

Introduction to xFitter from enthusiastic user

K. Wichmann

@ Low-x 2019

Nicosia Cyprus



xFitter PDF Fitting package

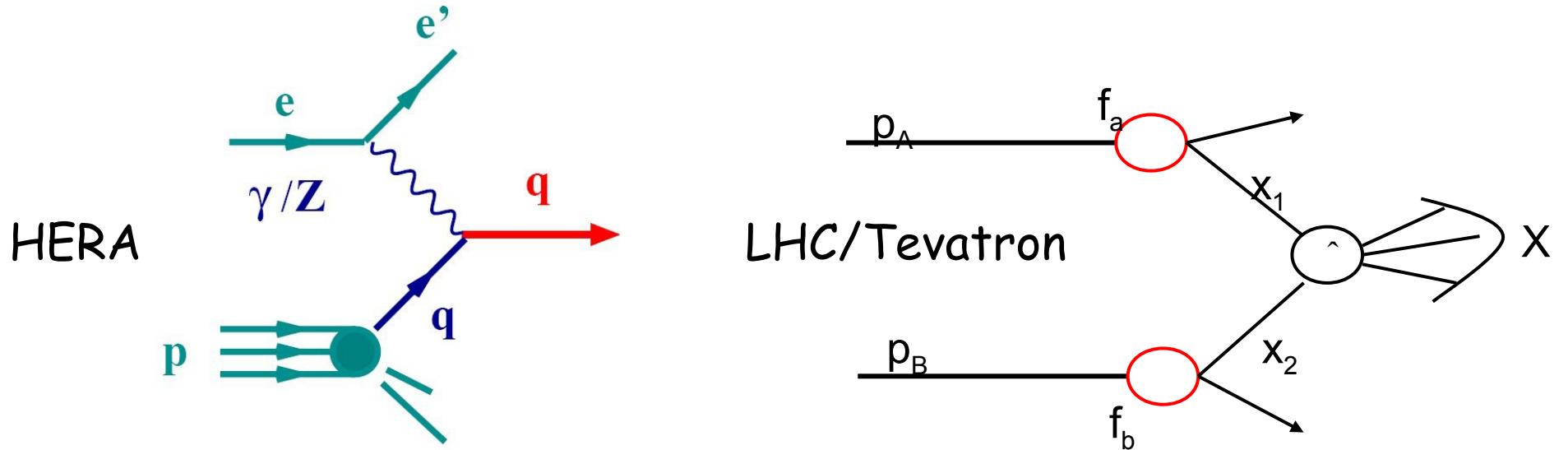


The determination of the proton patron distribution functions is a complex endeavor involving several physics processes. The main process is deep-inelastic scattering and the central data set covering most of the proton structure phase space is provided at the HERA ep collider. Further processes (fixed target DIS, ppbar collisions etc.) provide further constraints for particular aspects: flavor separation, very high Bjorken-x etc. In particular, the precise measurements obtained or to come from LHC will continue to improve the knowledge of the PDF. The xFitter project aim at providing a framework for QCD analyses related to proton structure in the context of multi-processes and multi-experiments. The framework includes modules or interfaces enabling a large number of theoretical and methodological options, as well as a large number of relevant data sets from HERA, Tevatron and LHC. This manual explains the theoretical input used in the QCD analysis, the fit methodology and the installation procedure of the program. More information and the package downloads can be found on the web site <http://xfitter.org>.

- xFitter, Open Source QCD Fit Project: <https://arxiv.org/abs/1410.4412>
- Manual:
<https://www.xfitter.org/xFitter/xFitter/DownloadPage?action=AttachFile&do=view&target=manual.pdf>
- xFitter releases:
<https://www.xfitter.org/xFitter/xFitter/DownloadPage>

Parton Distribution Functions (PDFs): $F_f(x)dx =$ prob. of finding constituent f with longitudinal momentum fraction x

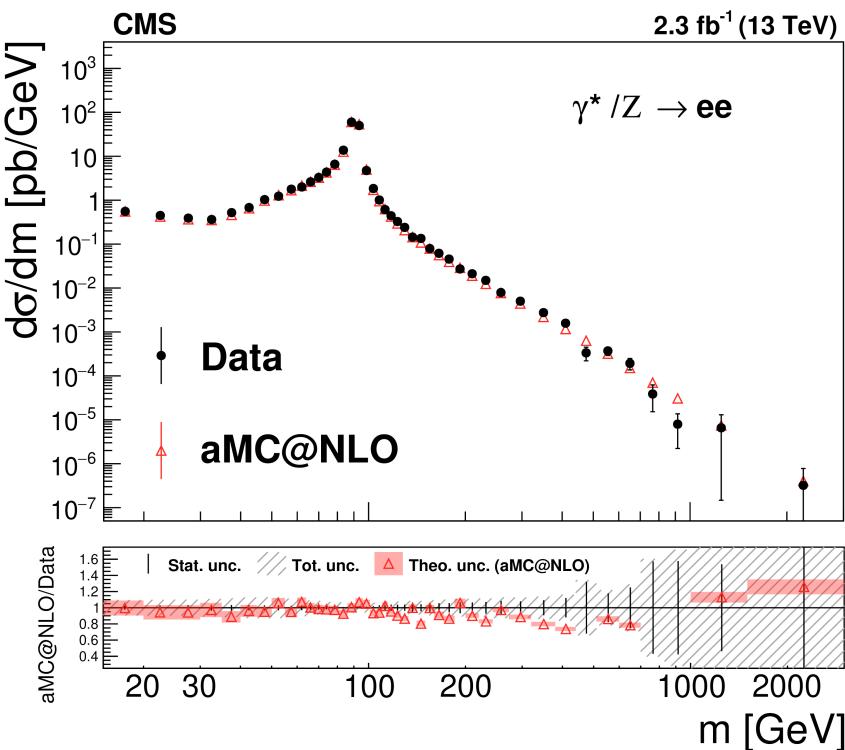
- They are necessary whenever we have interactions with proton



- Parton density functions necessary to compare data to theory/MC
- Different parton density functions result in different cross sections

Motivation for PDF studies

- Uncertainties of many variables often dominated by PDF uncertainty
→ limiting factor in theoretical predictions
 - Standard Model measurements
 - BSM Searches (high- x region)
 - Often very good statistical precision-improved PDF uncertainties help!



“Measurement of the differential Drell-Yan cross section in proton-proton collisions at $\sqrt{s} = 13$ TeV”, CMS, arXiv:1812.10529

m (GeV)	Eff. (%)	Det. resol. (%)	Bkgr. est. (%)	FSR (%)	Total (%)	Acceptance+PDF (%)
15–20	8.9	0.73	0.98	3.5	9.7	5.7
20–25	9.0	1.0	1.5	3.1	9.7	4.7
25–30	8.9	1.3	2.0	3.1	9.7	9.5
30–35	9.6	1.5	2.5	3.4	11	16
35–40	10	1.7	3.3	3.9	12	16
40–45	11	2.0	4.2	4.2	13	14
45–50	11	2.2	4.7	4.0	13	12
50–55	11	2.3	4.6	3.5	13	9.4
55–60	9.7	2.5	4.2	2.4	11	7.6
60–64	8.7	2.8	3.7	1.1	9.9	6.7
64–68	7.4	3.0	2.8	0.71	8.5	5.6
68–72	7.0	2.9	2.0	0.87	7.9	4.7
72–76	6.4	2.4	1.4	0.96	7.1	4.0
76–81	6.3	1.7	0.77	1.0	6.7	3.4
81–86	6.2	0.97	0.30	1.0	6.4	2.8
86–91	5.7	0.53	0.075	1.3	5.9	2.3
91–96	5.5	0.55	0.060	1.9	5.8	1.9
96–101	5.3	0.88	0.20	2.1	5.8	1.6
101–106	5.2	1.6	0.47	2.1	5.9	1.4

Better known PDFs → smaller uncertainties

xFitter - open source QCD framework

→ successful platform well integrated in HEP community

- xFitter infrastructure provides flexible environment for theory benchmarking
 - increased scientific output of both HERA and LHC data
- xFitter allows comparison of use of different data sets and different methodology (like choice of heavy flavour scheme) using common platform
- Stable release: xFitter-2.0.0/2.0.1 available at www.xfitter.org

Date	Version	Files	Remarks
 05/2019	2.0.1 OldFashioned	@xfitter-2.0.1.tgz	update/bug fix to 2.0.0 FrozenFrog
 03/2017	2.0.0 FrozenFrog	@xfitter-2.0.0.tgz	stable release with decoupled data and theory files

Global analysis of parton distributions

Goal: determination of the *input distributions* (for light quarks and gluons):

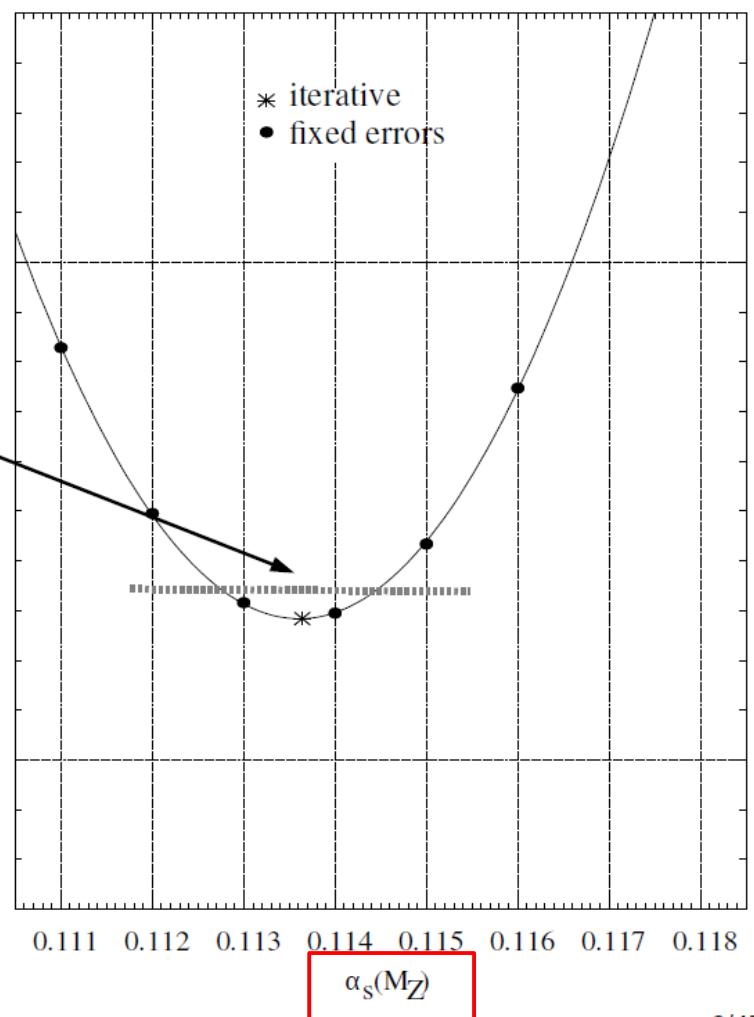
Method: Parametrizations $xf(x, Q_0^2) = Nx^a(1 - x)^b$ function(x)
and usual *statistical estimation* (fits):

$$\boxed{\chi^2(p)} = \sum_{i=1}^N \left(\frac{\text{data}(i) - \text{theory}(i, p)}{\text{error}(i)} \right)^2$$

Position of minimum gives the value
and curvature gives the error (region
within a certain “tolerance” $\Delta\chi^2 = 1$)

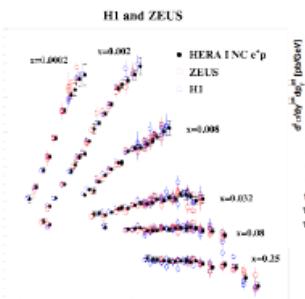
(Monte Carlo methods can also be used)

Usually the chi-square definition is
more sophisticated, experimental
correlations are also treated, etc.



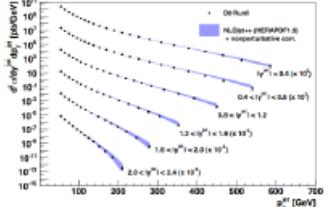
experimental input

H1 and ZEUS



• HERA I NC c'p
ZEUS
H1

Tevatron inclusive jet cross sections



experiments:
HERA, Tevatron,
LHC, fixed target

processes:

NC, CC DIS, jets, diffraction,
heavy quarks (c,b,t)
Drell-Yan, W production

theoretical calculations/tools

Heavy quark schemes: MSTW, CTEQ, ABM
 Jets, W, Z production: fastNLO, Applgrid
 Top production NNLO (Hathor)
 QCD Evolution DGLAP (QCDNUM)
 k_T factorisation
 Alternative tools NNPDF reweighting
 Other models Dipole model
 + Different error treatment models
 + Tools for data combination (HERAaverager)

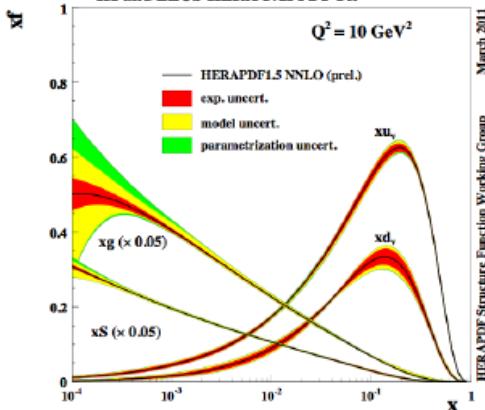
xFitter

H1 and ZEUS HERA I+II PDF Fit

$Q^2 = 10 \text{ GeV}^2$

March 2011

HERAPDF Structure Function Working Group



PDF or uPDF or DPDF

$\alpha_s(M_Z), m_c, m_b, m_t, f_s, \dots$

Theory predictions

Benchmarking

Comparison of schemes



xFitter super powers

Choice of input parametrisation

- HERAPDF/MSTW style
- CTEQ style
- Chebyshev polynomials
- Log-Normal distribution
- Fractal model

Choice of a wide range of data

- HERA data
- LHC: ATLAS, CMS, LHCb,
fixed target



Measurement at LHC	PDF sensitivity
Jets	high x quarks and gluons (alphas)
Inclusive W, Z and asymmetries	quark flavour separation (u,d,s)
Low and high mass Drell-Yan	quarks at low and high x (u,d)
W + charm	Direct sensitivity to s-quark
Isolated photons	medium - x gluons
Single top	u,d and b quark
ttbar (total, differential)	Medium- x gluon (alphas)
W,Z production with jets	Medium- x gluon
Z+b production	sensitive to b-quark



xFitter super powers



Various ways of treating experimental systematic uncertainties

- correlated systematic uncertainties in form of correlation matrix or in terms of correlated shifts for each systematic source
- used in the χ^2 minimisation
- Nuisance parameter representation
 - ▶ Simple form

$$\chi_{\text{exp}}^2(m, b) = \sum_i \frac{\left(m_i - \sum_j \gamma_j^i b_j - \mu_i \right)^2}{(\delta_{i;\text{stat}} \mu_i)^2 + (\delta_{i;\text{uncor}} \mu_i)^2} + \sum_j b_j^2$$

- ▶ Scaled form

$$\chi_{\text{exp}}^2(m, b) = \sum_i \frac{\left(m_i - \sum_j \gamma_j^i b_j - \mu_i \right)^2}{\delta_{i;\text{stat}}^2 \mu_i \left(m_i - \sum_j \gamma_j^i b_j \right) + (\delta_{i;\text{uncor}} m_i)^2} + \sum_j b_j^2 + \text{log penalty}$$

- Covariance matrix representation

$$\chi_{\text{exp}}^2(m) = \sum_{ij} (m_i - \mu_i) C_{ij}^{-1} (m_j - \mu_j)$$

- Mixed form (covariance matrix and nuisance parameters):

$$\chi_{\text{exp}}^2(m, b) = \sum_i \left(m_i - \sum_k \Gamma_k^i (m_j) b_k - \mu_i \right) C_{i;\text{stat}}^{-1} (m_i, m_j) \left(m_j - \sum_k \Gamma_k^j (m_j) b_k - \mu_j \right) + \sum_k b_k^2$$

Propagation of experimental errors to PDFs

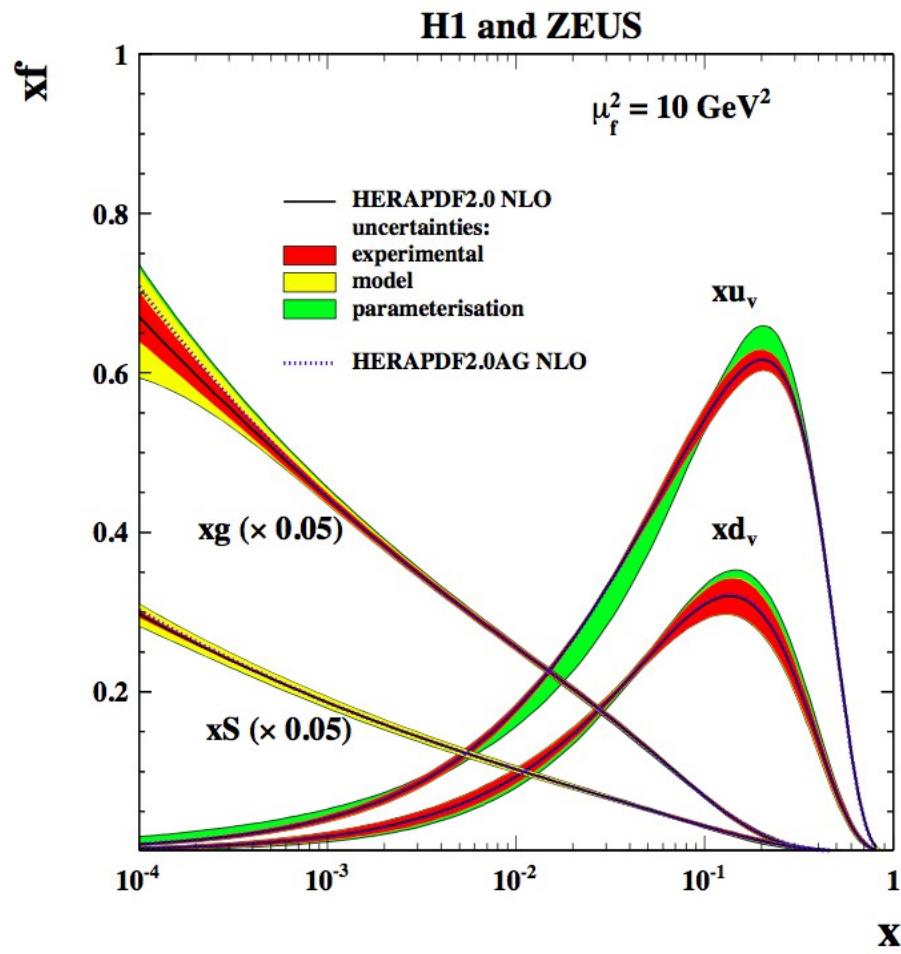
- Hessian method
 - nuisance parameters are fitted, χ^2 tolerance $T > 1$ can also be used to account for marginally compatible input data sets
- Offset method
 - nuisance parameters are applied as 1σ shifts
- MC-method
 - data points are shifted randomly within their 1σ limits to form MC replicas. this also allows for asymmetric uncertainties



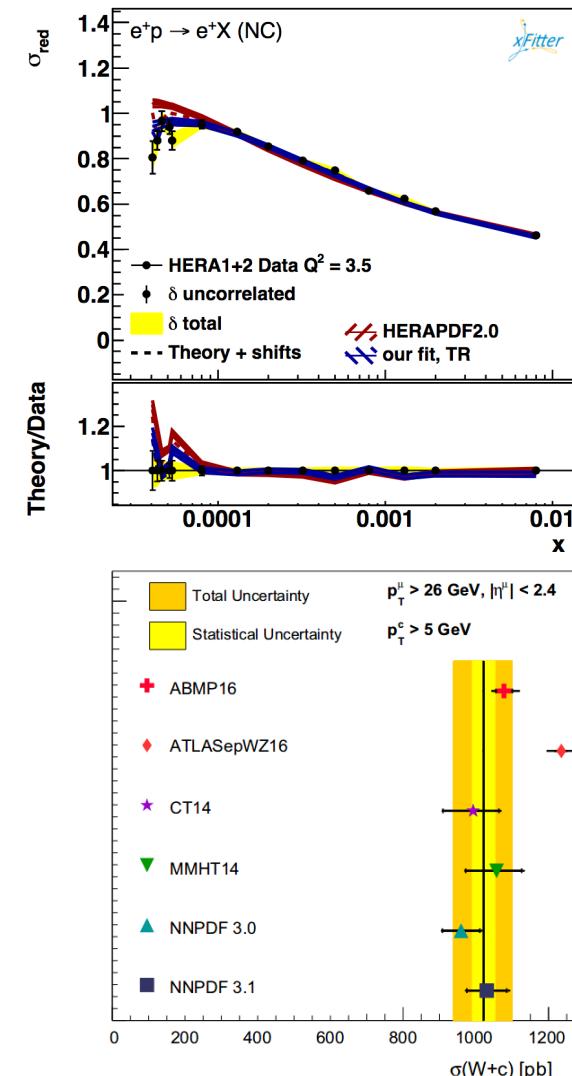
xFitter super powers



Estimating, studying and plotting
experimental, model & parameterisation
uncertainties for PDFs



Comparison of predictions to data
and between various PDF sets
→ full use of LHAPDF libraries



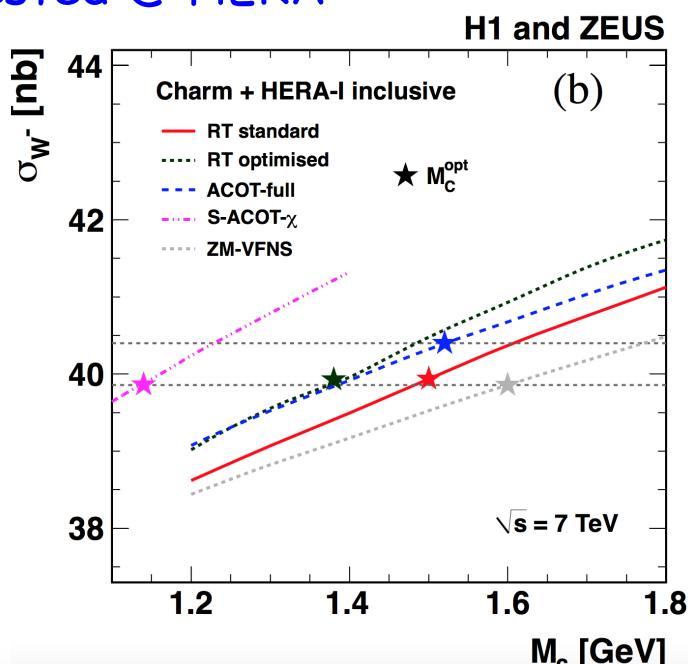
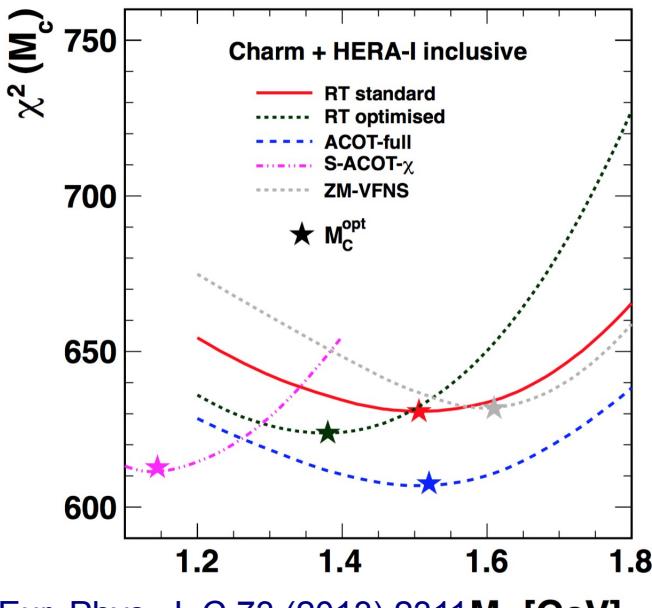


xFitter super powers

Comparison of different theoretical formalisms

- DGLAP evolution with QCDNUM, QCDRAD, APFEL, QEDEVOL
- Heavy quark schemes
 - Variable Flavour Number schemes a la MMHT (**TR**), CT (**ACOT**) or NNPDF (**FONLL**)
 - Fixed Flavour Number schemes as used by ABM (**FF**)

Various schemes tested @ HERA



Spread for W/Z production @ LHC reduced significantly when optimal charm mass evaluated from F_2^{charm} used

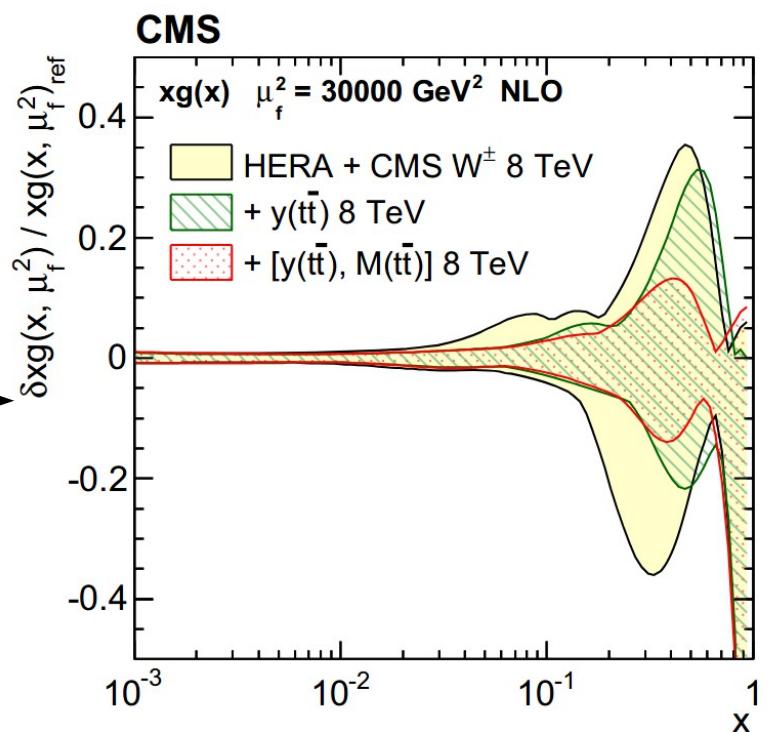
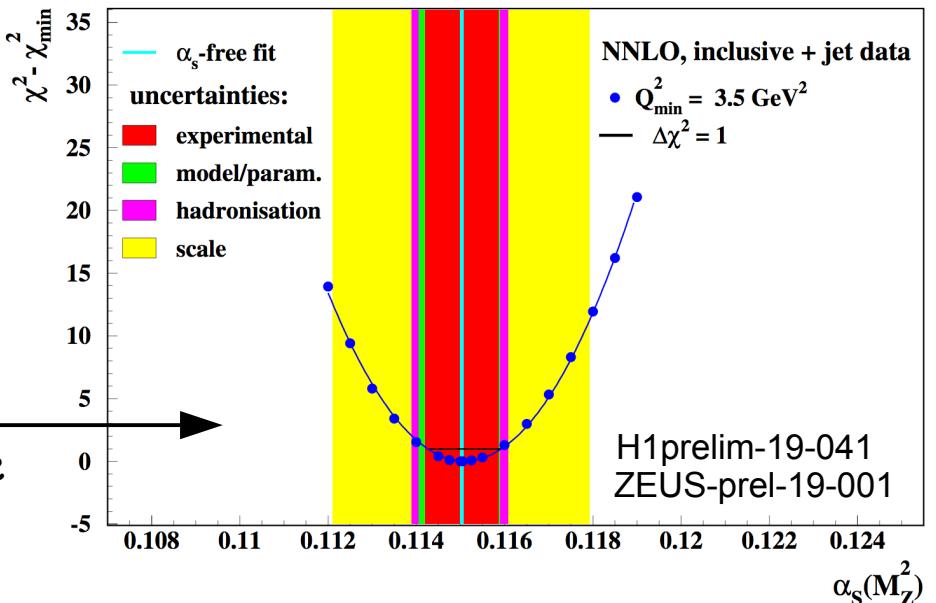


xFitter super powers

Integrates tools for fast calculation of
(N)NLO cross sections

- NLO jets
 - FastNLO or ApplGrid interfaced to NLOJet++
 - NNLO calculations are coming (possible now using master version)
- NLO Drell-Yan and W, Z production
 - Applgrid interfaced to MCFM
- NNLO Drell-Yan
 - DYNNLO should be available soon
- HATHOR
 - total top pair and single top up to NNLO
- FastNLO, MCFM/AppleGrid
 - differential ttbar or single top cross sections at NLO or NNLO
- Dipole model
- Diffractive PDFs

H1 and ZEUS preliminary





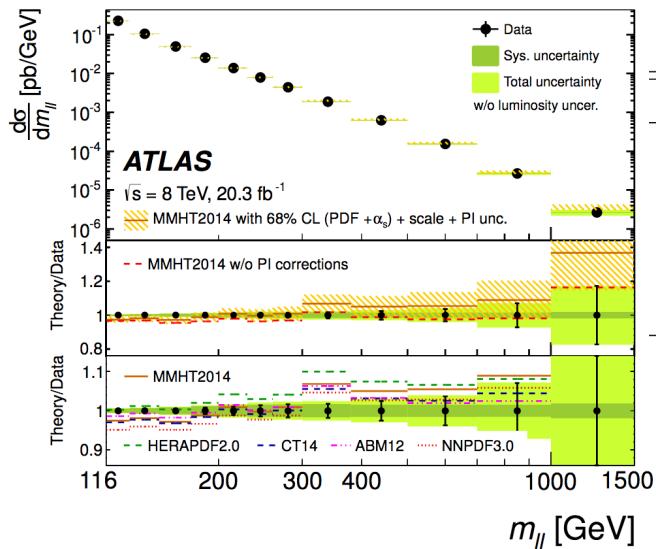
xFitter super powers

Quantitative comparison of agreement between data sets and predictions from various PDFs

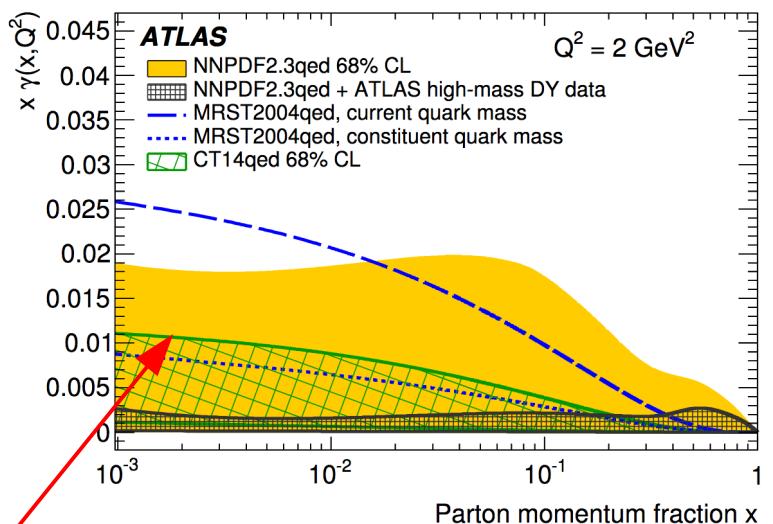
- Experimental and theory uncertainties taken into account

$$\chi^2(b_{\text{exp}}, b_{\text{th}}) =$$

$$\sum_{i=1}^{N_{\text{data}}} \frac{\left(\sigma_i^{\text{exp}} + \sum_{\alpha} \Gamma_{i\alpha}^{\text{exp}} b_{\alpha,\text{exp}} - \sigma_i^{\text{th}} - \sum_{\beta} \Gamma_{i\beta}^{\text{th}} b_{\beta,\text{th}} \right)^2}{\Delta_i^2} + \sum_{\alpha} b_{\alpha,\text{exp}}^2 + \sum_{\beta} b_{\beta,\text{th}}^2.$$



JHEP 08 (2016) 009



- Also NNPDF reweighting method, which weights replica PDFs according to their goodness of fit is implemented
 - Here data are used to reweight NNPDFqed photon

Basis of profiling method for determining impact of new data sets on current PDFs



xFitter super powers

Profiling technique

- Impact of new data on given PDFs can be quantitatively estimated with profiling of eigenvectors or reweighting of replicas, under **assumption** that new data are compatible with these PDFs
- Profiling is performed using χ^2 which includes both experimental and theoretical uncertainties arising from PDF variations:

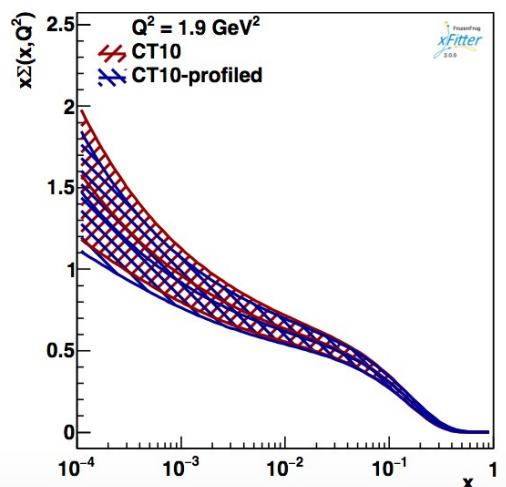
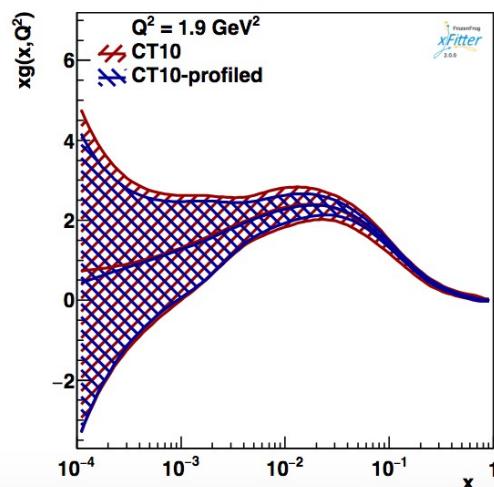
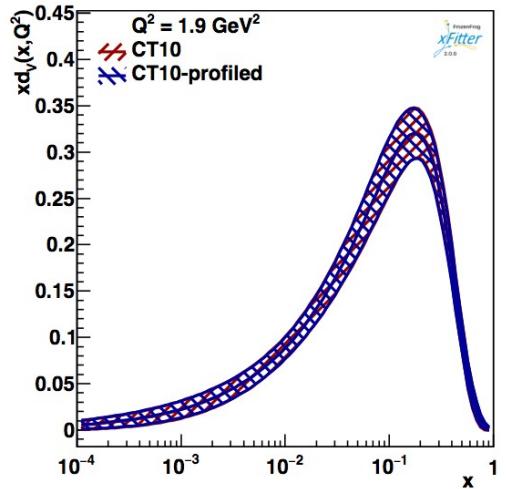
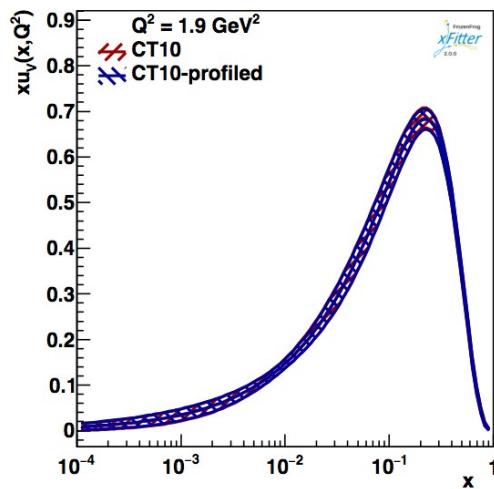
$$\begin{aligned} \chi^2(\mathbf{b}_{\text{exp}}, \mathbf{b}_{\text{th}}) = & \\ & \sum_{i=1}^{N_{\text{data}}} \frac{\left(\sigma_i^{\text{exp}} + \sum_{\alpha} \Gamma_{i\alpha}^{\text{exp}} b_{\alpha,\text{exp}} - \sigma_i^{\text{th}} - \sum_{\beta} \Gamma_{i\beta}^{\text{th}} b_{\beta,\text{th}} \right)^2}{\Delta_i^2} \\ & + \sum_{\alpha} b_{\alpha,\text{exp}}^2 + \sum_{\beta} b_{\beta,\text{th}}^2 \end{aligned}$$

Here $b_{\alpha,\text{exp}}$ and $b_{\beta,\text{th}}$ are nuisance parameters for correlated experimental ($\Gamma_{i\alpha}^{\text{exp}}$) and theoretical ($\Gamma_{i\beta}^{\text{th}}$) uncertainties, respectively

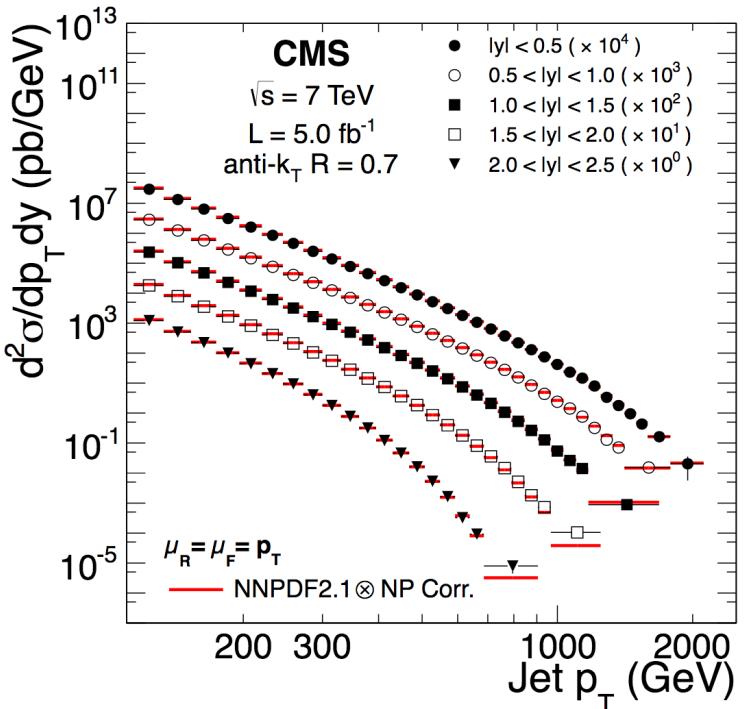
- $b_{\beta,\text{th}}$ which minimise χ^2 with new data are interpreted as optimisation (profiling) of PDFs

Profiling example: CMS jets

- Impact of CMS jet data at 7 TeV on PDFs
 - CT10 as reference
 - done as exercise for CMS data analysis school

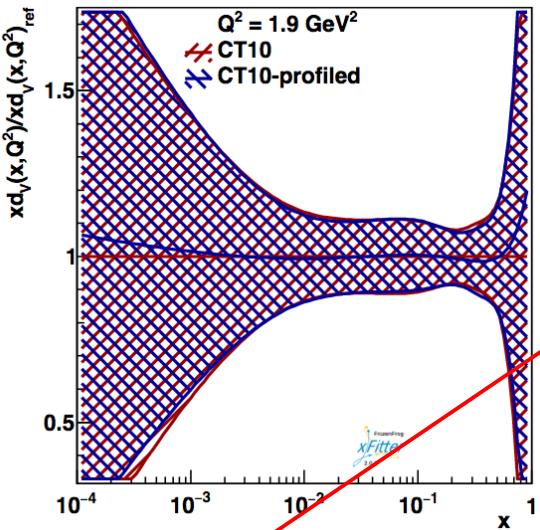
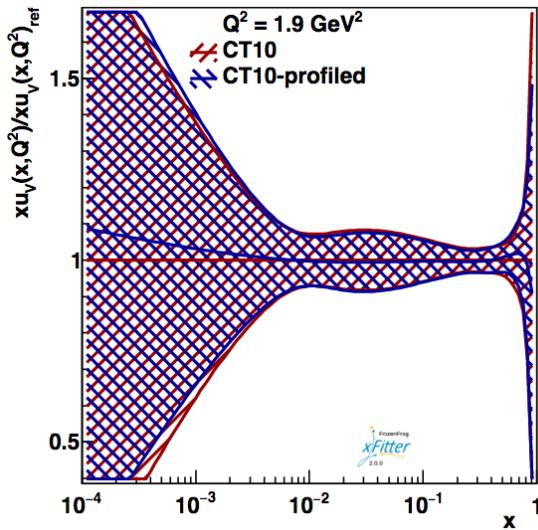


Phys. Rev. D 87 (2013) 112002

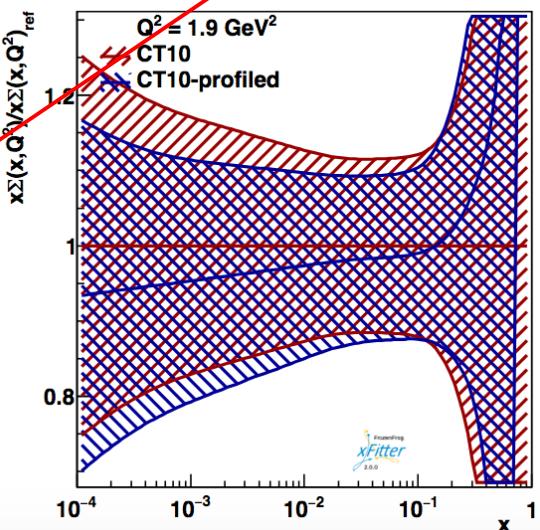
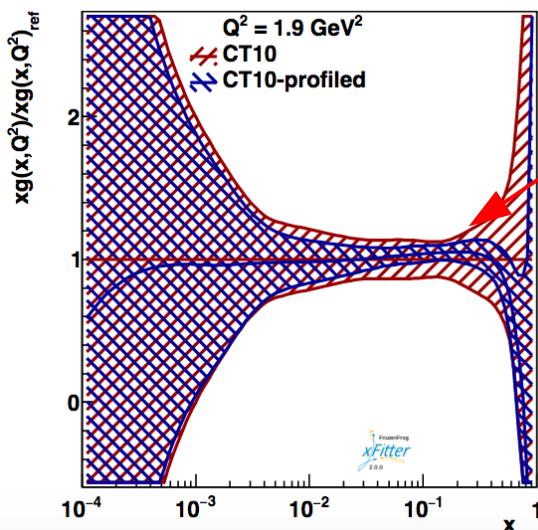


PDFs don't differ much

Experimental uncertainties



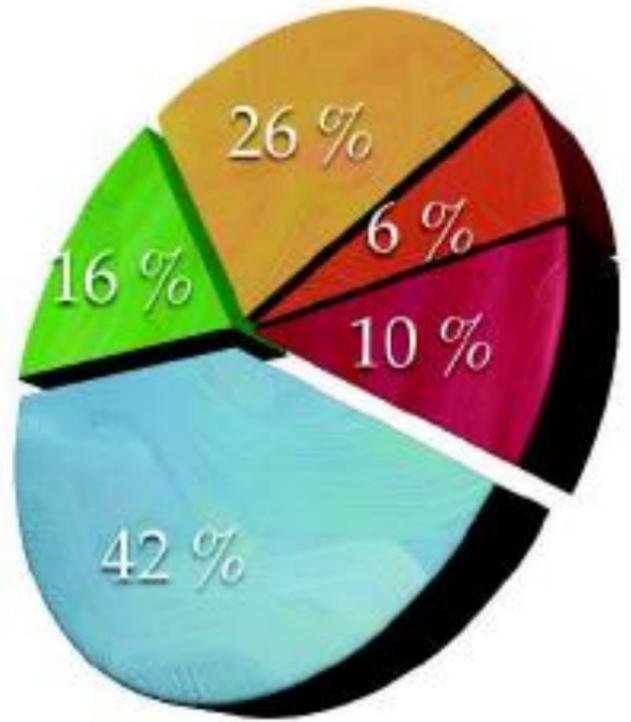
Gluon PDF experimental
uncertainties reduced
→ as expected



Profiling very
informative for "new
machine" studies

Results based on xFitter

- There are > 78 papers using xFitter package
- Talks at this conference (possibly more)
 - Recent PDF results from top and heavy flavour physics at CMS (kw)
 - NNLO QCD fits to jets and extraction of α_s (kw)
 - Determination of parton distribution functions using ATLAS data (L. Bellagamba)
 - low-x studies → example in this talk



Beautiful example of exploring xFitter super powers



**A new simple PDF parametrization:
improved description
of the HERA data**

M. Bonvini, F. Juli

[arXiv:1902.11125](https://arxiv.org/abs/1902.11125)

- HERAPDF-style parameterisation

$$xf(x, \mu_0^2) = Ax^B(1-x)^C[1 + Dx + Ex^2]$$

- HERA low- x low- Q^2 data poorly described - maybe more flexibility at low x helps?

New proposed parametrization

- To model small- x region we proposed polynomial in $\log(x)$

$$(1 + F \log(x) + G \log^2(x) + H \log^3(x) + \dots)$$
- Considered both a multiplicative and an additive option, and we chose the latter:

[arXiv:1902.11125](https://arxiv.org/abs/1902.11125) [Bonvini,FG]

$$xf(x, \mu_0^2) = Ax^B(1-x)^C[1 + Dx + Ex^2 + F \log(x) + G \log^2(x) + H \log^3(x)]$$

HERAPDF2.0 .vs. NEW parameterisation

$$xg(x, \mu_0^2) = A_g x^{B_g} (1-x)^{C_g} - A'_g x^{B'_g} (1-x)^{C'_g}$$

$$xu_v(x, \mu_0^2) = A_{u_v} x^{B_{u_v}} (1-x)^{C_{u_v}} [1 + E_{u_v} x^2]$$

$$xd_v(x, \mu_0^2) = A_{d_v} x^{B_{d_v}} (1-x)^{C_{d_v}}$$

$$x\bar{u}(x, \mu_0^2) = A_{\bar{u}} x^{B_{\bar{u}}} (1-x)^{C_{\bar{u}}} [1 + D_{\bar{u}} x]$$

$$x\bar{d}(x, \mu_0^2) = A_{\bar{d}} x^{B_{\bar{d}}} (1-x)^{C_{\bar{d}}}$$

$$xs(x, \mu_0^2) = x\bar{s}(x, \mu_0^2) = r_s x\bar{d}(x, \mu_0^2) \quad r_s = \frac{f_s}{1-f_s} \quad \text{with } f_s = 0.4 \text{ fixed.}$$

NEW parameterisation

$$xg(x, \mu_0^2) = A_g x^{B_g} (1-x)^{C_g} [1 + F_g \log x + G_g \log^2 x]$$

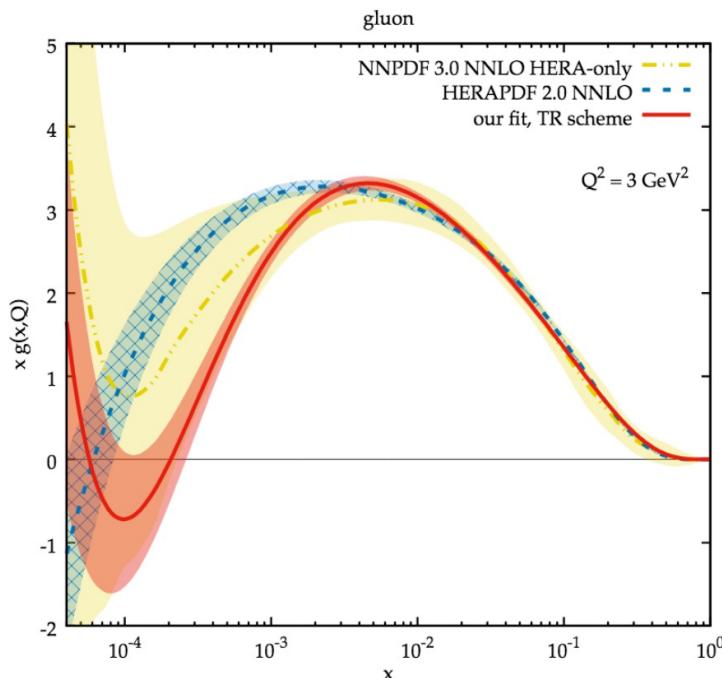
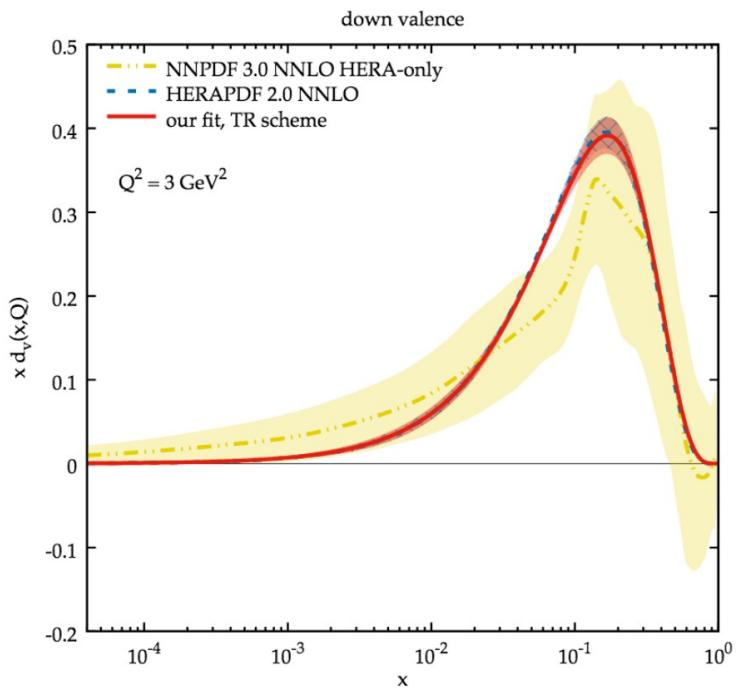
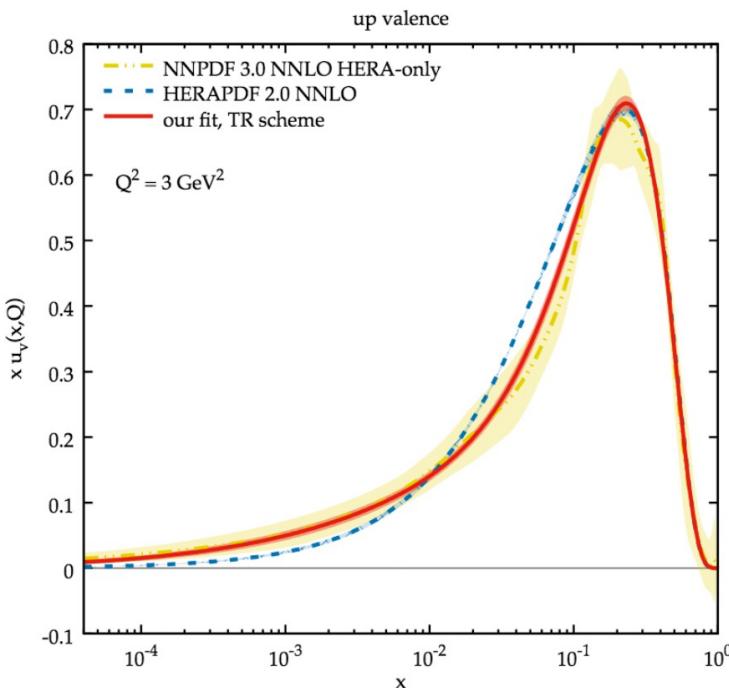
$$xu_v(x, \mu_0^2) = A_{u_v} x^{B_{u_v}} (1-x)^{C_{u_v}} [1 + E_{u_v} x^2 + F_{u_v} \log x + G_{u_v} \log^2 x]$$

$$xd_v(x, \mu_0^2) = A_{d_v} x^{B_{d_v}} (1-x)^{C_{d_v}}$$

$$x\bar{u}(x, \mu_0^2) = A_{\bar{u}} x^{B_{\bar{u}}} (1-x)^{C_{\bar{u}}} [1 + D_{\bar{u}} x + F_{\bar{u}} \log x]$$

$$x\bar{d}(x, \mu_0^2) = A_{\bar{d}} x^{B_{\bar{d}}} (1-x)^{C_{\bar{d}}} [1 + D_{\bar{d}} x + F_{\bar{d}} \log x],$$

**4 extra
parameters = 18**



- Most difference in u valence and gluon
→ shape agrees with NNPDF3.0
- More rapid decrease of gluon at $x \sim 0.01$ and rise for $x < 10^{-4}$

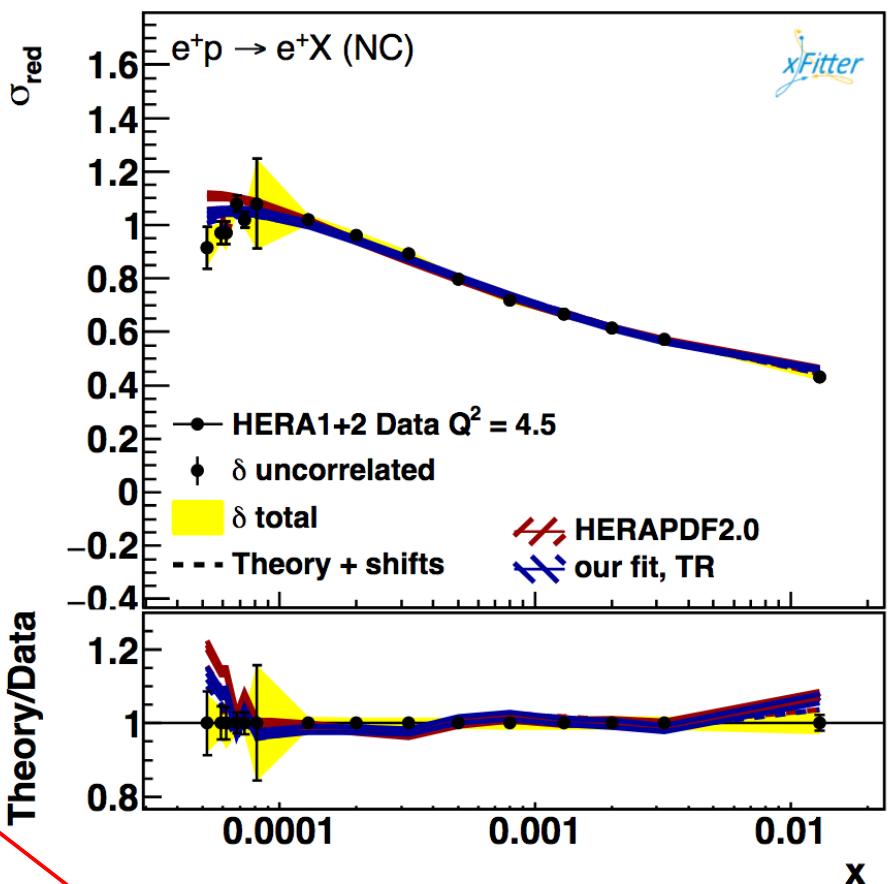
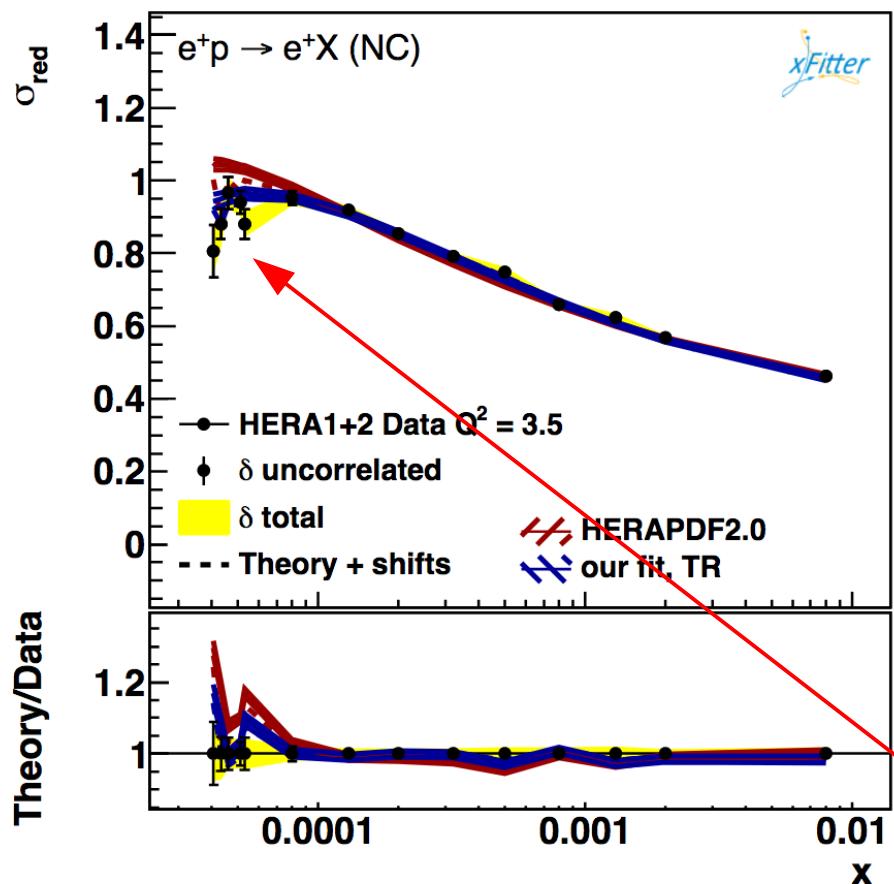
Comparing χ^2

- χ^2 dropped by 62 units - a lot!

Contribution to χ^2	HERAPDF2.0	Our fit (new parametrization)
subset NC e^+ 920 $\tilde{\chi}^2/\text{n.d.p.}$	444/377	403/377
subset NC e^+ 820 $\tilde{\chi}^2/\text{n.d.p.}$	66/70	74/70
subset NC e^+ 575 $\tilde{\chi}^2/\text{n.d.p.}$	219/254	221/254
subset NC e^+ 460 $\tilde{\chi}^2/\text{n.d.p.}$	217/204	222/204
subset NC e^- $\tilde{\chi}^2/\text{n.d.p.}$	219/159	220/159
subset CC e^+ $\tilde{\chi}^2/\text{n.d.p.}$	45/39	38/39
subset CC e^- $\tilde{\chi}^2/\text{n.d.p.}$	56/42	50/42
correlation term + log term	91 + 5	75 – 3
Total $\chi^2/\text{d.o.f.}$	1363/1131	1301/1127

χ^2 significantly improved with new parameterisation

Improved description of HERA data



- Improvement comes mostly from low- Q^2 and low- x region
 - Theoretical description clearly better
 - Impact of $\log(1/x)$ terms largest

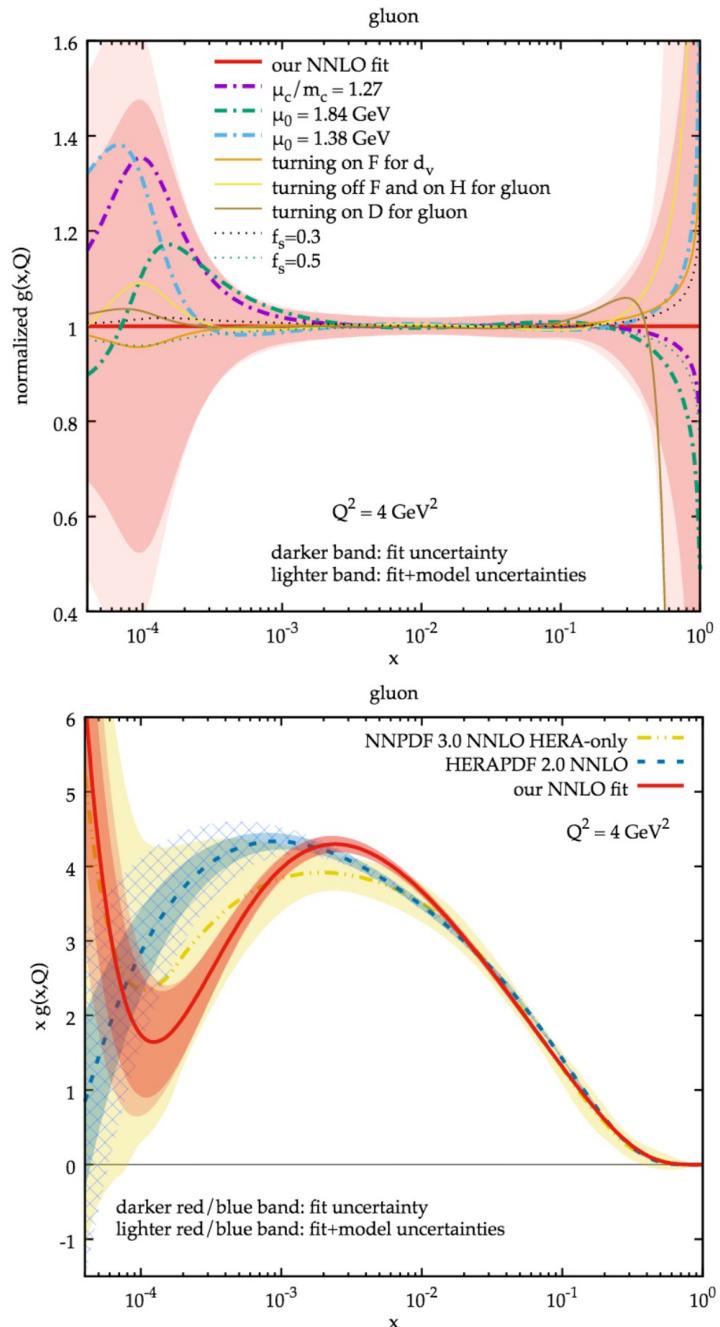
Is resummation really needed?

- χ^2 improvement of the same size as in xFitter small- x resummation paper
Eur. Phys. J. C78 (2018) 621
- Do we need resummation?
- Add resummation on top of new parameterisation using FONNL HF scheme
(and a bit different model settings)

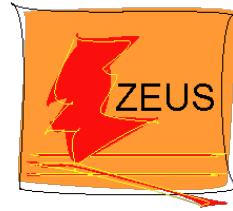
Contribution to χ^2	HELL3.0 (NLL)	New parametrization
subset NC e^+ 920 $\tilde{\chi}^2/\text{n.d.p.}$	402/377	406/377
subset NC e^+ 820 $\tilde{\chi}^2/\text{n.d.p.}$	70/70	74/70
subset NC e^+ 575 $\tilde{\chi}^2/\text{n.d.p.}$	219/254	222/254
subset NC e^+ 460 $\tilde{\chi}^2/\text{n.d.p.}$	223/204	225/204
subset NC e^- $\tilde{\chi}^2/\text{n.d.p.}$	219/159	217/159
subset CC e^+ $\tilde{\chi}^2/\text{n.d.p.}$	38/39	37/39
subset CC e^- $\tilde{\chi}^2/\text{n.d.p.}$	49/42	50/42
correlation term + log term	73 - 7	79 + 2
Total $\chi^2/\text{d.o.f.}$	1284/1127	1312/1127

Small- x resummation crucial for low- x (HERA/LHC/FCC) phenomenology

- Model uncertainties → varying model parameters within uncertainties
- Parameterisation uncertainties
 - Varying starting scale
 - Adding extra parameters
- Model uncertainties added in quadrature
- Parameterisation uncertainties - envelope taken
- Both added in quadrature to experimental uncertainty for total PDF errors



When do you start your
xFitter analysis?... :)



HERA combined DIS data are
core of every modern PDF
extraction

- 2927 data points combined to 1307
- impressive precision

