MSHT20 PDFs
Review and Updates

Thomas Cridge

Physikzentrum Bad Honnef - Forward Physics and QCD at the LHC and EIC

24th October 2023

WILHELM UND ELSE HERAEUS-STIFTUNG

DESY.

erc European Research Council

In collaboration with S. Bailey, L.A. Harland-Lang, A.D. Martin, J. McGowan, and R.S. Thorne.

MSHT20 PDFs:

Most accurate, precise PDF set yet, with reduced uncertainties.

- MSHT20 - New PDF set for precision LHC era - arXiv:2012.04684 .
- Global fit > 61 datasets - 10 Fixed Target Structure Function, 6 neutrino scattering, 2 fixed target DY, 8 HERA, 8 Tevatron, 27 LHC.
- 4000$+$ datapoints over (x, Q^2): $[10^{-4}, 0.8]$ and $[2, 10^6]$ GeV^2.
- Significant developments on all three fronts:
  1. Theoretical - Full NNLO QCD theory, with NLO EW where relevant.
  2. Experimental - Many new datasets, more precise, more channels, more differential.
  3. Methodological - Extended parameterisation to allow fitting accuracy to $< 1\%$ if data allows, better knowledge of central values (52 PDF parameters) and uncertainties (64 eigenvector directions).
- 3 years since then, we several additions/updates $\rightarrow$ this talk!

Approximate N3LO PDFs with Theory Uncertainties - MSHT20aN3LO

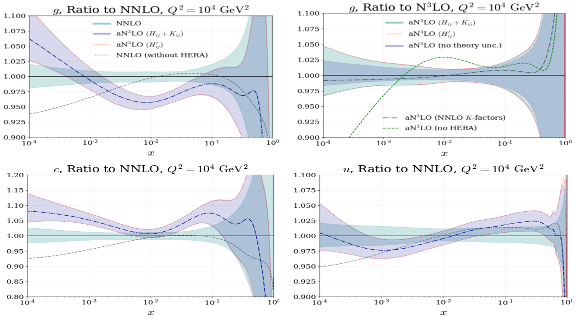
2207.04739 + updates.

aN3LO QCD PDFs with theory uncertainties: 2207.04739

- LHC precision program needs higher order PDFs + inclusion of PDF theory uncertainties for first time → **MSHT20aN3LO** PDFs.
- Idea - Include known N3LO info and *parametrise few unknown pieces* → aN3LO + theory uncertainties.
- Fit impact: perform aN3LO fit with identical dataset to NNLO:

Total Fit quality	LO	NLO	NNLO	aN3LO
χ^2/N_{pts}	2.57	1.33	1.17	1.14

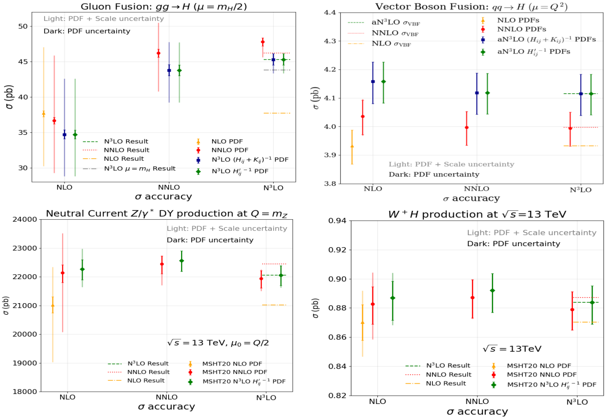
Smooth fit improvement with order. $\Delta\chi^2 = -154.4$ from NNLO to aN3LO.



- Gluon PDF raises and its uncertainty increases at low x.
- Heavy quarks also increase across x, purely perturbative.
- Other PDFs much more mildy affected.

How do aN3LO PDFs affect cross-sections?:

Updates



N.B. Results here only indicative, conclusions depend on scale and $\sqrt{s}$.

- Increase in ggH from xsec at N3LO partially compensated by reduction in PDFs at aN3LO $\Rightarrow$ improved perturbative convergence.
- Increase in VBF xsec with aN3LO PDFs from increased heavy quarks.
- Only small change from aN3LO PDFs for NC Drell-Yan or W^+H.

Strong Coupling $\alpha_S(M_Z^2)$ and Heavy quark Masses m_c, m_b and m_t

2106.10289, 2306.14885 + updates.

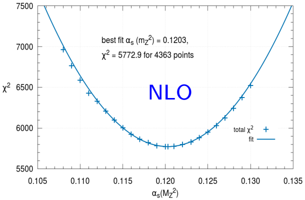
MSHT20 α_S and $m_{c,b}$ sensitivity:

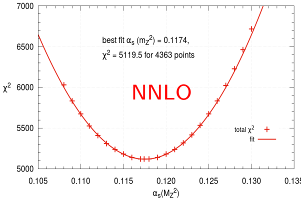
2106.10289

- Global PDF fit $\Rightarrow$ can provide precise, accurate α_S determination.
- The best fit values and bounds at NLO, NNLO are:

Order	NLO	NNLO
$\alpha_S(M_Z^2)$	0.1203 ± 0.0015	0.1174 ± 0.0013

Consistent with NNLO World Average 0.1179 ± 0.0009.

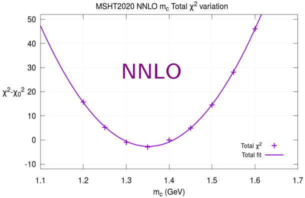




$\alpha_{S,\mathrm{NNLO}}(M_Z^2) < \alpha_{S,\mathrm{NLO}}(M_Z^2)$ as NNLO corrections +ve, so fitting same data $\Rightarrow$ lower α_S.

Nice Quadratic χ^2 profile ✓

- Also sensitive to heavy quark pole masses, e.g. m_c at NNLO:



- Global fit favours $m_c, m_b \approx 1.35, 4.5\mathrm{GeV}$. Much from HERA combined heavy flavour.
- Very low values of m_c, m_b disfavoured.
- What about at aN3LO?

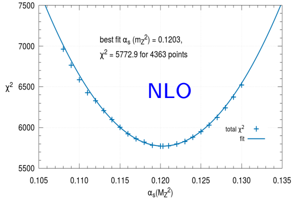
MSHT20 α_S and $m_{c,b}$ sensitivity, aN3LO:

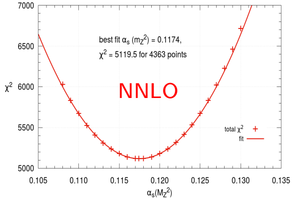
Upcoming!

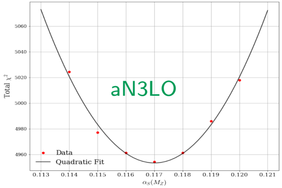
- Global PDF fit $\Rightarrow$ can provide precise, accurate α_S determination.
- The best fit values and bounds at NLO, NNLO and aN3LO are:

Order	NLO	NNLO	aN3LO
$\alpha_S(M_Z^2)$	0.1203 ± 0.0015	0.1174 ± 0.0013	≈ 0.1170

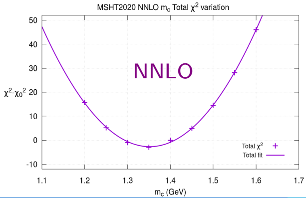
Consistent with NNLO World Average 0.1179 ± 0.0009.

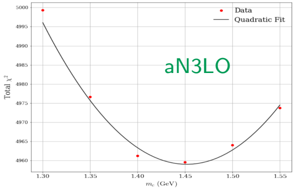






- Also sensitive to heavy quark pole masses, e.g. m_c at NNLO/aN3LO:



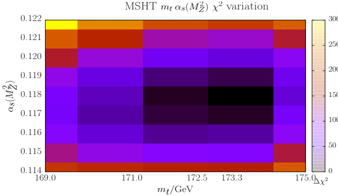


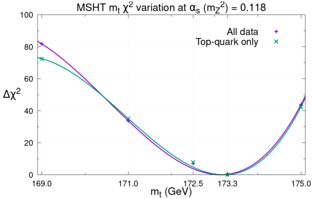
- $m_c \approx 1.45\text{GeV}$ at aN3LO.
- World average: $m_c = 1.5 \pm 0.2\text{GeV}$.

MSHT20 m_t dependence at NNLO:

2306.14885

- Global PDF fit also has sensitivity to m_t pole mass through top quark data - total $t\bar{t}$ xsec and ATLAS and CMS l+j at 8 TeV.

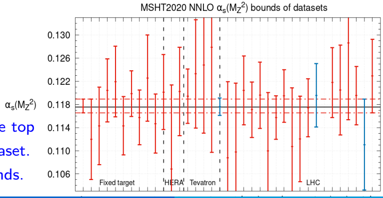


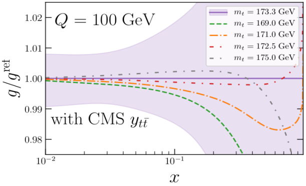


Consistent with PDG 172.5 ± 0.7 GeV.

- Best fit obtained at $m_t = 173.0$ GeV. Conservative bound ± 0.6 GeV.
- Analyse α_S - top quark data bounds α_S competitively (lower left).
- Effect of m_t on gluon well within PDF uncertainties (lower right).

Blue lines are top quark dataset. α_S bounds.





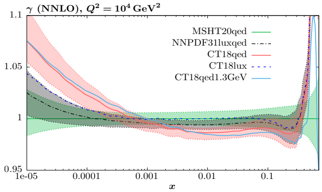
QED effects in MSHT20 - MSHT20qed

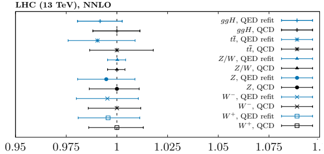
2111.05357 + updates.

Inclusion of QED effects + aN3LO:

2111.05357

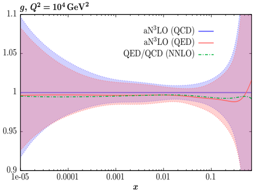
- With NNLO QCD standard, QED/EW important: $\alpha_{\text{QED}}(M_Z) \sim \alpha_S^2(M_Z)$.
- QED corrections via modifications to DGLAP, via photon PDF and photon-initiated processes.
- $\gamma(x, Q^2)$ PDF with $\mathcal{O}(\%)$ uncertainties.
- Effect of QED $\lesssim$ PDF uncertainties.
- Uncertainties similar to QCD only case.

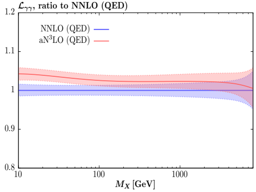




What happens when you consider aN3LO QCD + NLO QED? Upcoming!

- Results at aN3LO similar to QCD only vs QCD + QED comparison at NNLO.





New Data added on top of MSHT20

2309.11269 + updates.

Jet and Dijet data, NNLO and aN3LO:

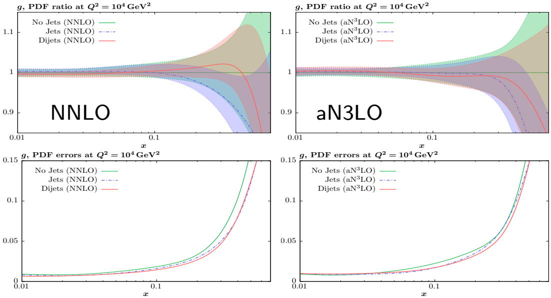
Upcoming!

- MSHT20 had ATLAS 7 TeV and CMS 7, 8 TeV inclusive jet data.
- Add ATLAS 8 TeV inclusive jet, or replace jets all with dijets data:

.	N_{pts}	χ^2/N_{pts} NNLO	χ^2/N_{pts} aN3LO	.	N_{pts}	χ^2/N_{pts} NNLO	χ^2/N_{pts} aN3LO
ATLAS 7 TeV jets	140	1.54	1.46	ATLAS 7 TeV dijets	90	1.06	1.12
CMS 7 TeV jets	158	1.29	1.32	CMS 7 TeV dijets	54	1.43	1.39
CMS 8 TeV jets	174	1.83	1.80	CMS 8 TeV dijets	122	1.05	0.82
ATLAS 8 TeV jets	171	1.96	1.90	-	-	-	-
Total (jets)	643	1.67	1.63	Total (dijets)	266	1.13	1.04
Total	4543	1.22	1.17	Total	4157	1.14	1.09

- Obtain better fit quality at NNLO and aN3LO with dijets than jets.

- Some difference in pull of jets/dijets at NNLO.
- More consistent at aN3LO.
- Clear reduction in high x gluon uncertainty, more for dijets.

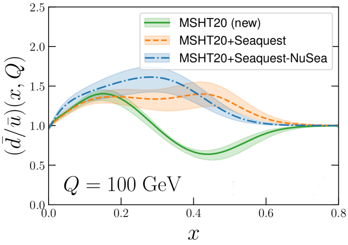


Seaquest Data at NNLO and aN3LO:

Preliminary!

- Seaquest (E906) fixed target DY data at high x: $\sigma_D/\sigma_H \sim 1 + \bar{d}/\bar{u}$.
- Raises high x $\bar{d}/\bar{u}$. Tension with NuSea(E866) which pulls it down.

NNLO:

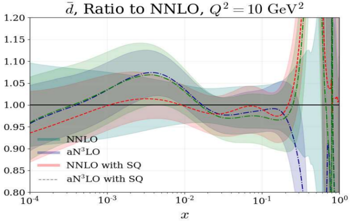


Dataset	$N_{\rm pts}$	MSHT20	New
Seaquest	6	-	8.2
NuSea	15	9.8	19.0
Total (without Seaquest or NuSea)	4348	5102.3	5112.1

- NuSea $\chi^2/N_{\rm pts}$: $0.65 \to 1.27$, when Seaquest added. Rest of data worsens by $\Delta\chi^2 = +9$.

aN3LO:

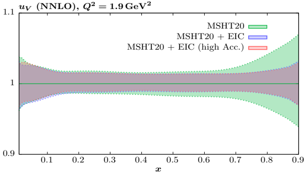
- Similar effect to at NNLO.
- Raises $\bar{d}$, lowers $\bar{u}$ such that they are very close at NNLO and aN3LO once Seaquest data is added.

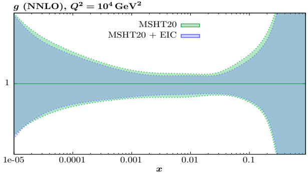


New data - EIC Pseudodata (at NNLO):

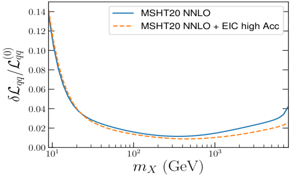
2309.11269
See also F. Giuli's talk!

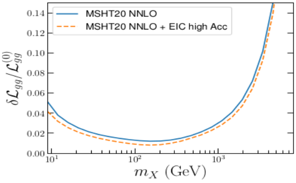
- Impact of EIC pseudodata on MSHT20 - *high* x *lower* Q^2 *sensitivity*.

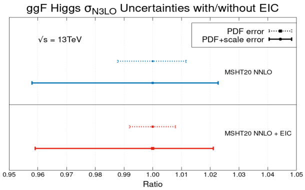




- Effect on up valence larger due to charge-squared γ coupling in DIS.
- Gluon uncertainty nonetheless reduced across range of x.
- Impact on luminosity uncertainties - $\delta\mathcal{L}_{qq}$ reduced at high x.
- $\delta\mathcal{L}_{gg}$ reduced across x causes smaller PDF uncertainty for $gg \to H$.







Availability and Summary

MSHT PDF sets available

All available at https://www.hep.ucl.ac.uk/msht/, and most also on LHAPDF.

- Overview of available MSHT20 PDF sets (this is a small selection!):

LHAPDF6 grid name	Order	n_f^{max}	N_{mem}	$\alpha_S(m_Z^2)$	Description
MSHT20nnlo_as118	NNLO	5	65	0.118	Default NNLO set
MSHT20nlo_as120	NNLO	5	65	0.118	Default NLO set
MSHT20lo_as130	NNLO	5	65	0.118	Default LO set
MSHT20nnlo_as_largerange	NNLO	5	23	0.108-0.130	$\alpha_S(M_Z^2)$ variation NNLO set
MSHT20nlo_as_largerange	NLO	5	23	0.108-0.130	$\alpha_S(M_Z^2)$ variation NLO set
MSHT20nnlo_mcrange_nf5	NNLO	5	9	0.118	Charm mass variation (1.2-1.6 GeV) NNLO set
MSHT20nnlo_mbrange_nf5	NNLO	5	7	0.118	Bottom mass variation (4.0-5.5 GeV) NNLO set
MSHT20nnlo_nf3,4	NNLO	3, 4	65	0.118	NNLO set with max. 3 or 4 flavours
MSHT20qed_nnlo	NNLO	5	77	0.118	NNLO set with QED effects and γ PDF
MSHT20qed_nnlo_(in)elastic	NNLO	5	77	0.118	NNLO set with QED effects and (in)elastic γ
MSHT20qed_nnlo_neutron	NNLO	5	77	0.118	NNLO neutron set with QED effects and γ
MSHT20an3lo_as118	aN3LO	5	105	0.118	Approximate N3LO set with theoretical uncertainties also included
MSHT20an3lo_as118_KCorr	aN3LO	5	105	0.118	Approximate N3LO set with theoretical uncertainties also included, K-factors correlated
PDF4LHC21	NNLO	5	901	0.118	Baseline PDF4LHC21 set
PDF4LHC21_mc	NNLO	5	101	0.118	Replica compressed PDF4LHC21 set
PDF4LHC21_40	NNLO	5	41	0.118	Hessian compressed PDF4LHC21 set

Selection of some of the MSHT PDF sets available in LHAPDF format. Many more online!

Key:

- Default - $\alpha_S, m_{c,b}$ - QED - aN3LO - PDF4LHC21

- Ongoing work and updates to come on several topics!

Conclusions

- MSHT20 was a significant step forward ⇒ our most accurate, precise PDF set yet.
- Many subsequent developments:
  - World-first approximate N3LO PDFs with theoretical uncertainties.
  - Strong coupling and heavy quark mass sensitivity - recent updates on m_t and on aN3LO.
  - MSHT20qed PDF sets with QED effects and photon PDF - updates on aN3LO.
  - New data examined - dijets, Seaquest, EIC often at NNLO and aN3LO.
- All PDFs available for public usage - LHAPDF and MSHT website.
- This will all be supplemented by further ongoing work driving our knowledge of PDFs forward.

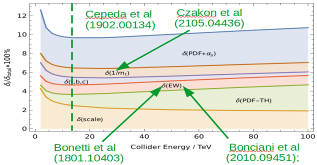
Backup Slides

Note: For some of the more recent work, this project (via TC) has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101002090 COLORFREE).

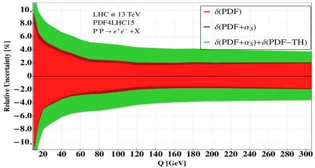
Motivation for MSHT20aN3LO PDFs

- LHC precision program - N3LO xsecs: Higgs (ggF, VBF, VH), DY(NC, CC).

Higgs - ggF:



Drell-Yan - NC:



- Large PDF uncertainties, also missing higher order uncertainties.

Therefore we require:

1. Higher order PDFs (N3LO).
2. Inclusion of theory uncertainties from missing higher orders.

$\Rightarrow$ we can address both in one go! $\Rightarrow$ **MSHT20aN3LO PDFs.**

- Include known N3LO info on $P_{ij}^{(3)}$, A_{ij}^{3}, C_{ij}^{3} etc $\rightarrow$ aN3LO PDFs.
- *Parametrise few unknown pieces* $\rightarrow$ aN3LO + theory uncertainties.

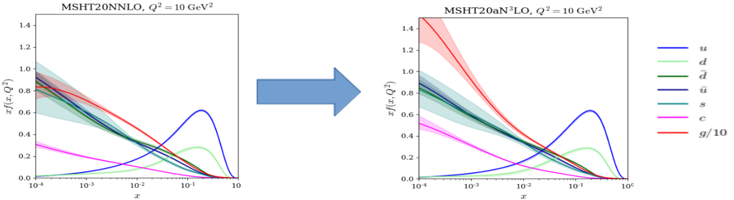
How does aN3LO affect the PDFs?:

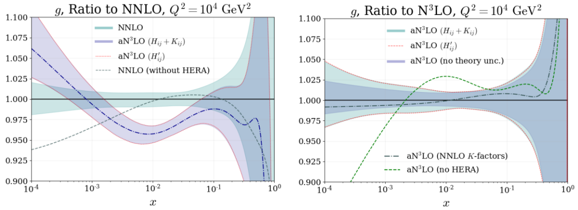
- Perform aN3LO fit with identical dataset to MSHT20 NNLO PDF fit:

Total Fit quality χ^2/N_{pts}	LO	NLO	NNLO	aN3LO
	2.57	1.33	1.17	1.14

Smooth fit improvement with order.

$\Delta\chi^2 = -154.4$ from NNLO to aN3LO.

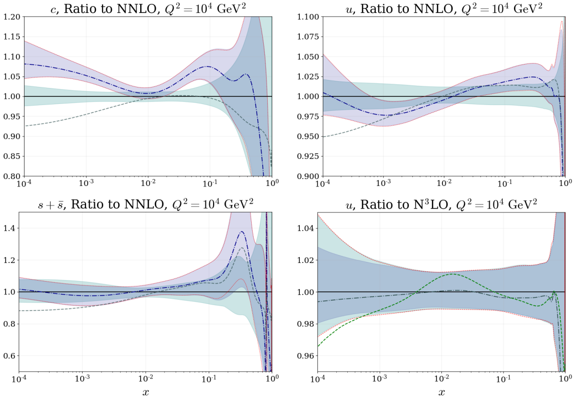




- Gluon PDF raises and its uncertainty increases at low x.
- Effects from large logs in splitting functions at low x and MHOUs.

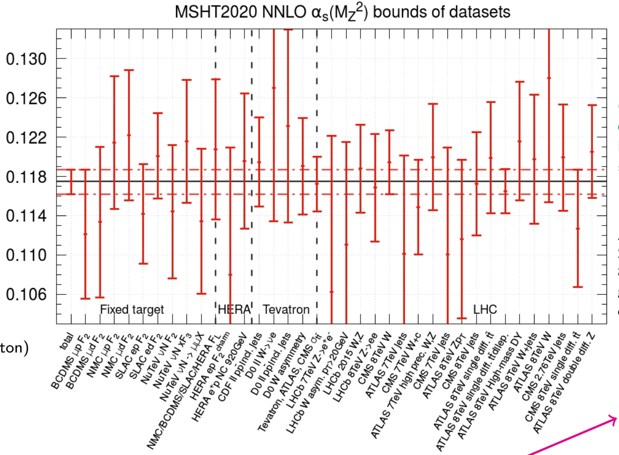
How does aN3LO affect the PDFs?:

- Milder effects on most other PDFs:



- Heavy quarks also increase across x, purely perturbative.
- Theory uncertainty from MHOUs included for first time in PDF fit - enlarges PDF uncertainty at small x.

MSHT20 α_S bounds - NNLO



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

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

CMS and ATLAS (dilepton) $t\bar{t}$ single diff. prefer lower α_S but no m_t dependence available so bound not used.

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

ATLAS 8 TeV Z, SLACd, ATLAS 8 TeV High-mass DY give lower bounds of $\Delta\alpha_S(M_Z^2) = -0.0017, -0.0018, -0.0019$

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

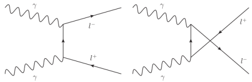
- Therefore upper and lower bounds are +0.0012 and -0.0013.

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

Consistent with World Average of 0.1179 ± 0.0009.

Inclusion of QED effects:

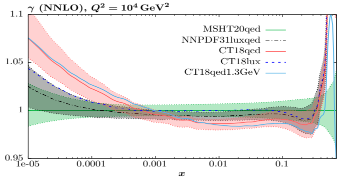
- With NNLO QCD now standard, noting that $\alpha_{\text{QED}}(M_Z) \sim \alpha_S^2(M_Z)$:
  $\Rightarrow$ important to consider EW effects, QED corrections are a key part.
- MSHT20 include EW corrections for:
  - Drell-Yan
  - inclusive jets
  - top
  - DIS.

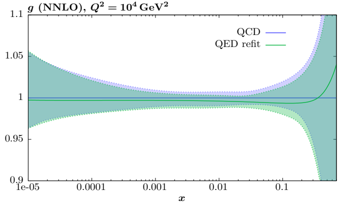


Photon-Initiated contributions to Drell-Yan.

- QED corrections via QED modifications to DGLAP, via photon PDF and photon-initiated processes.
- Obtain $\gamma(x, Q^2)$ with $\mathcal{O}(\%)$ uncertainties via LUXQED-related method.

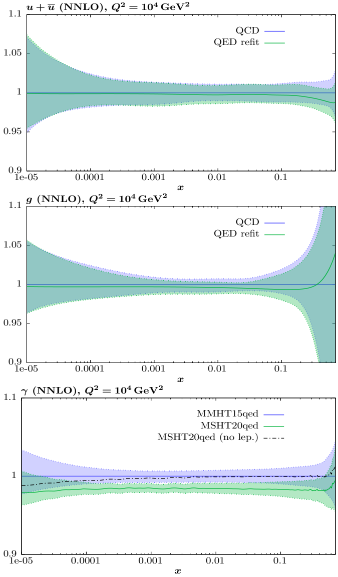
  Manohar et al, 1708.01256, *JHEP 12, 046 (2017)*.
- General consistency with NNPDF, CT.
- Quarks reduced at high x by $q \rightarrow q\gamma$, gluon reduced by momentum sum rule.





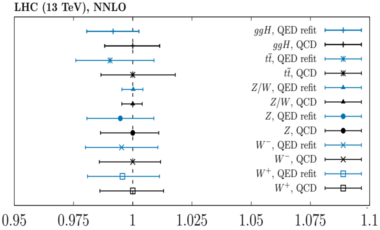
QED effects on PDFs:

- MSHT20qed has reduced $u + \bar{u}$ at high x from $q \to q\gamma$ splitting.
- Effect on down quarks (not shown) smaller due to smaller charge.
- Gluon reduced across almost entire x range due to momentum sum rule. $\Rightarrow$ Need to accommodate γ carrying extra momentum.
- Photon reduced relative to MMHT2015qed due to inclusion of lepton-loops in $P_{\gamma\gamma}$.
- Photon breakdown into elastic and inelastic components also provided, as are neutron PDFs (see backup).



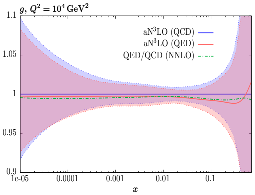
QED effects on Benchmark Cross-sections + aN3LO:

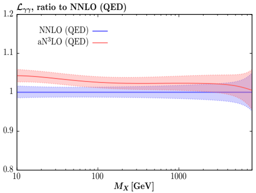
- Gluon-initiated processes, lower by $\sim 1\%$ in QED case.
- W, Z production reduced by $q \to q\gamma$ splitting, W/Z ratio stable.
- Effect of QED $\lesssim$ PDF uncertainties.
- Uncertainties similar to QCD only.



What happens when you consider aN3LO QCD + NLO QED? Upcoming!

- Results at aN3LO similar to QCD only vs QCD + QED comparison at NNLO.





New data - Dijets vs Inclusive Jets - Fit Quality (NNLO)

- Fit either 7+8 TeV inclusive jets or dijets on MSHT20 baseline.
- Inclusive jets have issues with systematic correlations and theoretical questions, e.g. scale choice, non-unitary nature, etc.
- Dijets may resolve some such issues, and triple differential measurement is more sensitive to PDF x-dependence.

Also investigated.
at aN3LO
⇒ see later!

Dijets:

Dataset	$N_{\rm pts}$	$\chi^2/N_{\rm pts}$
ATLAS 8 TeV Zp_T	104	1.65
Top differential data total	54	1.24
ATLAS 7 TeV dijets	90	1.05
CMS 7 TeV dijets	54	1.43
CMS 8 TeV dijets	122	1.04
Total dijets	266	1.12

Inclusive Jets:

Dataset	$N_{\rm pts}$	$\chi^2/N_{\rm pts}$
ATLAS 8 TeV Zp_T	104	1.85
Top differential data total	54	1.12
ATLAS 7 TeV jets	140	1.53
ATLAS 8 TeV jets	171	1.45
CMS 7 TeV jets	158	1.22
CMS 8 TeV jets	174	1.80
Total inclusive jets	643	1.50

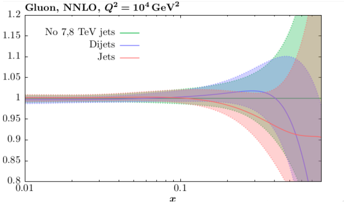
- Fit quality of dijets - 1.12, better than inclusive jets - 1.50.
- Clear improvement with order, NNLO needed for precise LHC data.

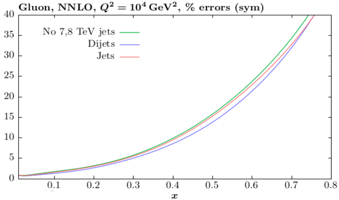
Dataset	$N_{\rm pts}$	NLO	NNLO
ATLAS 7 TeV dijets	90	1.10	1.05
CMS 7 TeV dijets	54	1.71	1.43
CMS 8 TeV dijets	122	5.30	1.04
Total dijets	266	3.15	1.12

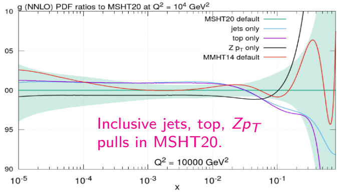
Dataset	$N_{\rm pts}$	NLO	NNLO
ATLAS 7 TeV jets	140	1.69	1.53
ATLAS 8 TeV jets	171	2.37	1.45
CMS 7 TeV jets	158	1.38	1.22
CMS 8 TeV jets	174	1.65	1.80
Total inclusive jets	643	1.78	1.50

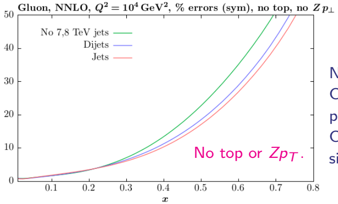
New data - Dijets vs Inclusive Jets - PDFs (NNLO)

- Impact on gluon PDF at high x, consistent but different pulls.
- Dijets have more impact on reducing gluon uncertainty at high x.







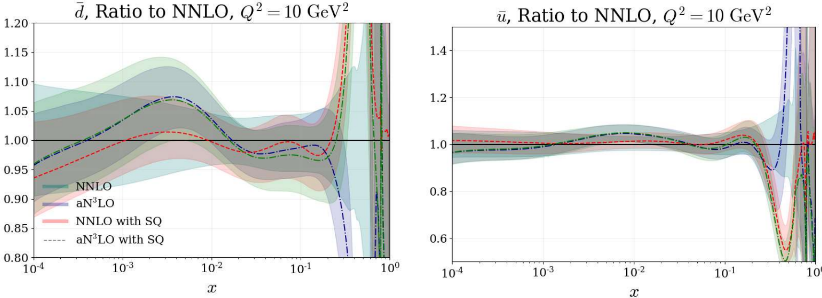


N.B. This is all Leading Colour, we have looked preliminarily at Full Colour and not found significant changes.

- Dijets increases high-x gluon, like Zp_T, inclusive jets reduces high x gluon, like top data. $\Rightarrow$ Interplay with other data.
- Without Zp_T or top, inclusive jets has greater impact on uncertainty.

MSHT20aN3LO PDFs - Seaquest

- At aN3LO, the $\bar{d}$ become negative above $x \sim 0.5$ with a minimum at $x \sim 0.6$. Nonetheless remains positive within uncertainties.
- Like at NNLO, adding the Seaquest data raises the $\bar{d}/\bar{u}$.



- Adding Seaquest $\Rightarrow$ NNLO and aN3LO $\bar{d}$, $\bar{u}$ again very similar.
- Effect on fit quality of adding Seaquest similar to NNLO, $\Delta\chi^2 = +6$ in rest of data, NuSea χ^2/N doubles from ~ 0.6 to ~ 1.3.

MSHT20 New data - Mainly LHC

- Significant new data in MSHT20 fit - Drell-Yan, inclusive jets, top, W+jets, $W + c$, HERA final combination and heavy quarks:

High x quarks - u_V, d_V.

Flavour Decomposition - e.g. strangeness.

High x gluon - jets, top, Zp_T.

Low/intermediate x - quarks, antiquarks, and gluon, e.g. LHCb and HERA data.

Data set	Points	NLO χ^2/N_{pts}	NNLO χ^2/N_{pts}
DØ W asymmetry	14	0.94 (2.53)	0.86 (14.7)
ATLAS 8 TeV W^+W^- + jets	30	1.13 (1.13)	0.60 (0.57)
CMS 7 TeV $W + c$	10	0.82 (0.85)	0.86 (0.84)
LHCb 7+8 TeV $W + Z$	67	1.71 (2.35)	1.48 (1.55)
LHCb 8 TeV $Z \to ee$	17	2.29 (2.89)	1.54 (1.78)
CMS 8 TeV W	22	1.05 (1.79)	0.58 (1.30)
ATLAS 7 TeV $W + Z$	61	5.00 (7.62)	1.91 (5.58)
ATLAS 8 TeV W^+W^-	22	3.85 (13.9)	2.61 (5.25)
ATLAS 8 TeV double differential Z	59	2.67 (3.26)	1.45 (5.16)
ATLAS 8 TeV high-mass DY	48	1.79 (1.99)	1.18 (1.26)
CMS 2.76 TeV jets	81	1.53 (1.59)	1.27 (1.39)
CMS 7 TeV jets $R = 0.7$	158	1.27 (1.32)	1.11 (1.17)
ATLAS 7 TeV jets $R = 0.6$	140	1.62 (1.59)	1.59 (1.68)
CMS 8 TeV jets $R = 0.7$	174	1.64 (1.73)	1.50 (1.59)
ATLAS 8 TeV Z p_T	104	2.26 (2.31)	1.81 (1.59)
$\sigma_{t\bar{t}}$	17	1.34 (1.39)	0.85 (0.87)
ATLAS 8 TeV $t\bar{t} \to l + j$ sd	25	1.56 (1.50)	1.02 (1.15)
ATLAS 8 TeV $t\bar{t} \to l^+l^-$ sd	5	0.94 (0.82)	0.68 (1.11)
CMS 8 TeV $(d\sigma_{\bar{t}t}/dp_{T,t}dy_t)/\sigma_{\bar{t}t}$	15	2.19 (2.20)	1.50 (1.48)
CMS 8 TeV $d\sigma_{\bar{t}t}/dy_t$	9	1.43 (1.02)	1.47 (2.14)
Total, LHC data in MSHT20	1328	1.79 (2.18)	1.33 (1.77)
Total, non-LHC data in MSHT20	3035	1.13 (1.18)	1.10 (1.18)
Total, all data	4363	1.33 (1.48)	1.17 (1.36)

New data χ^2/N_{pts} MSHT20 fit qualities (MMHT14 prediction central fit qualities).

More information to determine PDFs.

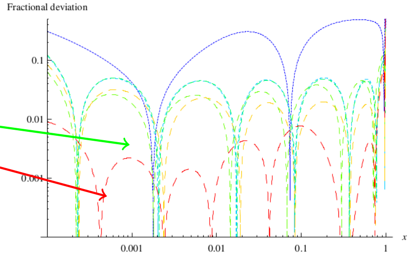
Clear preference for NNLO in new precision LHC data, NLO no longer sufficient.

- Overall good fit quality achieved, including for individual datasets.

More information in our MSHT20 paper: arXiv:2012.04684, *Eur.Phys.J.C 81 (2021) 4, 341*

MSHT20 extension of parameterisation

- MSHT use Chebyshev polynomials $T_i(1-2x^{0.5})$ to parameterise PDFs.
- MMHT used 4 Chebyshevs, MSHT now uses 6 Chebyshevs $\Rightarrow$ enables fitting to $< 1\%$ if data allows.
- Parameterise $\bar{d}/\bar{u}$ instead of $\bar{d}-\bar{u}$, with $\bar{d}/\bar{u} \to \mathrm{constant}$ as $x \to 0$.



MMHT: 1211.1215.

New parameterisation:

$u_v(x,Q_0^2) = A_u(1-x)^{\eta_u}x^{\delta_u}(1+\sum_{i=1}^{6} a_{i,u}T_i(1-2x^{\frac{1}{2}}))$; A_u fixed by $\int_0^1 u_v\,dx = 2$

$d_v(x,Q_0^2) = A_d(1-x)^{\eta_d}x^{\delta_d}(1+\sum_{i=1}^{6} a_{i,d}T_i(1-2x^{\frac{1}{2}}))$; A_d fixed by $\int_0^1 d_v\,dx = 1$

$sea(x,Q_0^2) = A_S(1-x)^{\eta_S}x^{\delta_S}(1+\sum_{i=1}^{6} a_{i,S}T_i(1-2x^{\frac{1}{2}}))$;

$s^+(x,Q_0^2) = A_s(1-x)^{\eta_s}x^{\delta_S}(1+\sum_{i=1}^{6} a_{i,s}T_i(1-2x^{\frac{1}{2}}))$; ($a_{i,s} \neq a_{i,S}, i = 5, 6$)

$(\bar{d}/\bar{u})(x,Q_0^2) = A_{\mathrm{rat}}(1-x)^{\eta_{\mathrm{rat}}}(1+\sum_{i=1}^{6} a_{i,\mathrm{rat}}T_i(1-2x^{\frac{1}{2}}))$;

$g(x,Q_0^2) = A_g(1-x)^{\eta_g}x^{\delta_g}(1+\sum_{i=1}^{4} a_{i,g}T_i(1-2x^{\frac{1}{2}})) - A_{g-}(1-x)^{\eta_{g-}}x^{\delta_{g-}}$;

$s^-(x,Q_0^2) = A_{s-}(1-x)^{\eta_{s-}}(1-x_o/x)x^{\delta_{s-}}$. x_0 fixed by $\int_0^1 s^-\,dx = 0$, δ_{s-} fixed.

51 parton parameters

(36 in MMHT14)

7 extra eigenvectors

- 1 extra in each of PDFs, except in s^-, 2 extra in s^+.

Net $\Delta\chi^2_{\mathrm{global}} = -73$.

More accurate and precise description.

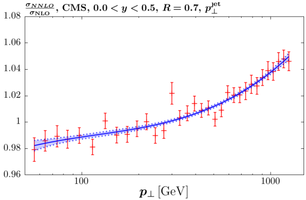
MSHT20: 2012.04684

Theoretical Developments - NNLO QCD

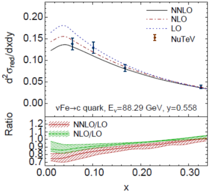
- Nearly all data now full NNLO in QCD, typically via k-factors relative to NLO grids. ↗Work on NNLO by Czakon et al, 2011.01011, JHEP 06 (2021) 100.
- Exception is CMS 7 TeV $W + c$ data only have NLO theory.
- Fit quality shows clear preference for NNLO over NLO now.

Data	N_{pts}	NLO χ^2/N_{pts}	NNLO χ^2/N_{pts}
Total, LHC data in MSHT20	1328	1.79	1.33
Total, non-LHC data in MSHT20	3035	1.13	1.10
Total, all data	4363	1.33	1.17

- K-factors smoothed with fit including adding MC error (MSHT20).
- Some data starting to be provided with NNLO grids - e.g. $t\bar{t}$.



Harland-Lang et al 1711.05757, Eur.Phys.J.C 78 (2018) 3, 248.

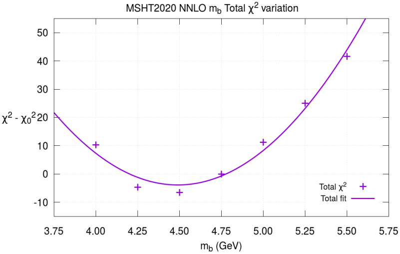


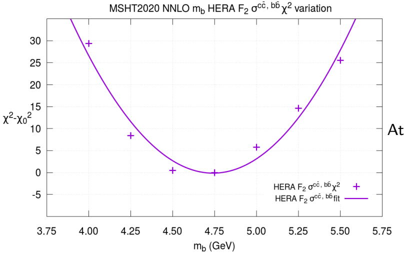
Thorne et al 1907.08147, PoS DIS2019 (2019) 036.

Greater theoretical accuracy.

MSHT20 m_b dependence at NNLO

- Default bottom (pole) mass $m_b = 4.75$ GeV, vary in steps of 0.25GeV in range $4.0\text{GeV} \leq m_b \leq 5.5\text{GeV}$ and examine fit qualities.





At fixed $\alpha_S(M_Z^2) = 0.118$

- Overall global fit dependence (left) centred on $m_b \approx 4.5\text{GeV}$.
- HERA heavy flavour combined charm and bottom (right) prefer bottom mass very close to our default $m_b = 4.75\text{GeV}$.
- Very low values of m_b clearly disfavoured, in contrast to MMHT14.

Motivation for inclusion of QED effects:

- With NNLO QCD now standard, noting that:

$$\alpha_{\mathrm{QED}}(M_Z) \sim \alpha_S^2(M_Z)$$

$\Rightarrow$ important to consider EW effects, QED corrections are a key part.

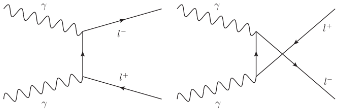
- QED corrections enter via QED modifications to DGLAP evolution:

$$P_{ij}^{\mathrm{QED}} = \frac{\alpha}{2\pi} P_{ij}^{0,1} + \frac{\alpha \alpha_S}{(2\pi)^2} P_{ij}^{1,1} + \frac{\alpha^2}{(2\pi)^2} P_{ij}^{0,2} + \ldots$$

$\Rightarrow$ Include $\mathcal{O}(\alpha)$, $\mathcal{O}(\alpha\alpha_S)$, $\mathcal{O}(\alpha^2)$ corrections.

- Requires also introduction of photon PDF, photon-initiated (PI) channels provide important QED corrections.

- MSHT20 include EW corrections for:
  - Drell-Yan
  - top
  - inclusive jets
  - DIS.



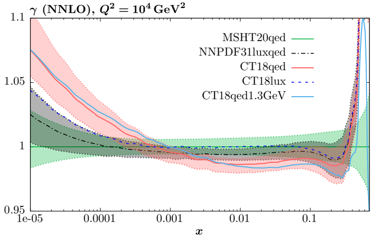
PI contributions to Drell-Yan.

Photon PDF in MSHT20qed:

- Obtain photon from experimentally well-measured NC proton structure functions, à la LUXQED. Manohar et al, 1708.01256, *JHEP 12, 046 (2017)*.

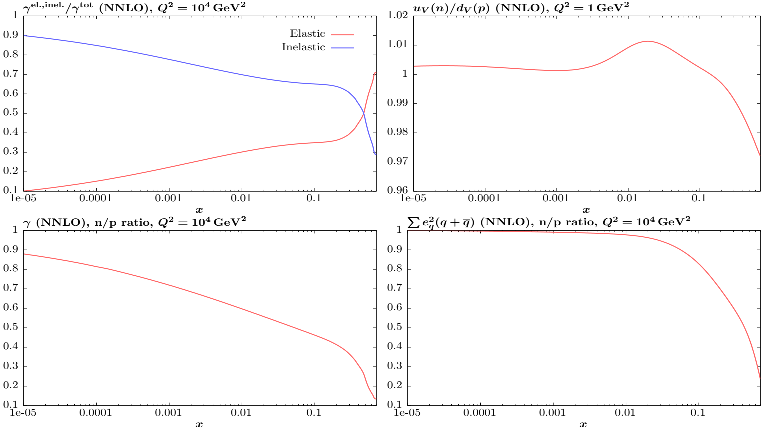
$$x\gamma(x,Q_0^2) = \frac{1}{2\pi\alpha(Q_0^2)}\int_x^1 \frac{dz}{z}\left\{\int_{\frac{x^2m_p^2}{1-z}}^{Q_0^2}\frac{dQ^2}{Q^2}\alpha^2(Q^2)\left[\left(zP_{\gamma,q}(z)+\frac{2x^2m_p^2}{Q^2}\right)F_2(x/z,Q^2)\right.\right.$$
$$\left.\left.-z^2F_L(x/z,Q^2)\right] - \alpha^2(Q_0^2)\left(z^2+\ln(1-z)zP_{\gamma,q}(z)-\frac{2x^2m_p^2z}{Q_0^2}\right)F_2(x/z,Q_0^2)\right\},$$

- $\gamma(x, Q_0^2)$ extracted from experimental data and then evolved in QED-modified DGLAP $\Rightarrow$ $\gamma(x, Q^2)$ with %-level uncertainties.
- General consistency compared to NNPDF, CT.
- Low x difference reflects differing charge-weighted singlet.
- High x difference may relate to inherent differences in methodology.



MSHT20qed - elastic/inelastic and neutron PDFs

- Breakdown of photon into elastic and inelastic pieces also provided, former dominates except at high x and low Q^2 (upper left).
- Neutron PDFs also provided as QED corrections lead to isospin violation: $u_V(p) \neq d_V(n)$, $u_V(n) \neq d_V(p)$, etc $\Rightarrow \gamma(p) \neq \gamma(n)$.



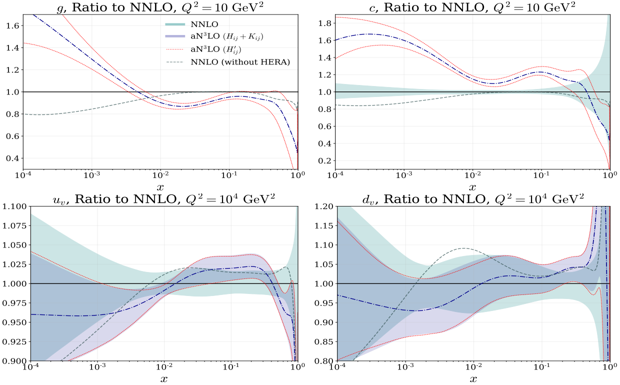
MSHT20aN3LO PDFs - Fit quality

- Smooth improvement and convergence in fit quality with increasing order.
- Fit quality improves by $\Delta\chi^2 = -172.5$ for 20 extra parameters.
- Reduction in tension between low and high x, HERA and fixed target fit better.
- ATLAS 8 TeV Zp_T improves significantly, reduction in tension with other data.
- Jets are only class of data with worsening of χ^2, looks better with dijet data (preliminary).

Order	LO	NLO	NNLO	aN3LO
χ^2/N_{pts}	2.57	1.33	1.17	1.13

Data set	Points	MSHT20aN3LO χ^2	$\Delta\chi^2$ from NNLO
HERA e^+p CC	39	51.8	-0.1
HERA e^-p CC	42	66.3	-3.8
HERA e^+p NC 820GeV	75	83.8	-6.0
HERA e^-p NC 460GeV	209	247.4	-0.9
HERA e^+p NC 920GeV	402	476.7	-36.0
HERA e^-p NC 575GeV	259	248.0	-15.0
HERA e^-p NC 920GeV	159	243.3	-1.0
CCFR $\nu N \to \mu\mu X$	86	69.2	+1.5
NuTeV $\nu N \to \mu\mu X$	84	55.3	-3.1
CMS double diff. DY	132	137.1	-7.4
ATLAS 7 TeV W, Z	61	110.5	-6.2
ATLAS 8 TeV W	22	55.1	-2.3
ATLAS 8 TeV Z	59	80.8	-4.8
ATLAS 8 TeV Zp_T	104	105.8	-82.7
CMS 7 TeV $W+c$	10	12.3	+3.7
ATLAS 8 TeV W+jets	30	19.1	+0.9
ATLAS 7 TeV jets	140	214.5	-7.1
CMS 7 TeV jets	158	189.8	+14.1
CMS 8 TeV jets	174	272.6	+11.3
CMS 2.76 TeV jets	81	113.9	+11.1
DIS data (total)	2375	2585.2	-86.4
Jets data (total)	739	972.9	+30.8
Top data (total)	71	73.4	-5.9
DY data (total)	864	1044.8	-43.1
Total	4363	4948.6	-172.5

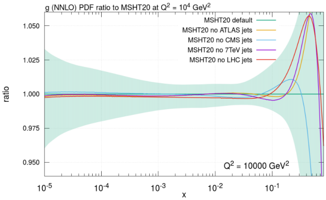
MSHT20aN3LO PDFs - PDF changes



- Small-x low-Q^2 gluon enhanced due to large logs included at N3LO.
- Enhanced charm via enlarged $A_{Hg}^{(3)}$ and increased small-x gluon.
- Reduced quarks at large/small-x accommodate small-x gluon.
- High-Q^2, intermediate/large-x light quarks largely follow NNLO no HERA fit, demonstrating eased tension with smaller x HERA data.

New data - Dijets - Introduction

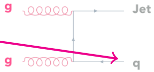
- High x gluon is of interest in PDFs, with tensions between datasets.
- MSHT20 - data on inclusive jets from ATLAS, CMS at 7 and 8 TeV, sensitive to high-x gluon. Different pulls.
- Known issues with systematic correlations in ATLAS 7, 8 TeV inclusive jets (latter therefore not included in MSHT20).



- Theoretical issues: scale choice, non-unitary nature of inclusive jets.
- Dijets also allow triple differential measurement, cf double differential for single inclusive jets. Schematically at LO:

$$x = \frac{p_T}{\sqrt{s}}(e^{y_j} + e^{y_{j'}})$$

Integrated over in inclusive jet case.



$\Rightarrow$ Single inclusive jets: $\frac{d\sigma}{dp_T^j d|y^j|}$, dijets: $\frac{d\sigma}{dp_T^{\rm avg} dy^* dy_b}$.

CMS 8 TeV dijets

Dijets when triple differential more sensitive to x-dependence.

New data - EIC Pseudodata

EIC: Future Constraints?

- Recent study presented at DIS22:

- Detailed simulation work to optimise resolutions throughout phase-space
→ 5 bins per decade in x and Q^2

- Kinematic coverage: $Q^2 > 1$ GeV^2, $0.01 < y < 0.95$, $W > 3$ GeV

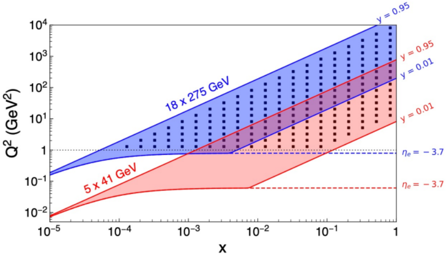
- Lower y accessible in principle, but easier to rely on overlaps between data at different $\sqrt{s}$

- Highest x bin centre at x=0.815

e-beam E	p-beam E	$\sqrt{s}$ (GeV)	inte. Lumi. (fb^{-1})
18	275	140	15.4
10	275	105	100.0
10	100	63	79.0
5	100	45	61.0
5	41	29	4.4

- CC data also included for highest $\sqrt{s}$

5



- Including sensible projections for main uncertainty sources.

→ 1.5-2.5% point-to-point uncorrelated
→ 2.5% normalisation (uncorrelated between different $\sqrt{s}$)

P. Newman, DIS22