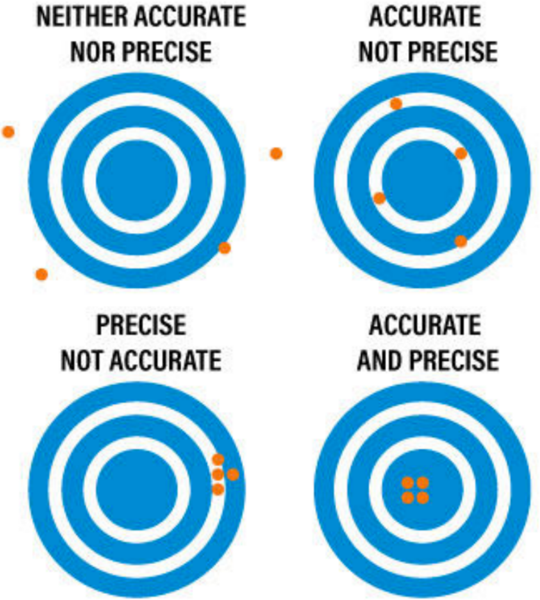
Open issues in the determination of collinear Parton Distribution and Fragmentation Functions

Transversity 2022

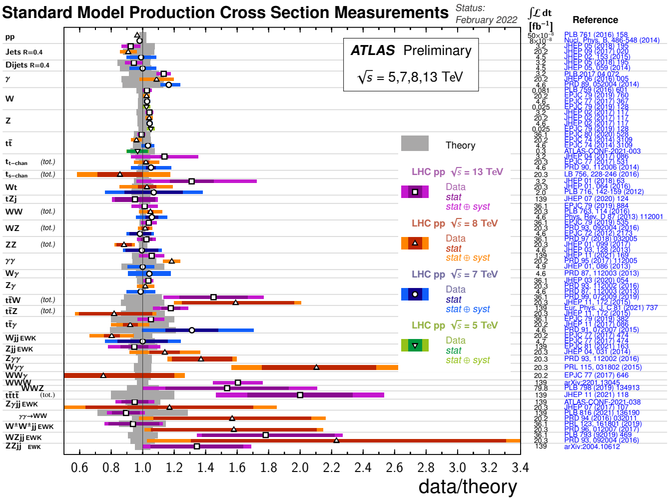
Emanuele R. Nocera
School of Physics and Astronomy, The University of Edinburgh

26 May 2022

Foreword: a tale of accuracy and precision



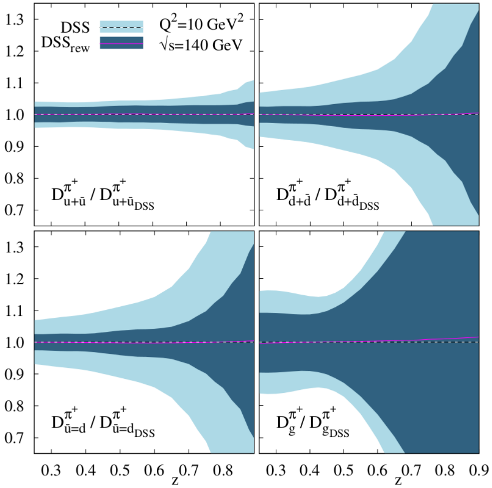
Foreword: a tale of accuracy and precision



[Plot from ATLAS Collaboration web page]

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

Foreword: a tale of accuracy and precision

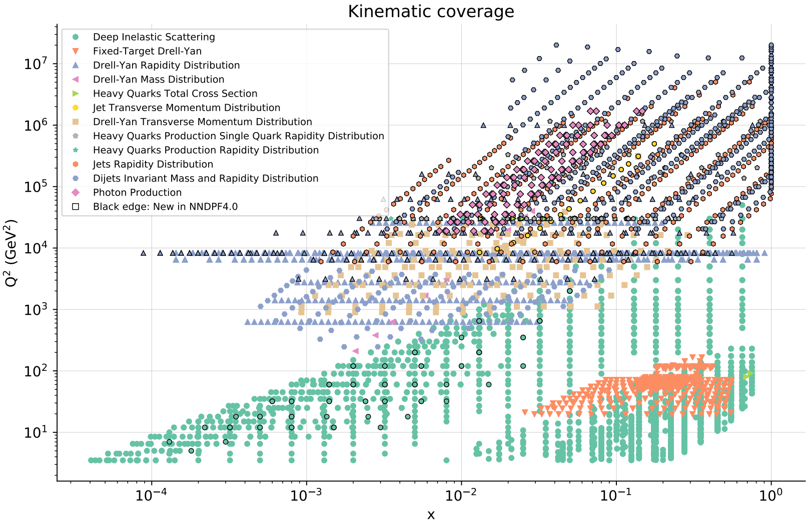


[Plot from the EIC Yellow Report]

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

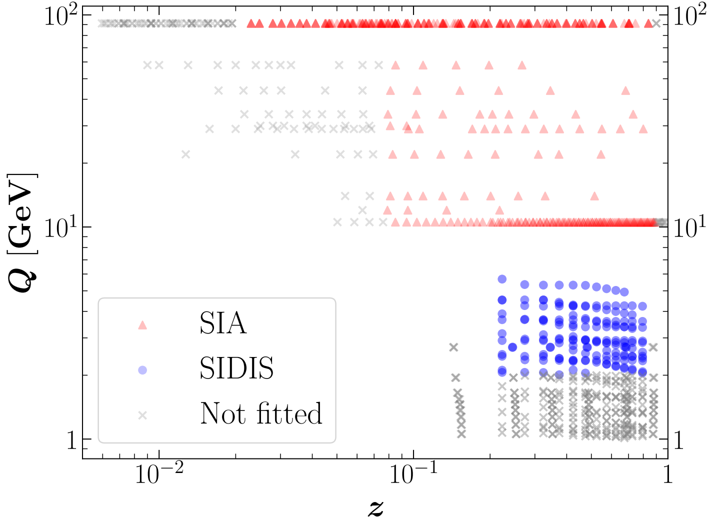
1. Data

Data: kinematic coverage



$$N_{\rm dat} \sim 4500$$

Data: kinematic coverage



$N_{\rm dat} \sim 700$

Data inconsistency: tensions between data sets

Give more weight to a data set p

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

Refit: the total χ^2 will increase

Which data sets get worse? How much?

Refit: the data set χ_p^2 will decrese

Self-consistency? Inconsistency?

Examples: ATLAS W, Z and $t\bar{t}$

Inconsistency clearly spotted

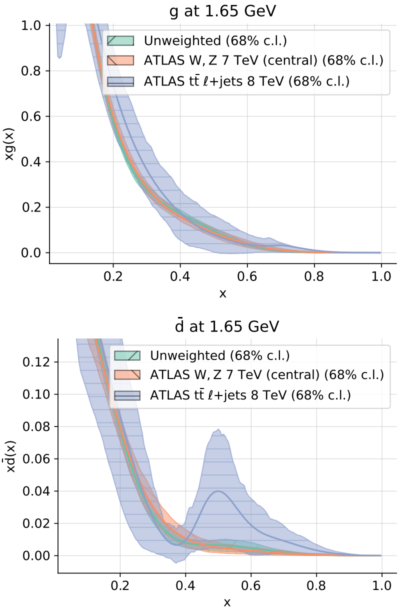
unnatural PDF shapes appear

error in other data sets increases

Otherwise global fit quality

and PDFs remain unaltered

Data set	baseline	rw W, Z	rw $t\bar{t}$
ATLAS W, Z 7 TeV	1.86	1.23	—
ATLAS $t\bar{t}$ 8 TeV	4.11	—	1.21
Total	1.20	1.21	1.73



Data inconsistency: experimental correlations

Single inclusive jet data from ATLAS 7 TeV

default correlations: terrible χ^2

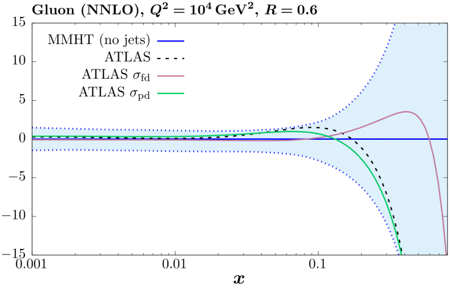
(correlations across rapidity bins)

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 χ^2

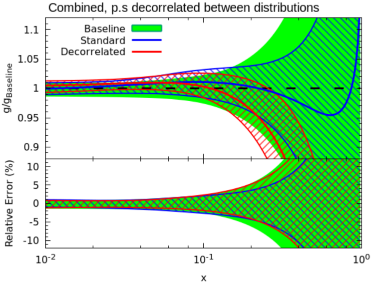
(correlations across different spectra)

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]

Good knowledge of experimental correlations is important

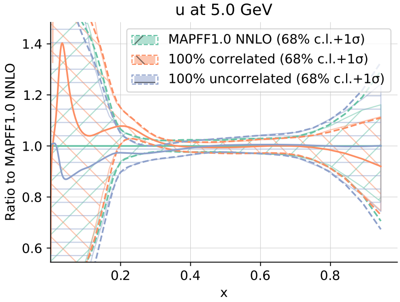
Consider the COMPASS $\pi^{\pm}$ multiplicities

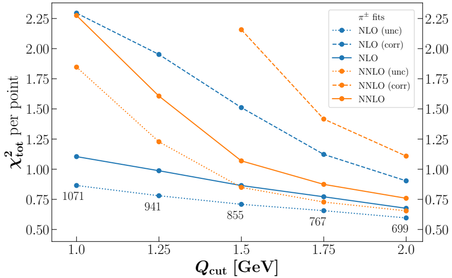
[PLB 764 (2017) 1]

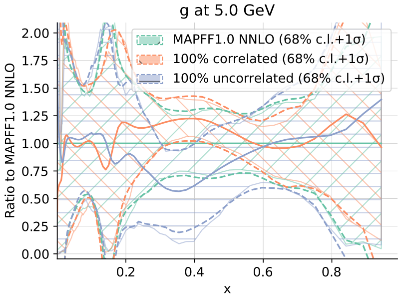
Only 80% of the systematic uncertainty is bin-by-bin correlated

What if you incorporate a different piece of information in a FF fit?

Consider two cases: [arXiv:2204.10331]
full correlation; full decorrelation







2. Theory

Theory uncertainties in PDF determination

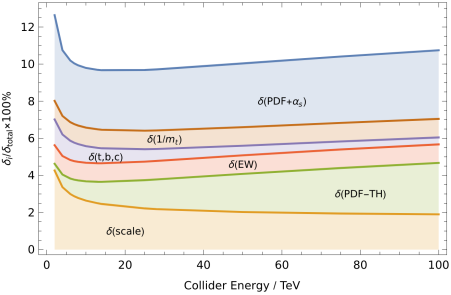
NNLO is the precision frontier for PDF determination

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

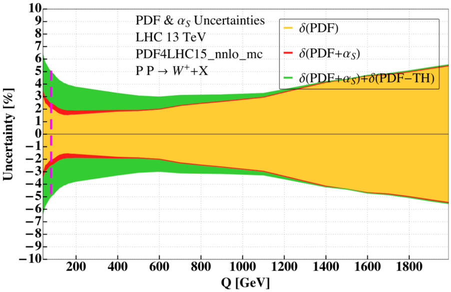
$$\hat{\sigma} = \alpha_s^p \hat{\sigma}_0 + \alpha_s^{p+1} \hat{\sigma}_1 + \alpha_s^{p+2} \hat{\sigma}_2 + \mathcal{O}(\alpha_s^{p+3}) \qquad \delta(\mathrm{PDF-TH}) = \frac{1}{2} \left| \frac{\sigma^{(2)}_{\mathrm{NNLO-PDFs}} - \sigma^{(2)}_{\mathrm{NLO-PDFs}}}{\sigma^{(2)}_{\mathrm{NNLO-PDFs}}} \right|$$

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

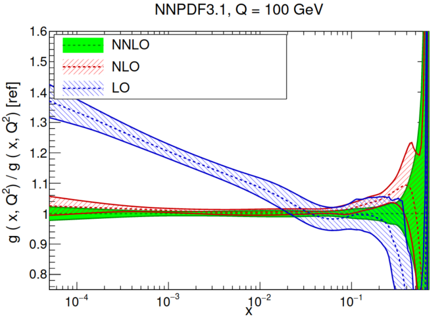
NNLO is the precision frontier for PDF determination

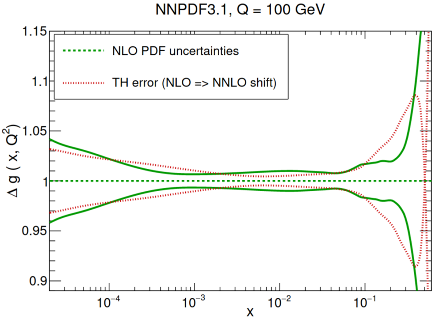
N3LO is the precision frontier for partonic cross sections

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]

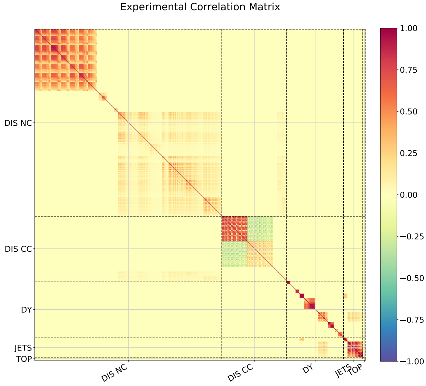
Theory uncertainties in PDF determination

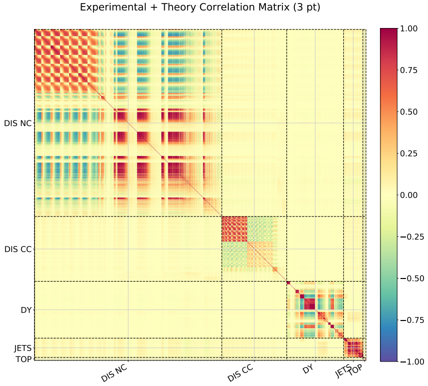
Assuming that theory uncertainties are (a) Gaussian and (b) independent from experimental uncertainties, modify the figure of merit to account for theory errors

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

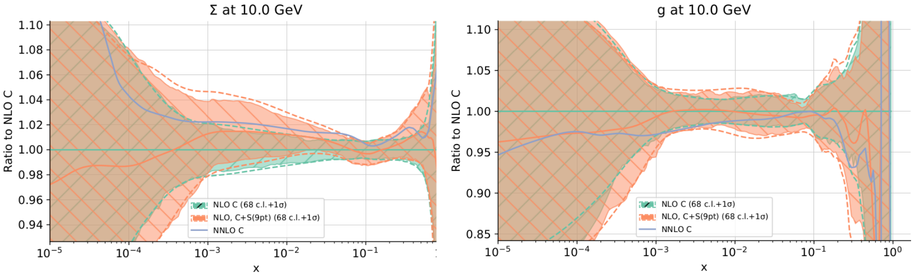
Problem reduced to estimate the th. cov. matrix, *e.g.* in terms of nuisance parameters

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





Theory uncertainties in PDF determination



PDF uncertainty increase encapsulates NLO-NNLO shift

Overall (rather small) increase in uncertainties

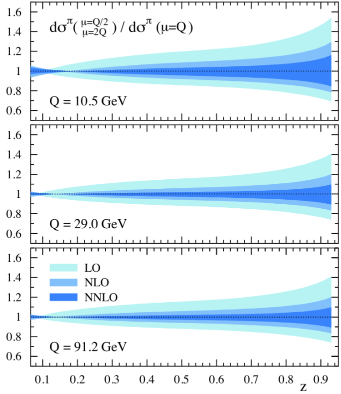
Increase in PDF uncertainties due to replica generation
is counteracted by extra correlations in fitting minimisation

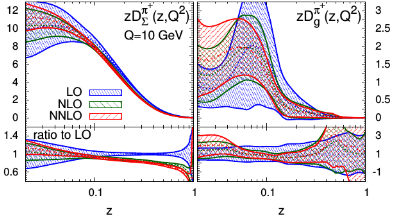
Tensions relieved: improvement in χ^2
exp only: $\chi^2/N_{\rm dat} = 1.139$ exp+th: $\chi^2/N_{\rm dat} = 1.110$

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]

FFs at NNLO: SIA [PR D92 (2015) 114017; EPJ C77 (2017) 516]

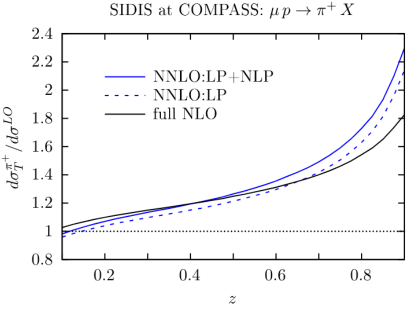


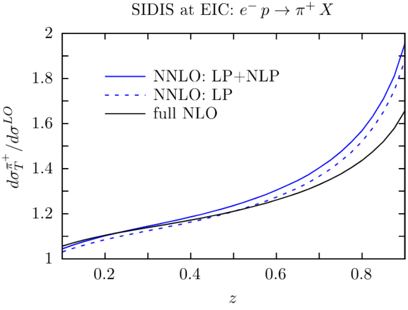


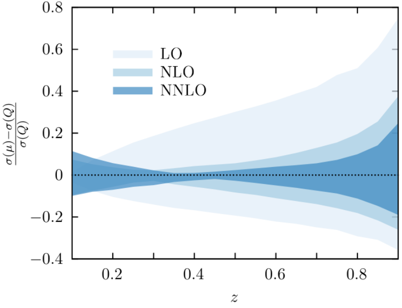
Exp.	$N_{\rm dat}$	LO $\chi^2/N_{\rm dat}$	NLO $\chi^2/N_{\rm dat}$	NNLO $\chi^2/N_{\rm dat}$
BELLE	70	0.60	0.11	0.09
BABAR	40	1.91	1.77	0.78
TASSO12	4	0.70	0.85	0.87
TASSO14	9	1.55	1.67	1.70
TASSO22	8	1.64	1.91	1.91
TPC	13	0.46	0.65	0.85
TPC-UDS	6	0.78	0.55	0.49
TPC-C	6	0.55	0.53	0.52
TPC-B	6	1.44	1.43	1.43
TASSO34	9	1.16	0.98	1.00
TASSO44	6	2.01	2.24	2.34
TOPAZ	5	1.04	0.82	0.80
ALEPH	23	1.68	0.90	0.78
DELPHI	21	1.44	1.79	1.86
DELPHI-UDS	21	1.30	1.48	1.54
DELPHI-B	21	1.21	0.99	0.95
OPAL	24	2.29	1.88	1.84
SLD	34	2.33	1.14	0.83
SLD-UDS	34	0.95	0.65	0.52
SLD-C	34	3.33	1.33	1.06
SLD-B	34	0.45	0.38	0.36
TOTAL	428	1.44	1.02	0.87

Excellent perturbative convergence
FFs almost stable from NLO to NNLO
LO FF uncertainties larger than HO
Effects less evident for $K^{\pm}$ and $p/\bar{p}$

FFs at NNLO: SIDIS [PRD 104 (2021) 094046; arXiv:2203.07928]







At k-th order, there are terms of the form:

$$\alpha_s^k \delta(1-x)\left(\frac{\ln^m(1-z)}{(1-z)}\right)_+ \alpha_s^k \delta(1-z)\left(\frac{\ln^m(1-x)}{(1-x)}\right)_+$$
$$\alpha_s^k \left(\frac{\ln^m(1-x)}{(1-x)}\right)_+ \left(\frac{\ln^n(1-z)}{(1-z)}\right)_+$$

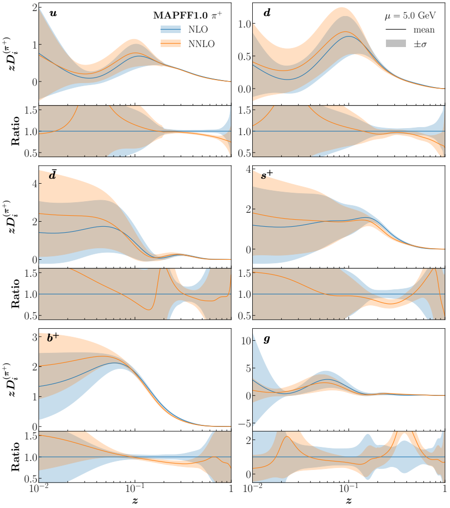
$m \leq 2k-1$ (non-mixed) $m+n \leq 2k-2$ (mixed)

Approximate NNLO corrections obtained from threshold resummation

Use similarity between DY and SIDIS to expand NNLL results to fixed-order NNLO

FFs at NNLO: combining SIA and SIDIS [arXiv:2202.05060; arXiv:2204.10331]

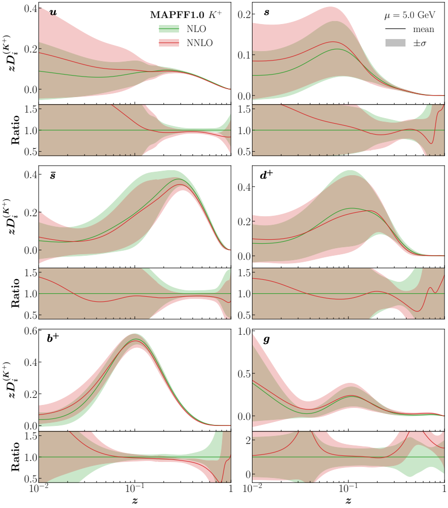
PIONS



Data set	$N_{\rm dat}$	NLO	NNLO
BELLE $h^\pm$	70	0.14	0.13
BABAR $h^\pm$	39	0.91	0.76
TASSO 12 GeV $h^\pm$	4	0.90	0.92
TASSO 14 GeV $h^\pm$	9	1.33	1.35
TASSO 22 GeV $h^\pm$	8	1.65	1.81
TPC $h^\pm$	13	0.23	0.25
TASSO 30 GeV $h^\pm$	2	0.30	0.34
TASSO 34 GeV $h^\pm$	9	1.08	1.48
TASSO 44 GeV $h^\pm$	6	1.13	1.37
TOPAZ $h^\pm$	5	0.24	0.37
ALEPH $h^\pm$	23	1.24	1.46
DELPHI (inclusive) $h^\pm$	21	1.31	1.25
DELPHI (uds tagged) $h^\pm$	21	2.68	2.89
DELPHI (b tagged) $h^\pm$	21	1.58	1.73
OPAL $h^\pm$	24	1.63	1.79
SLD (inclusive) $h^\pm$	34	1.05	1.13
SLD (uds tagged) $h^\pm$	34	1.59	2.16
SLD (b tagged) $h^\pm$	34	0.55	0.68
HERMES h^- d	2	0.41	0.32
HERMES h^+ p	2	0.01	0.02
HERMES h^- d	2	0.17	0.11
HERMES h^+ p	2	0.35	0.32
COMPASS h^-	157	0.48	0.55
COMPASS h^+	157	0.62	0.72
Global data set	**699**	**0.68**	**0.76**

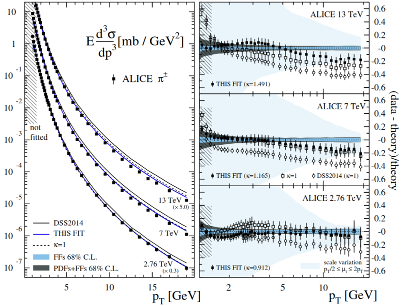
FFs at NNLO: combining SIA and SIDIS [arXiv:2202.05060; arXiv:2204.10331]

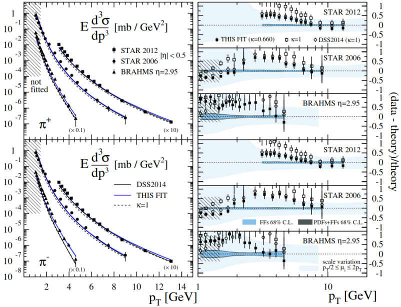
KAONS



Data set	$N_{\rm dat}$	NLO	NNLO
BELLE $h^\pm$	70	0.39	0.41
BABAR $h^\pm$	28	0.36	0.25
TASSO 12 GeV $h^\pm$	3	0.85	0.87
TASSO 14 GeV $h^\pm$	9	1.24	1.22
TASSO 22 GeV $h^\pm$	6	0.89	0.90
TPC $h^\pm$	13	0.38	0.40
TASSO 30 GeV $h^\pm$	—	—	—
TASSO 34 GeV $h^\pm$	5	0.07	0.06
TASSO 44 GeV $h^\pm$	—	—	—
TOPAZ $h^\pm$	3	0.10	0.11
ALEPH $h^\pm$	18	0.49	0.48
DELPHI (inclusive) $h^\pm$	23	0.97	0.99
DELPHI (uds tagged) $h^\pm$	23	0.44	0.38
DELPHI (b tagged) $h^\pm$	23	0.42	0.45
OPAL $h^\pm$	10	0.39	0.36
SLD (inclusive) $h^\pm$	35	0.83	0.67
SLD (uds tagged) $h^\pm$	35	1.37	1.52
SLD (b tagged) $h^\pm$	35	0.75	0.77
HERMES h^- d	2	0.18	0.13
HERMES h^+ p	2	0.05	0.04
HERMES h^- d	2	0.58	0.48
HERMES h^+ p	2	0.56	0.43
COMPASS h^-	156	0.74	0.59
COMPASS h^+	156	0.76	0.67
Global data set	**659**	**0.62**	**0.55**

FFs at NNLO: pp collisions [PRD 105 (2022) L031502]



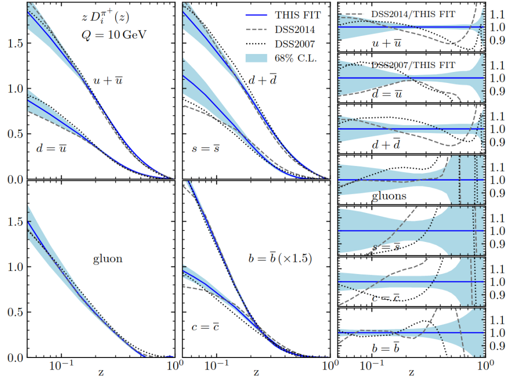


$$\mu_R = \mu_{F,I} = \mu_{F,F} = \kappa\epsilon \qquad 1/2 < \kappa < 2$$

Treat scale variations factors ϵ as nuisance parameters and fit them to the data

Excellent statistical quality of the fit
$\chi^2/N_{\rm dat} = 1.17,\ N_{\rm dat} = 1460$

Significant impact on FFs



3. Methodology

Validation of PDF uncertainties

Data region: closure tests

Fit PDFs to pseudodata generated assuming a known underlying law

Define bias and variance

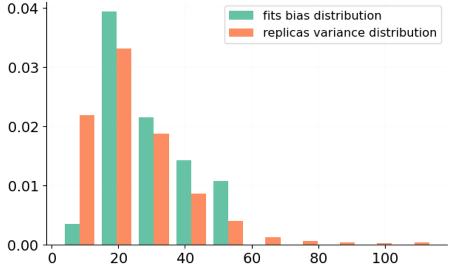
bias difference of central prediction and truth

variance uncertainty of replica predictions

If PDF uncertainty faithful, then

E[bias] = variance

25 fits, 40 replicas each



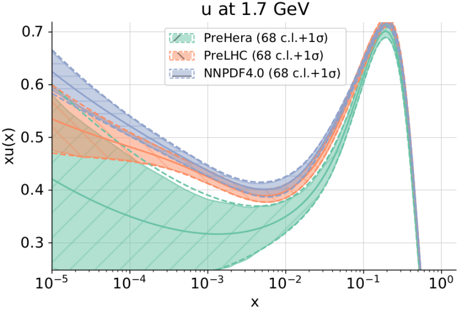
[EPJ C77 (2017) 663; EPJ C82 (2022) 330]

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]

Validation of PDF uncertainties

Data region: closure tests

Fit PDFs to pseudodata generated
assuming a known underlying law

Define bias and variance

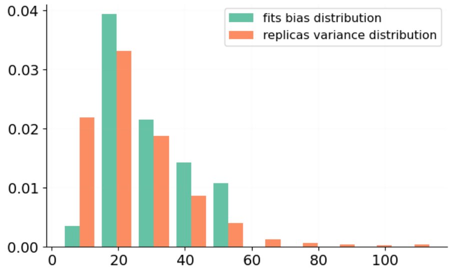
bias difference of central prediction and truth

variance uncertainty of replica predictions

If PDF uncertainty faithful, then

E[bias] = variance

25 fits, 40 replicas each



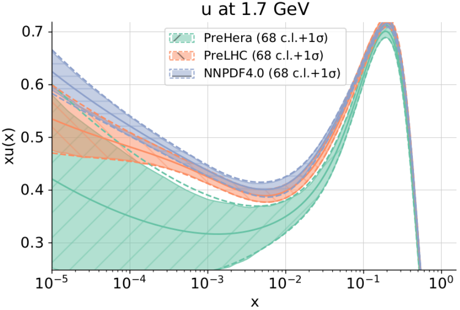
[EPJ C77 (2017) 663; EPJ C82 (2022) 330]

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

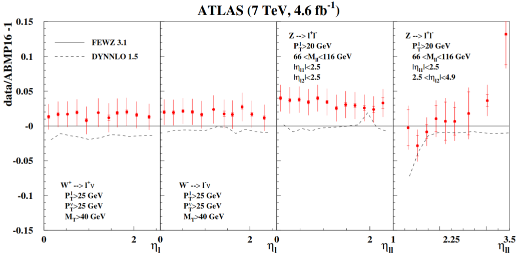
Exp+PDF cov. matrix



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

Benchmarks: PDFs

Benchmark of the theory



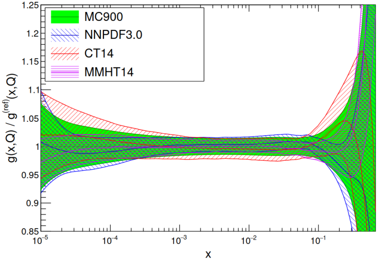
Be careful about the use of different NNLO codes for DY production

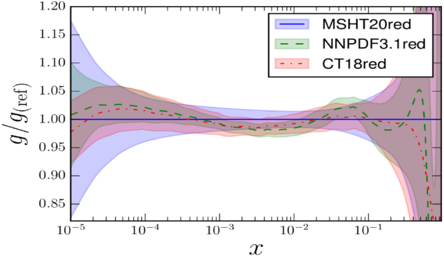
in particular when experiments use non-optimal fiducial cuts [EPJ C81 (2021) 573]

NNLO corrections usually implemented via K-factors

NNLOJet/ApplFast provide NNLO lookup tables for a limited set of data

Benchmark of PDF sets

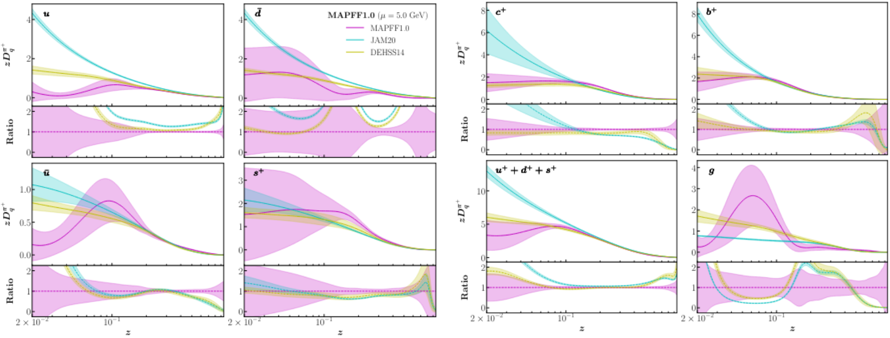




[PDF4LHC15 benchmark, JPG 43 (2016) 023001]

[PDF4LHC21 benchmark, arXiv:2203.05506]

Benchmarks: FFs



[PRD 104 (2021) 034007]

Significant differences are observed across different FF sets

A joint effort to understand the origin of these differences is possibly useful in a spirit similar to the PDF4LHC activities (FF4EIC ?)

Public codes: PDFs [EPJ C81 (2021) 958]

NNPDF

Tests passing | DOI 10.5281/zenodo.6542572

NNPDF: An open-source machine learning framework for global analyses of parton distributions

The NNPDF collaboration determines the structure of the proton using Machine Learning methods. This is the main repository of the fitting and analysis frameworks. In particular it contains all the necessary tools to reproduce the NNPDF4.0 PDF determinations.

Documentation

The documentation is available at https://docs.nnpdf.science/

Install

See the NNPDF installation guide for the conda package, and how to build from source.

Please note that the conda based workflow described in the documentation is the only supported one. While it may be possible to set up the code in different ways, we won't be able to provide any assistance.

We follow a rolling development model where the tip of the master branch is expected to be stable, tested and correct. For more information see our releases and compatibility policy.

`https://github.com/NNPDF`

NNPDF

Search docs

- Getting started
- Fitting code: `n3fit`
- Code for data: validphys
- Handling experimental data: Buildmaster
- Storage of data and theory predictions
- Theory
- Chi square figures of merit
- Contributing guidelines and tools
- Releases and compatibility policy
- Continuous integration and deployment
- Servers
- External codes
- Tutorials

Public codes: FFs [PRD 104 (2021) 034007]

M.A.P. Collaboration

Multi-dimensional Analyses of Partonic distributions

Amsterdam, Edinburgh, Paris, Pavia

Overview | Repositories 3 | Packages | People 11 | Teams | Projects | Settings

Pinned — Customize your pins

- **NangaParbat** (Public) — Nanga Parbat: a fitting framework for the determination of the non-perturbative component of TMD distributions — HTML ☆ 3, forks 3
- **MontBlanc** (Public) — A code for the determination of collinear distributions — Jupyter Notebook

Repositories

Find a repository… | Type | Language | Sort | New

- **MontBlanc** (Public) — A code for the determination of collinear distributions — Jupyter Notebook ☆ 0, forks 0, issues 0, PRs 0 — Updated 2 days ago
- **NangaParbat-Beta** (Private) — HTML ☆ 1, MIT, forks 0, issues 0, PRs 0 — Updated 2 days ago
- **NangaParbat** (Public) — Nanga Parbat: a fitting framework for the determination of the non-perturbative component of TMD distributions — HTML ☆ 3, MIT, forks 3, issues 0, PRs 0 — Updated on 28 Jun

Members

- Alessandro Bacchetta — abacchetta
- Andrea Signori — asignori
- Chiara Bissolotti — ChiaBis
- Emanuele R. Nocera — enocera
- Giuseppe Bozzi — giuboz
- lorenzorossi97
- marcoradici
- mcerutti996
- Rabah Abdul Khalek — rabah-khalek
- Simo_tat_96 — Simotat96
- Valerio Bertone — vbertone

`https://github.com/MapCollaboration`

4. Conclusions

Summary

A precise and accurate determination of PDFs and FFs is key to good phenomenology.

LHC and EIC measurements are (will be) instrumental
to reduce PDF uncertainties to few percent.

Achieving percent precision in PDF/FF determinations is a big challenge.

Understand experimental systematic uncertainties and their correlations.

Refine the theoretical accuracy of PDF/FF determinations.

Represent theory uncertainties into PDF/FF uncertainties.

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

Benchmark efforts are crucial.
They may greatly benefit from public releases of PDF/FF codes and inputs.

Summary

A precise and accurate determination of PDFs and FFs is key to good phenomenology.

LHC and EIC measurements are (will be) instrumental
to reduce PDF uncertainties to few percent.

Achieving percent precision in PDF/FF determinations is a big challenge.

Understand experimental systematic uncertainties and their correlations.

Refine the theoretical accuracy of PDF/FF determinations.

Represent theory uncertainties into PDF/FF uncertainties.

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

Benchmark efforts are crucial.
They may greatly benefit from public releases of PDF/FF codes and inputs.

Thank you