



Good agreement with both $t\bar{t} + X$ and $(t + X) + (\bar{t} + X)$ data (included in fit)

# ATLAS $t + X$, $\bar{t} + X$ data and their ratio vs NNLO theory predictions




*from ATLAS collaboration*

$(t + X)$ mainly probes $u$ distribution, $(\bar{t} + X)$ mainly probes $d$ distribution.

Still to be investigated:

- Why is ABMP16 performing quite badly w.r.t. the ATLAS ratio ?
  - Are the input PDFs, $\alpha_s(M_Z)$, $m_t$ used in the ATLAS computations fully consistently ?
  - Are the uncertainties on $\sigma(t + X)/\sigma(\bar{t} + X)$ computed correctly ?
  - Are systematic uncertainties well under control ?
  - Are there issues with the $u$ and $d$-quark distributions from ABMP16 fit (related to e.g. target mass corrections) ?
- What is happening when using as input ABMPtt ?

# CMS $\sigma(t+X)/\sigma(\bar{t}+X)$ data vs. NNLO theory predictions

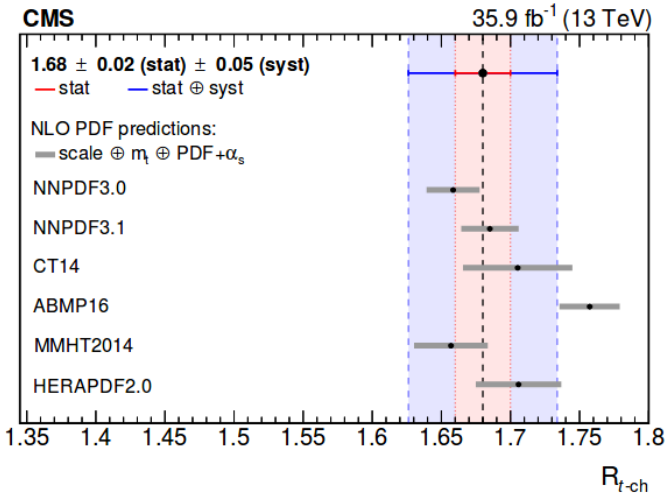
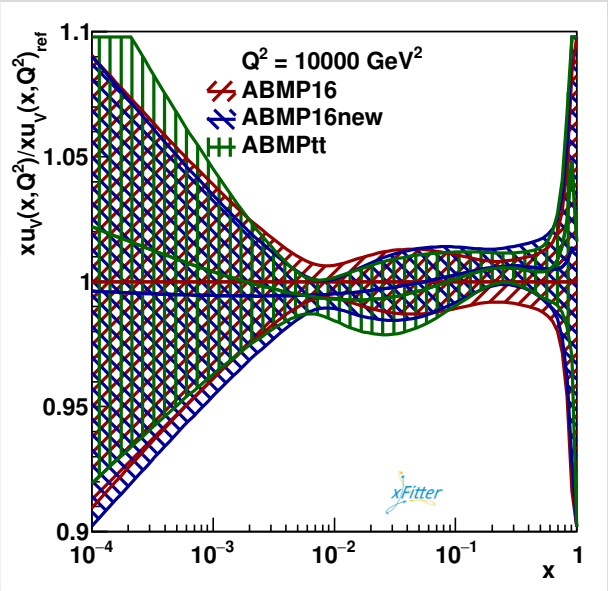


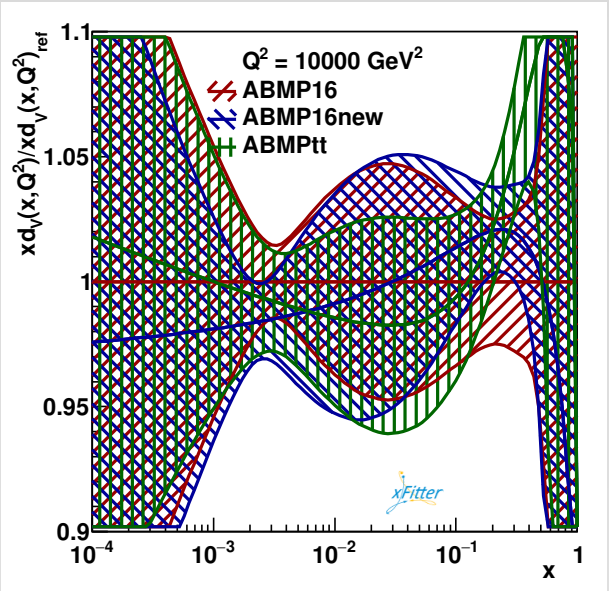

Figure 8: Comparison of the measured $R_{t\text{-ch}}$ (central dashed line) with the NLO predictions from different PDF sets, provided by LHAPDF 6.2.1 [65]: NNPDF3.0 [39], NNPDF3.1 [66], CT14 [67], ABMP16 [68, 69], MMHT2014 [70], HERAPDF2.0 [71]. The HATHOR 5FS calculation is used with the nominal values for the top quark pole mass and $\alpha_S$ set to the best values of each PDF set. The uncertainty bars for the different PDF sets include the uncertainty due to the factorization and renormalization scales, the uncertainty in the top quark pole mass, and the combined internal PDF+$\alpha_S$ uncertainty. For the measurement, the statistical and total uncertainties are indicated individually by the inner and outer uncertainty bars.

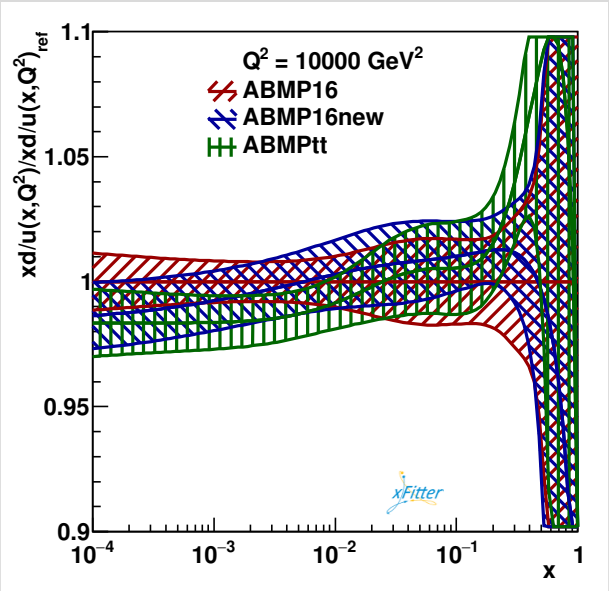
*from CMS collaboration, [arXiv:1812.10514] PLB 800 (2020) 135042*

- CMS ratio shifted towards higher values with respect to the ATLAS case.
  $\Rightarrow$ The disagreement with ABMP16 is less important, but still present.
- Smaller integrated luminosity.

# The ABMP16 and ABMPtt $u_v$, $d_v$-quark distributions and the $d/u$ ratio




- The $u_v$ distribution of ABMPtt fully compatible with the one from ABMP16
- The $d_v$ distribution of ABMPtt larger than ABMP16 at large $x$.
- This difference has some implication for the $d/u$ ratio:
  - in the "right" direction, but not enough to solve the discrepancy with ATLAS $\sigma(t+X)/\sigma(\bar{t}+X)$ data.
  - useful to compare to old data at smaller $\sqrt{S}$:
    past work by [S. Alekhin et al. Phys.Rev.D 94 (2016) 11, 114038] shows that ABM12 $R_t$ was well compatible with ATLAS data at $\sqrt{S}$= 7 TeV and CMS data at $\sqrt{S}$= 8 TeV

# Photons in the proton and the preliminary ABMPttgam fit

- At increasingly high orders, it is increasingly important to know the photon content of the proton and to include photon effects in PDF fits.
- Many PDF collaborations have already studied the role of the photon distribution in the proton. This has also been done in the case of the first aN3LO QCD PDF fits.
- In the following we present work in progress in the framework of the ABMP collaboration, which lead to the ABMPttgam fit. All results are preliminary and work is in progress.

# Photon in PDF fits

## Two approaches have been considered so far:

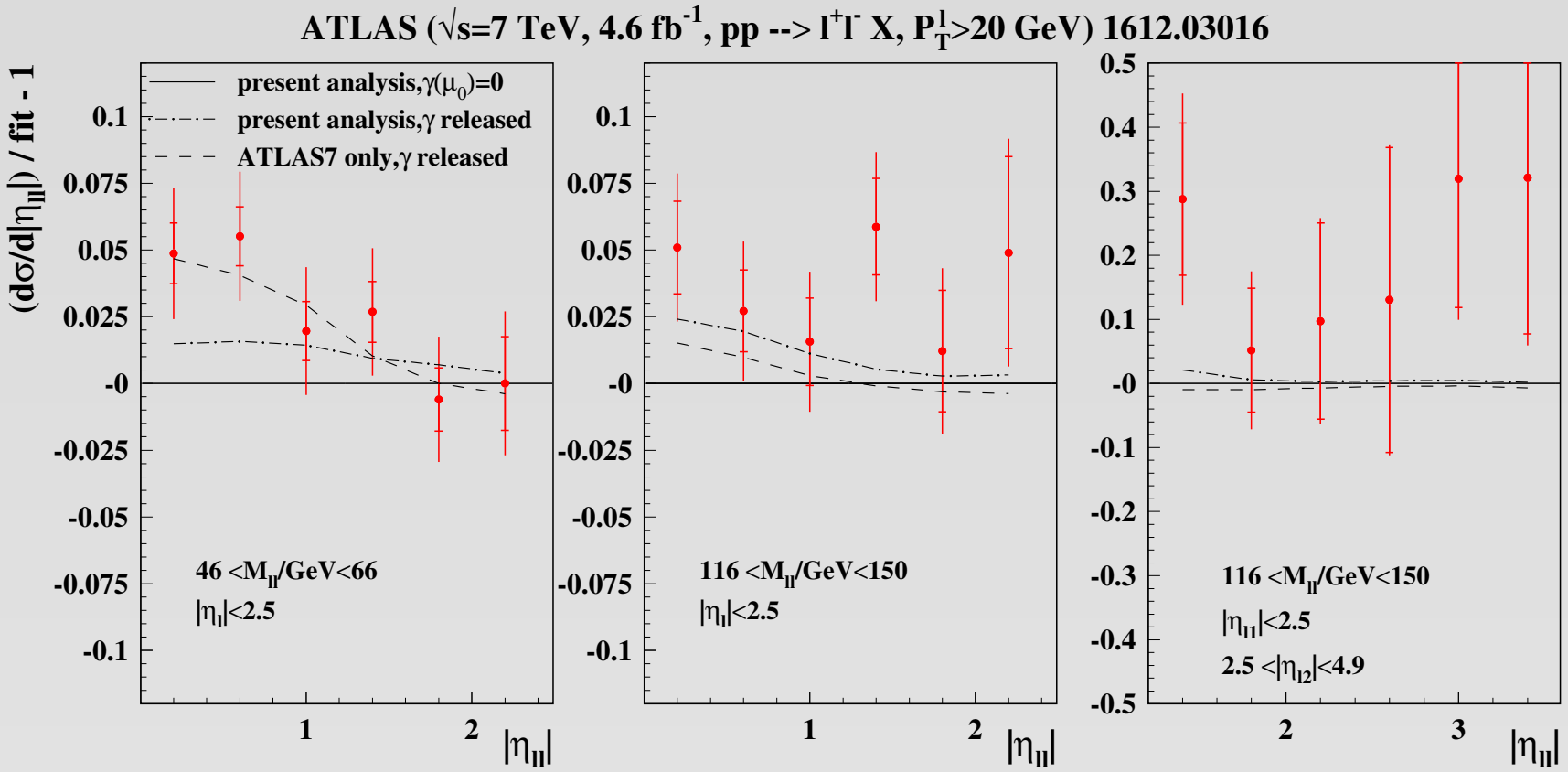
- Photon according to the LUXQed approach
  - implemented in most modern PDF fits (MSHT, NNPDF), basically following the guidelines in the LUXQed papers, with some variations. Photon distributions are computed by first principles, however relying on assumptions on the proton structure functions down to low scales $Q^2$ and/or low hadronic invariant mass $W^2$ and on elastic contributions

- Photon treated similarly to partons
  - photon distribution parameterized at a low scale and then evolved
  - initial condition fixed at such a scale (difficult to establish, because the available experimental data are hardly constraining photons at low scales).
  - photon evolves with standard evolution equations (resummation effects included)
  - approach used in “old” PDFs (i.e. PDF fits before the LUXQed approach was introduced) and by the ABMP collaboration in its new fit ABMPgam

## Selected preliminary results

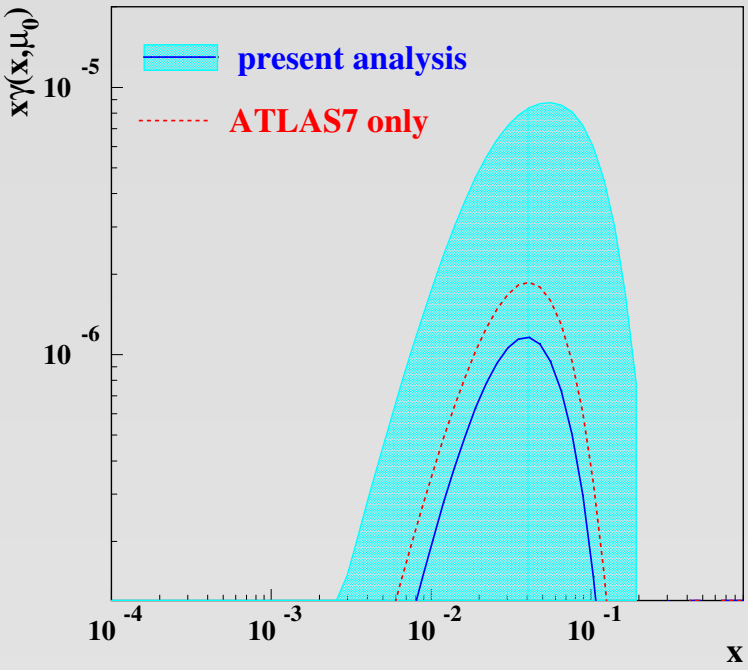
- Effects of ATLAS and CMS non-resonant dilepton data in shaping the initial condition for the photon distribution are presented in the following.
- Performances of the ABMPttgam fit with respect to top-quark data are also presented in the following, in comparison to the ABMPtt case.

# ABMPttgam fit variants with different initial conditions for the photon distribution vs. ATLAS non-resonant dilepton data at $\sqrt{S}$ = 7 TeV




- $\gamma\gamma \to \ell^+\ell^-$ process included in the theory predictions compared to the data.
- The ATLAS data at low ($M_{\ell^+\ell^-}$, $y_{\ell^+\ell^-}$) are in better agreement with central predictions obtained by assuming the presence of a photon component already at the initial starting scale $\mu_0$.
- They are compatible with central predictions with $\gamma(\mu_0) = 0$ within $\sim 2.3\sigma$.

# The “bump” in photon distribution

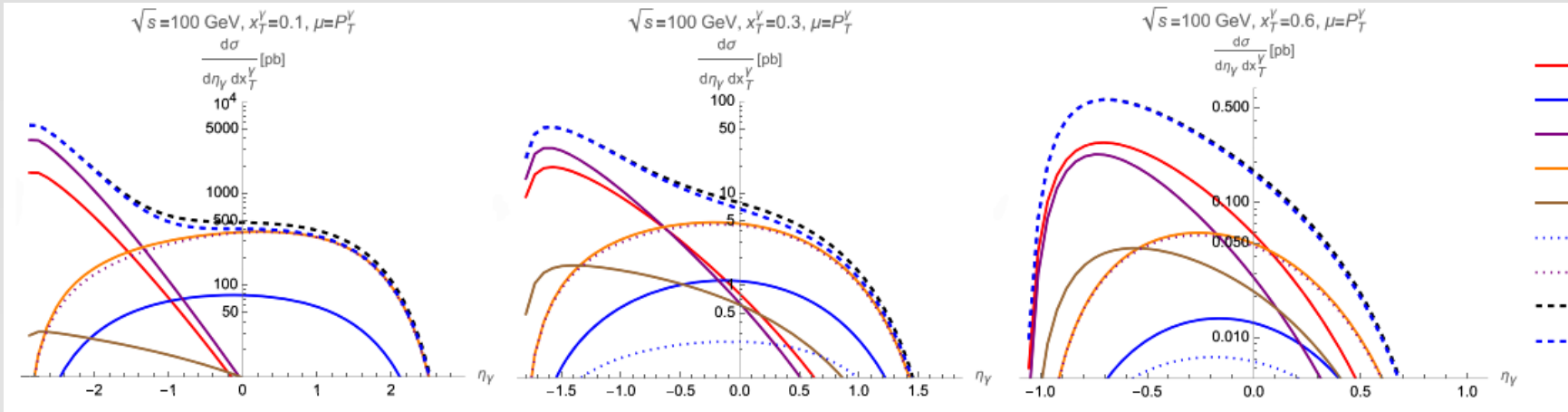



- ATLAS dilepton data favour the presence of a bump in $\gamma$ distribution peaked at $x \sim 5 \cdot 10^{-2}$ at the initial scale $\mu_0$. This bump is present both in the variant of the fit without other dilepton data, and, attenuated, in the global fit where the CMS dilepton and other data are also included.
- However ATLAS data, referring to a scale $\mathcal{O}(100\ \text{GeV})$, are not enough to impose a finite constraint on the $\gamma$ at scale $\mu_0$. Considering that the CMS data are instead well compatible with an initial condition $\gamma(\mu_0) = 0$, and that data capable of directly constraining photons at such low scales are missing, the uncertainty on the photon in the region of the bump is extremely large.
- We can conclude that the photon in the global ABMPgam fit is compatible with $\gamma(\mu_0) = 0$ at all $x$ values, i.e. with the hypothesis of being generated fully perturbatively. There is no need for an intrinsic photon component at low scales.
- This analysis should be repeated adding other data (e.g. the 13 TeV ones).

# Photon distributions at the EIC

Processes useful to constrain $\gamma$ distributions at "lower" scales (as compared to the LHC scales) could be measured at the EIC:

- $ep \to e\gamma + X$ [Pisano, Gluck, Muckherjee et al. 2002 - 2004]
- $ep \to \gamma + X$ [Vogelsang et al. PRD 110 (2024) 014041]:
  - analytical formulas for partonic cross-sections up to NLO
  - $\gamma$ PDF contributions and $q, g \to \gamma$ fragmentation contributions are enhanced/suppressed in different kinematic regions.
  - $\gamma$ fragmentation contribution suppressed by photon isolation cuts.



# Preliminary: the ABMPttgam fit

| Experiment | Dataset | $\sqrt{s}$ (TeV) | *NDP* | $\chi^2$ | |
|---|---|---|---|---|---|
| | | | | ABMPtt | ABMPttgam |
| ATLAS | $ATLAS13_{ljet}$ | 13 | 19 | 34.0 | 32.9 |
| | $ATLAS13_{had}$ | 13 | 10 | 11.9 | 12.4 |
| CMS | $CMS13_{ll}$ | 13 | 15 | 20.7 | 22.7 |
| | $CMS13_{ljet}$ | 13 | 34 | 44.3 | 36.3 |

**Table:** The values of $\chi^2$ obtained for various $t\bar{t}+X$ datasets included in the present ABMPtt and ABMPttgam analyses

$*$ The fit including photons (ABMPttgam) allows to accomodate the $t\bar{t}+X$ normalized double-differential ATLAS data with $\chi^2$ values within statistical uncertainties of the case of the fit without photons (ABMPtt).

$*$ For the CMS $\ell+j$ analysis the $\chi^2$ turns out to improve in the fit including photons.

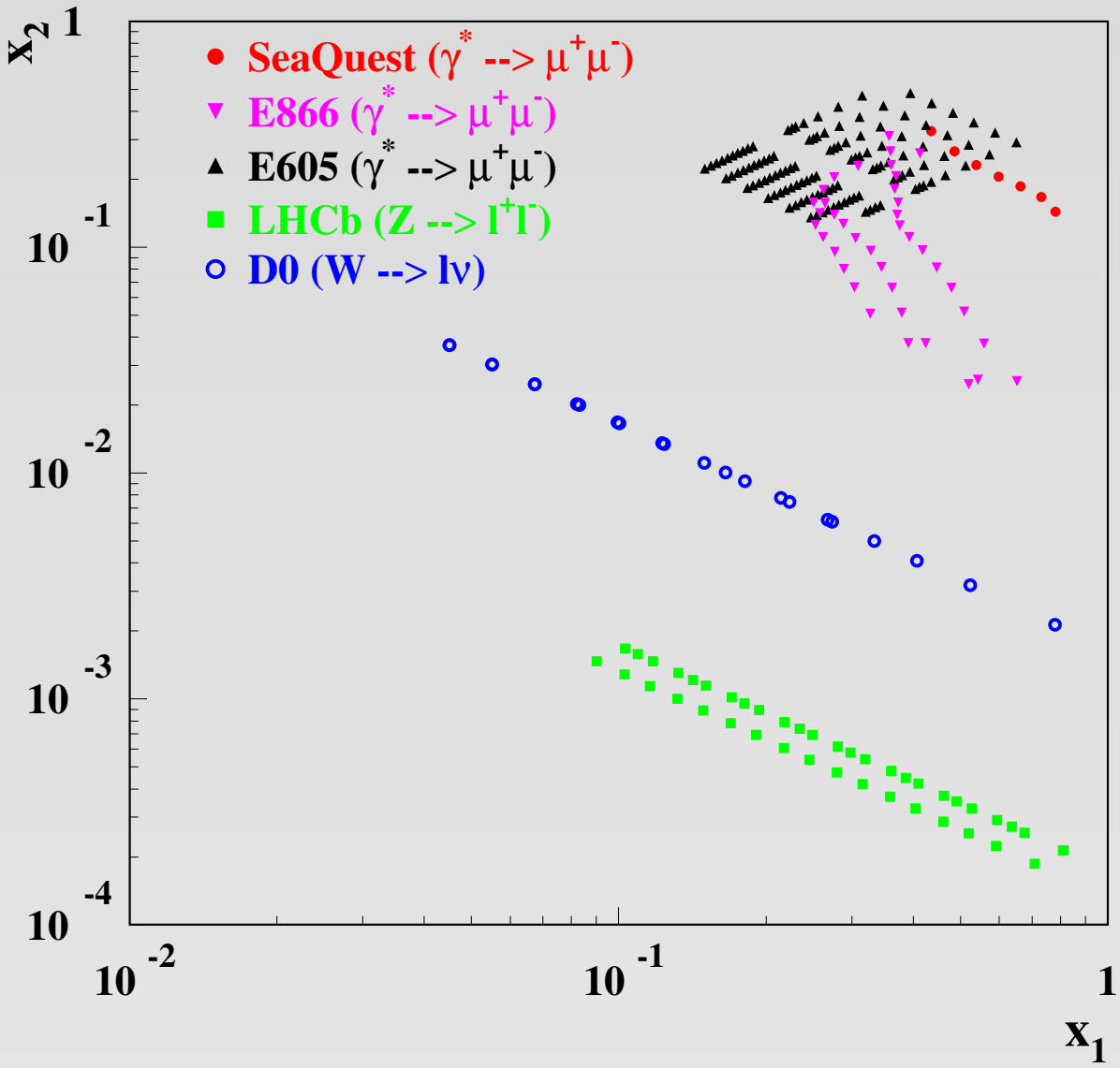
# Conclusions from the ABMPtt/ABMPttgam studies

- Double-differential $M(t\bar{t})$, $y(t\bar{t})$ cross sections included in the ABMPtt PDF + $\alpha_s(M_Z)$ + $m_t(m_t)$ fit make it possible to reduce gluon PDF uncertainties at large $x$ and $m_t(m_t)$ uncertainties by a factor $\sim 2$ with respect to ABMP16 fit, retaining consistency, with no impact on the $\alpha_s(M_Z)$ value and uncertainty.
- $m_t(m_t)$ fitted value from different variants of the fit agree among each other within uncertainties.
- correlations between $m_t(m_t)$ and $\alpha_s(M_Z)$ reduced by the inclusion of double-differential data in the fit w.r.t. to the case of total cross sections, where the effects of correlations are much larger.
- ATLAS $(\ell + \mathrm{j})$ data characterized by the worst theory description, in tension with all other data. A new ATLAS $(\ell + \mathrm{j})$ analysis producing normalized double-differential distributions with larger statistics (full Run 2 statistics) is needed.
- We encourage combinations of ATLAS and CMS data!
- Single top production is still an open problem: discrepancy with ATLAS data on ratio of $(t + X)$ and $(\bar{t} + X)$ cross sections is due to issues in our $d$ and $u$ quark distributions, or to experimental systematics in the categorization of events ?
- Accomodating a photon in the ABMPtt fit and LHC non-resonant dilepton data does not create tensions with the top-quark data.

# Possible improvements and extensions

- Inclusion of NLO EW corrections to $pp \to t\bar{t} + X$: available, but expected to play a minor role, at least considering the PDF groups who already included them in their NNLO fit. However, NLO EW corrections become indispensable when going to higher-order fits.
- Inclusion of resummation effects related to soft-gluon emission and Coulomb-gluon exchange around threshold: partly available, expected to play a role, first of all on the extraction of $m_t$ from bins close to threshold. This will become more relevant when small bin size will be reached in the experiments.
- Fit of 3-differential cross-sections in $M_{t\bar{t}}$, $y_{t\bar{t}}$, $N_j$: experimentally already available (at least from CMS), relevant for better constraints on $\alpha_s(M_Z)$ and further decreasing correlations with the $m_t$ value: NNLO theory predictions for $t\bar{t}j$ (still missing) are necessary!
- Use data at more differential level, e.g. information on leptons. This requires the implementation of $pp \to l^+\nu b l^-\bar{\nu}\bar{b}$ at NNLO in publiclty available codes. Inclusion of top-decays is work in progress within the `MATRIX` collaboration.
- aN$^3$LO computations of single and double-differential distributions in aN$^3$LO PDF fits: theory work allowed already to obtain some of these distributions (in particular those differential in $p_T$ and $y$) from fixed-order expansion of soft-gluon resummation formulas. The theory work needs to be extended to the consideration of other differential distributions.
- Improved unfolding procedures
- More info on correlations between different datasets (experimental effort)

# $(x_1, x_2)$ coverage of DY data in the ABMP fits

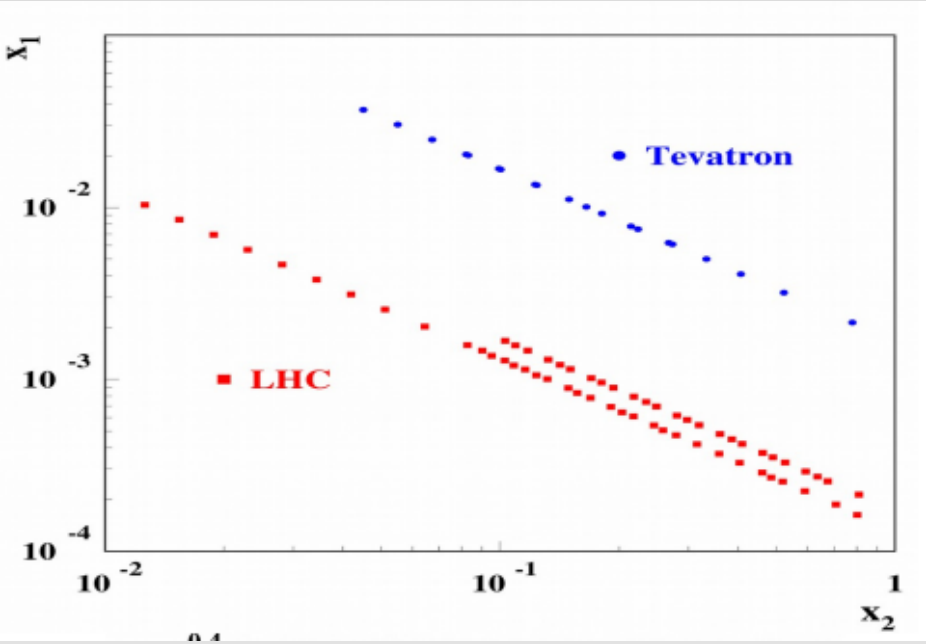



∗ Each point corresponds to one bin in the experimental data

∗ $(x_1, x_2)$ correlation more evident for collider DY data than for fixed-target DY data, in relation to the mass interval of the off-shell boson propagator ($Z$, $W^{\pm}$ vs. $\gamma^*$)

∗ coverage of LHCb $W^{\pm}$ data not shown but similar to LHC $Z$ ones.

# Forward DY at hadron colliders




In the forward region $x_2 >> x_1$

$\sigma(W^+) \sim u(x_2)\,\bar{d}(x_1)$

$\sigma(W^-) \sim d(x_2)\,\bar{u}(x_1)$

$\sigma(Z) \sim Q_U^2 u(x_2)\,\bar{u}(x_1) + Q_D^2 d(x_2)\,\bar{d}(x_1)$

$\sigma(DIS) \sim q_u^2 u(x_2) + q_d^2 d(x_2)$

*Forward W&Z production probes small/large x and is complementary to the DIS ⇒ good quark disentangling*

# The light-flavour sea isospin asymmetry

∗ The light-flavour sea is isospin symmetric or asymmetric ? And why ?

$$\bar{u}(x) = \bar{d}(x) = \bar{s}(x) \text{ vs. } \bar{u}(x) \neq \bar{d}(x) \neq \bar{s}(x) \text{ ?}$$

⇒ models for the flavour structure of the nucleon sea
(see reviews, e.g. [Geeseman and Reimer, arXiv:1812.10372])

∗ Using DY $\sigma_{pd}/\sigma_{pp}$ to constrain $\bar{d}/\bar{u}$: [Ellis and Stirling, PLB 256 (1991) 258]

$$\frac{\sigma_{pd}}{2\sigma_{pp}} \simeq \frac{1}{2}\left(1 + \frac{\sigma_{pn}}{\sigma_{pp}}\right) \simeq \frac{1}{2}\left(1 + \frac{u_p(x_1)\bar{u}_n(x_2)}{u_p(x_1)\bar{u}_p(x_2)}\right) \simeq \frac{1}{2}\left(1 + \frac{\bar{d}_p(x_2)}{\bar{u}_p(x_2)}\right)$$

valid under condition $x_1 > x_2$ and of negligible nuclear effects,
exploited in CERN NA51, FNAL-E866 (NuSea) and FNAL-E906 (SeaQuest) experiments.

# SeaQuest dataset [arXiv:2212.12160]

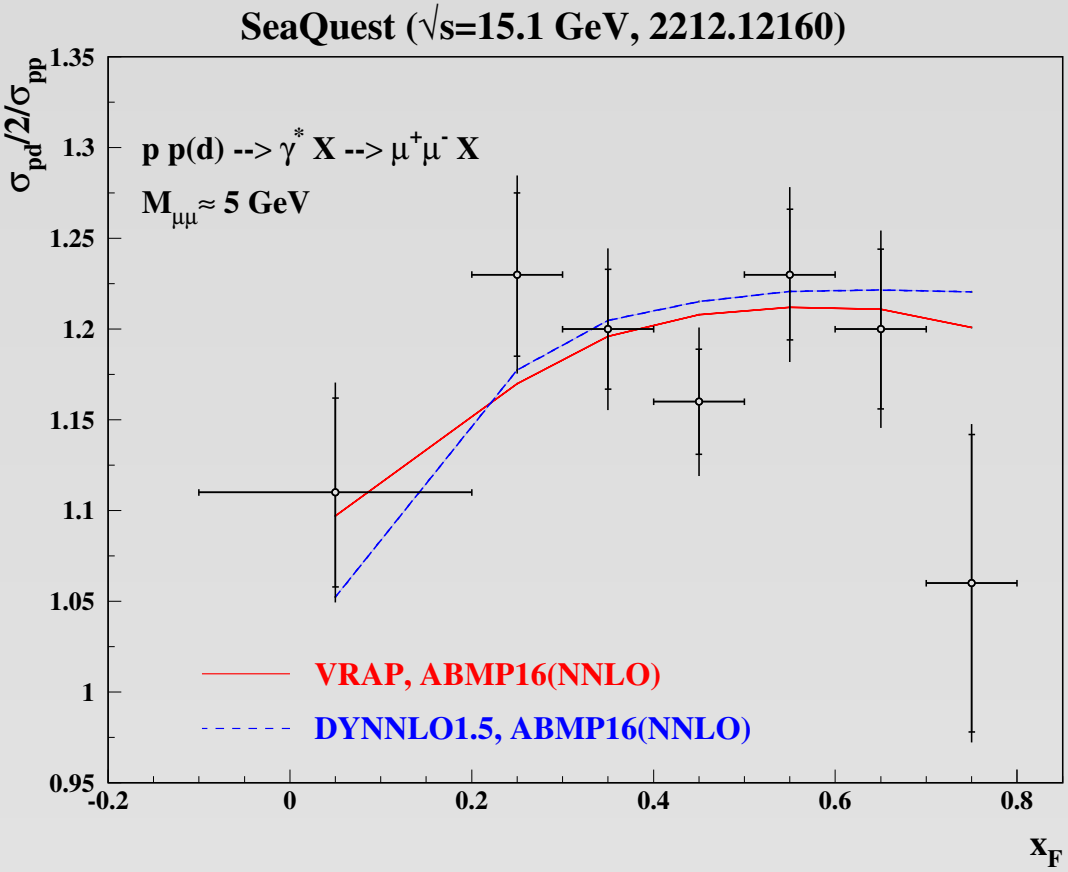
non-resonant $pd, pp \to \gamma^* \to \mu^+\mu^-$

| $x_F$ range | $\langle x_F \rangle$ | $\langle x_1 \rangle$ | $\langle x_2 \rangle$ | $\langle P_T \rangle$ (GeV) | $\langle M_{\mu^+\mu^-} \rangle$ (GeV) | $\sigma^{pd}/2\sigma^{pp}$ MF |
|---|---|---|---|---|---|---|
| -0.100–0.200 | 0.124 | 0.435 | 0.327 | 0.674 | 5.624 | $1.113 \pm 0.052 \pm 0.031$ |
| 0.200–0.300 | 0.255 | 0.486 | 0.266 | 0.714 | 5.340 | $1.228 \pm 0.045 \pm 0.031$ |
| 0.300–0.400 | 0.352 | 0.538 | 0.232 | 0.704 | 5.242 | $1.201 \pm 0.033 \pm 0.030$ |
| 0.400–0.500 | 0.449 | 0.598 | 0.206 | 0.719 | 5.201 | $1.163 \pm 0.029 \pm 0.029$ |
| 0.500–0.600 | 0.546 | 0.663 | 0.186 | 0.717 | 5.192 | $1.227 \pm 0.036 \pm 0.032$ |
| 0.600–0.700 | 0.644 | 0.730 | 0.167 | 0.707 | 5.146 | $1.196 \pm 0.044 \pm 0.032$ |
| 0.700–0.800 | 0.721 | 0.781 | 0.144 | 0.665 | 4.946 | $1.061 \pm 0.082 \pm 0.031$ |

$*$ $x_F = (x_1 - x_2)$, forward $x_F \Rightarrow x_1 > x_2$.
1 = projectile (proton), 2 = target (deuteron, proton),
largest $x_2$ for the smallest $x_F$.

$*$ $x_2$ is large, but not terribly large.... $x_{2,max} = 0.45$ in the extreme of the first bin, limited by the experimental coverage

# DY theory predictions vs. SeaQuest exp. data




∗ DYNNLO allows for exact computations of $d\sigma/dx_F$, with phase-space point by phase-space point information on $p_T$, $M$, etc.

∗ VRAP allows to compute in a fast way rapidity distributions $d\sigma/dy$, by just a 2-dim integration. At this purpose the combination of $\langle x_F \rangle$, $\langle p_T \rangle$, $\langle M \rangle$ values corresponding to the center of each SeaQuest bin (see SeaQuest Table) is mapped to a $y$ rapidity value.

∗ Differences in the first bin, due to its extension, not significant for our fits, considering the big experimental errorbars.

## Total $\chi^2$ of the ABMP16 and ABMP16+SeaQuest fits

| fit | NDP | $\chi^2$ | |
|---|---|---|---|
| | | NLO | NNLO |
| ABMP16 | 2861 | 3428.9 | 3377.6 |
| present analysis (ABMP16 + SeaQuest) | 2868 | 3438.4 | 3384.7 |

$*$ $\chi^2$ with/without SeaQuest data compatible among each other within statistical uncertainties.

$*$ $\chi^2$ for NNLO fits better than for NLO fits
(better description of DY data at NNLO than at NLO).

$*$ Systematic uncertainties assumed correlated bin-by-bin. SeaQuest does not report exact information on the degree of correlation. When assumption of complete uncorrelation is done, the $\chi^2$ slightly decreases.

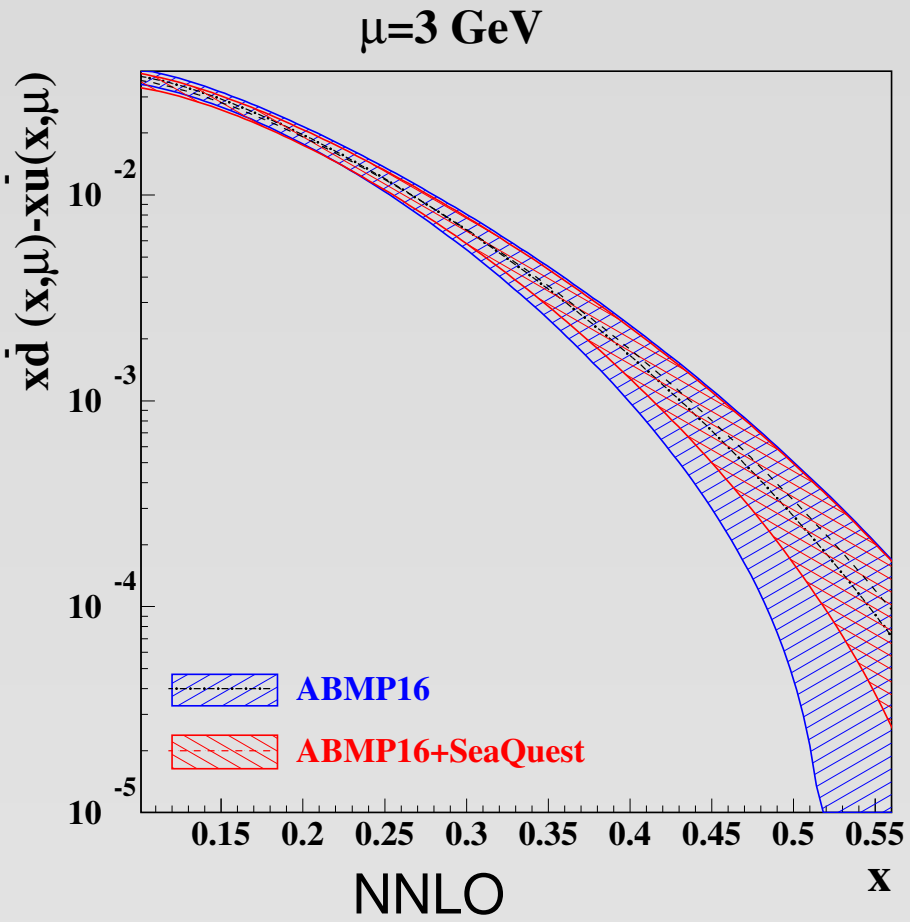
# Partial $\chi^2$ for variants of the ABMP16+SeaQuest fits

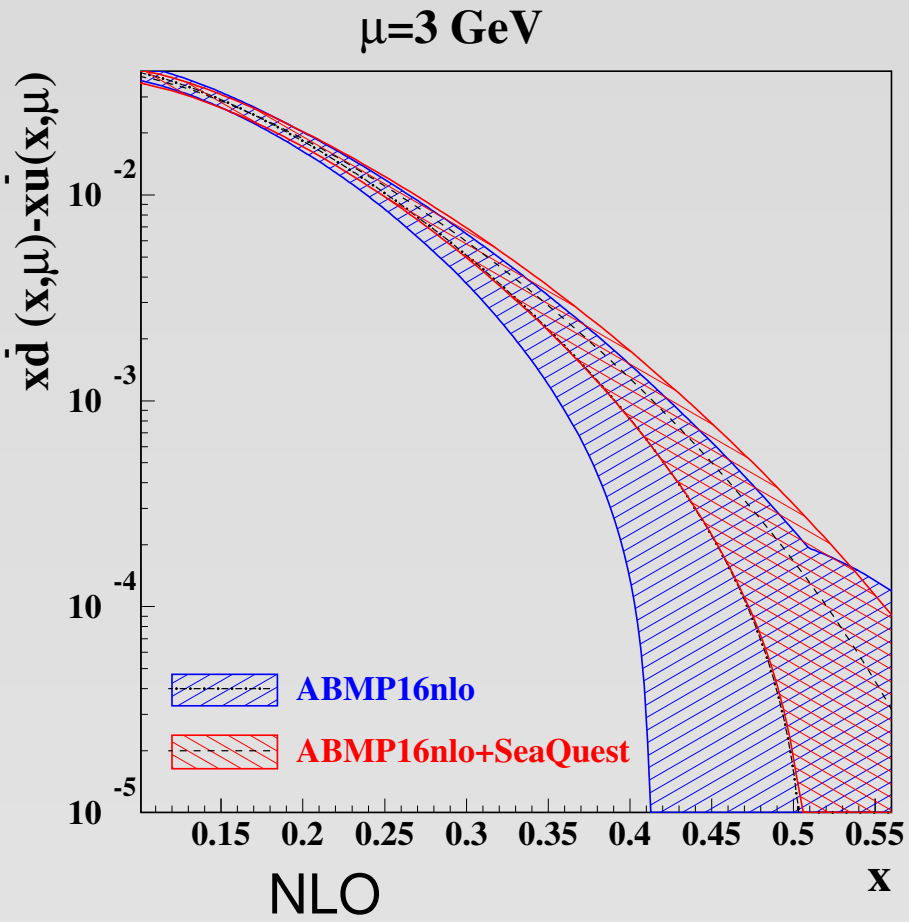
| Experiment | Process | $\sqrt{s}$ (TeV) | Ref. | NDP | $\chi^2$ | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | I | II | III | IV |
| SeaQuest | $pp \to \gamma^* X \to \mu^+\mu^- X$<br>$pd \to \gamma^* X \to \mu^+\mu^- X$ | 0.0151 | [24] | 7 | – | 7.3 | 8.1 | 7.6 |
| NuSea | $pp \to \gamma^* X \to \mu^+\mu^- X$<br>$pd \to \gamma^* X \to \mu^+\mu^- X$ | 0.0388 | [25] | 39 | 52.8 | 54.3 | 52.5 | 53.0 |
| D0 | $\bar{p}p \to W^\pm X \to \mu^\pm \overset{(-)}{\nu} X$ | 1.96 | [45] | 10 | 17.6 | 17.6 | – | 14.5 |
| | $\bar{p}p \to W^\pm X \to e^\pm \overset{(-)}{\nu} X$ | 1.96 | [39] | 13 | 19.0 | 19.0 | – | 15.9 |
| LHCb | $pp \to W^\pm X \to \mu^\pm \overset{(-)}{\nu} X$<br>$pp \to ZX \to \mu^+\mu^- X$ | 7 | [33] | 31 | 45.1 | 43.9 | 35.0 | – |
| | $pp \to W^\pm X \to \mu^\pm \overset{(-)}{\nu} X$<br>$pp \to ZX \to \mu^+\mu^- X$ | 8 | [34] | 32 | 40.0 | 39.6 | 38.2 | – |
| | $pp \to ZX \to e^+e^- X$ | 8 | [46] | 17 | 21.7 | 21.9 | 21.9 | – |

TABLE II: The values of $\chi^2$ obtained for the data sets probing the large-$x$ PDFs, which are included in various analyses (column I: NNLO ABMP16 PDF fit [32], column II: present analysis, column III: a variant of present analysis with D0 DY data excluded, column IV: a variant of present analysis with LHCb DY data excluded).

* $\chi^2$ with/without SeaQuest data compatible within statistical uncertainties for all datasets (full compatibility of SeaQuest with all)
* Slight tension between D0 and LHCb data
* NuSea equally compatible with both.

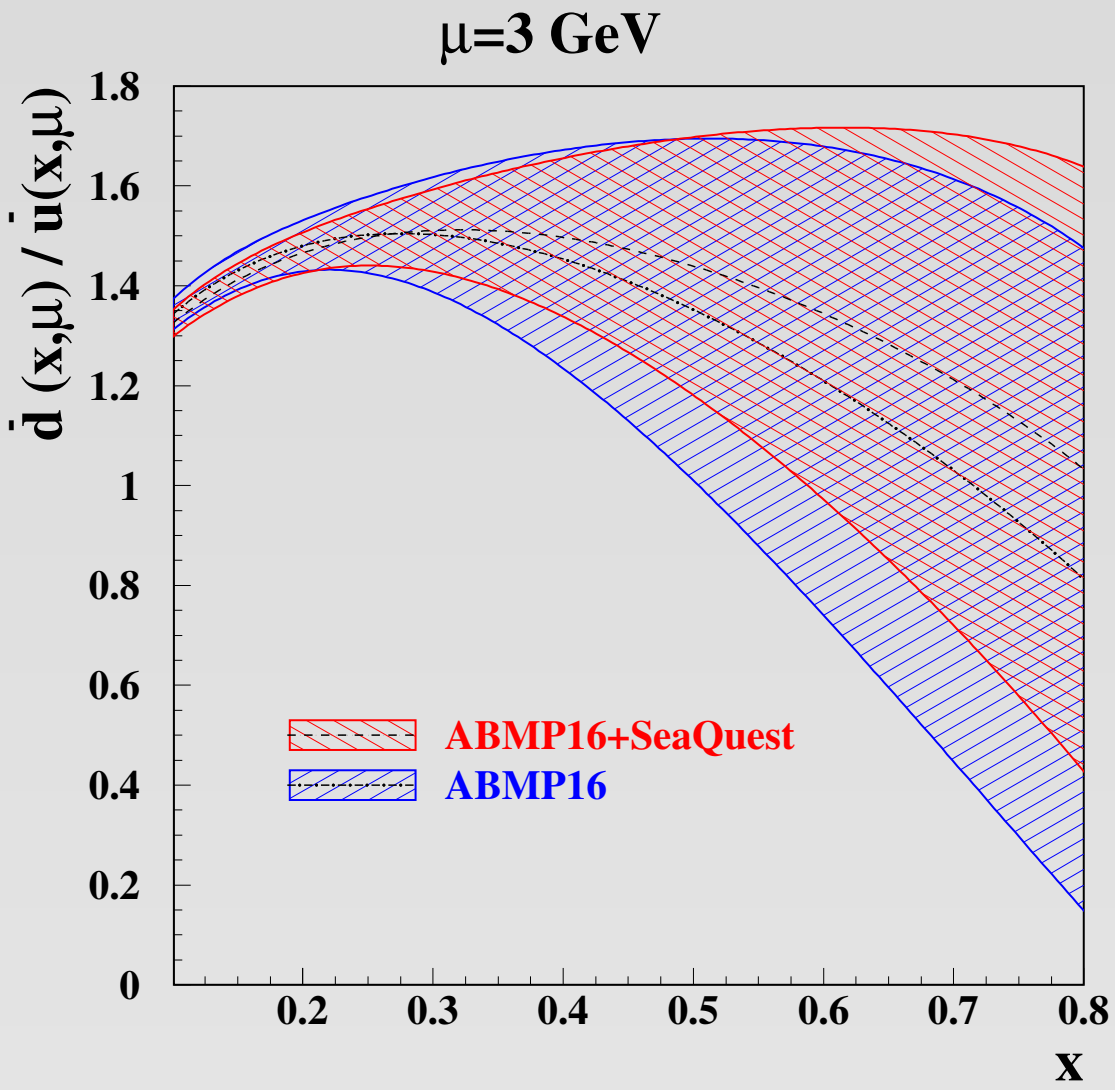
# SeaQuest constraints on $\bar{d}(x) - \bar{u}(x)$




* Uncertainty bands at NNLO smaller than at NLO.
* SeaQuest data play a constraining role for $x \gtrsim 0.3$.
* The asymmetry becomes small at large $x$.

# SeaQuest constraints on $\bar{d}(x)/\bar{u}(x)$ ratio

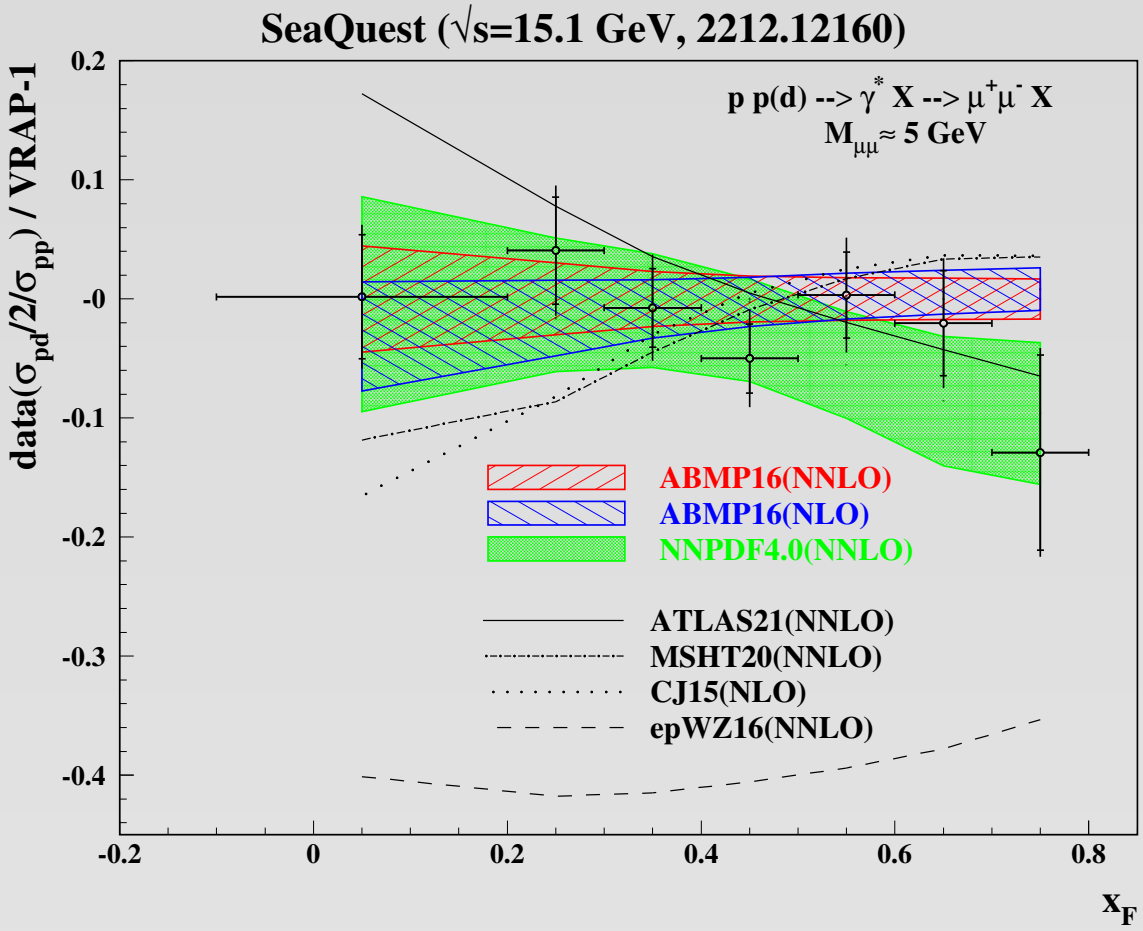



* $\bar{d}(x)/\bar{u}(x) > 1$ up to $x \lesssim 0.5 - 0.6$.

* $\bar{d}(x)/\bar{u}(x) < 1$ at both small ($10^{-4}$) (small $x$ behaviour from LHCb DY) and very large $x$.

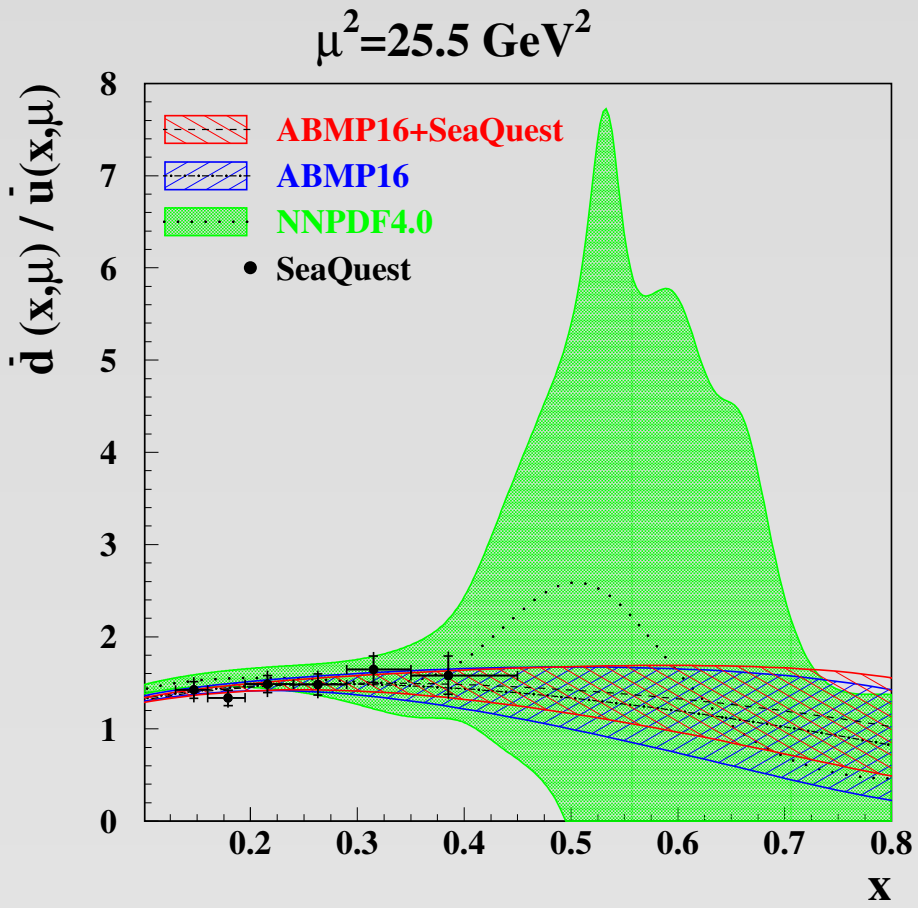
* SeaQuest data play a constraining role for $x \gtrsim 0.3$ (slight decrease of the uncertainties and modification of the central set).

# Compatibility of SeaQuest data with different PDF fits




* NNPDF4.0 is the only one already including SeaQuest data in their nominal fit.
* See also separate analyses of SeaQuest data by the ATLAS, CT18, CJ15, and MSHT collaborations, partially ongoing.
* Symmetric sea ruled out.

# SeaQuest constraints on $\bar{d}(x)/\bar{u}(x)$ ratio

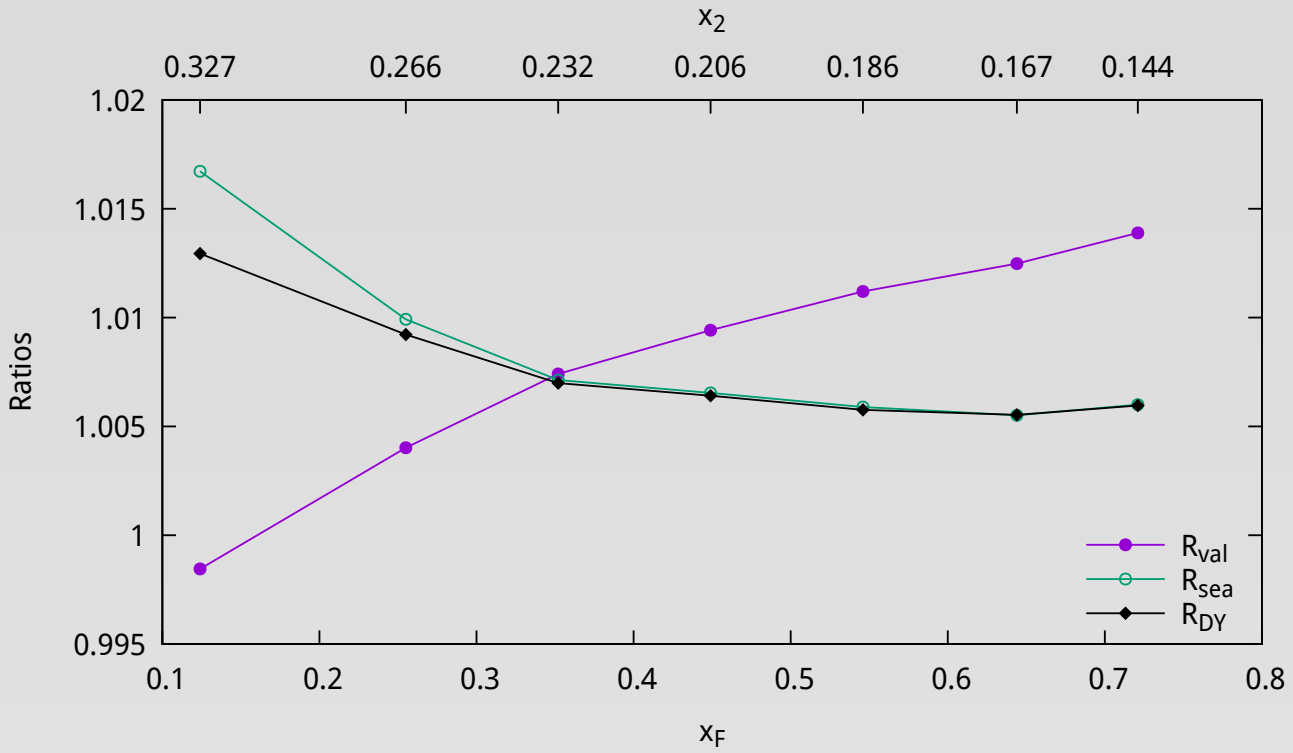



∗ NNPDF4.0 includes SeaQuest data, but its uncertainties in the region $0.3 < x < 0.45$ start to blow up

∗ ABMP16 $\bar{d}(x)/\bar{u}(x)$ in agreement with the one extracted by SeaQuest (on the basis of CT18 PDFs).

∗ uncertainties for $x > 0.5$ driven by extrapolation (lack of data).

# Effects of deuteron nuclear corrections

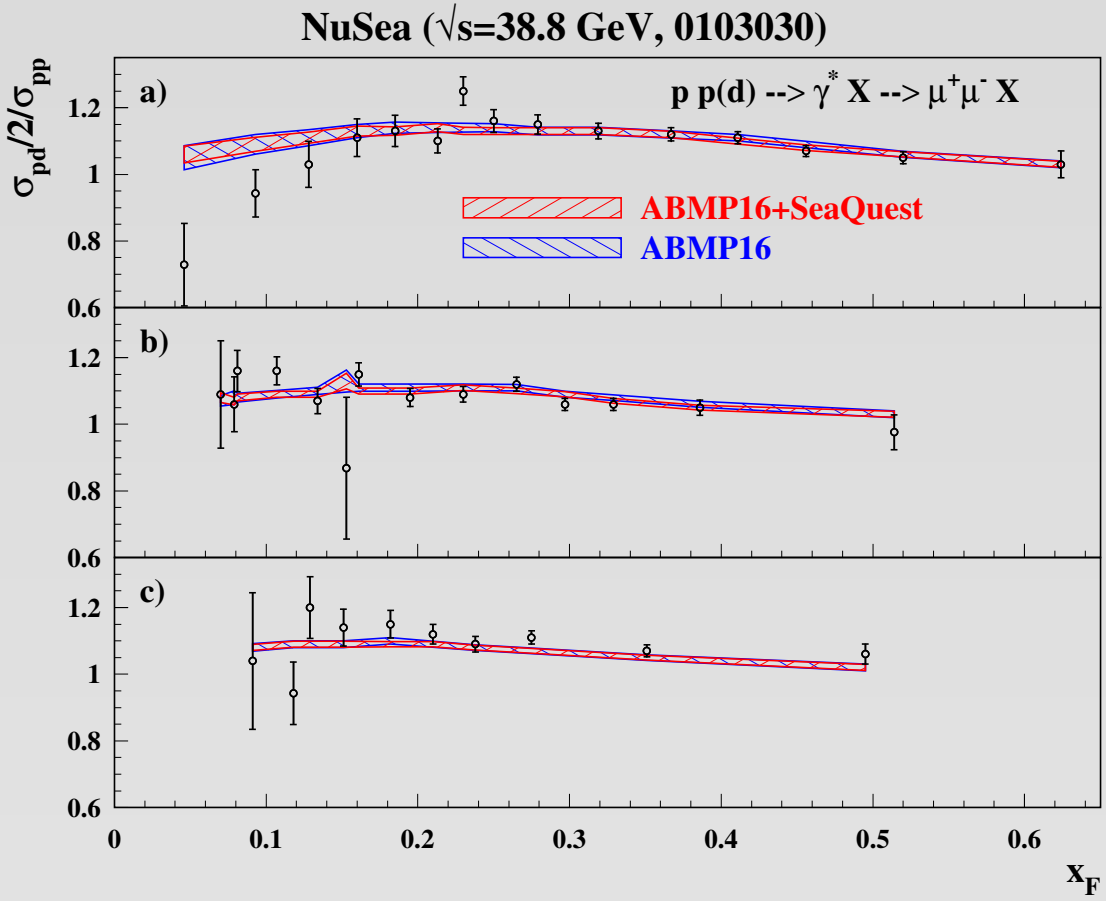



$R_{val} = u_{val,d}/(u_{val,p} + u_{val,n})$, $R_{sea} = \bar{u}_d/(\bar{u}_p + \bar{u}_n)$, $R_{DY} = \sigma_{pd}/(\sigma_{pp} + \sigma_{pn})$.

$*$ Sea quarks in the deuteron (instead of the nucleon) are subject to nuclear corrections effects, mainly due to Fermi motion, nuclear binding and off-shell modifications of bound nucleon PDFs

$*$ Nuclear effects on DY $\sigma_{pd}/\sigma_{pp}$ ratio follows those on the sea quarks.

# Compatibility of NuSea data with AMBP PDF fits




∗ three different spectrometer settings, emphasizing the role of different dimuon invariant mass intervals.

∗ NuSea data important for constraining $(\bar{d}/\bar{u})(x)$ for $0.015 < x < 0.12$, not covered by SeaQuest (complementarity).

∗ Predictions diverge from ABMP16 or ABMP16+SeaQuest at large $x_2$, where however NuSea data uncertainties increase. Data still compatible with ABMP16(+ SeaQuest) PDF fits within $2\,\sigma$.

# Effects of threshold resummation

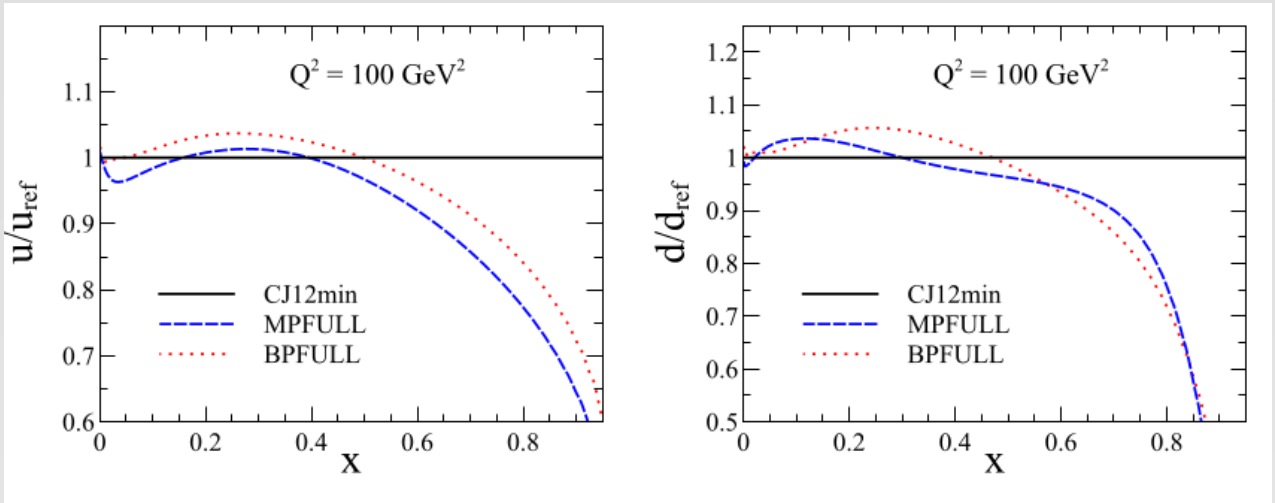
* Relevant at large $x$

* Phase-space for multiple photon emission, which is a convolution in Mellin space, becomes a product in Mellin space $\Rightarrow$ computation performed in Mellin space.

* Various prescriptions for inverse Mellin transform, to avoid spurious effects of the Landau pole.

* When applied to DIS and DY: effects of 10-20% on quark distributions for $0.5 < x < 0.7$.

| | No. points | $\chi^2$ MP | $\chi^2$ BP |
|---|---|---|---|
| E866 (pp) | 136 | 176.0 | 163.1 |
| E866 (pd) | 144 | 219.0 | 169.8 |
| Total | 3655 | 3635.1 | 3530.6 |
| Total + norm | | 3674.2 | 3580.6 |




*from D. Westmark and J. Owens, PRD 95 (2017) 056024*

# Conclusions - constraints from SeaQuest

* SeaQuest $\sigma_{pd}/2\sigma_{pp}$ DY data relevant for constraining $(\bar{d}-\bar{u})(x)$ and $(\bar{d}/\bar{u})(x)$ at relatively large $x$ target, $0.12 < x < 0.45$.
* Effects on the ABMP16 NNLO and NLO fits visible for $x \gtrsim 0.3$, mainly consisting in a reduction of uncertainties on these quantities.
* Better theoretical description of data at NNLO than NLO.
* PDF constraints from SeaQuest and NuSea DY data well compatible with those from LHC DY data.
* Complementarity of E866 and E906 $\sigma_{pd}/2\sigma_{pp}$ DY data, constraining $(\bar{d}/\bar{u})(x)$ ratio in different $x$ regions.
* Tension in the region $0.24 < x < 0.35$ (last two bins of E866) is not significant for the AMBP16+SeaQuest fit, due to the large E866 uncertainty bands in that region.
* So far, only $\sim 1/2$ of SeaQuest already collected data have been published.
* Therefore we are optimistic that data uncertainties $\sim 5\%$ can be decreased.
* next step: (ABMP16ttgam + SeaQuest) analysis.

**Thank you for your attention!**