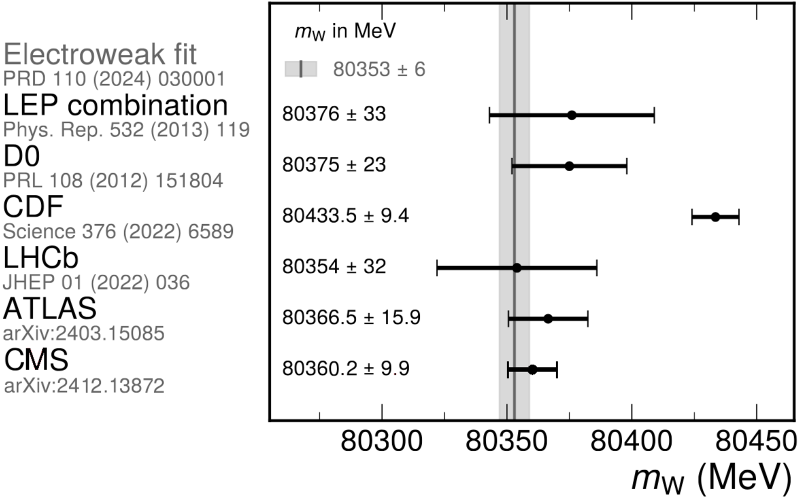
Collinear PDFs for M_W determination

GDR QCD workshop on W mass

Emanuele R. Nocera
Università degli Studi di Torino and INFN, Torino

30 June 2025

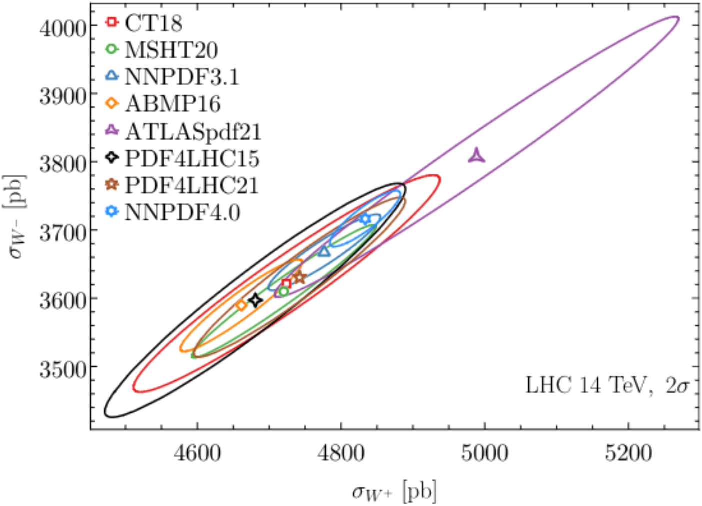
PDFs and recent M_W determinations



[Adapted from arXiv:2412.13872]

PDF uncertainty non-negligible on the final M_W uncertainty
e.g. on the CMS result $\delta_{\rm tot} = 9.9$ MeV and $\delta_{\rm PDF} = 4.4$ MeV

Making predictions with PDFs

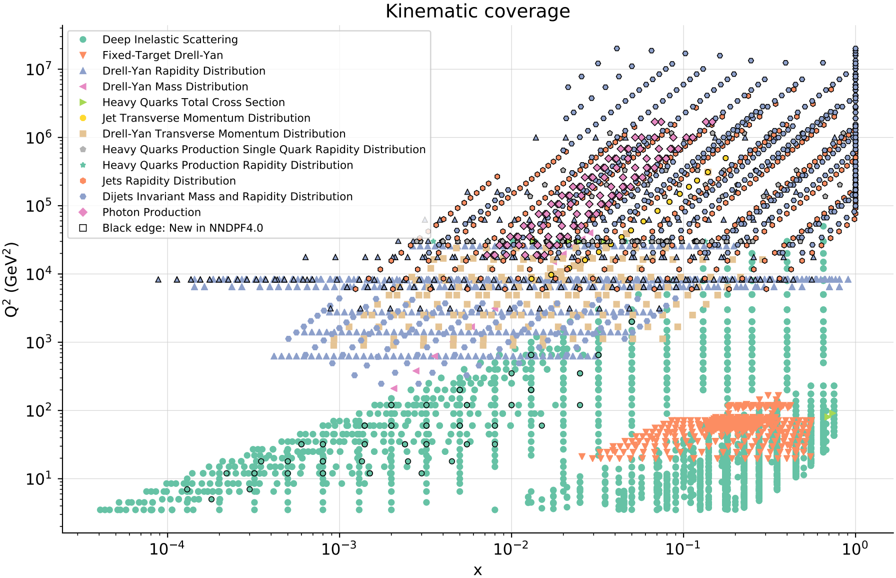


[Acta Phys.Polon.B 53 (2022) 12]

Predictions obtained with different PDF sets display significantly different uncertainties
Where do these come from (data, theory, methodology)?

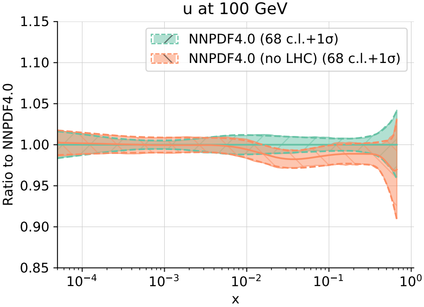
1. Data

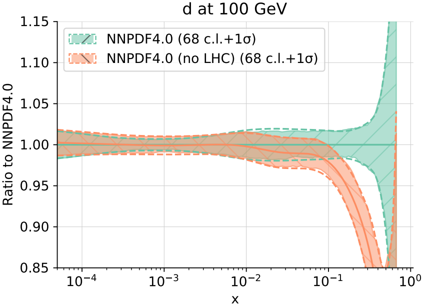
Available cross setions and their kinematic coverage

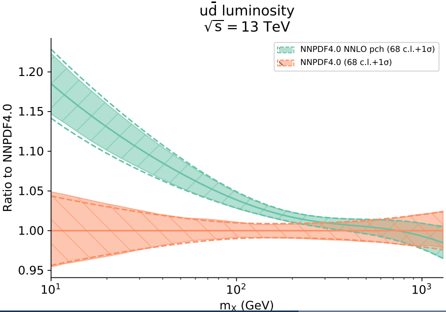


Increasing relevance of LHC data in PDF determination

Quark flavour separation







Perturbative charm alters the flavour decomposition and deteriorates the fit

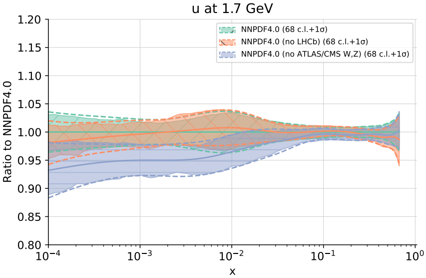
$$\chi^2_{\text{fitted charm}} = 1.17 \rightarrow \chi^2_{\text{pert. charm}} = 1.19$$

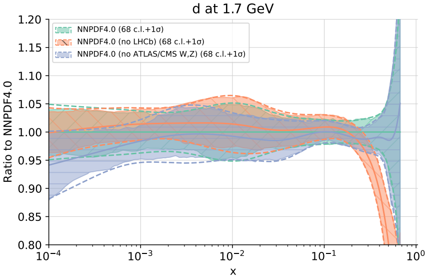
mainly due to a worsening of the LHC W, Z and top pair data sets

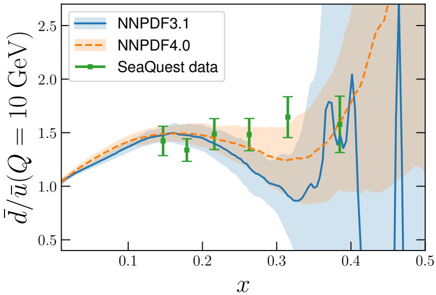
fitting charm reduces the dependence from m_c

[EPJ C76 (2016) 647; C77 (2017) 663; C82 (2022) 428]

Quark flavour separation







Relative impact of ATLAS/CMS/LHCb gauge boson production

LHCb is at forward rapidity

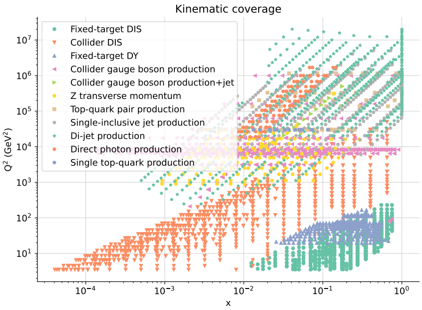
New constraint on $\bar{d}/\bar{u}$ ratio from SeaQuest

[Nature 590 (2021) 561]

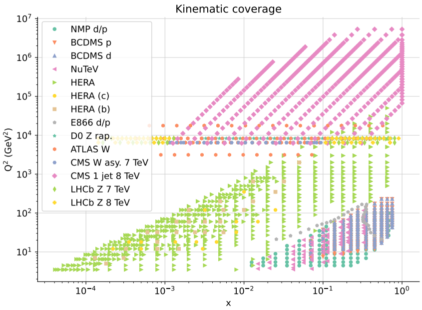
Studied by CT, MSHT, NNPDF, ABMP
Some tension with NuSea found

The PDF4LHC21 benchmark [JPG 49 (2022) 080501]

Global data set ($N_{\rm dat} \sim 5000$)



Reduced data set ($N_{\rm dat} \sim 2000$)



Same reduced data set

Same heavy quark mass values: $m_c = 1.4$ GeV; $m_b = 4.75$ GeV; $\alpha_s(M_Z) = 0.118$

Same strong coupling value: $\alpha_s(M_Z) = 0.118$

No strangeness asymmetry at input scale: $(s - \bar{s})(Q_0) = 0$

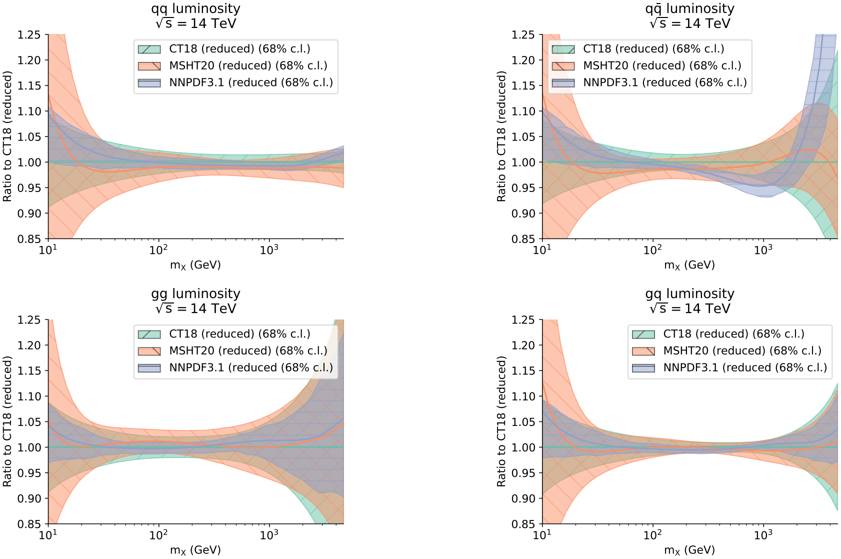
Charm perturbatively generated

Positive-definite quark distributions

No deuteron or nuclear corrections to analyse νN cross sections

Fixed branching ratio for charm hadrons to muons; NNLO corrections for dimuon data

The PDF4LHC21 benchmark [JPG 49 (2022) 080501]

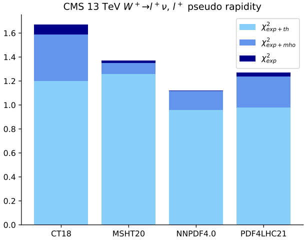


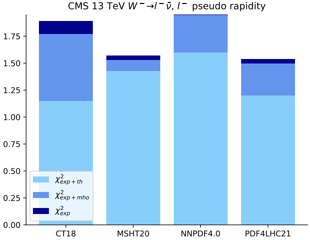
Very good agreement within uncertainties; similar size of uncertainties in the data region
Remaining differences reflect methodological choices

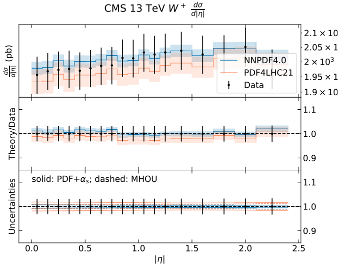
How well do different PDF sets generalise on unseen data?

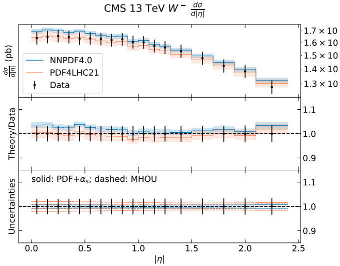
Example: the CMS W rapidity measurement [PRD 102 (2020) 092012]

The covariance matrix include exp, PDF, MHOU, and α_s uncertainties [arXiv:2501.10359]









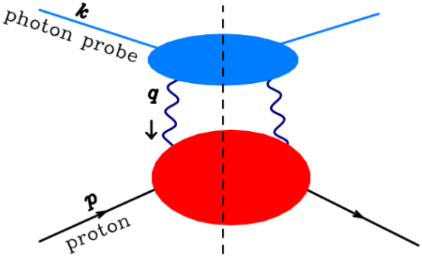
2. Theory

QED corrections in PDF determination: LuxQED photon

LUXQED [PRL 117 (2016) 242002]

View the $ep \to e + X$ process as an electron scattering off the photon field of the proton

Consider a BSM process, *e.g.* production of a heavy supersymmetric lepton L in ep collision, write the cross section in terms of structure functions and of f_γ, and equate the two to obtain f_γ



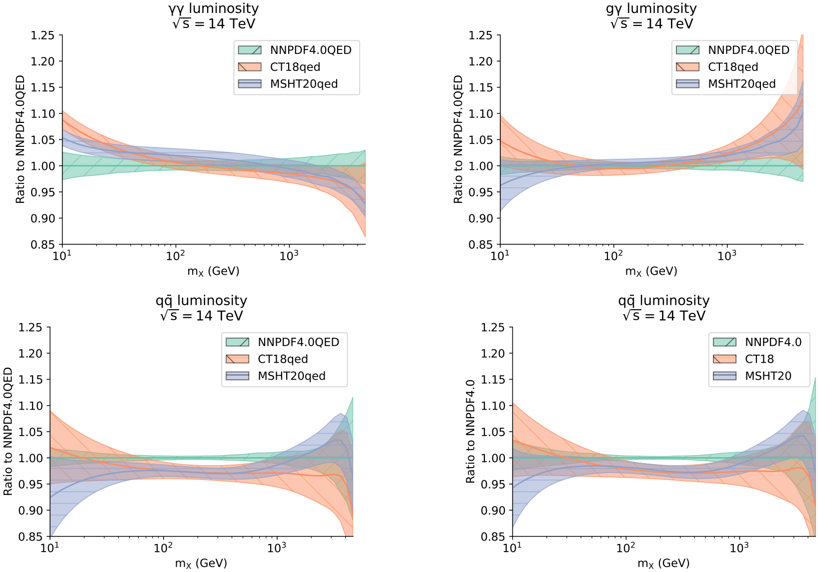
$$\sigma = c_0 \sum_a \int_x^1 \frac{dz}{z}\, \hat{\sigma}_a(z,\mu^2)\, \frac{M^2}{zs} f_{a/p}\left(\frac{M^2}{zs},\mu^2\right)$$

$$\sigma = \frac{c_0}{2\pi} \int_x^{1-\frac{2xm_p}{M}} \frac{dz}{z} \int_{Q^2_{\min}}^{Q^2_{\max}} \frac{dQ^2}{Q^2} \alpha^2_{\mathrm{ph}}(-Q^2) \Bigg[\Bigg(2-2z+z^2 + \frac{2x^2m_p^2}{Q^2} + \frac{z^2Q^2}{M^2} - \frac{2zQ^2}{M^2} - \frac{2x^2Q^2m_p^2}{M^4}\Bigg) F_2(x/z,Q^2) + \left(-z^2 - \frac{z^2Q^2}{2M^2} + \frac{z^2Q^4}{2M^4}\right) F_L(x/z,Q^2)\Bigg]$$

$$xf_{\gamma/p}(x,\mu^2) = \frac{1}{2\pi\alpha(\mu^2)} \int_x^1 \frac{dz}{z} \Bigg\{ \int_{\frac{x^2m_p^2}{1-z}}^{\frac{\mu^2}{1-z}} \frac{dQ^2}{Q^2} \alpha^2(Q^2) \left[\left(zp_{\gamma q}(z) + \frac{2x^2m_p^2}{Q^2}\right) F_2(x/z,Q^2) - z^2 F_L\left(\frac{x}{z},Q^2\right)\right] - \alpha^2(\mu^2) z^2 F_2\left(\frac{x}{z},\mu^2\right) \Bigg\}$$

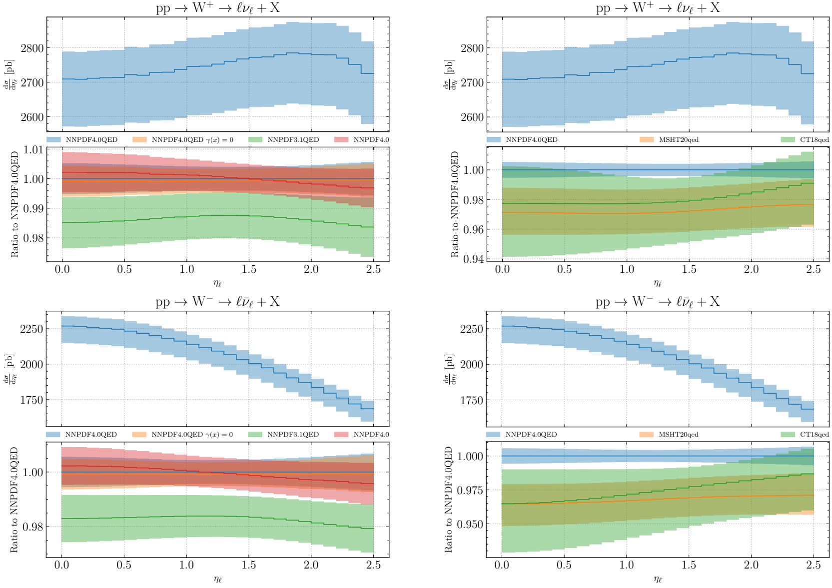
Iterate a QCD fit including f_γ in DGLAP and in the momentum sum rule

QED corrections in PDF determination: LuxQED photon



General agreement in $\mathcal{L}_{\gamma\gamma}$ and $\mathcal{L}_{\gamma g}$; differences in $\mathcal{L}_{q\bar{q}}$ independent of photon PDF

Photon PDF: impact on $W^{\pm}$ differential cross sections



PDF4LHC benchmark and combination ongoing

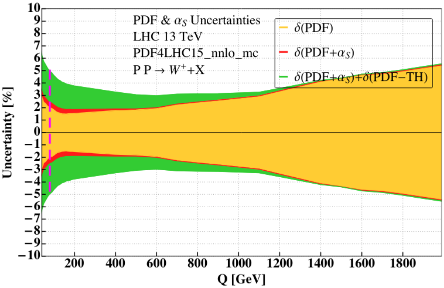
N^{3}LO QCD corrections in PDF determination

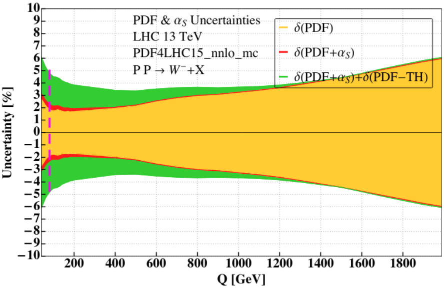
NNLO is the precision frontier for PDF determination

N3LO is the precision frontier for partonic cross sections

Mismatch between perturbative order of partonic cross sections and accuracy of PDFs is becoming a significant source of uncertainty

$$\hat{\sigma} = \alpha_s^p \hat{\sigma}_0 + \alpha_s^{p+1}\hat{\sigma}_1 + \alpha_s^{p+2}\hat{\sigma}_2 + \mathcal{O}(\alpha_s^{p+3}) \qquad \delta(\mathrm{PDF}-\mathrm{TH}) = \frac{1}{2}\left|\frac{\sigma^{(2)}_{\mathrm{NNLO-PDFs}} - \sigma^{(2)}_{\mathrm{NLO-PDFs}}}{\sigma^{(2)}_{\mathrm{NNLO-PDFs}}}\right|$$





[JHEP 11 (2020) 143]

N^{3}LO QCD corrections in PDF determination

Splitting Functions

Singlet (P_{qq}, P_{gg}, P_{gq}, P_{qg})

– large-n_f limit [NPB 915 (2017) 335; arXiv:2308.07958]

– small-x limit [JHEP 06 (2018) 145]

– large-x limit [NPB 832 (2010) 152; JHEP 04 (2020) 018; JHEP 09 (2022) 155]

– 5 (10) lowest Mellin moments [PLB 825 (2022) 136853; ibid. 842 (2023) 137944; ibid. 846 (2023) 138215]

Non-singlet ($P_{NS,v}$, $P_{NS,+}$, $P_{NS,-}$)

– large-n_f limit [NPB 915 (2017) 335; arXiv:2308.07958]

– small-x limit [JHEP 08 (2022) 135]

– large-x limit [JHEP 10 (2017) 041]

– 8 lowest Mellin moments [JHEP 06 (2018) 073]

DIS structure functions (F_L, F_2, F_3)

– DIS NC (massless) [NPB 492 (1997) 338; PLB 606 (2005) 123; NPB 724 (2005) 3]

– DIS CC (massless) [Nucl.Phys.B 813 (2009) 220]

– massive from parametrisation combining known limits and damping functions [NPB 864 (2012) 399]

PDF matching conditions

– all known except for $a^3_{H,g}$ [NPB 820 (2009) 417; NPB 886 (2014) 733; JHEP 12 (2022) 134]

Coefficient functions for other processes

– DY (inclusive) [JHEP 11 (2020) 143]; DY (y differential) [PRL 128 (2022) 052001]

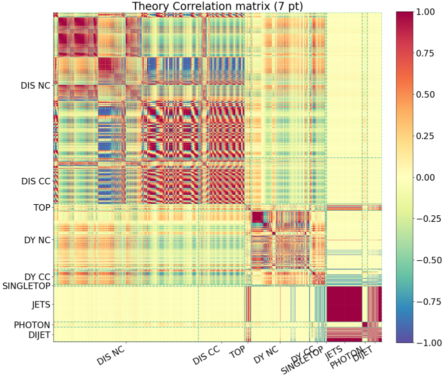
Theory uncertainties in PDF determination

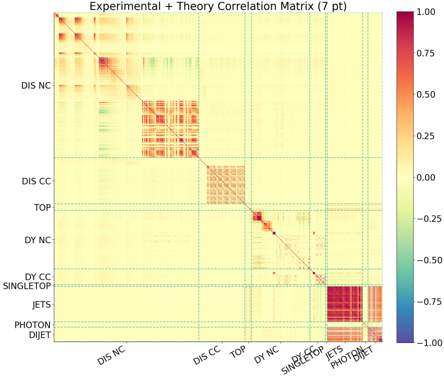
Assuming that theory uncertainties are (a) Gaussian and (b) independent from experimental uncertainties, modify the figure of merit to account for theory errors

$$\chi^2 = \sum_{i,j}^{N_{\rm dat}} (D_i - T_i)(\mathrm{cov}_{\rm exp} + \mathrm{cov}_{\rm th})^{-1}_{ij}(D_j - T_j);\ (\mathrm{cov}_{\rm th})_{ij} = \frac{1}{N}\sum_k^N \Delta_i^{(k)}\Delta_j^{(k)};\ \Delta_i^{(k)} \equiv T_i^{(k)} - T_i$$

Problem reduced to estimate the th. cov. matrix, *e.g.* in terms of nuisance parameters

$$\Delta_i^{(k)} = T_i(\mu_R, \mu_F) - T_i(\mu_{R,0}, \mu_{F,0});\ \text{vary scales in } \frac{1}{2} \leq \frac{\mu_F}{\mu_{F,0}}, \frac{\mu_R}{\mu_{R,0}} \leq 2$$





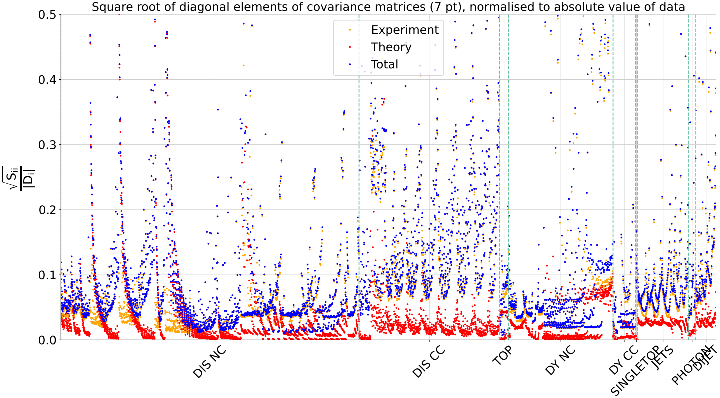
Theory uncertainties in PDF determination

Assuming that theory uncertainties are (a) Gaussian and (b) independent from experimental uncertainties, modify the figure of merit to account for theory errors

$$\chi^2 = \sum_{i,j}^{N_{\rm dat}} (D_i - T_i)(\mathrm{cov}_{\rm exp} + \mathrm{cov}_{\rm th})^{-1}_{ij}(D_j - T_j);\ (\mathrm{cov}_{\rm th})_{ij} = \frac{1}{N}\sum_k^N \Delta_i^{(k)}\Delta_j^{(k)};\ \Delta_i^{(k)} \equiv T_i^{(k)} - T_i$$

Problem reduced to estimate the th. cov. matrix, *e.g.* in terms of nuisance parameters

$$\Delta_i^{(k)} = T_i(\mu_R, \mu_F) - T_i(\mu_{R,0}, \mu_{F,0});\ \text{vary scales in } \frac{1}{2} \le \frac{\mu_F}{\mu_{F,0}}, \frac{\mu_R}{\mu_{R,0}} \le 2$$



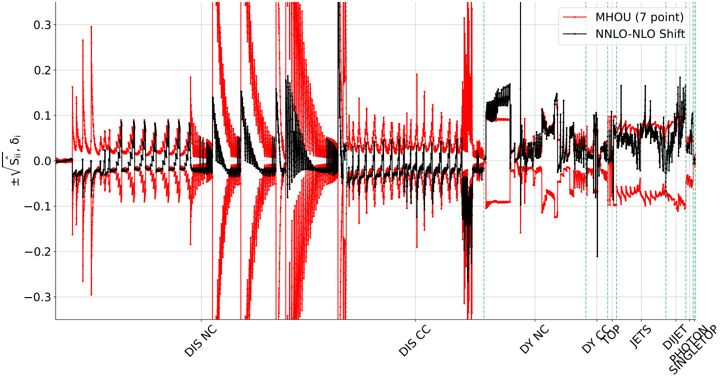
Theory uncertainties in PDF determination

Assuming that theory uncertainties are (a) Gaussian and (b) independent from experimental uncertainties, modify the figure of merit to account for theory errors

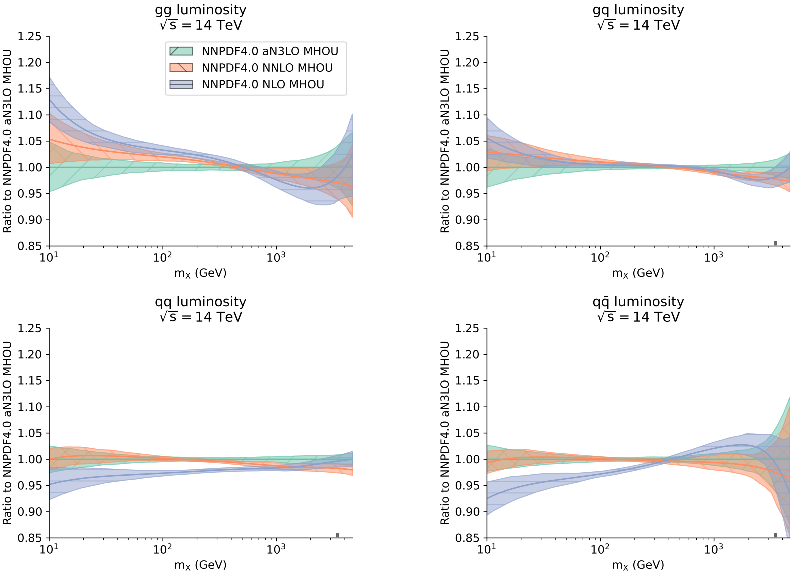
$$\chi^2 = \sum_{i,j}^{N_{\rm dat}} (D_i - T_i)(\mathrm{cov}_{\rm exp} + \mathrm{cov}_{\rm th})^{-1}_{ij}(D_j - T_j);\ (\mathrm{cov}_{\rm th})_{ij} = \frac{1}{N}\sum_k^N \Delta_i^{(k)}\Delta_j^{(k)};\ \Delta_i^{(k)} \equiv T_i^{(k)} - T_i$$

Problem reduced to estimate the th. cov. matrix, *e.g.* in terms of nuisance parameters

$$\Delta_i^{(k)} = T_i(\mu_R, \mu_F) - T_i(\mu_{R,0}, \mu_{F,0});\ \text{vary scales in } \frac{1}{2} \leq \frac{\mu_F}{\mu_{F,0}}, \frac{\mu_R}{\mu_{R,0}} \leq 2$$

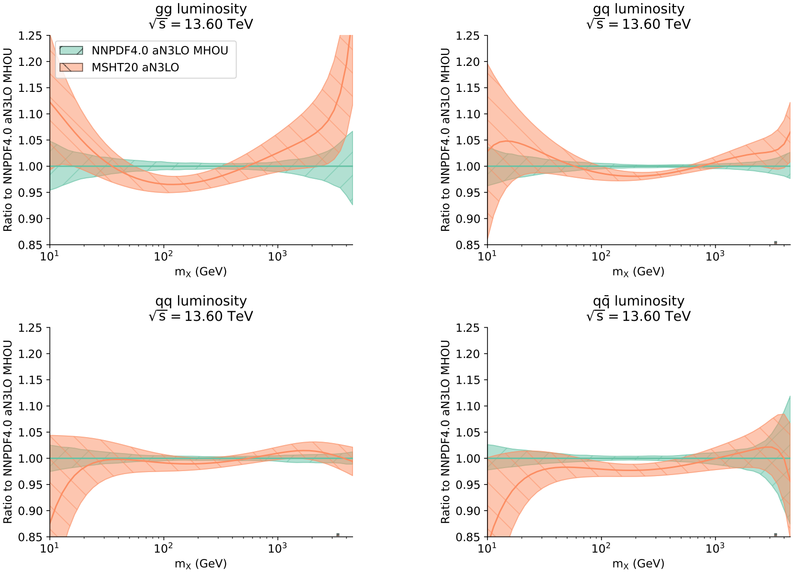


Impact of aN^{3}LO corections on partonic luminositites



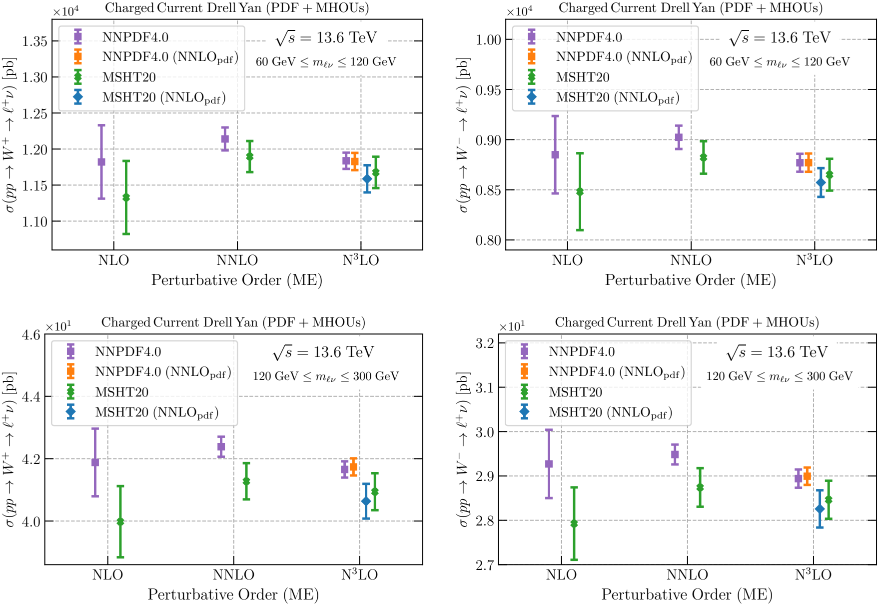
Trend observed for NNPDF4.0 [EPJ C84 (2024) 659] similar to MSHT20 [EPJ C83 (2023) 185]

aN^{3}LO PDFs: MSHT20 *vs* NNPDF4.0

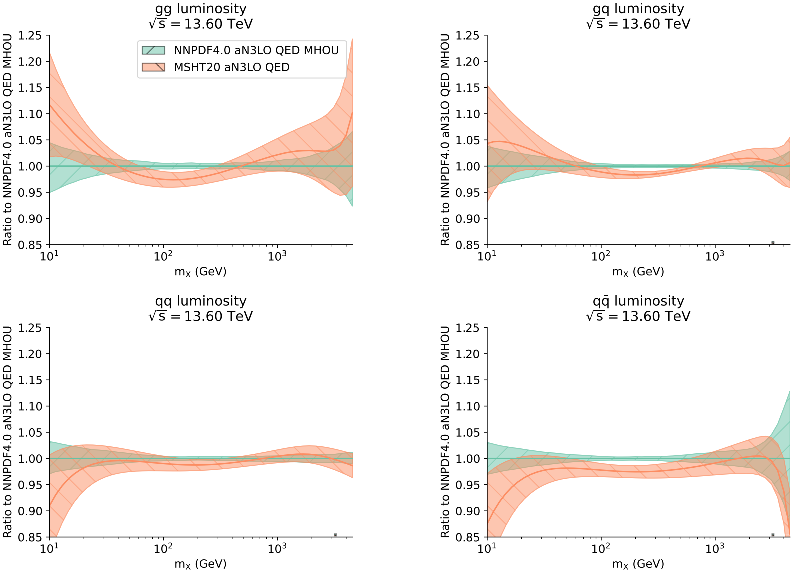


A 3% shift for $M_X \sim 100$ GeV, already at NNLO, partly sensitive to higher moments

aN^{3}LO PDFs: impact on $W^\pm$ total cross sections

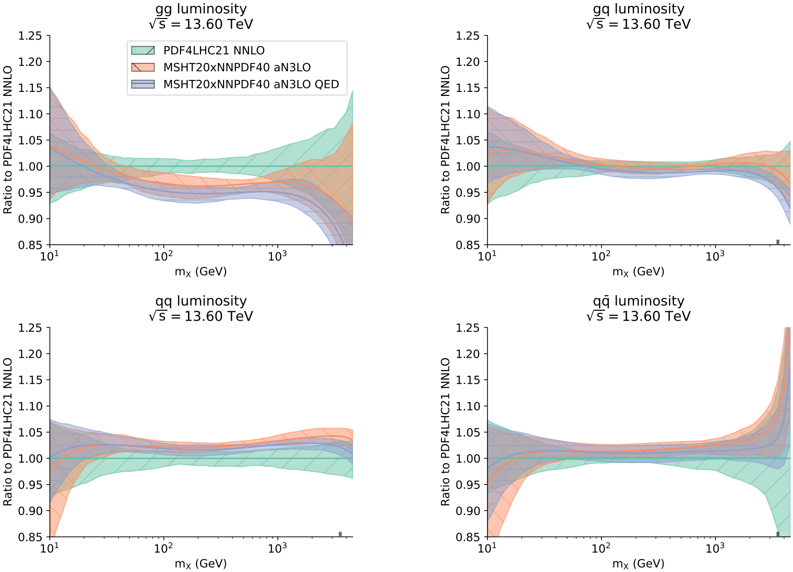


Combining QED and QCD higher order corrections



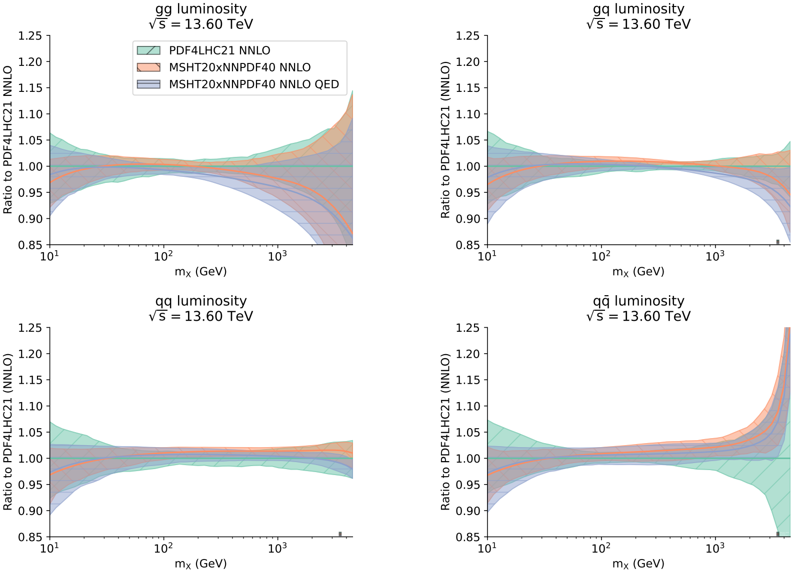
QED corrections slightly reduce discrepancies between the MSHT20 and NNPDF4.0

Combining MSHT20 and NNPDF4.0 aN^{3}LO QED PDFs



Statistical combination performed as in PDF4LHC21 [JPG 52 (2025) 065002]

Combining MSHT20 and NNPDF4.0 NNLO QED PDFs



Percent impact of aN^{3}LO corrections; comparatively small impact of QED corrections

3. Methodology

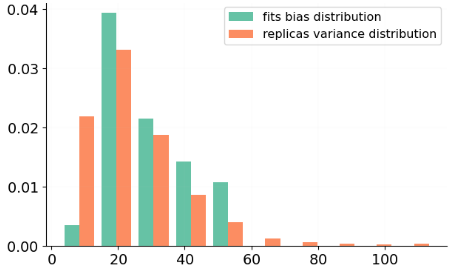
Validate PDF uncertainties: closure tests

Fit PDFs to pseudodata generated assuming a known underlying law

NNPDF

bias difference of central prediction and truth
variance uncertainty of replica predictions

If model complexity optimal,
generalisation error minimum
$(\mathrm{E[bias]})^2 = \text{variance}$
25 fits, 40 replicas each

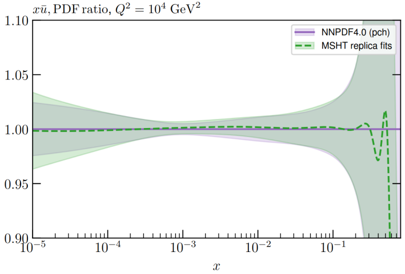


[EPJ C77 (2017) 663; EPJ C82 (2022) 330]

MSHT

parameterisation flexible enough to give faithful description of global pseudodata

NNPDF *vs* MSHT comparison
similar fit quality
similar errors, with MSHT unit tolerance



[EPJ C85 (2025) 316]

Inconsistent closure tests

Generate pseudodata with statistical and systematic uncertainties

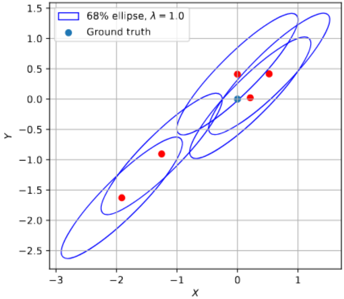
$C = C^{\rm stat} + C^{\rm syst}$ $\quad C^{\rm syst}_{ij} = \sum_k \Delta^k_i \Delta^k_j$ $\quad \Delta^k_i$ is the k-th sys. unc. for the i-th point

Assume systematic uncertainties are underestimated and perform a fit with

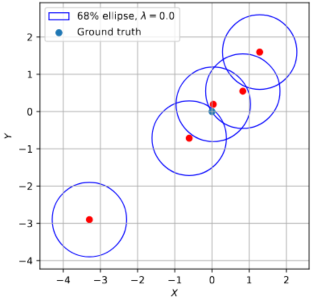
$\Delta^k_i \to \lambda \Delta^k_i$ $\quad \lambda = 1$ consistency $\quad \lambda = 0$ extreme inconsistency

PREDICTED UNCERTAINTY ON GENERATED DATA

CONSISTENT



EXTREME INCONSISTENCY



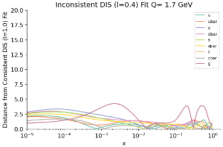
[arXiv:2503.17447]

Inconsistent closure tests

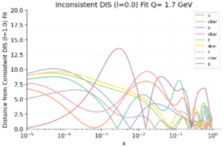
DISTANCES: $\dfrac{f^{\text{consistent}} - f^{\text{inconsistent}}}{\sigma^{\text{PDF}}/\sqrt{N_{\text{rep}}}}$; $d \sim 1 \Rightarrow$ STATISTICAL EQUIVALENCE

DIS: BULK INCONSISTENCY

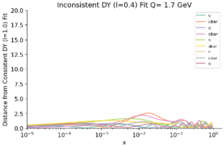
$\lambda = 0.4$: MODEL CORRECTS

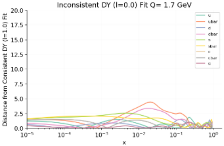


$\lambda = 0$: MODEL FAILS

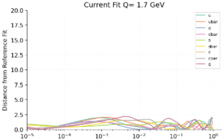


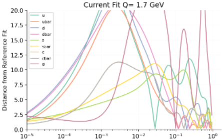
DY: SINGLE DATASET INCONSISTENCY





JETS: HIGH-IMPACT INCONSISTENCY





The ML model corrects for inconsistency except in extreme cases [arXiv:2503.17447]

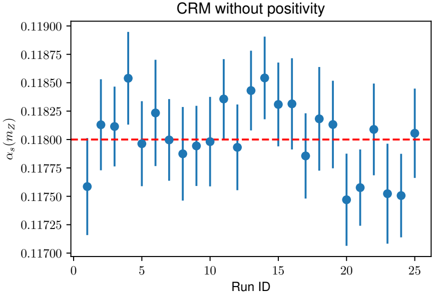
Simultaneous determination of PDFs and SM parameters

Example: α_s and PDFs are correlated, hence looking at the χ^2 profile as a function of α_s with fixed PDFs results in a bias [EPJ C80 (2020) 182]

Simultaneous fits of SM parameters and PDFs are costly

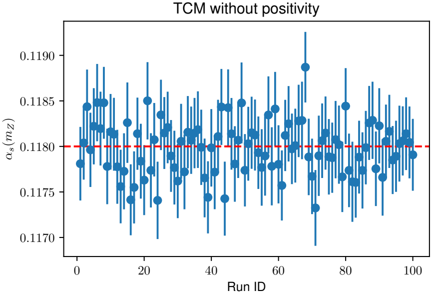
Correlated replica method [EPJ C78 (2018) 408]

Perform multiple fits of the same data replica, changing the value of α_s in each fit, thereby correlating PDFs and α_s



Theory covariance method [EPJ C81 (2021) 830]

Perform a single fit with a theory covariance matrix encoding a prior α_s distribution, and determine α_s from the Bayesian posterior



Closure test passed by both methodologies. Apply them to real data to get $\alpha_s(M_Z)$

$\alpha_s(M_Z) = 0.1194^{+0.0007}_{-0.0014}$ at aN^{3}LO$_{\rm QCD}\otimes$NLO$_{\rm QED}$ [arXiv:2506.13871]

4. Conclusions

Summary

A precise and accurate determination of PDFs is key to do precision phenomenology.

LHC measurements are being instrumental to reduce PDF uncertainties to few percent.

This is not enough.

The goal of achieving PDF determinations accurate to 1% opens up some challenges.

Understand the interplay between data, theory, and methodology into PDF uncertainties.

Refine the theoretical accuracy of a PDF determination.

Represent theory uncertainties into PDF uncertainties.

Deploy a robust fitting methodology and good statistical tests of it.

Benchmark efforts may benefit from public releases of PDF codes and inputs.

Summary

A precise and accurate determination of PDFs is key to do precision phenomenology.

LHC measurements are being instrumental to reduce PDF uncertainties to few percent.

This is not enough.

The goal of achieving PDF determinations accurate to 1% opens up some challenges.

Understand the interplay between data, theory, and methodology into PDF uncertainties.

Refine the theoretical accuracy of a PDF determination.

Represent theory uncertainties into PDF uncertainties.

Deploy a robust fitting methodology and good statistical tests of it.

Benchmark efforts may benefit from public releases of PDF codes and inputs.

Thank you