Deep structure of proton

Oleksandr Zenaiev
(Hamburg University, II. Institute for Theoretical Physics)

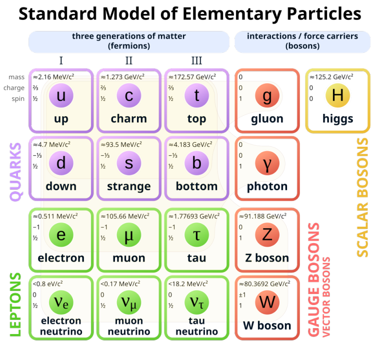
DESY Theory Workshop: Synergy Towards the Future Standard Model
26 Sep 2025

Perturbative and non-perturbative QCD

- Quantum Chromodynamics (QCD) governs the strong force: interaction of quarks and gluons that confines hadrons and forms “ordinary” baryon matter around us
- Solution of renormalization group equation at LO:

$$\alpha_S(Q^2) = \frac{\alpha_S(\mu^2)}{1+\alpha_S(\mu^2)\beta_0 \ln(\frac{Q^2}{\mu^2})} = \frac{1}{\beta_0 \ln(\frac{Q^2}{\Lambda^2})}$$

$\beta_0 = \frac{33-2N_f}{12\pi}$, $\Lambda \sim$ a few hundred of MeV

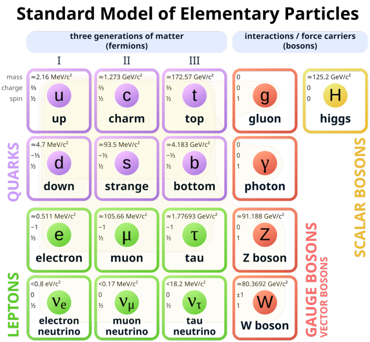


Perturbative and non-perturbative QCD

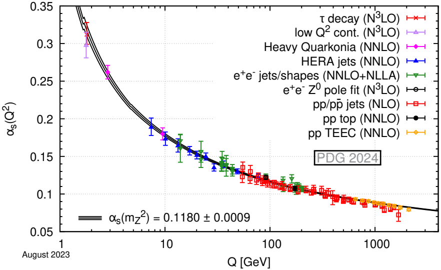
- Quantum Chromodynamics (QCD) governs the strong force: interaction of quarks and gluons that confines hadrons and forms “ordinary” baryon matter around us
- Solution of renormalization group equation at LO:

$$\alpha_S(Q^2) = \frac{\alpha_S(\mu^2)}{1+\alpha_S(\mu^2)\beta_0 \ln(\frac{Q^2}{\mu^2})} = \frac{1}{\beta_0 \ln(\frac{Q^2}{\Lambda^2})}$$

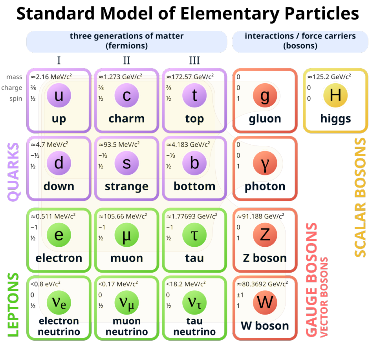
$\beta_0 = \frac{33-2N_f}{12\pi}$, $\Lambda \sim$ a few hundred of MeV



perturbative QCD at $Q \gtrsim 1$ GeV: asymptotic freedom



Perturbative and non-perturbative QCD



- Quantum Chromodynamics (QCD) governs the strong force: interaction of quarks and gluons that confines hadrons and forms "ordinary" baryon matter around us
- Solution of renormalization group equation at LO:

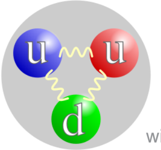
$$\alpha_S(Q^2) = \frac{\alpha_S(\mu^2)}{1+\alpha_S(\mu^2)\beta_0 \ln(\frac{Q^2}{\mu^2})} = \frac{1}{\beta_0 \ln(\frac{Q^2}{\Lambda^2})}$$

$\beta_0 = \frac{33-2N_f}{12\pi}$, $\Lambda \sim$ a few hundred of MeV

non-perturbative QCD: confinement

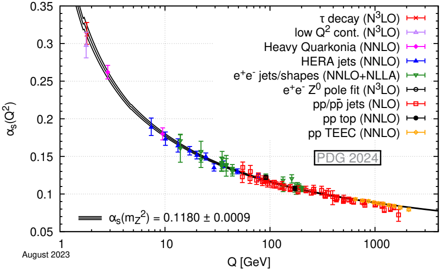
- if $Q \lesssim 1$ GeV, perturbation theory is not applicable anymore: QCD potential grows linearly with larger distances

→ no free quarks or gluons: hadrons are described by **parton distribution functions**

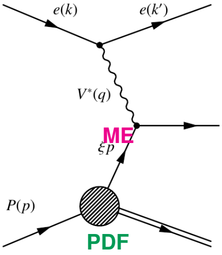


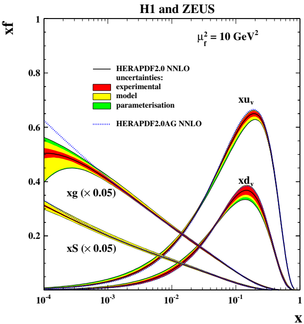
wikipedia.org

perturbative QCD at $Q \gtrsim 1$ GeV: asymptotic freedom



- Collinear factorization theorem: $\boldsymbol{\sigma} = \text{PDF} \otimes \text{ME}$
- **Parton distribution functions (PDFs)** $\boldsymbol{f(x, \mu_f)}$ describe distribution of quarks and gluons in hadrons
- **Matrix elements (ME)** are partonic cross sections calculated in pQCD $\sigma = \Sigma_{i=0}^{n} \sigma_i \alpha_S^i$ requiring $\alpha_S(\mu_r) < 1$ ($\mu_r \gg \Lambda_{QCD}$)
- At low scales $\sim$ 1 GeV non-perturbative QCD effects are parametrised by PDFs **which are extracted using data**
  - typically shaped like $x^a(1-x)^b$ with a few tens of parameters
- At higher scales $>$ 1 GeV PDF evolution is predicted by pQCD
- $\alpha_S(M_Z)$, m_c, m_b, m_t are free parameters of SM
  - can be fitted or fixed in global QCD analyses
- **Challenges:**
  - choose suitable PDF parametrization
  - select **PDF sensitive** and **consistent** data sets
  - use appropriate statistical method (typically minimizing χ^2)
  - **most challenging are PDF uncertainties:** very much depend on all above





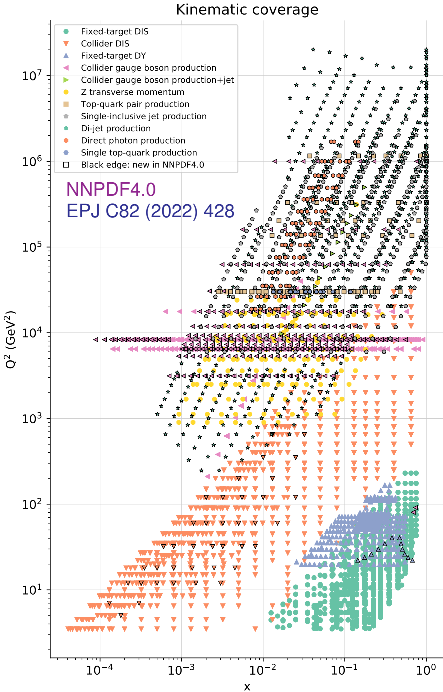
Determination of PDFs

- PDFs are functions of energy scale (Q) and parton momentum fraction (x):

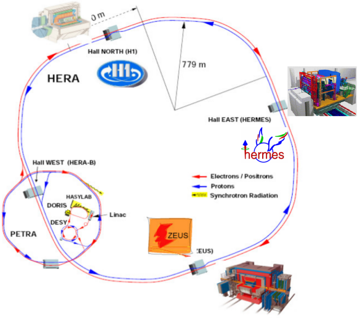
$$f_{q,g,\gamma,\ldots}(Q^2, x)$$

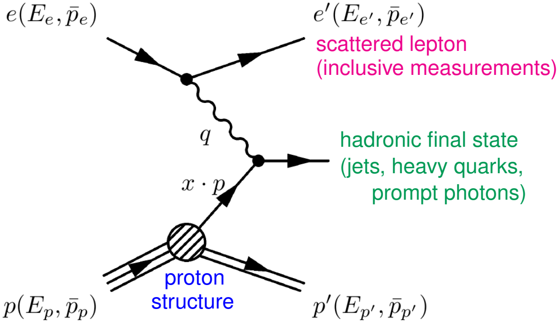
  - assumed to be process independent (universal)
- NNLO QCD is state of the art
  - approximate N^{3}LO fits started to appear
- Directly constrained by experimental data at $3 \lesssim Q^2 \lesssim 10^7$ GeV^2, $10^{-4} \lesssim x \lesssim 1$
  - Q^2 dependence calculable in pQCD
  - extrapolation at $x \lesssim 10^{-4}$ and $x \to 1$
- Various processes constrain various distributions and are sensitive to various SM parameters

process	constrain
$ep(N)$ DIS	core of PDF determination
ep and pp jets	gluon PDF & α_S
$pp\ t\bar{t}$	gluon PDF & α_S, m_t
$pp\ W, Z$	q flavour separation
$pp\ p_T(Z)$	gluon PDF and α_S
...	...



ep Deep Inelastic Scattering $Q^2 \gtrsim 1\ GeV^2$





$$Q^2 = -q^2 = -(e - e')^2$$

$$x_{Bj} = \frac{Q^2}{2q \cdot p}, \; y = \frac{q \cdot p}{q \cdot e}$$

$$s = (e + p)^2, Q^2 = s x_{Bj} y$$

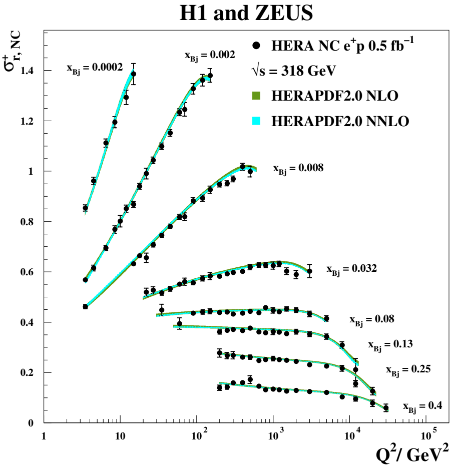
$$\sigma_{\mathrm{red}}^{e^\pm p} = \frac{\mathrm{d}^2 \sigma^{e^\pm p}}{\mathrm{d}x_{\mathrm{Bj}} \mathrm{d}Q^2} \cdot \frac{x_{\mathrm{Bj}} Q^4}{2\pi\alpha^2 \, (1+(1-y)^2)}$$

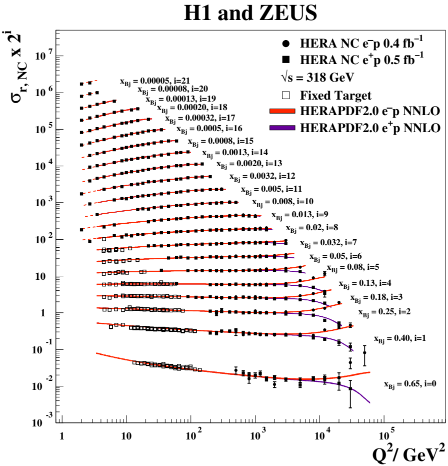
$$= F_2(x, Q^2) - \frac{y^2}{Y_+} F_L(x, Q^2) \mp \frac{Y_-}{Y_+} x F_3(x, Q^2),$$

$$Y_\pm = 1 \pm (1 - y)^2$$

- At LO QCD, only γ exchange: $F_2(x, Q^2) = x \sum_q e_q^2 [f_q(x, Q^2) + f_{\bar{q}}(x, Q^2)]$, $F_L = xF_3 = 0$

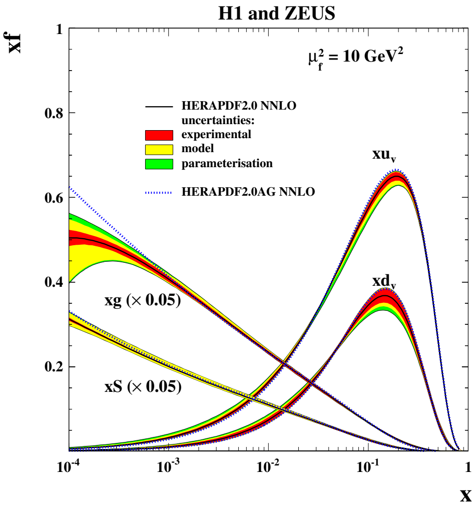
  $\rightarrow$ direct probe of quark PDFs





- HERA data on *ep* DIS scattering are a backbone of all global QCD analyses
  - direct constraints on valence and sea quark PDFs in a wide kinematic range
  - however only indirect sensitivity to gluon PDF and α_S (scaling violation and F_L)

HERA DIS data H1&ZEUS, EPJ C 75 (2015) 580



- HERA data on *ep* DIS scattering are a backbone of all global QCD analyses
  - direct constraints on valence and sea quark PDFs in a wide kinematic range
  - however only indirect sensitivity to gluon PDF and α_S (scaling violation and F_L)

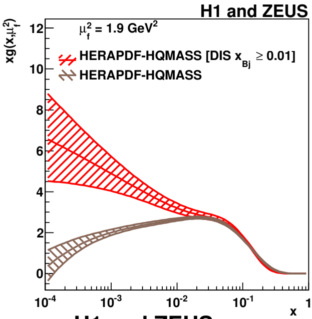
Charm and beauty at HERA H1&ZEUS, EPJ C78 (2018) 473

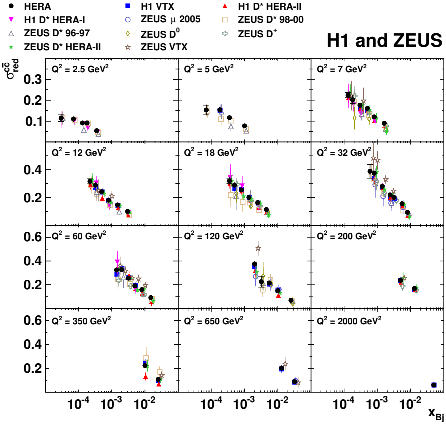
- Legacy data on charm and beauty production from HERA
- Enable **precise** determination of m_c and m_b

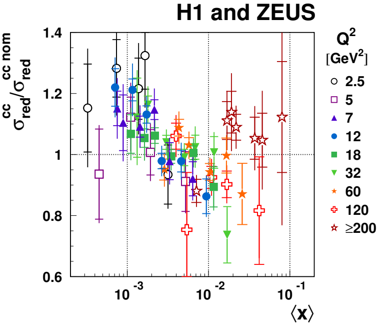
$$m_c(m_c) = 1290^{+46}_{-41}(\text{fit})^{+62}_{-14}(\text{mod})^{+3}_{-31}(\text{par})\ \text{MeV}$$

$$m_b(m_b) = 4049^{+104}_{-109}(\text{fit})^{+90}_{-32}(\text{mod})^{+1}_{-31}(\text{par})\ \text{MeV}$$

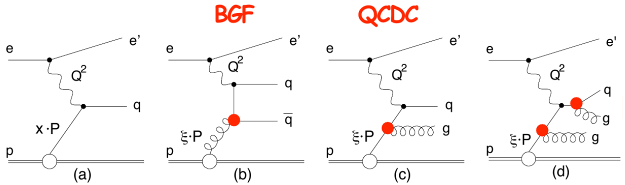
- Revealed tension in describing simultaneously heavy-quark and inclusive HERA data



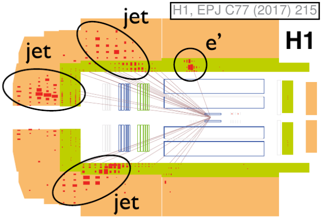




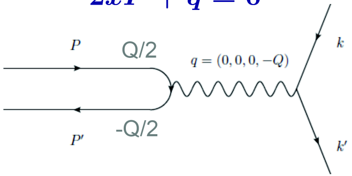
Jet production at HERA H1&ZEUS, EPJ C 82 (2022) 3, 243



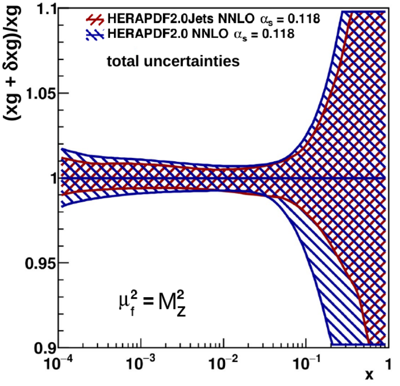
elweak coupling $\propto \alpha_S$ $\propto \alpha_S$ $\propto \alpha_S^2$

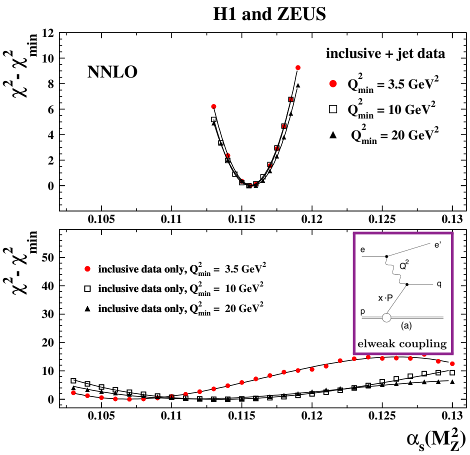


$$2xP + q = 0$$



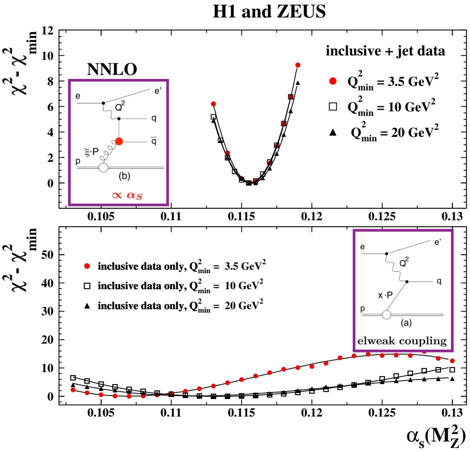
- Breit frame: separates QCD from EW
- QCD jets at LO produced via Boson-Gluon Fusion (BGF) and QCD Compton (QCDC): $\rightarrow$ *probe* $\boldsymbol{g \cdot \alpha_S}$ *at LO*



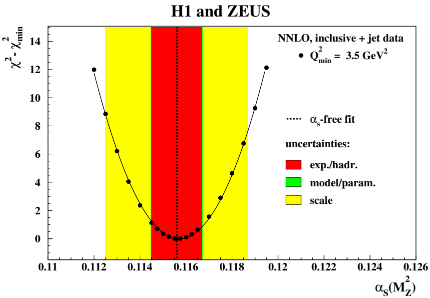


- NNLO analyses of inclusive data **cannot sufficiently constrain** α_S
  - only indirectly sensitive to α_S via scaling violation and F_L
  - extracted α_S is very sensitive to the cut on photon virtuality Q^2: higher twist effects, higher-order corrections, etc.

Jet production at HERA H1&ZEUS, EPJ C 82 (2022) 3, 243



- NNLO analyses of inclusive data **cannot sufficiently constrain** α_S
  - only indirectly sensitive to α_S via scaling violation and F_L
  - extracted α_S is very sensitive to the cut on photon virtuality Q^2: higher twist effects, higher-order corrections, etc.
- However, measurements of jet production are **directly** sensitive to α_S

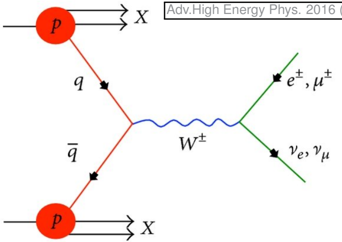


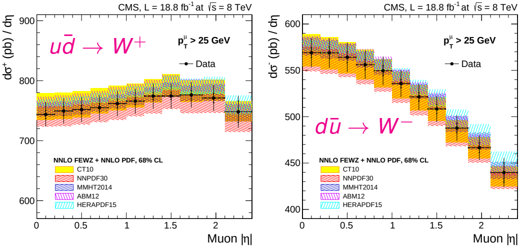
Combined PDF+α_S NNLO fit to the HERA inclusive+jet data:

$$\alpha_S(M_Z) = 0.1156 \pm 0.0011(\text{exp}) ^{+0.0001}_{-0.0002}(\text{model} + \text{param.}) \pm 0.0029(\text{scale})$$

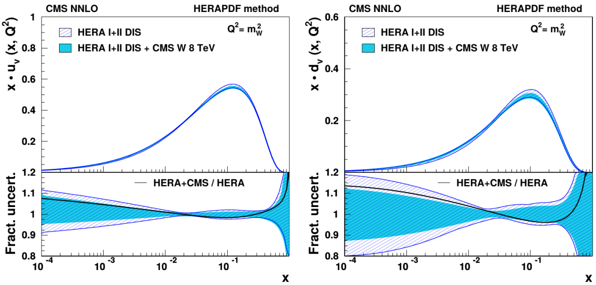
W production at LHC CMS, EPJ C76 (2016) 469

Adv.High Energy Phys. 2016 (2016) 7865689

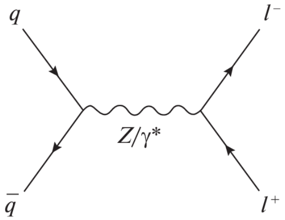


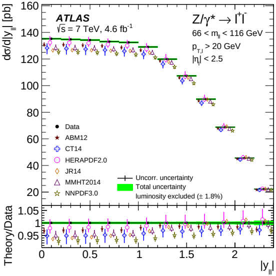


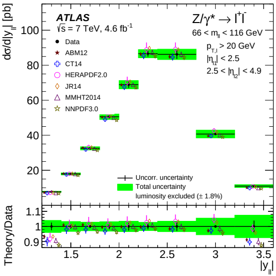
→ Flavor separation: improves valence distributions at $10^{-3} < x < 10^{-1}$



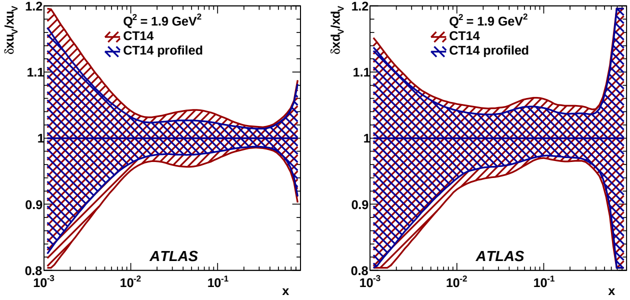
Z production at LHC ATLAS, EPJ C77 (2017) 367



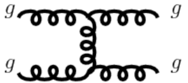


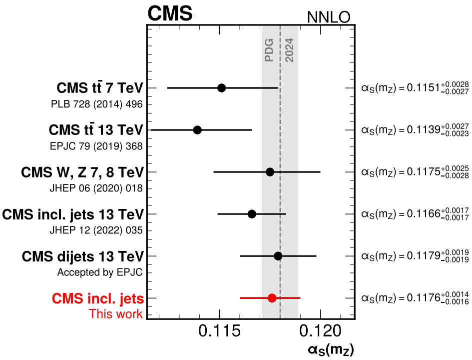


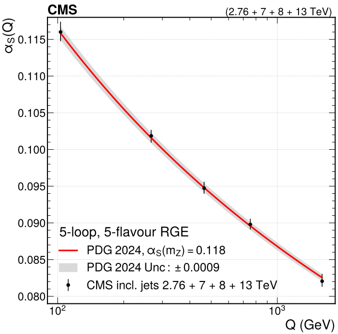
→ Flavor separation: improves valence distributions at $10^{-3} < x < 10^{-1}$

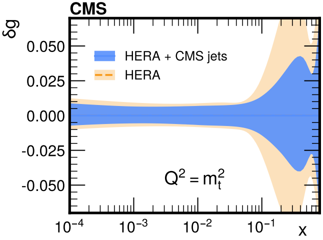


Jet production at LHC CMS, PLB 868 (2025) 139651



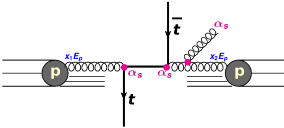




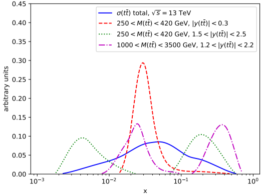


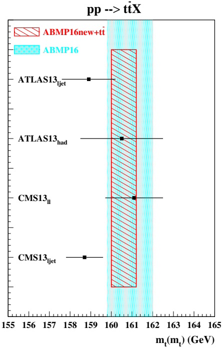
- **Simultaneous fit** of PDFs and α_S
- Used data on inclusive jet production $d^2\sigma/dp_T dy$ at $\sqrt{s} = 2.76, 7, 8$ and 13 TeV and **their correlations**
- **Significant constraints on gluon PDF, especially at large** x
- $\alpha_S(M_Z) = 0.1176^{+0.0014}_{-0.0016}$ + test of α_S running from $Q = 100$ to 1600 GeV

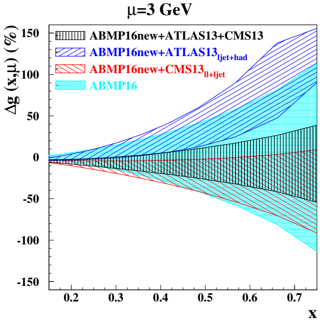
Top quark production at LHC [Alekhin et al., EPJ C 85 (2025) 162]

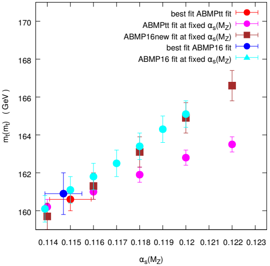


- Top quark pair production is sensitive to g PDF, α_S and m_t:
  $\sigma^{tot}(t\bar{t}) \sim g \times g \times \alpha_S^2/m_t^2$ (LO)
- Differential $t\bar{t}$ cross sections are sensitive to these parameters in different way
- e.g. double-differential $t\bar{t}$ data $d^2\sigma/M(t\bar{t})y(t\bar{t})$:
  - $M(t\bar{t}) \sim \sqrt{sx_1 x_2}$ is sensitive to m_t, g
  - $y(t\bar{t}) \sim \frac{1}{2} \ln \frac{x_1}{x_2}$ is sensitive to g
  - ideally one needs $d^3\sigma/M(t\bar{t})y(t\bar{t})N_{jet}$ [CMS,EPJ C80 (2020) 658], however, theory is still NLO only for $t\bar{t}$ + jets









Modern proton PDF fits: brief summary of methodology

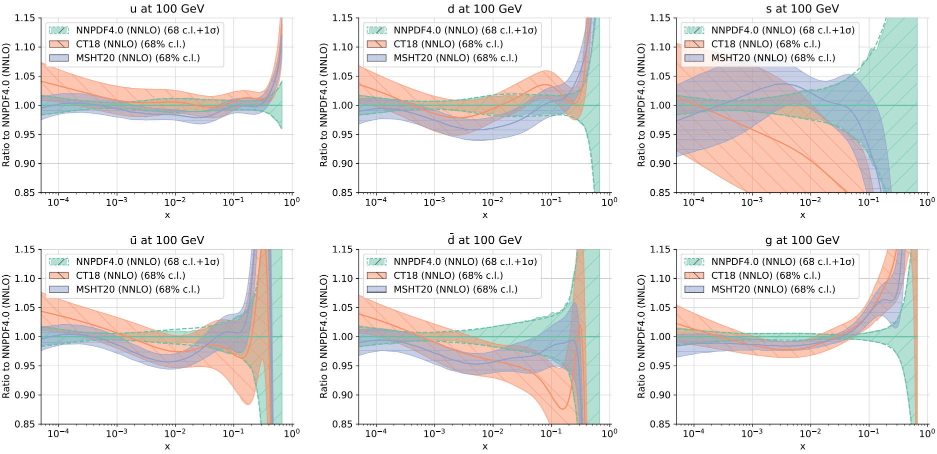
	ABMPtt	CT18	MSHT20	NNPDF40
Year	2023	2019	2020	2021
Data:				
DIS	$+$	$+$	$+$	$+$
W,Z	$+$	$+$	$+$	$+$
$t\bar{t}$	$+$	$+$	$+$	$+$
single t	$+$	$-$	$-$	$+$
jets	$-$	$+$	$+$	$+$
...				
Parametrisation	functional	functional	functional	NN
Uncertainties	$\Delta\chi^2 = 1$	dyn. $\Delta\chi^2 > 1$	dyn. $\Delta\chi^2 > 1$	MC method
α_S, $m_{c,b,t}$	fitted	fixed	fixed	fixed

Remarks:

- **disclaimer: this is not an exhaustive summary**
- these are latest "general-purpose" proton PDF fits at NNLO QCD available on LHAPDF
  - do not consider e.g. QCD+QED, or approximate N^{3}LO QCD fits
- there is a variety of data sets in each group (e.g. various $t\bar{t}$ differential distributions)
  - additional data on $p_T(Z)$, $W + c$, direct photon production, ...
- CT, MSHT, NNPDF determine α_S, $m_{c,b,t}$ in separate fits, but fix them in their nominal ones

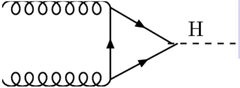
Modern proton PDF fits: comparison

EPJ C82 (2022) 428



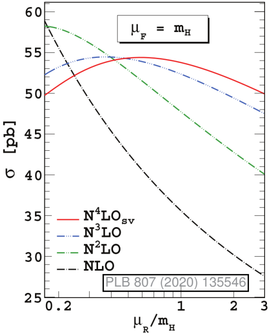
- *u*, *ū*, *d*, $\bar{d}$, *g*: 1% to a few % uncertainties in the bulk of *x* region
    - larger uncertainties at $x \lesssim 10^{-4}$ or $x \to 1$; *s*: $\sim$ 5–10% uncertainty
- However, different PDF fits do not always agree within their uncertainties
- ABMP is not shown here because of different α_S: larger differences w.r.t other fits

Application: Higgs x-section with modern PDFs



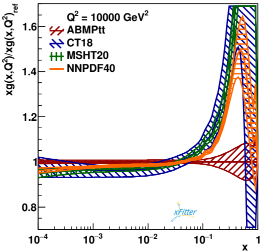
PDF Name	N2LO	N3LO	N4LOsv
ABMP16	$(45.4\pm4.6)^{+0.7}_{-0.7}$	$(49.6\pm2.6)^{+0.8}_{-0.8}$	$(50.8\pm1.9)^{+0.9}_{-0.9}$
ABMPtt	$(45.0\pm4.6)^{+0.6}_{-0.6}$	$(49.2\pm2.6)^{+0.7}_{-0.7}$	$(50.4\pm1.9)^{+0.8}_{-0.8}$
CT18NNLO	$(47.4\pm5.1)^{+1.3}_{-1.7}$	$(52.0\pm2.9)^{+1.4}_{-1.9}$	$(53.4\pm2.1)^{+1.5}_{-1.9}$
MMHT2014nnlo68cl	$(47.7\pm5.1)^{+0.6}_{-0.8}$	$(52.3\pm2.9)^{+0.7}_{-1.0}$	$(53.8\pm2.2)^{+0.7}_{-1.0}$
MSHT20nnlo_as118	$(47.4\pm5.1)^{+0.5}_{-0.6}$	$(52.0\pm2.9)^{+0.6}_{-0.6}$	$(53.4\pm2.1)^{+0.6}_{-0.6}$
NNPDF40_nnlo_as_01180	$(47.8\pm5.1)^{+0.3}_{-0.3}$	$(52.4\pm2.9)^{+0.3}_{-0.3}$	$(53.8\pm2.2)^{+0.3}_{-0.3}$
PDF4LHC21_40	$(47.6\pm5.1)^{+0.8}_{-0.8}$	$(52.3\pm2.9)^{+0.9}_{-0.9}$	$(53.7\pm2.2)^{+0.9}_{-0.9}$
MSHT20an3lo_as118	$(45.0\pm4.8)^{+0.8}_{-0.7}$	$(49.4\pm2.8)^{+0.9}_{-0.8}$	$(50.7\pm2.0)^{+0.9}_{-0.8}$

Table 1: Higgs cross-section along with the absolute error obtained from seven-point scale variation around $(\mu_R{}^c, \mu_F{}^c) = (1,1)m_H$ as well as intrinsic PDF uncertainty using LHAPDF. $\sqrt{S} = 14$ TeV, α_S from LHAPDF (NNLO value).



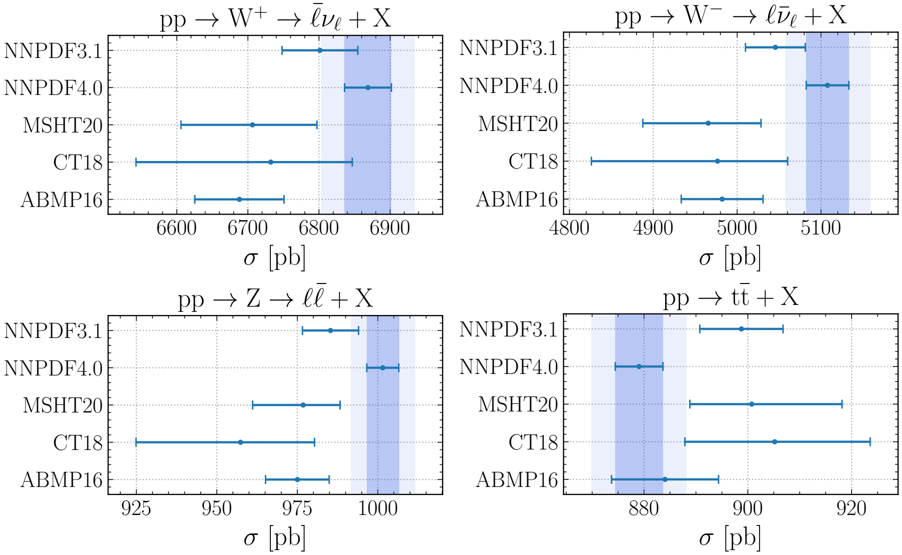
- Das, Moch, Vogt, Phys.Lett.B 807 (2020) 135546:
  - N4LOsv: soft virtual ggF corrections at 4 loops
  - N3LO: effective theory for $m_t \gg m_H$
  - N2LO: full theory for $m_H \lesssim m_t$
  - → apparent convergence of perturbative series
- N4LOsv estimates missing higher-order corrections: 2%
- Larger differences originate from PDF + α_S sets:

  7% (1995) → 12% (2020) → **7% (2024)** → **???**



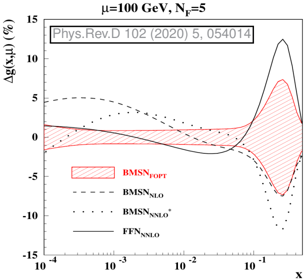
Application: other processes at LHC with modern PDFs

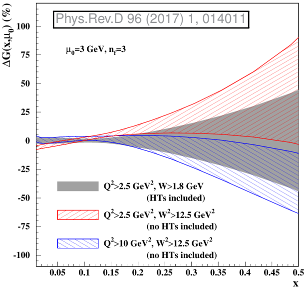
EPJ C82 (2022) 428



Some reasons for differences between global fits

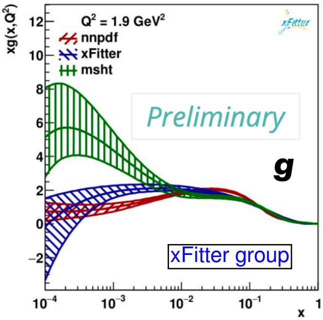
- Heavy-flavor PDF evolution and heavy flavour scheme for DIS
  Phys.Rev.D 102 (2020) 5, 054014
  - fixed-flavour number scheme in ABMP fits is accurate to NNLO for light and heavy quark production (exact or to the best available approximation) and works very well for HERA kinematics
  - CT, MSHT and NNPDF fits use different variable-flavour number schemes which miss some NNLO corrections for heavy quark production and have further theoretical uncertainties
- Higher-twist (HT) corrections for DIS
  Phys.Rev.D 96 (2017) 1, 014011
  - ABMPtt/ABMP16 fits HT terms
  - other fits apply cuts to reduce impact of HT
- Correlation between gluon PDF, α_S and m_t:
  - fully accounted for in ABMPtt/ABMP16
  - α_S and m_t are fixed in other fits

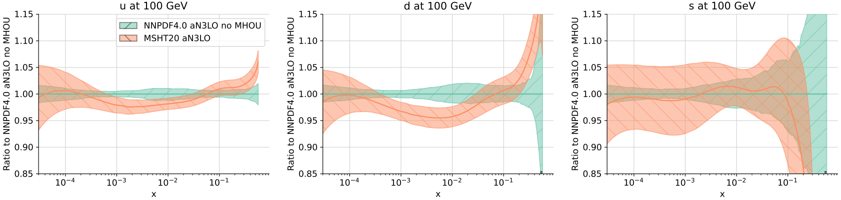




Towards PDF extraction at N^{3}LO

- PDF determination is being pushed to N^{3}LO now
  - N^{3}LO theory for PDF evolution and DIS is almost there (see e.g. arXiv:2406.16188)
  - some processes at hadron colliders, e.g. DY production, are also known at N^{3}LO
- Recent N^{3}LO PDF fits:
  - MSHT group: **includes α_S variation study**
  - NNPDF group
  - xFitter group [preliminary]
- There is a large difference for the gluon PDF among the different groups:
  - different approximations of missing N^{3}LO pieces
  - *g* is strongly correlated with α_S

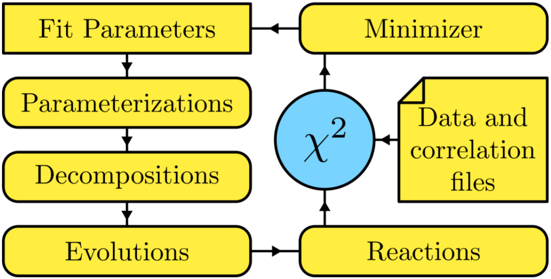




→ careful benchmark is desired: **need public codes**

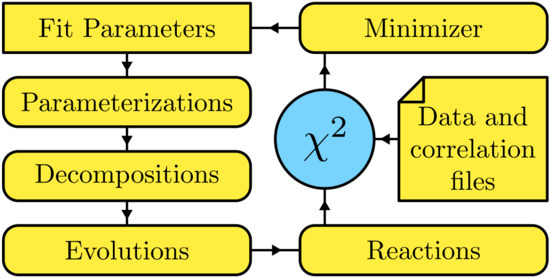
Overview of xFitter [gitlab.com/fitters/xfitter]

I. Novikov



- **xFitter (HERAfitter before 2015) is a unique open-source QCD fit framework:**
  - extract PDFs and theory parameters
  - assess impact of new experimental (pseudo-)data and check consistency
  - test different theoretical assumptions
  - . . . any exercise which involves data vs. theory
- It is widely used by LHC experiments and theorists ($>$ 100 publications)

Flexibility of xFitter (1)

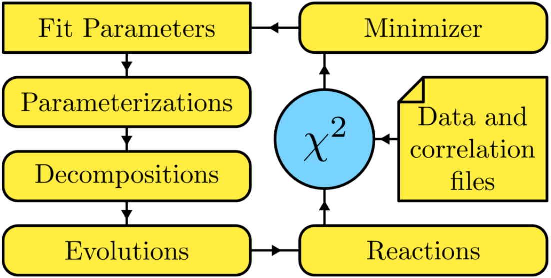


Why is xFitter UNIQUE and so VERSATILE/FLEXIBLE/ADAPTABLE?
Because it is fully modular.

E.g., hadron interactions are realized as:

- PDF parametrisation at starting scale: it is enough to type your favourite formulas
- PDF decomposition: construct valence, sea and gluon, apply sum rules (automatic numerical integration is available)
- PDF evolution: interfaced various codes (QCDNUM, OPENQCDRAD, APFEL, LHAPDF, **APFEL++, HOPPET (new!)** for PDF evolution up to **approximate N^{3}LO** (TMD PDF evolution is also available in (unofficial) branch)
- hard scattering ("reaction"): very many processes are available

Flexibility of xFitter (2)

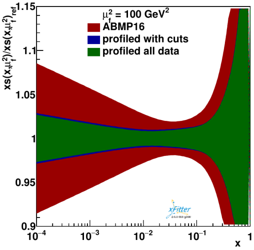


- hard scattering (“reaction”): very many processes are available
  - various heavy-quark schemes for *ep* DIS
  - some “simple” calculations, e.g. LO DY
  - interfaced external packages, e.g. HATHOR (NNLO total heavy-quark and single *t* hadroproduction) and HVQMNR (NLO heavy-quark differential hadroproduction)
  - but main emphasis is put on interfaces to **fast interpolation tables**, such as **fastNLO, ApplGrid, PineAppl**: allows us to get recent higher-order calculations (e.g. MCFM, MATRIX etc.) “for free”
- Flexible χ^2 implementation (treatment of experimental uncertainties)
- χ^2 minimisation: MINUIT, CERES
- **. . . and one can change/mix all these ingredients freely!**

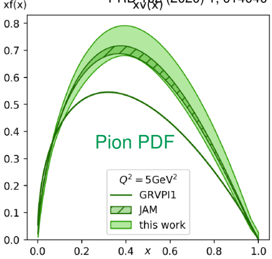
Selected studies by the xFitter team

- "Probing the strange content of the proton with charm production in charged current at LHeC" [Eur. Phys. J. C 79, 864 (2019)]
- "Parton Distribution Functions of the Charged Pion Within The xFitter Framework" [Phys.Rev.D 102 (2020) 1, 014040]
- (preliminary) N^{3}LO fit of DIS HERA data (extension with fixed-target data in progress)

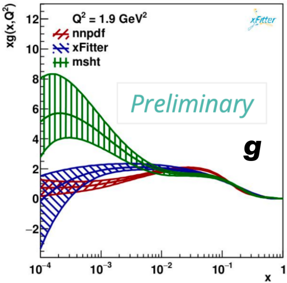
Eur. Phys. J. C 79, 864 (2019)



PRD 102 (2020) 1, 014040



EPS 2025



Basic ingredient for any data vs. theory fit: χ^2 expression

$$\chi^2_{\text{exp}}(\boldsymbol{m}, \boldsymbol{b}) = \sum_{ij} \left(m_i - \sum_\alpha \Gamma^i_\alpha b_\alpha - \mu_i \right) C^{-1}_{\text{stat},\ ij} \left(m_i - \sum_\alpha \Gamma^i_\alpha b_\alpha - \mu_i \right) + \sum_\alpha b^2_\alpha$$

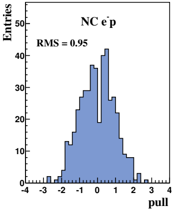
- m_i: data
- μ_i: theory
- $C_{\text{stat},\ ij}$: statistical covariance matrix
- b_α: nuisance parameters for correlated systematic uncertainties
- Γ_α: scaled correlated systematic uncertainties; might depend on m_i, μ_i:

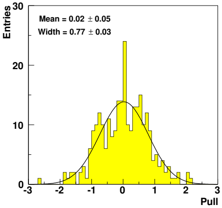
Treatment	Scaling rule (Γ^i_α)
Poisson	$\sqrt{m_i \mu_i}$
Multiplicative	m_i
Additive	μ_i

- Correlated uncertainties can be supplied as covariance matrix or source-by-source
- Also uncertainties can be included with offset method (external variations)

$\Longrightarrow$ Need to know what are uncorrelated and correlated uncertainties, and how they scale (input from experiments which is not always available)

HERA DIS data: discussion which might be relevant for future machines





Q^2 GeV^2	x_{Bj}	$\sigma^{+}_{r,\mathrm{NC}}$	δ_{stat} %	δ_{uncor} %	δ_{cor} %	δ_{rel} %	$\delta_{\gamma p}$ %	δ_{had} %	δ_1 %	δ_2 %	δ_3 %	δ_4 %	δ_{tot} %
70	0.922×10^{-3}	1.385	2.57	5.67	1.10	0.12	−0.16	−0.29	−0.04	−0.29	0.01	−0.80	6.39
70	0.100×10^{-2}	1.434	2.18	2.55	0.86	−0.08	−0.11	−0.20	−0.02	0.08	0.01	−0.93	3.60
70	0.110×10^{-2}	1.506	1.88	2.10	0.81	−0.08	−0.11	−0.17	−0.01	0.48	0.01	−0.97	3.14
70	0.124×10^{-2}	1.445	1.48	1.80	0.80	−0.15	−0.10	−0.16	−0.01	0.24	0.01	−0.95	2.66
70	0.130×10^{-2}	1.414	1.32	1.32	0.82	−0.16	−0.08	−0.12	0.00	0.45	0.00	−1.09	2.37
70	0.200×10^{-2}	1.246	1.15	1.36	0.78	−0.03	−0.09	−0.12	0.00	0.22	0.01	−0.96	2.19
70	0.250×10^{-2}	1.190	0.96	0.97	0.77	0.03	−0.09	−0.12	0.00	0.13	0.01	−0.91	1.82
70	0.320×10^{-2}	1.084	0.91	0.78	0.74	0.26	−0.10	0.00	0.01	−0.13	0.03	−0.52	1.53
70	0.500×10^{-2}	0.958	0.83	0.74	0.75	0.23	−0.09	−0.09	−0.01	−0.49	0.05	−0.43	1.51
70	0.800×10^{-2}	0.819	1.62	0.48	0.77	0.26	−0.05	−0.11	0.01	−0.06	0.07	−0.05	1.88
70	0.130×10^{-1}	0.716	1.89	0.51	0.86	0.46	−0.07	−0.62	0.01	−0.06	0.05	−0.04	2.27
70	0.200×10^{-1}	0.637	1.77	0.67	0.80	0.37	−0.11	−0.46	−0.01	−0.06	0.07	−0.05	2.14
70	0.320×10^{-1}	0.561	2.10	0.92	0.85	−0.02	−0.05	0.26	−0.04	−0.05	0.07	−0.08	2.46
70	0.500×10^{-1}	0.512	1.61	0.90	1.19	0.35	−0.16	−0.37	−0.12	−0.03	0.00	0.02	2.27
90	0.130×10^{-2}	1.479	0.66	1.49	0.87	0.06	−0.04	−0.13	0.00	−0.05	0.15	0.24	1.87
90	0.150×10^{-2}	1.418	1.12	1.03	1.33	1.02	−1.18	−0.21	0.01	−0.04	0.08	0.12	2.56
90	0.200×10^{-2}	1.328	0.79	0.95	0.76	0.53	−0.25	−0.01	−0.01	0.06	0.07	−0.10	1.57
90	0.320×10^{-2}	1.165	0.66	0.63	0.75	0.45	−0.10	−0.08	0.00	0.08	−0.01	−0.20	1.29
90	0.500×10^{-2}	1.023	0.71	0.58	0.74	0.46	−0.09	−0.06	0.00	0.04	−0.03	−0.13	1.27
90	0.800×10^{-2}	0.883	0.87	0.56	0.78	0.38	−0.07	−0.12	0.00	−0.07	0.06	−0.16	1.37
90	0.130×10^{-1}	0.754	0.91	0.69	0.80	0.33	−0.08	−0.22	−0.01	−0.05	−0.22	−0.09	1.47
90	0.200×10^{-1}	0.649	0.95	0.62	0.81	0.36	−0.09	−0.19	0.00	−0.04	0.19	−0.08	1.47

- HERA DIS data are **final combined** H1 and ZEUS data
  - essentially provided as a single data set (no overlap)
  - combination served as a data consistency test
- Very complete description of correlated uncertainties
- Bin-by-bin unfolding (very good resolution of kinematic variables Q^2,x_{Bj})
  - however, sometimes at phase space corners a coarse binning had to be used
- Data are reported at (Q^2,x_{Bj}) **values**
  - although experimental measurements were done in *intervals* of Q^2,x_{Bj}
  - these intervals were different in H1 and ZEUS measurements
  - interpolation procedure (**swimming**) was applied to provide data at (Q^2,x_{Bj}) values
  - → potential model dependence, however, corresponding uncertainties were estimated

LHC data: discussion

Typical description of correlated systematic uncertainties

LHCb 5 TeV JHEP06 (2017) 147

Table 2: Fractional systematic uncertainties, in percent. Uncertainties that are computed bin-by-bin are expressed as ranges giving the minimum to maximum values. Ranges for the correlations between p_{T}-y bins and between modes are also given, expressed in percent.

	Uncertainties (%)				Correlations (%)	
	D^0	D^+	D_s^+	D^{*+}	Bins	Decay modes
Luminosity		3.8			100	100
Tracking	3–5	5–7	4–7	5–7	90–100	90–100
Branching fractions	1.2	2.1	5.8	1.5	100	0–95
Simulation sample size	0–10	0–10	2–9	1–10	0	0
Simulation modelling	0.3	0.7	0.6	2	0	0
PID sample size	0–1	0–1	0–2	0–2	0–100	0–100
PID binning	0–30	0–10	0–20	0–20	0	0
Fit model shapes	0–3	0–3	0–3	0.0–1.0	0	0

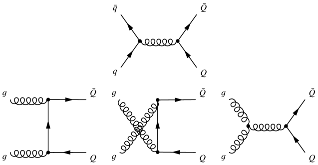
This information is not really sufficient:

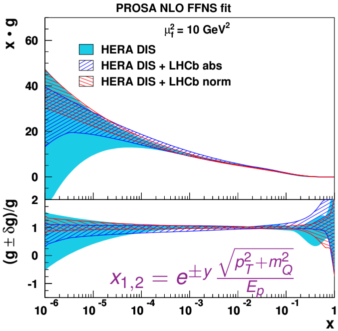
- need to know contributions of different systematic uncertainties for each bin (not just ranges)
- need to know correlation between different *D* and energies
- *total* covariance matrices were provided for some LHCb data sets, but
  - such matrices do not allow one to properly correlate different data sets
  - some of them appeared to be not positive definite (issue of rounding?)

  $\rightarrow$ **similar problems with other LHC data**

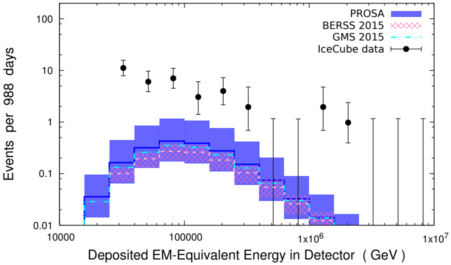
Charm production at LHCb $\rightarrow$ gluon at low x $\rightarrow$ atmosphere ν fluxes

- LHCb measured:
  - charm $0 < p_T < 8$ GeV, $2 < y < 2.5$ [NPB871 (2013) 1]
  - beauty $0 < p_T < 40$ GeV, $2 < y < 2.5$ [JHEP 1308 (2013) 1
- first NLO QCD analysis of these data: Eur. Phys. J. C75 (2015)
- Improved gluon and sea-quark distributions up to $x \gtrsim 5 \times 10^{-6}$ (not covered by other experimental data)
  - predict IceCube background for very high energy cosmic ν [JHEP05 (2017) 004]



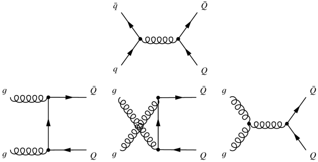


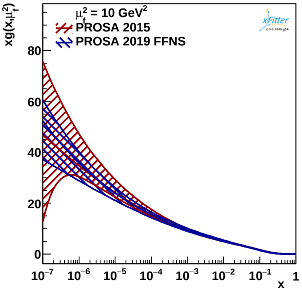
$\Rightarrow$



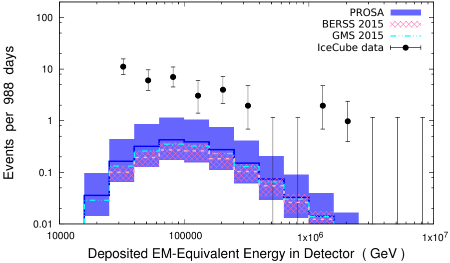
Charm production at LHCb → gluon at low x → atmosphere ν fluxes

- LHCb measured:
  - charm $0 < p_T < 8$ GeV, $2 < y < 2.5$ [NPB871 (2013) 1]
  - beauty $0 < p_T < 40$ GeV, $2 < y < 2.5$ [JHEP 1308 (2013) 1
- first NLO QCD analysis of these data: Eur. Phys. J. C75 (2015)
- Improved gluon and sea-quark distributions up to $x \gtrsim 5 \times 10^{-6}$ (not covered by other experimental data)
  - predict IceCube background for very high energy cosmic ν [JHEP05 (2017) 004]
  - further update with ALICE and new LHCb data [JHEP04 (2020) 118]





$\Rightarrow$



Summary and outlook

- Presently proton PDFs can only be extracted from data:
  - comparing data to theoretical predictions for well understood and well measured processes
  - NNLO QCD is state of the art: available for many processes today
  - first steps towards N^3LO have been done recently
- Different methodology is employed by different groups:
  - parametric or NN approach
  - different selection of data
  - different schemes to treat heavy quarks
  - different calculation codes
  - different methodology to estimate uncertainties

  → sizeable differences which need to be understood independently of moving to N^3LO
- Most of the distributions in the bulk of the kinematic space are claimed to be known with a few % uncertainty today
- Public tools are in place to further benchmark and improve agreement between different PDF extractions

BACKUP

xFitter community

- $\sim$ 10 active developers (both experimentalists and theorists) + a few tens of users
- bi-weekly meetings via zoom + physical room at DESY (Wednesday 3pm)
- (almost) annual external meetings:
  - 2020: DESY
  - 2022: Paris-Saclay University
  - 2023: CERN
- CERN Gitlab `https://gitlab.cern.ch/fitters/xfitter`
  - read access for everyone (also for development branches)
  - needed CERN account to commit new code, or use mirror at `https://gitlab.com`
- DESY support: `naf-xfitter` machine + access to DESY NAF BIRD computing cluster (need DESY computing account)
- every winter/summer school at DESY $\sim$ 1 student (remote/on site) for xFitter
  - very successful projects (e.g. A. Anataichuk et al., EPJ C84 (2024) 1277)
  - contact us if you have students willing to work on phenomenology topics matching xFitter scope
- mailing list: `xfitter-users@googlegroups.com`