# Hot topics in PDF fits
# (The path towards 1% PDF uncertainties)

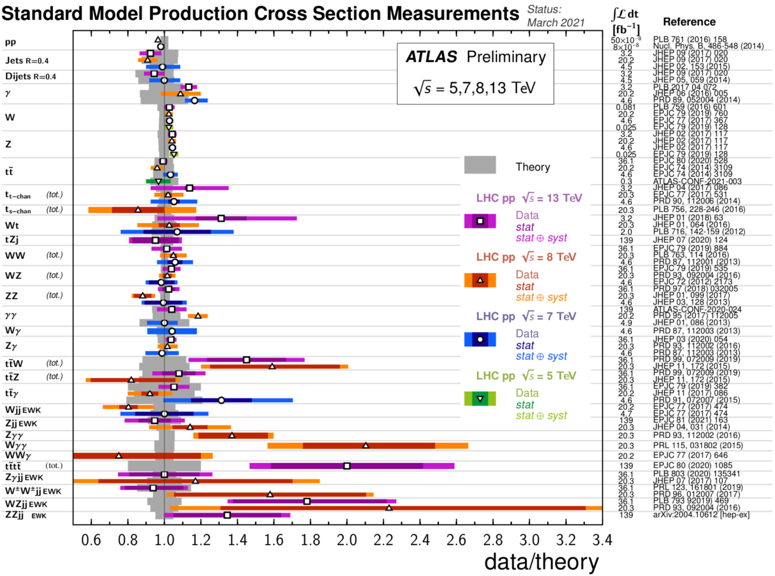
Standard Model at the LHC 2021

Emanuele R. Nocera

School of Physics and Astronomy, The University of Edinburgh

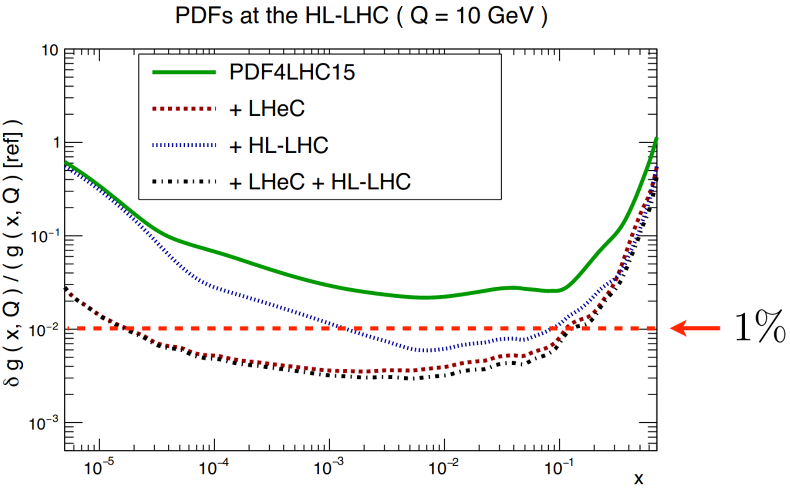
April 26, 2021

# Physics at the LHC as Precision Physics




[Plot from ATLAS Collaboration web page]

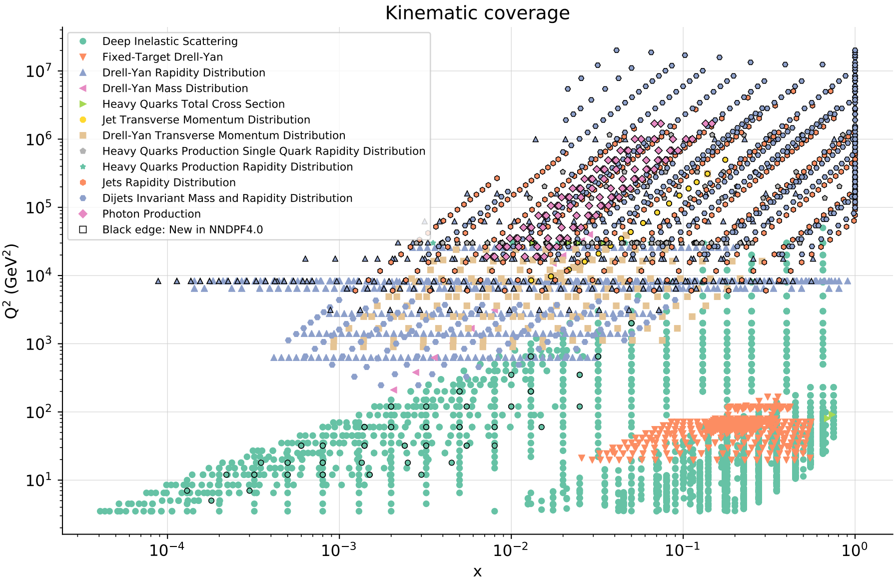
# Towards 1% PDF uncertainties




[SciPost Phys. 7 (2019) 051]

The path towards 1% PDF uncertainties goes through data, theory and methodology

# Overview of experimental data




Precision of the data of the order of percent; mostly from correlated systematic uncertainties

# Data consistency: neutrino DIS vs gauge boson production

| Process | Dataset | $n_{\rm dat}$ | $\chi^2_{\rm base}$ | $\chi^2_{\rm pr}$ | $\chi^2_{\rm str}$ |
|---|---|---|---|---|---|
| $\nu$DIS ($\mu\mu$) | | 76/76/95 | 0.70 | 0.71 | 0.53 |
| | NuTeV | 76/76/76 | 0.70 | 0.71 | 0.53 |
| | NOMAD | —/—/19 | [9.0] | [8.8] | 0.55 |
| $W$, $Z$ (incl.) | | 327/418/418 | 1.38 | 1.40 | 1.40 |
| | ATLAS | —/61/61 | 3.22 | 1.65 | 1.67 |
| $W+c$ | | —/37/37 | [0.76] | 0.68 | 0.60 |
| | CMS | —/15/15 | [1.10] | 0.98 | 0.96 |
| | ATLAS | —/22/22 | [0.53] | 0.48 | 0.42 |
| $W$+jets | ATLAS | —/32/32 | [1.58] | 1.18 | 1.18 |
| **Total** | | 3917/4077/4096 | 1.17 | 1.17 | 1.17 |

Satisfactory description of all datasets
no evidence for tensions

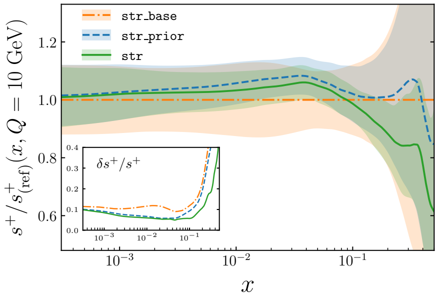
Sizeable constraint from NOMAD data
consistent with collider data

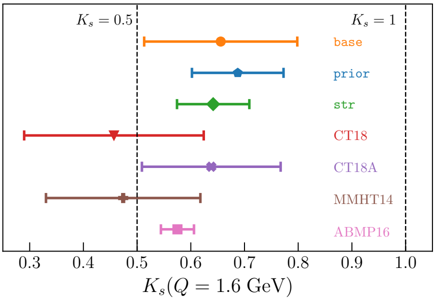
Moderate suppression of strange PDF

Good consistency of $K_s$ across PDF sets

$$K_s(Q^2) = \frac{\int_0^1 dx[s(x,Q^2)+\bar{s}(x,Q^2)]}{\int_0^1 dx[\bar{u}(x,Q^2)+\bar{d}(x,Q^2)]}$$

[EPJ C80 (2020) 1168]



# Data consistency: single-inclusive vs di-jet production

| Process | Dataset | $n_{\rm dat}$ | $\chi^2_{\rm base}$ | $\chi^2_{\rm 1j}$ | $\chi^2_{\rm 2j}$ |
|---|---|---|---|---|---|
| sin.-inc. jets | | | | | |
| | ATLAS 7 TeV | –/31/– | [1.87] | 1.59* | [1.63] |
| | ATLAS 8 TeV | –/171/– | [5.01] | 3.22* | [3.36] |
| | CMS 7 TeV | –/133/– | [1.06] | 1.09 | [1.06] |
| | CMS 8 TeV | –/185/– | [1.59] | 1.25 | [1.61] |
| di-jets | | | | | |
| | ATLAS 7 TeV | –/–/90 | [2.47] | [1.95] | 1.76 |
| | CMS 7 TeV | –/–/54 | [2.40] | [2.08] | 1.60 |
| | CMS 8 TeV | –/–/122 | [3.81] | [2.21] | 1.58 |
| | **Total** | | 1.20 | 1.18 | 1.22 |

*Become 1.22 and 0.98 with decorrelation models

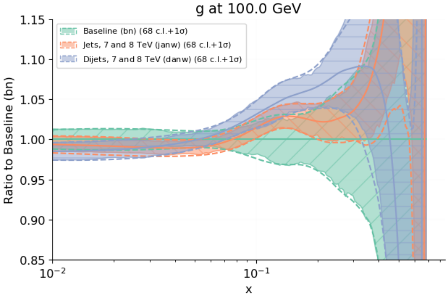
Good consistency of the two observables
similar impact on the gluon PDF

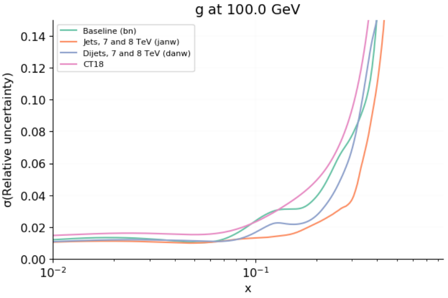
Single-inclusive jets: smaller uncertainties

Di-jets: larger enhancement at large $x$

Inclusion of di-jet is preferred over
single-inclusive jet measurements
given their greater theoretical accuracy and
the avoidance of decorrelation models

[EPJ C80 (2020) 797]



# Data inconsistency: experimental correlations

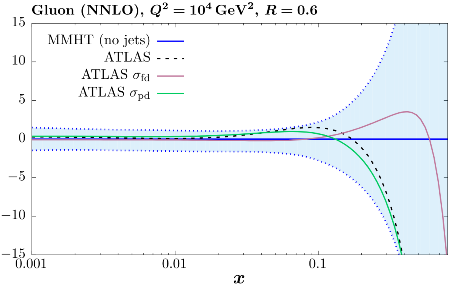
Single inclusive jet data from ATLAS 7 TeV

default correlations: terrible $\chi^2$
(correlations across rapidity bins)

decorrelation models: improve the fit a lot

| $n_{\rm dat}$ | default | part. decorr. | full decorr. |
|---|---|---|---|
| 140 | 1.89 | 1.28 | 0.83 |

no significant effect on the extracted gluon
similar gluon irrespective of the rapidity bin




[EPJ C78 (2018) 248; EPJ C80 (2020) 797]

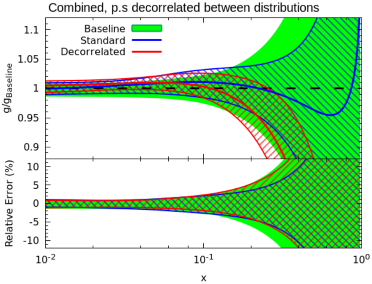
Top pair production from ATLAS 8 TeV

default correlations: terrible $\chi^2$
(correlations across different spectra)

decorrelation models: improve the fit a lot

| $n_{\rm dat}$ | default | stat. uncorr. | p.s. uncorr |
|---|---|---|---|
| 25 | 7.00 | 3.28 | 1.80 |

appreciable effect on the extracted gluon
different gluon depending on the top spectrum




[EPJ C80 (2020) 1; Les Houches proceedings, 2019]

# Data inconsistency: tensions between data sets

| Data set | baseline | rw D0 el.asy. |
|---|---|---|
| D0 e asy. | 5.42 | 1.73 |
| D0 $\mu$ asy. | 2.01 | 5.44 |
| Total | 1.17 | 1.29 |

Give more weight to a data set $p$

$$\chi^2 \to \chi^2 + w\chi^2_p$$

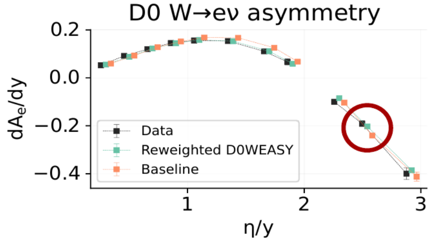
Refit: the total $\chi^2$ will increase
Which data sets get worse? How much?

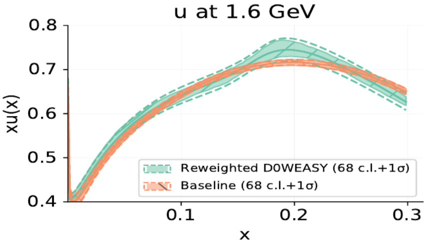
Refit: the data set $\chi^2_p$ will decrese
Self-consistency? Inconsistency?

Example: D0 el. asy.; $w = 411$

Can lift the downward prediction but
unnatural PDF shapes appear
error in other data sets increases

Fit quality for D0 el. asy. remains poor



# Theory uncertainties in PDF determination

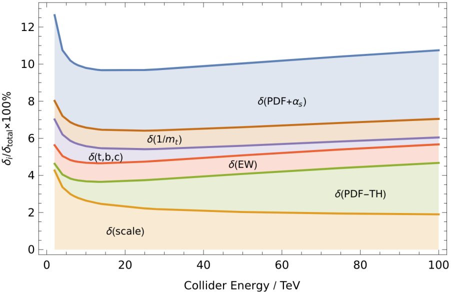
NNLO is the precision frontier for PDF determination

N3LO is the precision frontier for partonic cross sections [See Fabrizio Caola's talk]

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

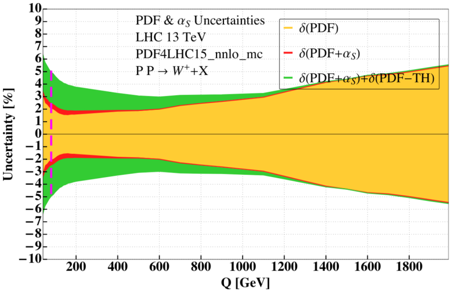
$$\hat{\sigma} = \alpha_s^p \hat{\sigma}_0 + \alpha_s^{p+1}\hat{\sigma}_1 + \alpha_s^{p+2}\hat{\sigma}_2 + \mathcal{O}(\alpha_s^{p+3}) \qquad \delta(\mathrm{PDF-TH}) = \frac{1}{2}\left|\frac{\sigma^{(2)}_{\mathrm{NNLO-PDFs}} - \sigma^{(2)}_{\mathrm{NLO-PDFs}}}{\sigma^{(2)}_{\mathrm{NNLO-PDFs}}}\right|$$

Higgs production in gluon-gluon fusion




[CERN Yellow Rep.Monogr. 7 (2019) 221]

$W^+$ boson production in CC Drell-Yan




[JHEP 11 (2020) 143]

# Theory uncertainties in PDF determination

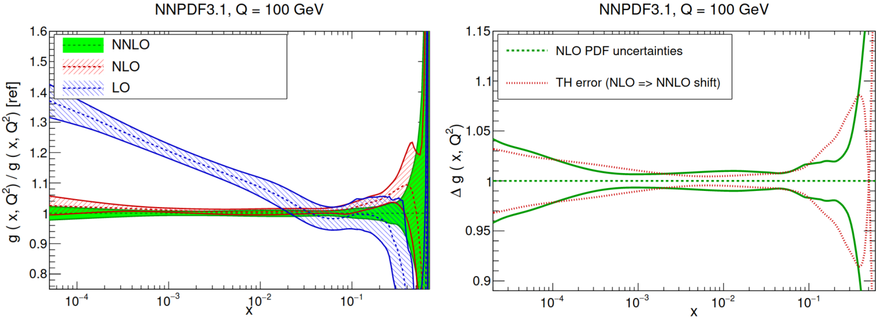
NNLO is the precision frontier for PDF determination

N3LO is the precision frontier for partonic cross sections [See Fabrizio Caola's talk]

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

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

Perturbative stability and uncertainty of the gluon PDF




[EPJ C77 (2017) 663]

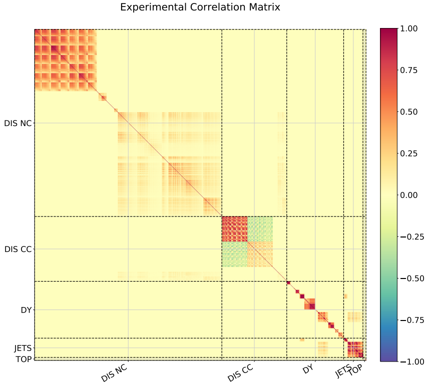
# Theory uncertainties in PDF determination

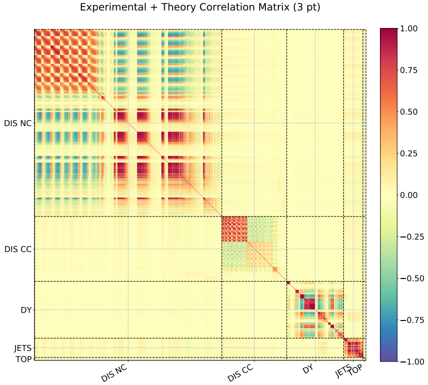
Assuming that theory uncertainties are (a) Gaussian and (b) independent from experimental uncertainties, modify the figure of merit to account for theory errors

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

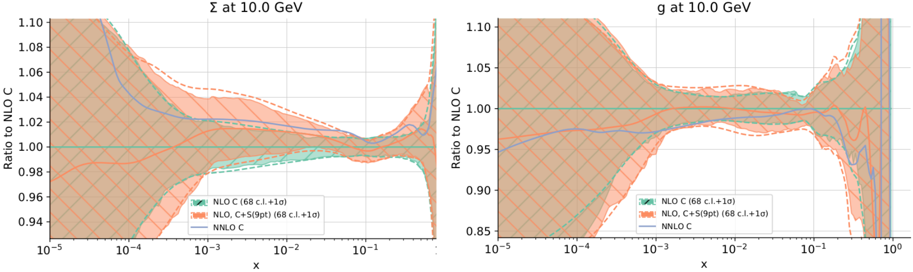
Problem reduced to estimate the th. cov. matrix, *e.g.* in terms of nuisance parameters

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



# Theory uncertainties in PDF determination




PDF uncertainty increase encapsulates NLO-NNLO shift

Overall (rather small) increase in uncertainties

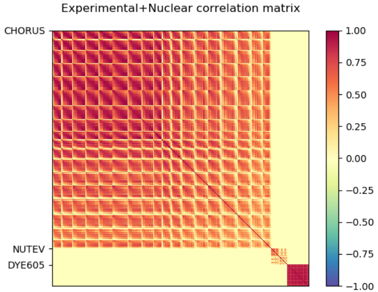
Increase in PDF uncertainties due to replica generation
is counteracted by extra correlations in fitting minimisation

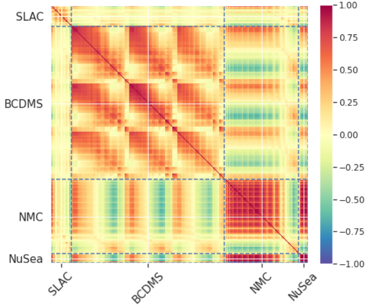
Tensions relieved: improvement in $\chi^2$
exp only: $\chi^2/N_{\rm dat} = 1.139$ exp+th: $\chi^2/N_{\rm dat} = 1.110$

Data whose theoretical descrition is affected by large scale uncertainties
are deweighted in favour of more perturbatively stable data

[EPJ C79 (2019) 838; ibid. 931]

# Theory uncertainty in PDF determination




Effect of nuclear uncertainties relevant at large $x$

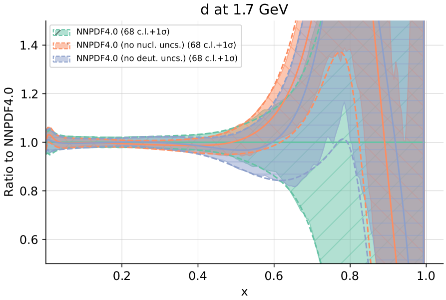
to reconcile FT DIS with LHC DY data

$\chi^2_{\rm tot} = 1.17 \to \chi^2_{\rm tot} = 1.26$ (no nucl. uncs.)

$\chi^2_{\rm LHCb} = 1.54 \to \chi^2_{\rm tot} = 1.76$ (no nucl. uncs.)

The bulk of the effect is due to nuclear uncertainties for heavy nuclei

deuteron uncertainties have a comparatively smaller effect at inermediate values of $x$




[EPJ C79 (2019) 282; EPJ C81 (2021) 37]

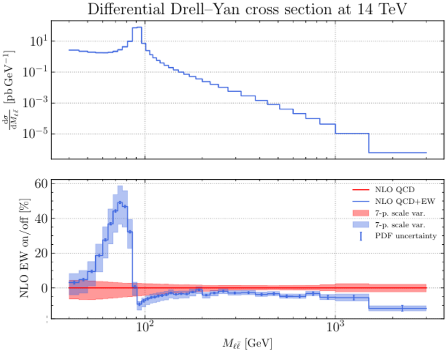
# NLO EW corrections in PDF determination

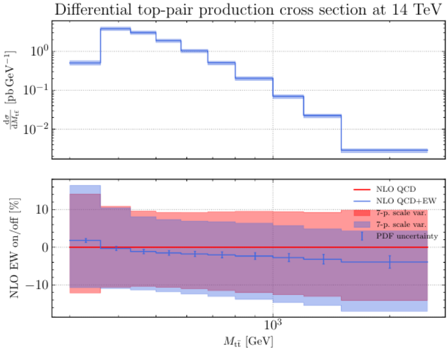
If we aim to PDF accurate to 1% NLO EW corrections do matter

especially as higher invariant mass and transverse momentum regions are accessed

Different approaches taken in general-purpose PDF fits

NLO EW $K$-factors (MSHT20); no NLO EW corrections by default (NNPDF4.0)




QED corrections in DGLAP evolution

[Com.Phys.Commun. 185 (2014) 1647]

Photon PDF

[PRL 117 (2016) 242002; JHEP 12 (2017) 046]

Photon PDF fits à la LuxQED

[SciPost Phys. 5 (2019) 1; JHEP 79 (2019) 10]

Automation of NLO EW corrections

[JHEP 07 (2018) 185]

Fast interpolation grids: PineAPPL

[JHEP 12 (2020 108]

Careful scrutiny of data

(no FSR nor photon-initiated subtraction)

# Methodology: validation of PDF uncertainties

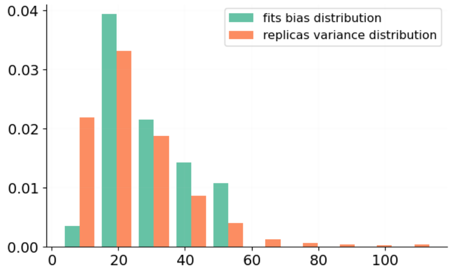
Data region: closure tests

Fit PDFs to pseudodata generated
assuming a known underlying law

Define bias and variance

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

If PDF uncertainty faithful, then
E[bias] = variance
25 fits, 40 replicas each



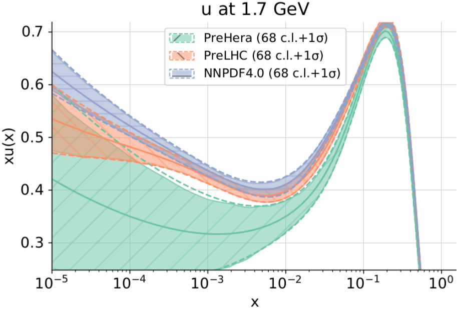

[EPJ C77 (2017) 663; Del Debbio and Wilson, in preparation]

Extrapolation regions: future test

Test PDF uncertainties on data sets
not included in a given PDF fit
that cover unseen kinematic regions

| Data set | NNPDF4.0 | pre-LHC | pre-HERA |
|---|---|---|---|
| pre-HERA | 1.09 | 1.01 | 0.90 |
| pre-LHC | 1.21 | 1.20 | 23.1 |
| NNPDF4.0 | 1.29 | 3.30 | 23.1 |

Only exp. cov. matrix




[Acta Phys.Polon. B52 (2021) 243]

# Methodology: validation of PDF uncertainties

Data region: closure tests

Fit PDFs to pseudodata generated assuming a known underlying law

Define bias and variance

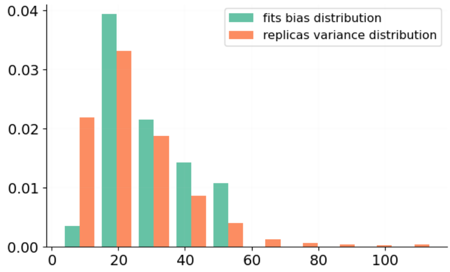
**bias** difference of central prediction and truth

**variance** uncertainty of replica predictions

If PDF uncertainty faithful, then

E[bias] = variance

25 fits, 40 replicas each



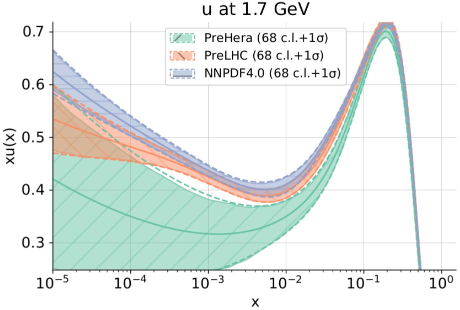

[EPJ C77 (2017) 663; Del Debbio and Wilson, in preparation]

Extrapolation regions: future test

Test PDF uncertainties on data sets not included in a given PDF fit that cover unseen kinematic regions

| Data set | NNPDF4.0 | pre-LHC | pre-HERA |
|---|---|---|---|
| pre-HERA | | | 0.86 |
| pre-LHC | | 1.17 | 1.22 |
| NNPDF4.0 | 1.12 | 1.30 | 1.38 |

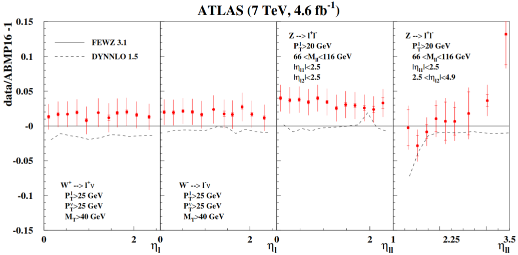
Exp+PDF cov. matrix




[Acta Phys.Polon. B52 (2021) 243]

# Methodology: benchmarks

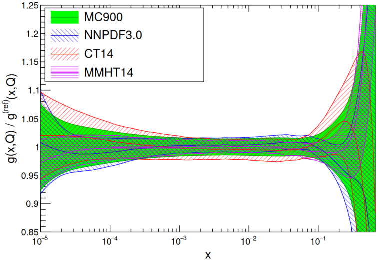
## Benchmark of the theory



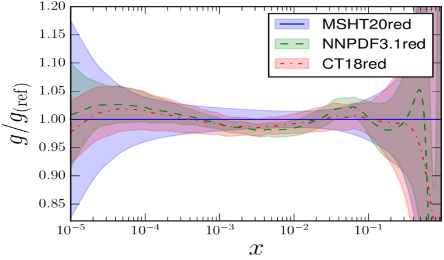

Be careful about the use of different NNLO codes for DY production in particular when experiments use non-optimal fiducial cuts [arXiv:2104.02400]

NNLO corrections usually implemented via $K$-factors
NNLOJet/ApplFast provide NNLO lookup tables for a limited set of data

## Benchmark of PDF sets




[PDF4LHC15 combination, JPG 43 (2016) 023001]




[PDF4LHC21 benchmark, see T. Cridge's talk at DIS2021]

# 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.

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

Understand experimental systematic uncertainties and their correlations.

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.

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

Understand experimental systematic uncertainties and their correlations.

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**