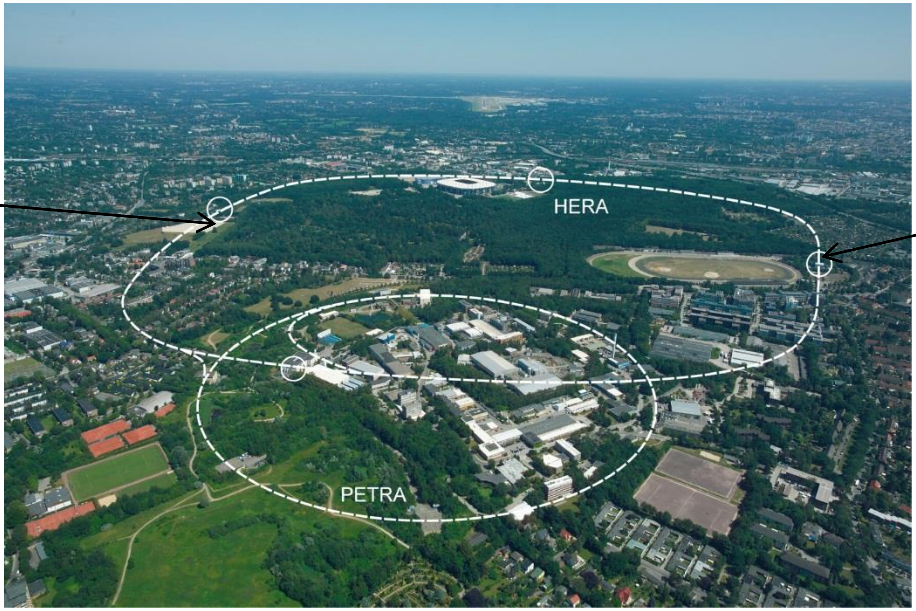


HERA PDFs and lessons for EIC?

AM Cooper-Sarkar, Oxford

There were two general purpose detectors H1 and ZEUS

At the end of running there was about 500pb^{-1} of data per experiment split ~equally between e^{+} and e^{-} beams

PDF fits to HERA combined data can be found in ArXivs:0911.0884, 1506.06042, 2112.01120,
These are all published in EPJC, but you may find the arXiV easier to find

HERA ran from 1992 to 2000 and again from 2002 to 2007

HERA-I	1992-2000	Ep=820,920 GeV
HERA-II	2003-2007	Ep=920, 460,575 GeV

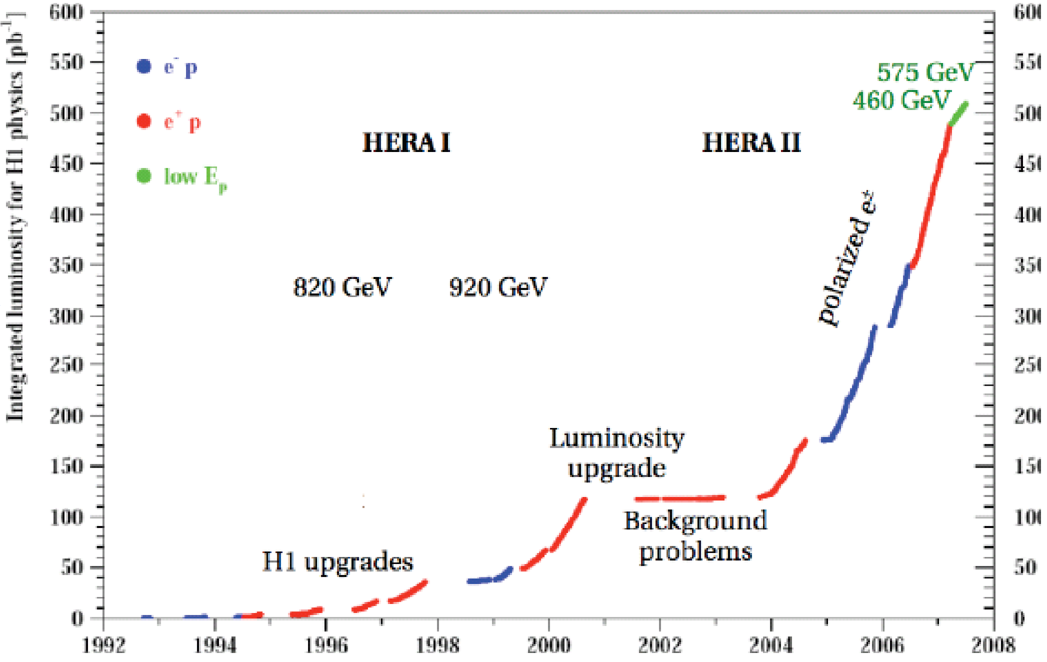
HERA- I combination ~250 pb^{-1} of data

HERA-II gives 4 times as much data in total

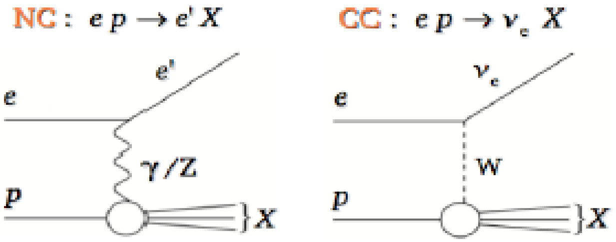
Registered ~1fb^{-1} of integrated luminosity of physics data.

Running at Ep = 920, 820, 575, 460 GeV
$\sqrt{s}$ = 320, 300, 251, 225 GeV

The lower proton beam energies allow a measurement of F_L and thus give more information on the gluon.



Deep Inelastic Scattering (DIS) is the best tool to probe proton structure



- Kinematic variables:

$Q^2 = -q^2 = -(k-k')^2$ Virtuality of the exchanged boson

$x = \frac{Q^2}{2p \cdot q}$ Bjorken scaling parameter

$y = \frac{p \cdot q}{p \cdot k}$ Inelasticity parameter

$s = (k+p)^2 = \frac{Q^2}{xy}$ Invariant c.o.m.

Neutral current:

$$\frac{d^2\sigma^{\pm}_{NC}}{dxdQ^2} = \frac{2\alpha\pi^2}{xQ^4}\left(Y_+ F_2 \mp Y_- xF_3 - y^2 F_L\right)$$

$F_2 \propto \sum_i e_i^2 (xq_i + x\bar{q}_i)$ quark distributions

$xF_3 \propto \sum_i (xq_i - x\bar{q}_i)$ valence quarks

$F_L \propto \alpha_s \times g$ gluon at NLO

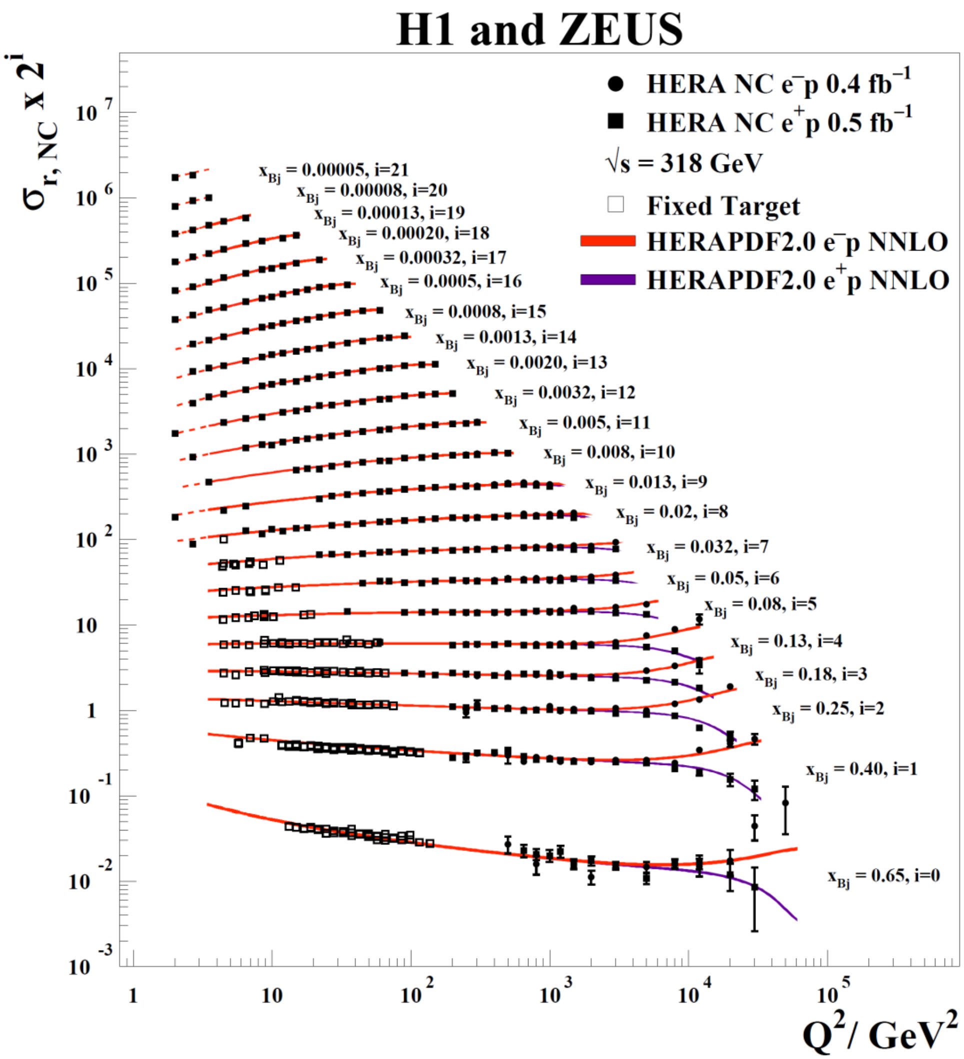
LO expressions

Charged current:

$$\frac{d^2\sigma^{-}_{CC}}{dxdQ^2} = \frac{G_F^2}{2\pi}\frac{M_W^2}{M_W^2+Q^2}\left(u+c+(1-y^2)(\bar{d}+\bar{s})\right)$$

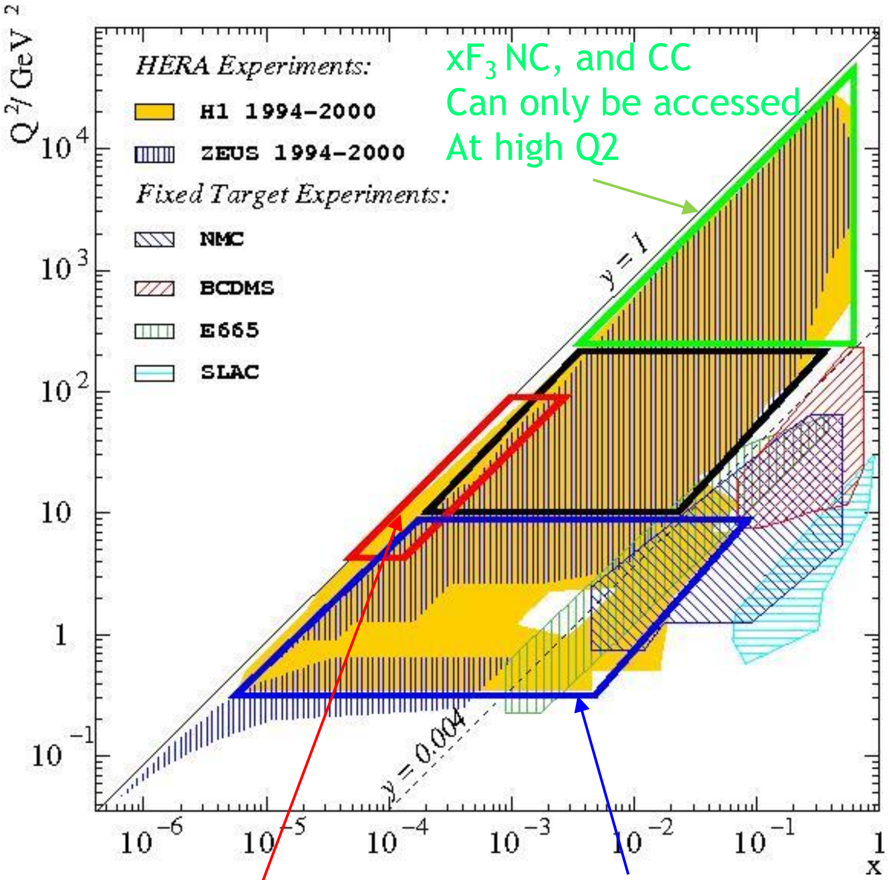
$$\frac{d^2\sigma^{+}_{CC}}{dxdQ^2} = \frac{G_F^2}{2\pi}\frac{M_W^2}{M_W^2+Q^2}\left(\bar{u}+\bar{c}+(1-y^2)(d+s)\right)$$

flavour decomposition



Gluon from the scaling violations: DGLAP equations tell us how the partons evolve

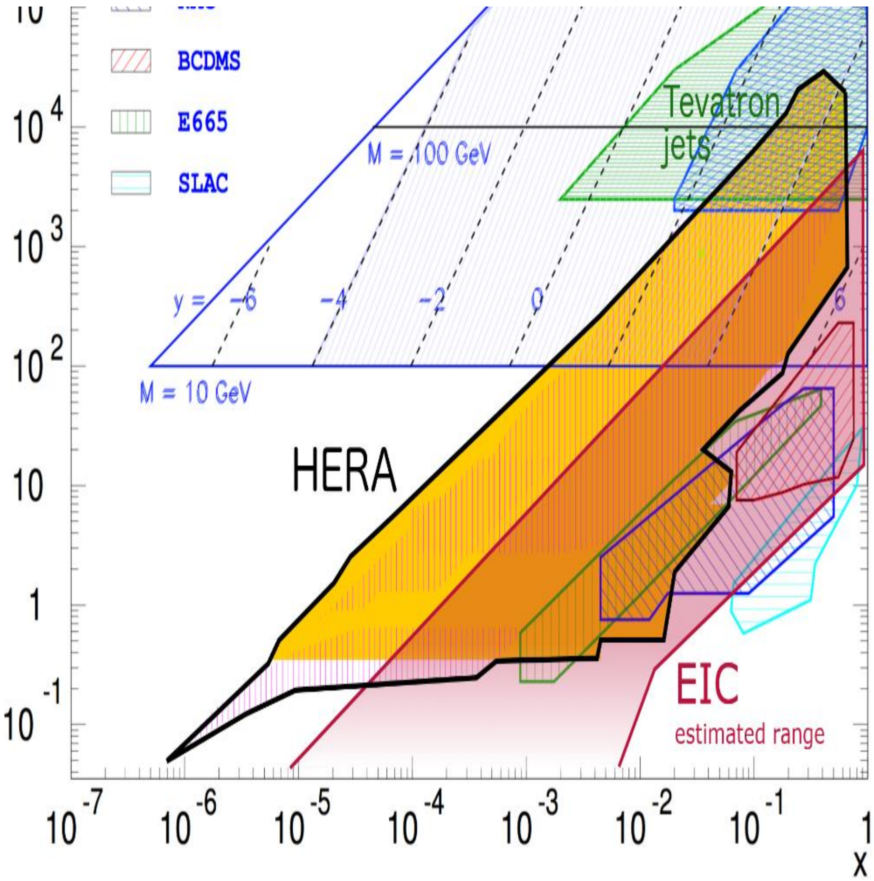
The DIS kinematic plane



xF_3 NC, and CC
Can only be accessed
At high Q2

F_L has been
Measured by
changing the
beam energies

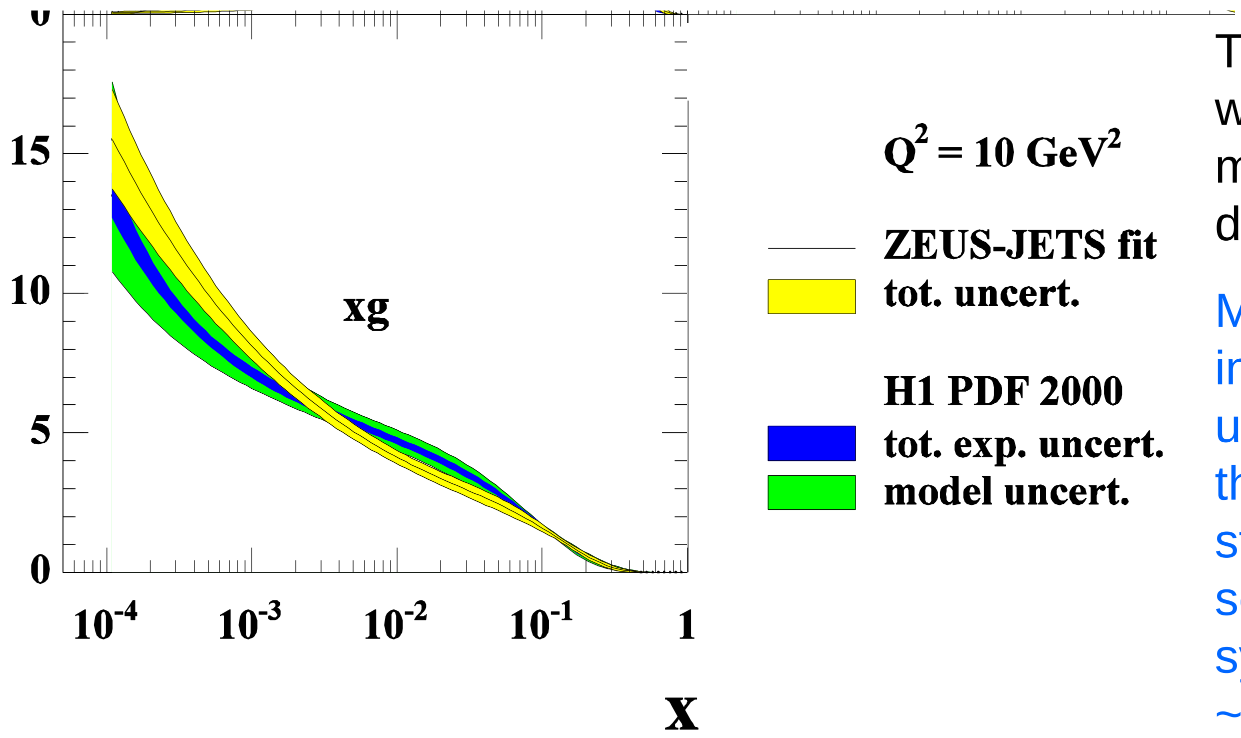
F2 is accessed in all regions but beware of low Q^2 for HERA, because it accesses low-x



So EIC will not access high Q2 so efficiently but the higher luminosity could compensate?

By contrast the low Q^2 region does not access such low-x, so that Q^2 > ~3 GeV_2 should not access low-x effects for a proton target

Both ZEUS and H1 made their own PDF fits starting with H1PDF2000, ZEUS-S in 2001 and ZEUS-JETS in 2005. The results were not always fully consistent.



There are many assumptions which go into a PDF fit. The most important of these is data consistency.

Most experimental data used in PDF fits today have uncertainties dominated by their systematics not by the statistics. For H1 and ZEUS separate data sets systematic uncertainties were ~3 times statistical error

The situation with large systematic uncertainties and tensions between data sets only gets worse when many LHC data sets are added to the HERA data…
But that is not a story for today

This motivated the H1 and ZEUS collaborations to combine their data, investigating their correlated systematic uncertainties in great detail. Some of these are also correlated between ZEUS and H1 and some are not.

The resulting combination is much better than expected just from the increased statistics of combining two experiments.

The expriments cross calibrate each other

The combination of the HERA data yields a very accurate and consistent data set for 4 different processes: e+p and e-p Neutral and Charged Current reactions.

The post-combination systematic errors are smaller than the statistical across a large part of the kinematic plane

Combination procedure

- Swim all points to a common x-Q^2 grid
- Calculate comnibed values and uncertainties as follows.

This is done by making a χ2 fit to the data points of both experiments which simply assumes that for each process (NC or CC, e+ or e-) and each x, Q^2 point (i) there is only one ‘true’ value of the cross-section- these are the predictions m_i _ whereas there can be several measurements of this value, from ZEUS and H1 and from different years of running- these are the measurements μ_i

$$\chi^2_{\exp}(\boldsymbol{m},\boldsymbol{b}) = \sum_i \frac{\left[m^i - \sum_j \Gamma^i_j b_j - \mu^i\right]^2}{\Delta_i^2} + \sum_j b_j^2.$$

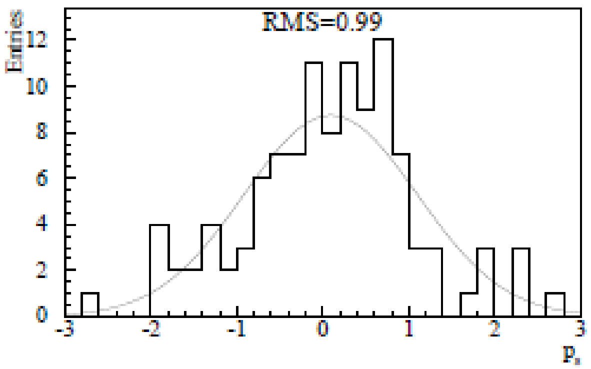
- The chisq accounts for the correlated systematic uncertainties of the data points- each data point can have several such uncertainties Γ_j, hence sum over j for each data point i, but these uncertainties are correlated between all data points for large sub-sets of data. The fit determines the value of the cross-sections m_i and the nuisance parameters b_j . It also evaluates their uncertainties.
- Evaluate further uncertainties due to choices in combination procedure,e.g. Correlations between ZEUS and H1

HERAPDF1.0 averaged 1402 data points to 741 combined data points
χ2/ndf =637/656 for this HERA-I combination

HERAPDF2.0 averaged 2927 data points to 1307 combined data points
χ2/ndf =1687/1620 for this HERA-II combination

Some examples from HERA-I (just because we gave more detail for HERA-I)

arXiv:0911/0884

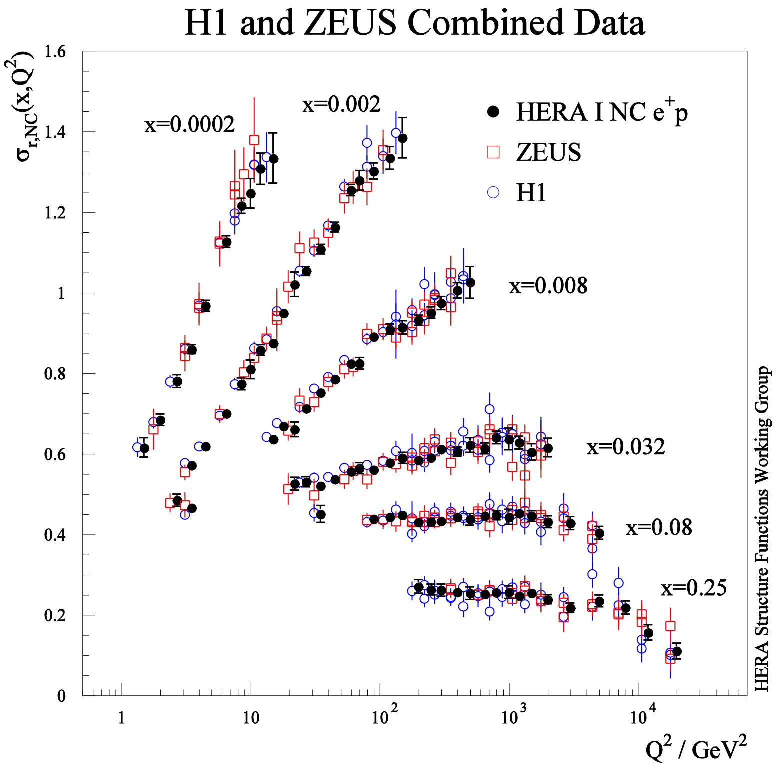


Distributions of the nuisance parameters b, most systematic shifts < 1 std deviation

The fit also determines uncertainties on the shift parameters Δb, some of these are much reduced e.g

ZEUS γp background uncertainty is reduced by a factor of 3

H1 LAr hadron calorimeter energy scale uncertainty is halved



Hence this is not just statistical improvement from combination . Each experiment has been used to calibrate the other since they have rather different sources of experimental systematics

Before combination the systematic errors are ~3 times the statistical for $Q^2 < 100$

After combination systematic errors are < statistical

Correlated systematic uncertainties, χ2 and Δχ2

The data combination results in a data set which not only has improved statistical uncertainty, but also improved systematic uncertainty.

Even though there are >100 sources of correlated systematic uncertainty on the data points these uncertainties are small. The total systematic uncertainty is significantly smaller than the statistical uncertainty across the kinematic region used in the QCD fits

This very consistent HERA data set was used as sole input to a Parton Distribution Function fit known as HERAPDF1.0

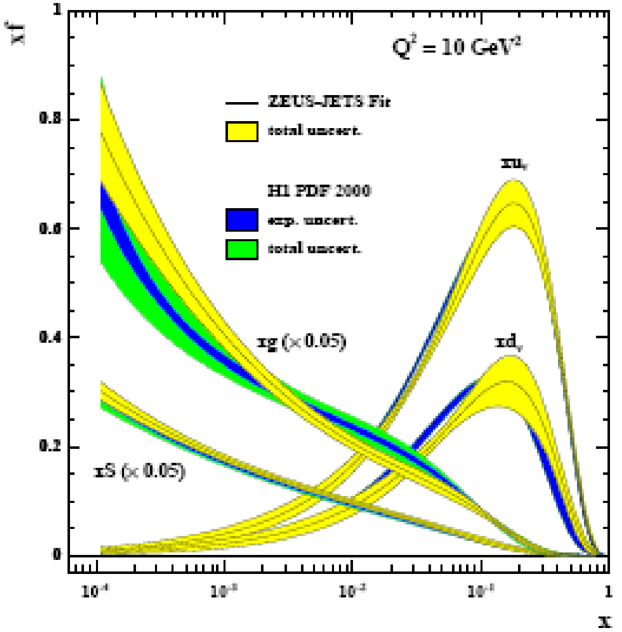
We set the experimental uncertainties on our PDFs at 68% CL by the conventional χ2 tolerance

$$\Delta\chi 2 = 1$$

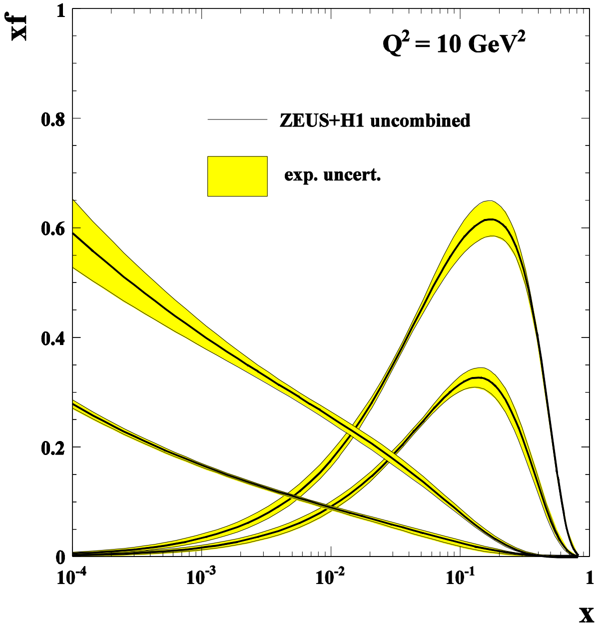
When there are data inconsistencies, larger tolerances are considered—

as in the PDF fits of CT and MSHT

variation included in model error, ZEUS does not.

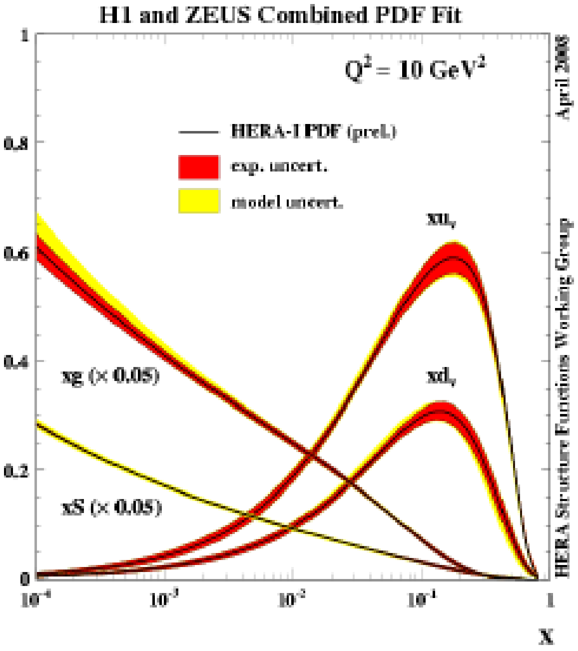


PDFs from separate QCD analyses of separate ZEUS and H1 data



PDFs from same QCD analysis of separate ZEUS and H1 data sets - before 'smart' combination

Experimental error only

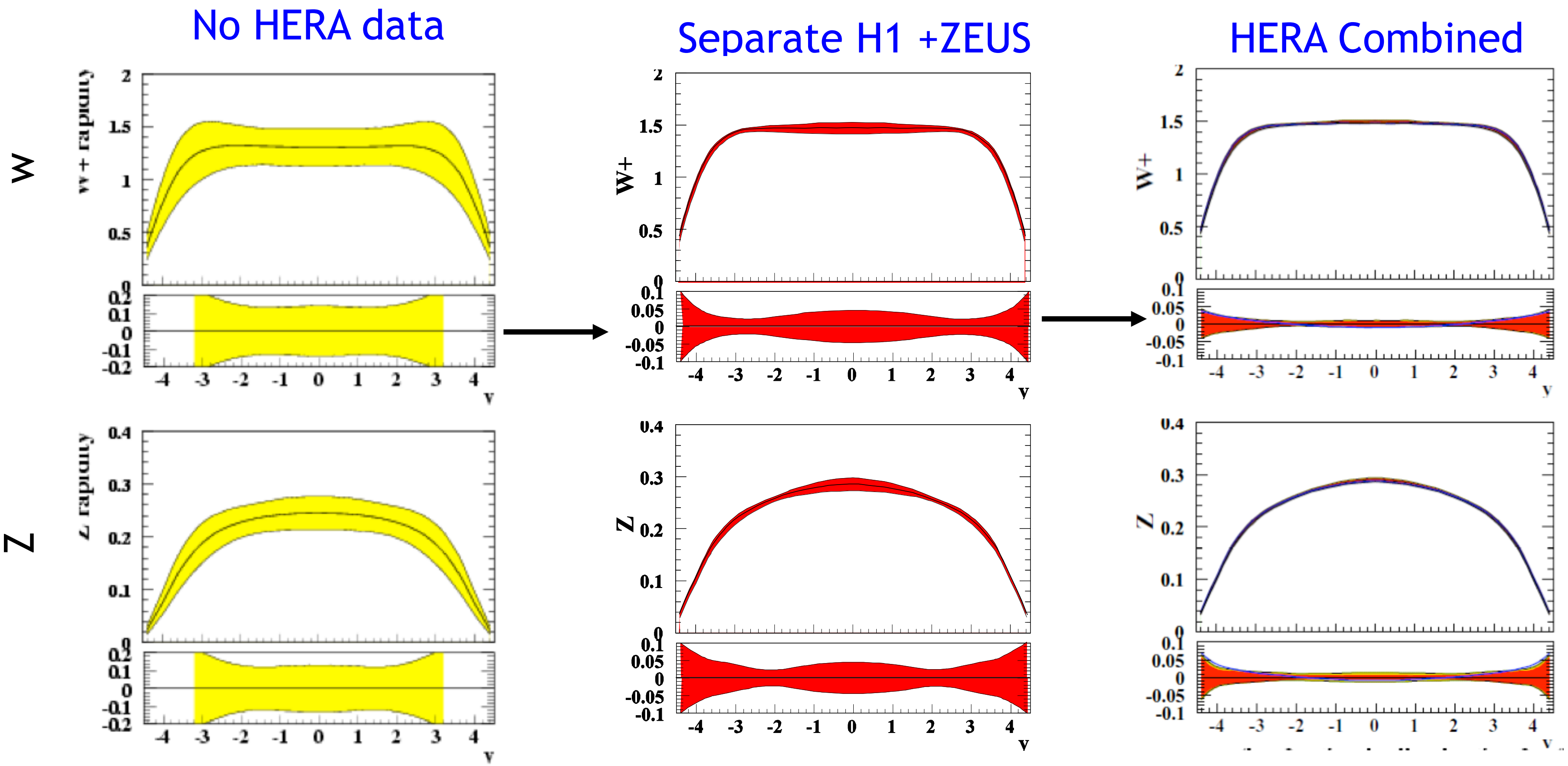


PDFs from same QCD analysis of combined HERA data - after 'smart' combination

HERAPDF0.1 has small experimental errors and modest model errors

Impact of HERA data on the LHC

W and Z rapidity distributions as predicted by PDFs before and after HERA



No HERA data

Separate H1 +ZEUS

HERA Combined

W

Z

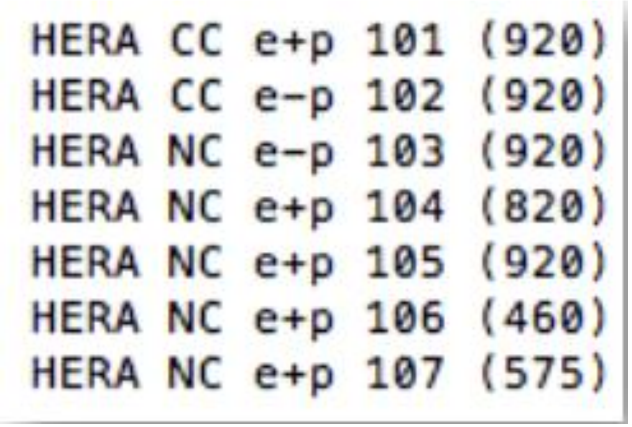
WHY? -It's due to the improvement in the low-x gluon

At the LHC the q-qbar which make the boson are mostly sea-sea partons at low-x

And at high scale, $Q^2 \sim M_Z^2$, the sea is driven by the gluon by g→q qbar splitting

Now let’s move to the HERA-II +I combination arxiv:1506.06042

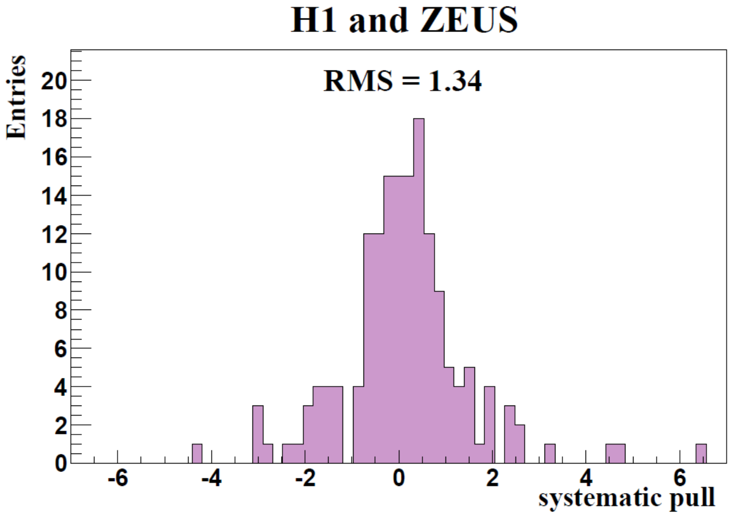
The HERA-I+II combination is more ambitious
41 input data files to 7 output files with 169 sources of correlated uncertainty

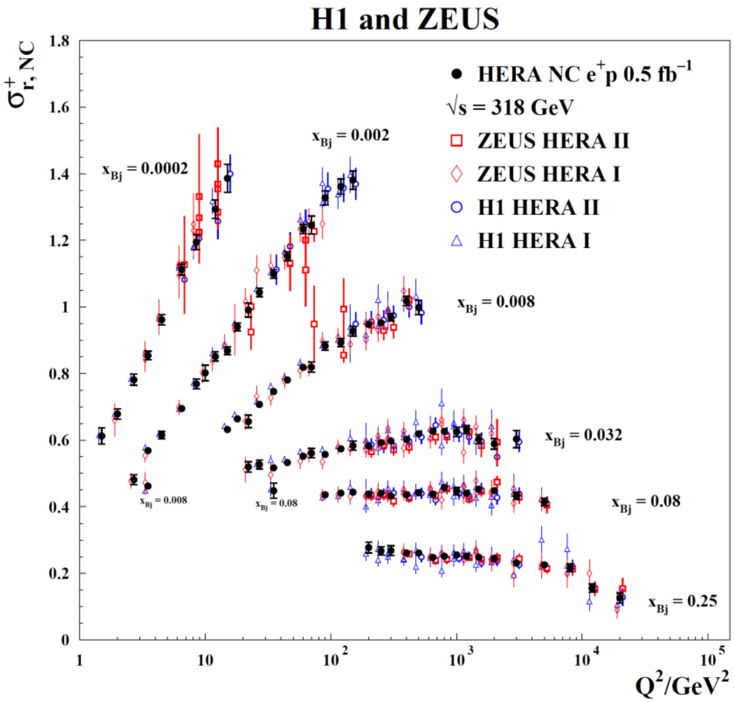


```
HERA CC e+p 101 (920)
HERA CC e-p 102 (920)
HERA NC e-p 103 (920)
HERA NC e+p 104 (820)
HERA NC e+p 105 (920)
HERA NC e+p 106 (460)
HERA NC e+p 107 (575)
```

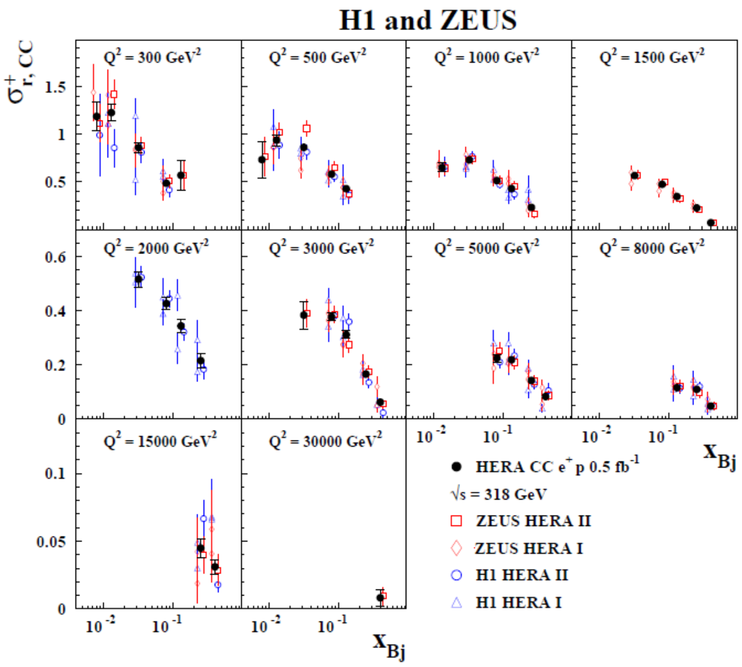
$0.045 < Q^2 < 50000\ \mathrm{GeV}^2$ $6.\ 10^{-7} < x_{Bj} < 0.65$

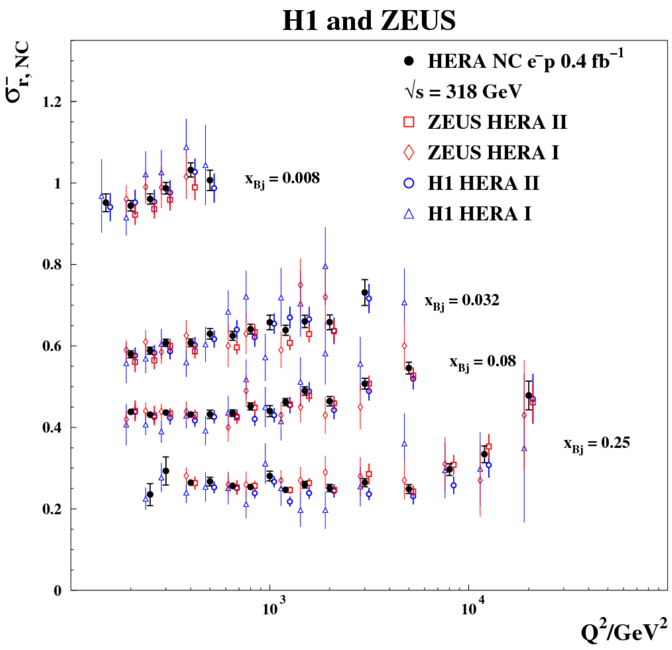
Uncertainties are less than 1.5% for $3 < Q^2 < 500\ \mathrm{GeV}^2$ and below 3% up to $Q^2 \sim 3000\ \mathrm{GeV}^2$





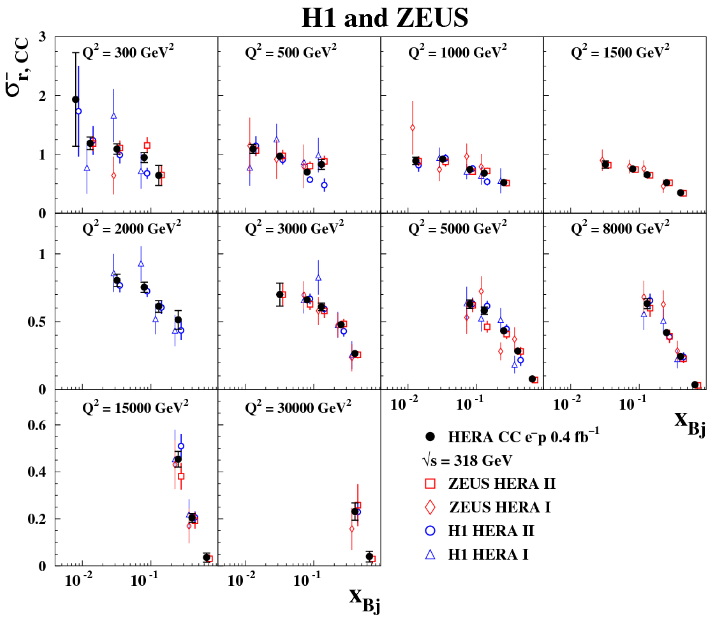
NC and CC e$^+$ vs H1 and ZEUS inputs

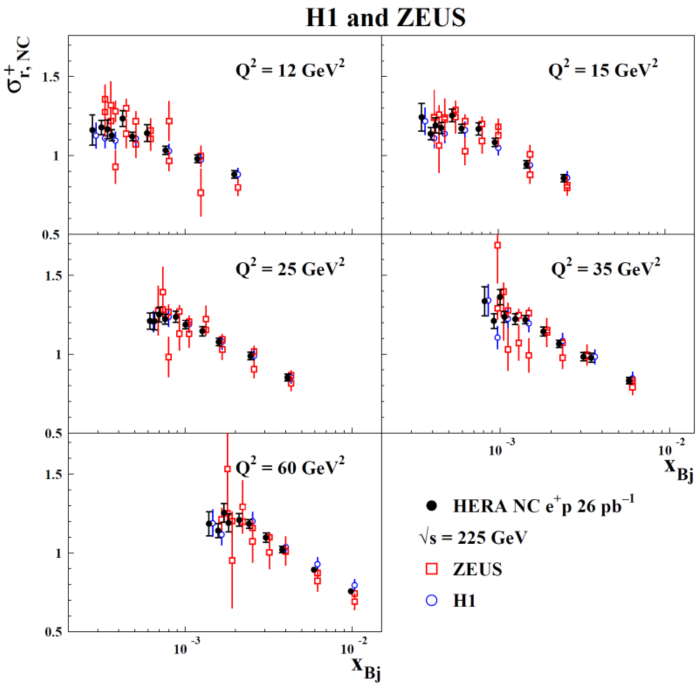




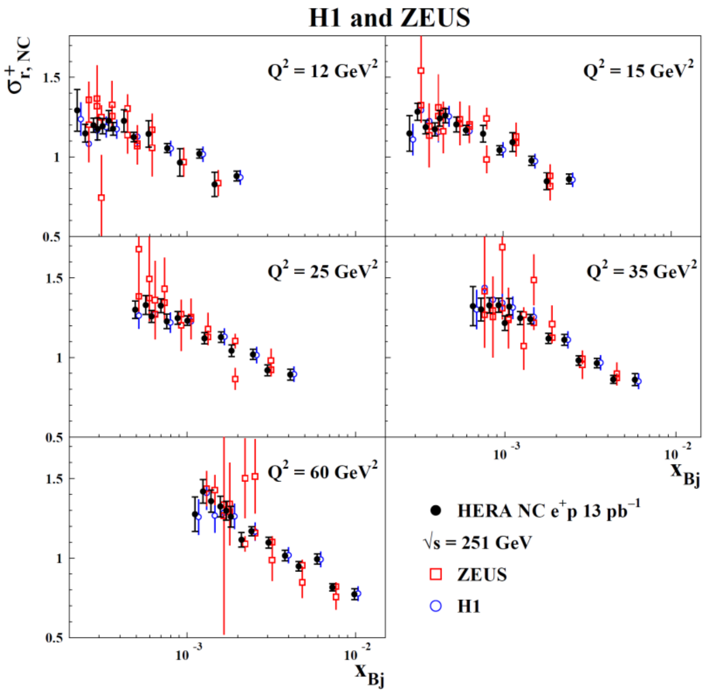
NC and CC e$^-$ vs H1 and ZEUS inputs

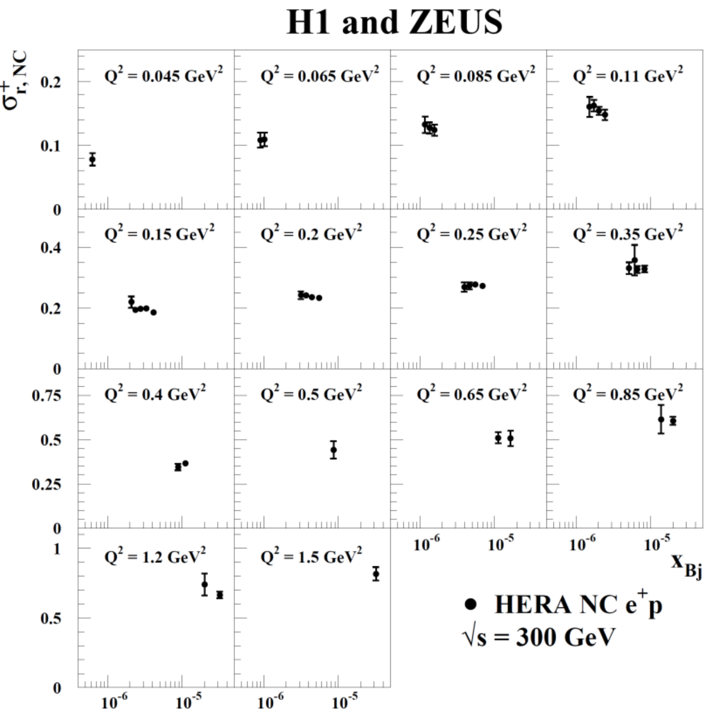
10 fold increase in e$^-$ statistics compared to old HERA-1 combination



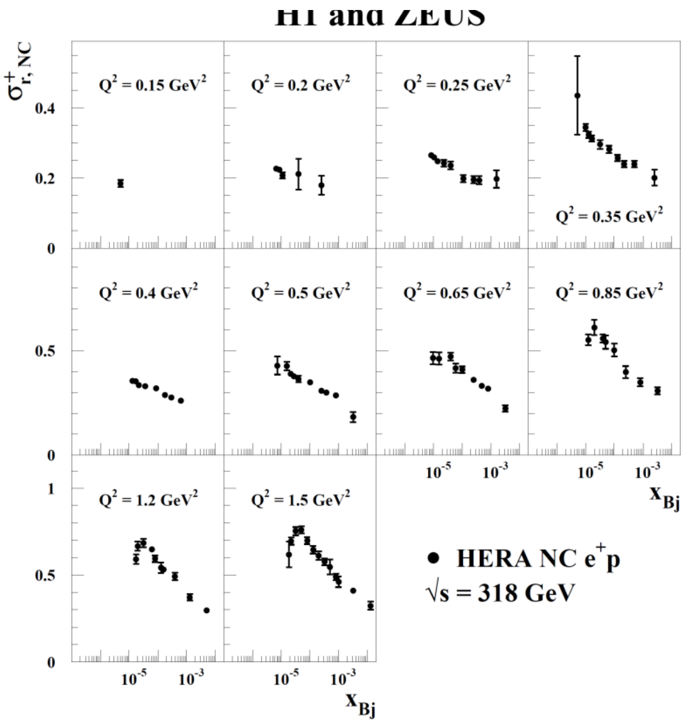


New for this combination is the data at different beam energies





And let's not forget that there is data at very low Q^2



The HERAPDF approach uses only HERA data

The combination of the HERA data yields a very accurate and consistent data set for 4 different processes: e^+p and e^-p Neutral and Charged Current reactions and for e^+p Neutral Current at 4 different beam energies
The use of the single consistent data set allows the usage of the conventional χ2 tolerance Δχ2 = 1 when setting 68%CL experimental errors

NOTE the use of a pure proton target means no need for heavy target/deuterium corrections.

d-valence is extracted from CC e^+p without assuming d in proton= u in neutron

All data are at high W (> 15 GeV), so high-x, higher twist effects are negligible.

These are the only PDFs for which this is true

HERAPDF evaluates model uncertainties and parametrisation uncertainties in addition to experimental uncertainties

- HERAPDF1.0 was based on the combination of HERA-I data (LO, NLO)
- HERAPDF1.5 included preliminary HERA-II data (LO,NLO,NNLO)
- HERAPDF2.0 is based on the final combination of HERA-I and HERA-II data which supersedes the HERA-I combination and supersedes all previous HERAPDFs (LO,NLO and NNLO)
  HERAPDF2.0 also has fits including combined HERA charm and beauty data and including H1 and ZEUS separate jet data (arXiv:2112.01120 for NNLO)

Theoretical framework

Fits are made in the DGLAP formalism –using the xFitter framework with QCDNUM

The Thorne-Roberts optimised massive variable-flavour number scheme is used

The staring scale Q^2_0 (= 1.9 GeV^2) is below the charm mass^2 (mc=1,4 GeV) and charm and beauty (mb=4.75) are generated dynamically

A minimum Q^2 cut $Q^2 > 3.5\ GeV^2$ is applied to stay within the supposed region of validity of leading twist pQCD (no data are at low W^2)

The choices of values in green are varied and the results added as model uncertainties

Parametrisation

We chose to fit the PDFs for:

gluon, u-valence, d-valence and the Sea u and d-type flavours:

Ubar = ubar, Dbar = dbar+sbar (below the charm threshold)

To the functional form $xf(x,Q_0^2) = Ax^B(1-x)^C(1+Dx+Ex^2)$

The normalisations of the gluon and valence PDFs are fixed by the momentum and number sum-rules resp.

Parameters D and E are added until the χ2 no longer improves, ‘saturation of χ2’ but further D,E parameters which change shape not χ2 are used as parametrisation variations

For the NLO and NNLO fits the resulting central parametrisation at $Q^2_0 = 1.9\ \mathrm{GeV}^2$ is

$$\begin{aligned}
xg(x) &= A_g x^{B_g}(1-x)^{C_g} - A'_g x^{B'_g}(1-x)^{C'_g}, \\
xu_v(x) &= A_{u_v} x^{B_{u_v}}(1-x)^{C_{u_v}}\left(1 + E_{u_v}x^2\right), \\
xd_v(x) &= A_{d_v} x^{B_{d_v}}(1-x)^{C_{d_v}}, \\
x\bar{U}(x) &= A_{\bar{U}} x^{B_{\bar{U}}}(1-x)^{C_{\bar{U}}}(1 + D_{\bar{U}}x), \\
x\bar{D}(x) &= A_{\bar{D}} x^{B_{\bar{D}}}(1-x)^{C_{\bar{D}}}.
\end{aligned}$$

QCD sum-rules constrain A_g, A_{uv}, A_{dv}
$x\bar{s} = f_s x\bar{D}$ sets the size of the strange PDF and the constraints $B_{\bar{U}} = B_{\bar{D}}$ and $A_{\bar{U}} = A_{\bar{D}}(1 - f_s)$ ensure $x\bar{u} \to x\bar{d}$ as $x \to 0$.

- There are 14 free parameters in the central fit determined by saturation of the χ2
- $\alpha_S(M_Z) = 0.118$ for central fits
- PDFs are evolved using the DGLAP equations and convoluted with coefficient functions to evaluate structure functions and hence measurable cross sections
- An LO fit with $\alpha_S(M_Z) = 0.130$ is also provided with an alternative gluon (AG) parametrisation without the negative term
- The form of the χ2 accounts for 169 correlated uncertainties, 162 from the input data sets and 7 from the procedure of combination

$$\chi^2_{\mathrm{exp}}(\boldsymbol{m}, \boldsymbol{s}) = \sum_i \frac{\left[m^i - \sum_j \gamma^i_j m^i s_j - \mu^i\right]^2}{\delta^2_{i,\mathrm{stat}}\,\mu^i m^i + \delta^2_{i,\mathrm{uncor}}\,(m^i)^2} + \sum_j s^2_j + \sum_i \ln \frac{\delta^2_{i,\mathrm{stat}}\mu^i m^i + (\delta_{i,\mathrm{uncor}} m^i)^2}{(\delta^2_{i,\mathrm{stat}} + \delta^2_{i,\mathrm{uncor}})(\mu^i)^2}$$

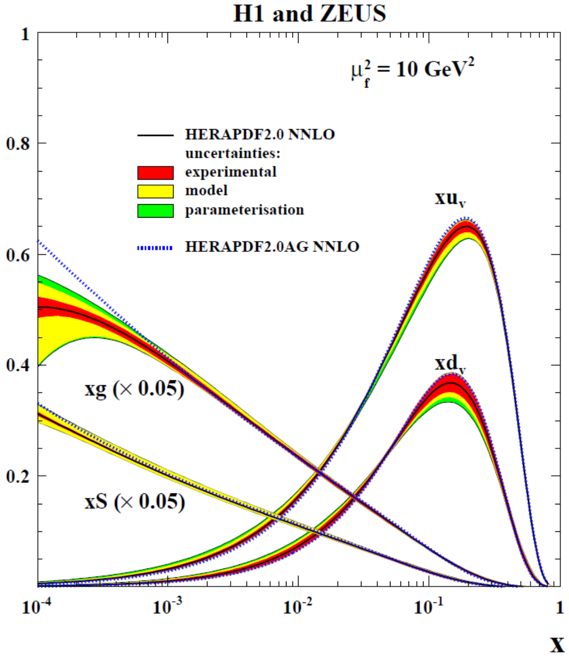
m_i is the theoretical prediction
μ_i is the measured cross section

δ_{i} stat, $\delta_{i,\mathrm{unc}}$ statistical and uncorrelated systematic uncertainty
γ^i_j correlated systematic uncertainties
b_j shifts

<u>Experimental</u>
Hessian uncertainties: 14 eigenvector pairs, evaluated with Δχ2 = 1
Cross checked uncertainties evaluated from the r.m.s. of MC replicas

<u>Model</u>: Variation of input assumptions
Variation of charm mass and beauty mass parameters is restricted using HERA charm and beauty data

Variation	central	Upper	lower
f_s size and shape	0.4	0.5	0.3
M_c (NLO) GeV	1.43	1.49	1.37
M_c (NNLO) GeV	1.47	1.53	1.41
M_b GeV	4.5	4.25	4.75
Q^2_{min} GeV^2	3.5	2.5	5.0
Q^2_{min}(HiQ2)	10.0	7.5	12.5

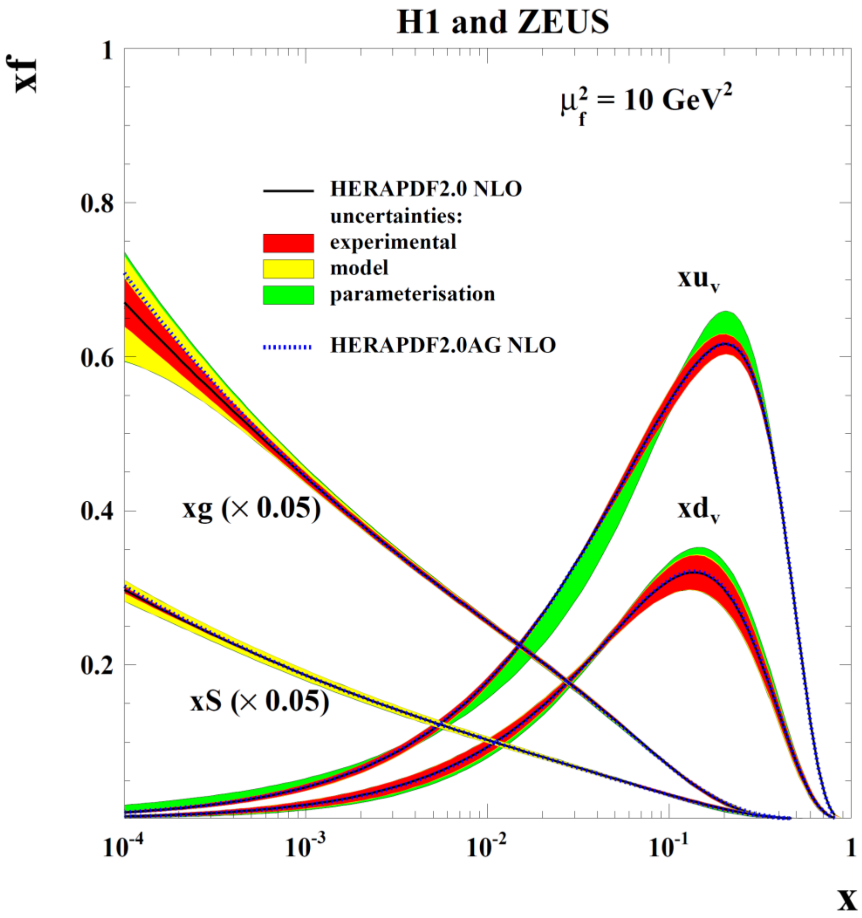


<u>Parametrisation</u>
Variation of Q^2_0 = 1.9 ± 0.3 GeV^2 and addition of 15th parameters

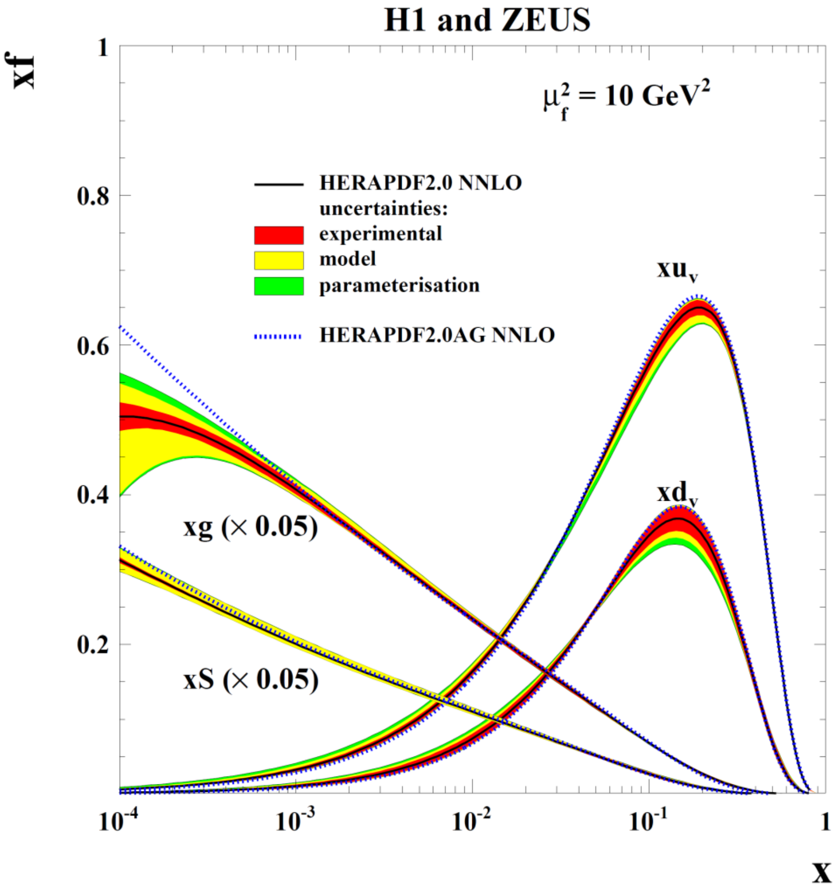
The value of $\alpha_S(M_Z)$ is not treated as an uncertainty. The central value is $\alpha_S(M_Z) = 0.118$
But PDFs are supplied for $\alpha_S(M_Z)$ values from 0.110 to 0.130 in steps of 0.001

HERAPDF2.0: NLO and NNLO fits

NLO

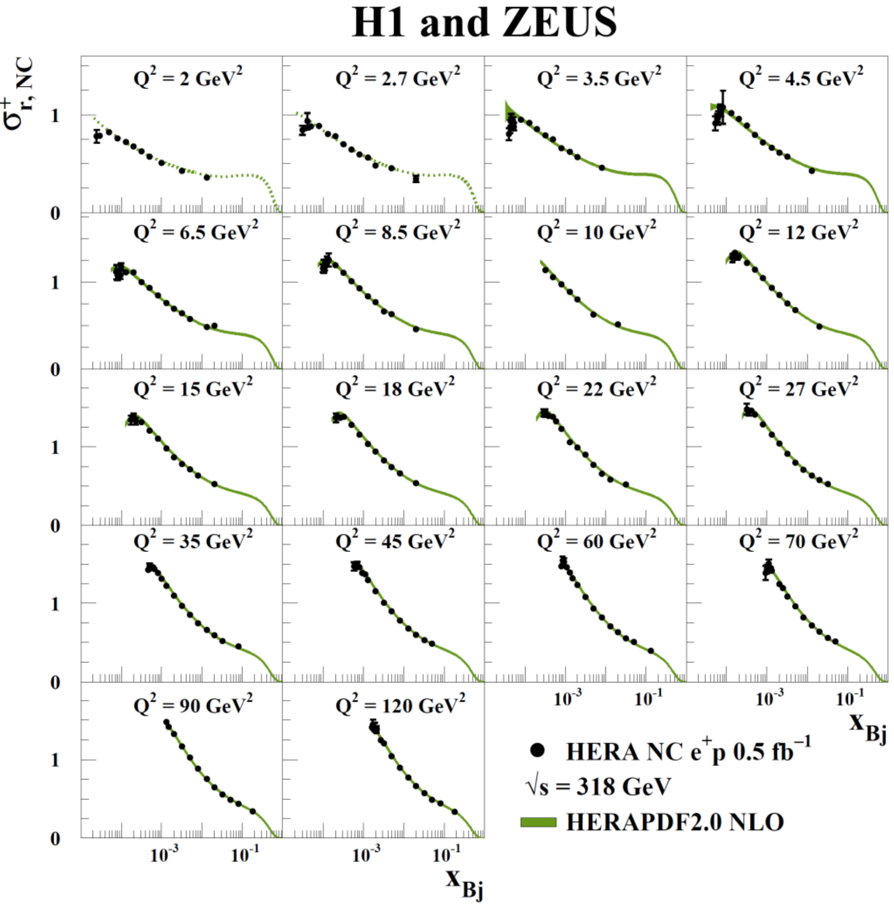


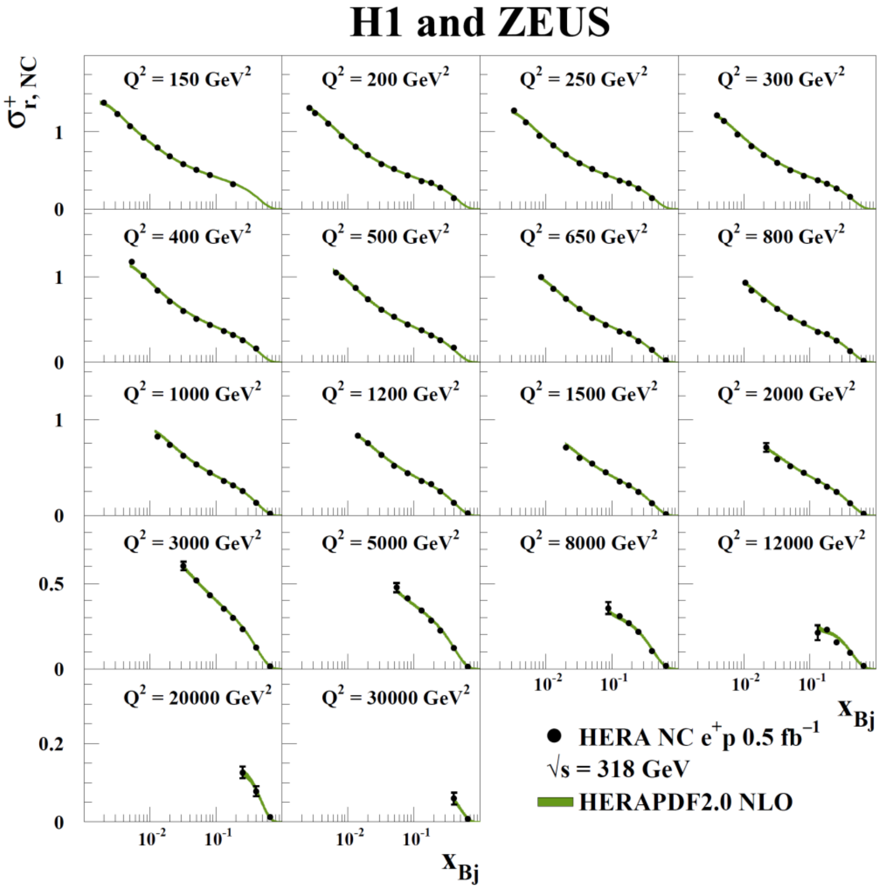
NNLO



The HERAPDF2.0AG is an alternative gluon parametrisation which is positive definite for all x and all $Q^2 > Q^2_0$

HERAPDF2.0 compared to data



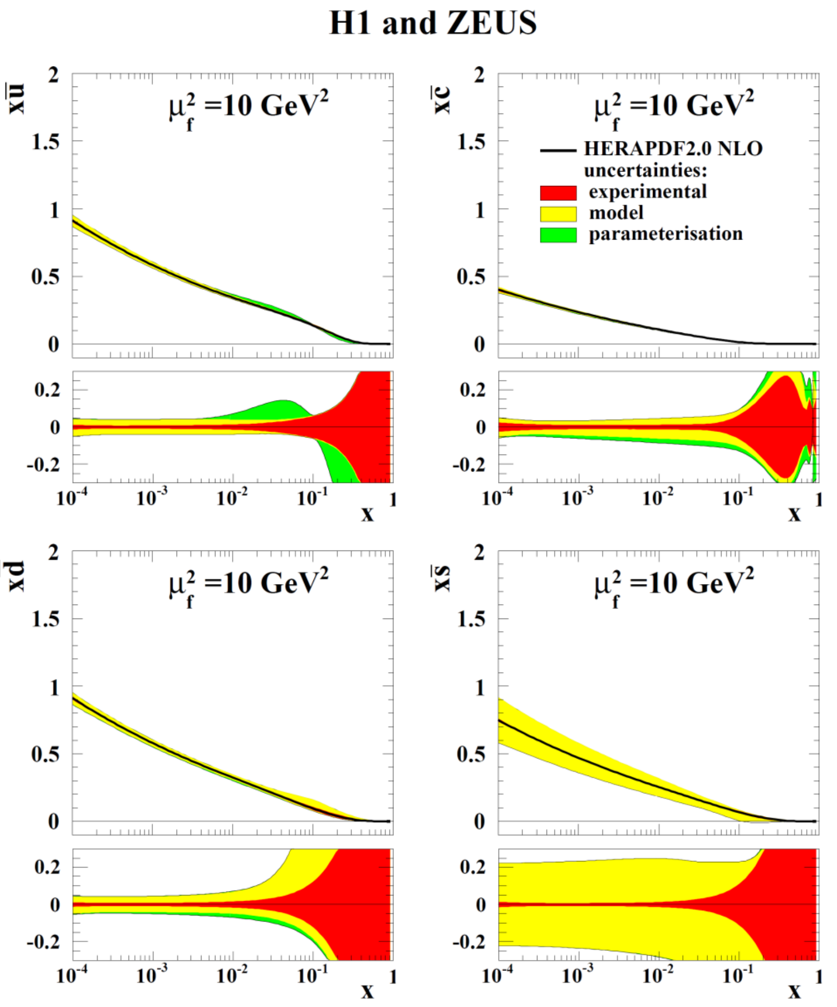


Here is the comparison to the NC e$^+$ data for $2 < Q^2 < 30000$ GeV^2

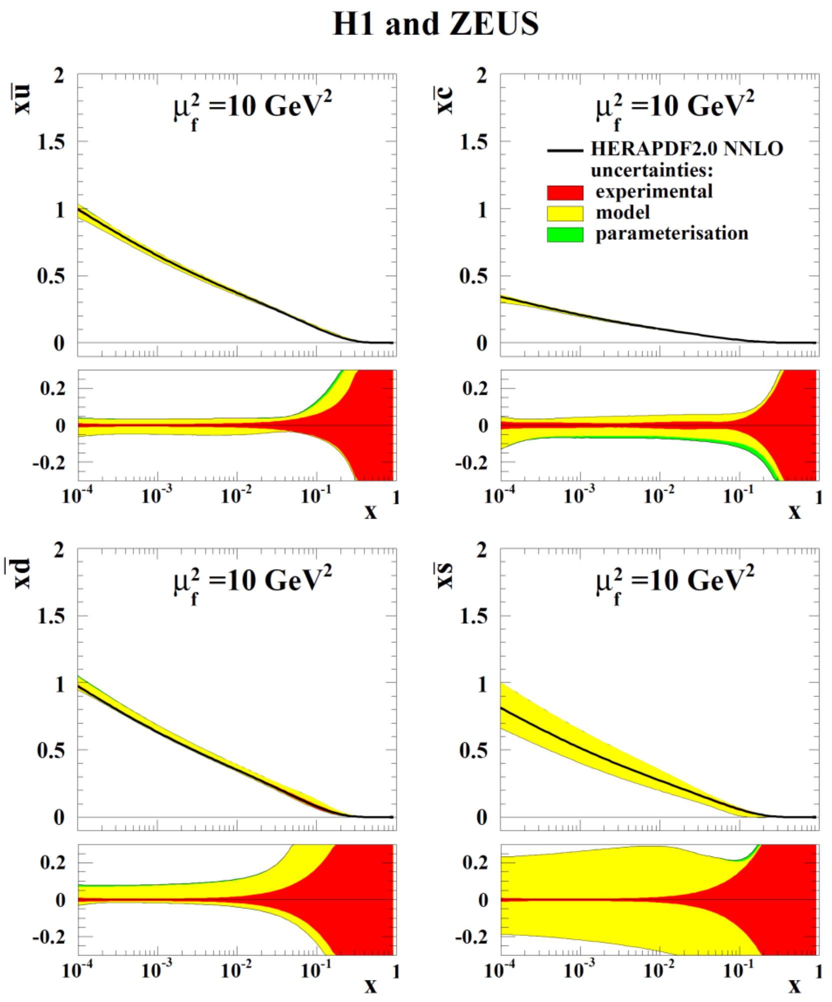
NLO and NNLO fits look very similar

HERAPDF2.0: NLO and NNLO fits

NLO

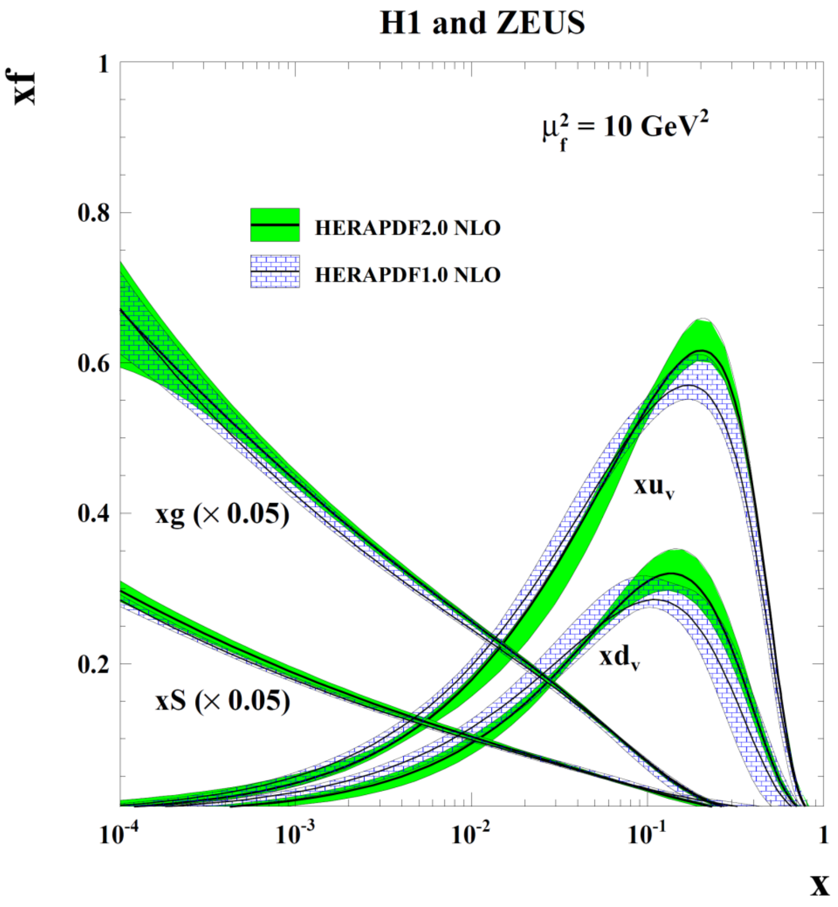


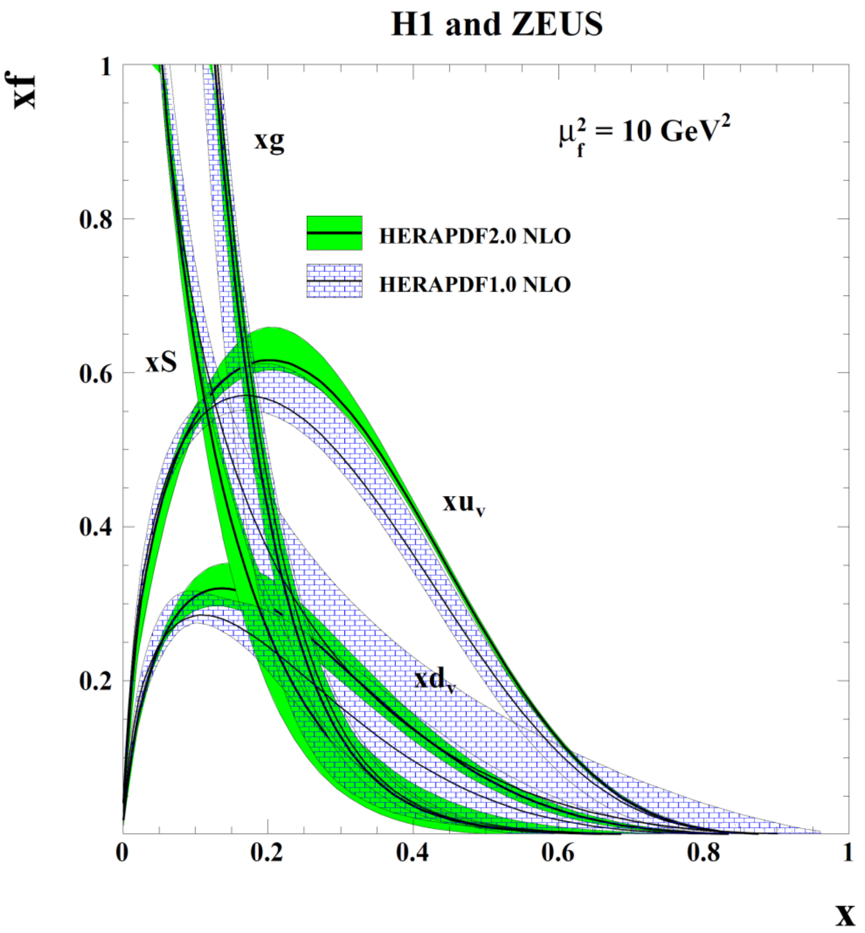
NNLO



Flavour break-up of the sea

Compare HERAPDF2.0 to HERAPDF1.0 at NLO

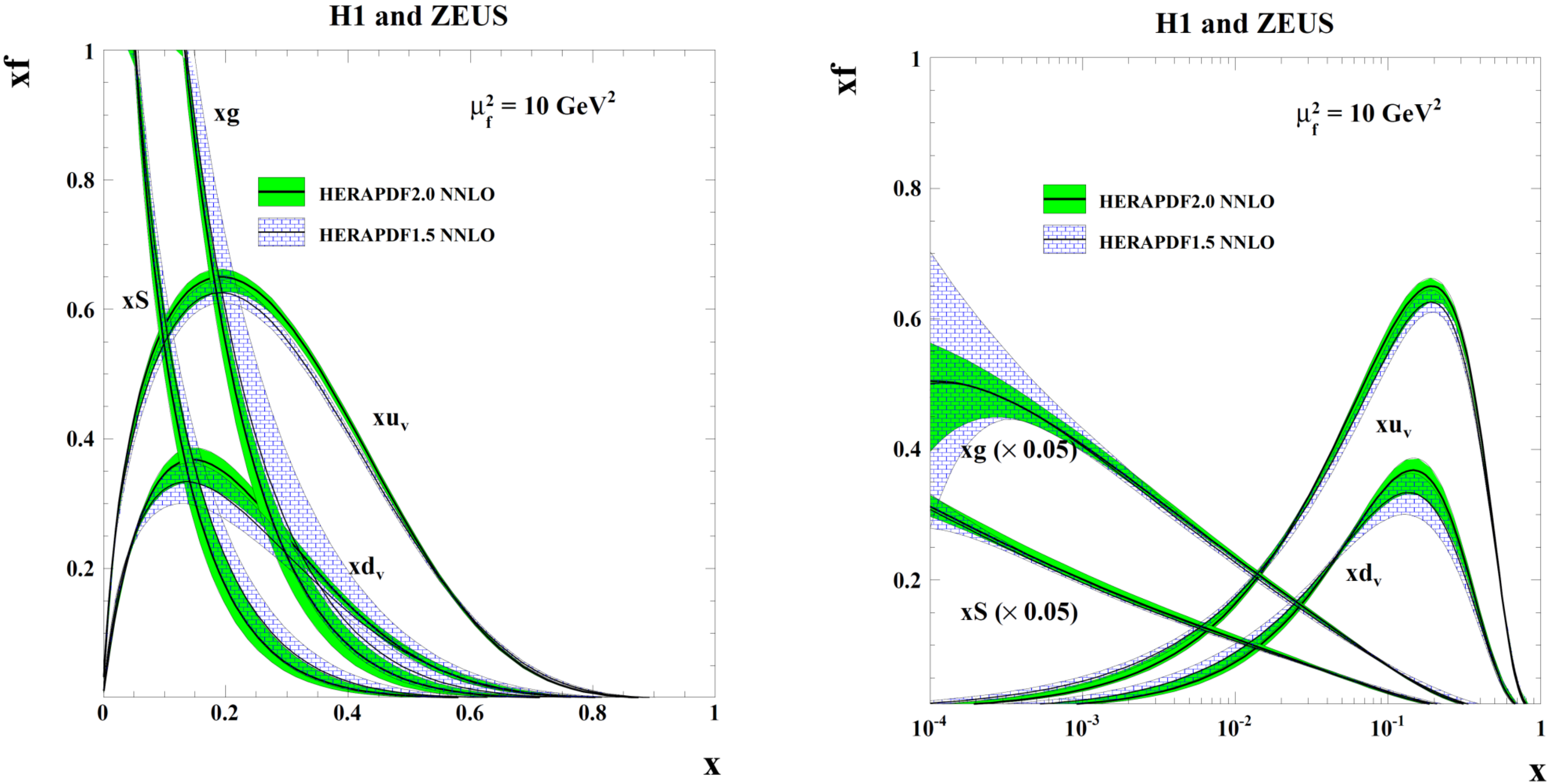




Much more high-x data
Substantial reductions in high-x uncertainty
Some change in valence shape

- HERAPDF1.0 (and 1.5) had rather hard high-x sea, harder than the gluon (within large uncertainties).This is no longer the case and uncertainties are much reduced
- HERAPDF1.0 and 1.5 had a soft high-x gluon this moves to the top of its previous error band- but is still soft (at NLO)

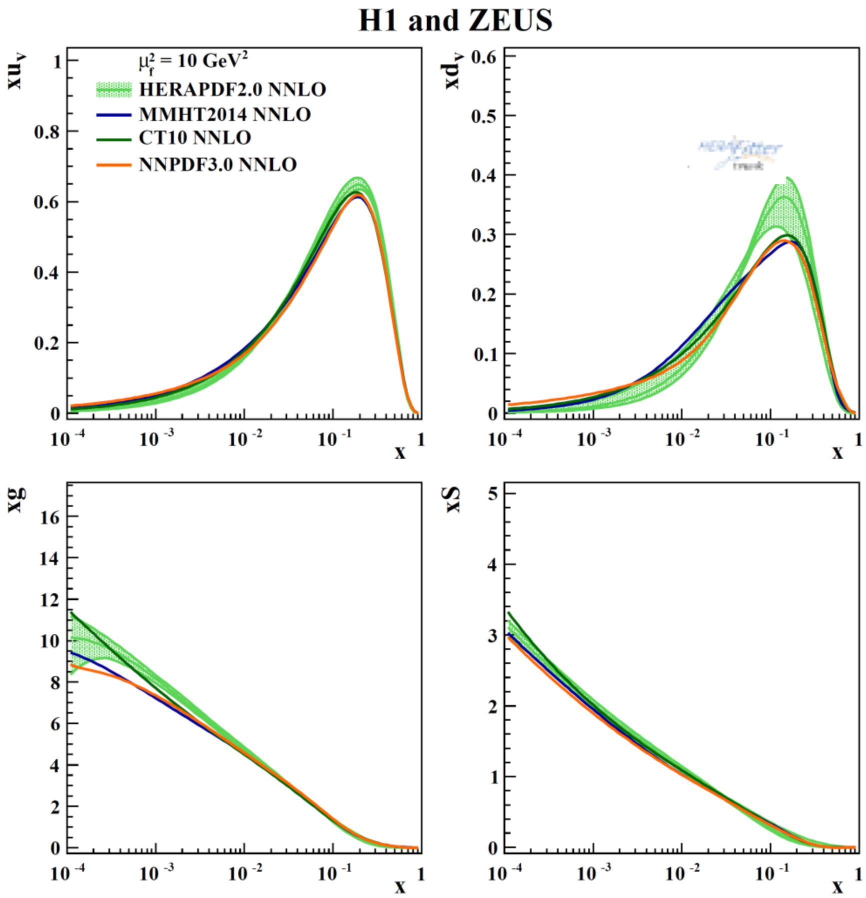
Compare HERAPDF2.0 to HERAPDF1.5 at NNLO

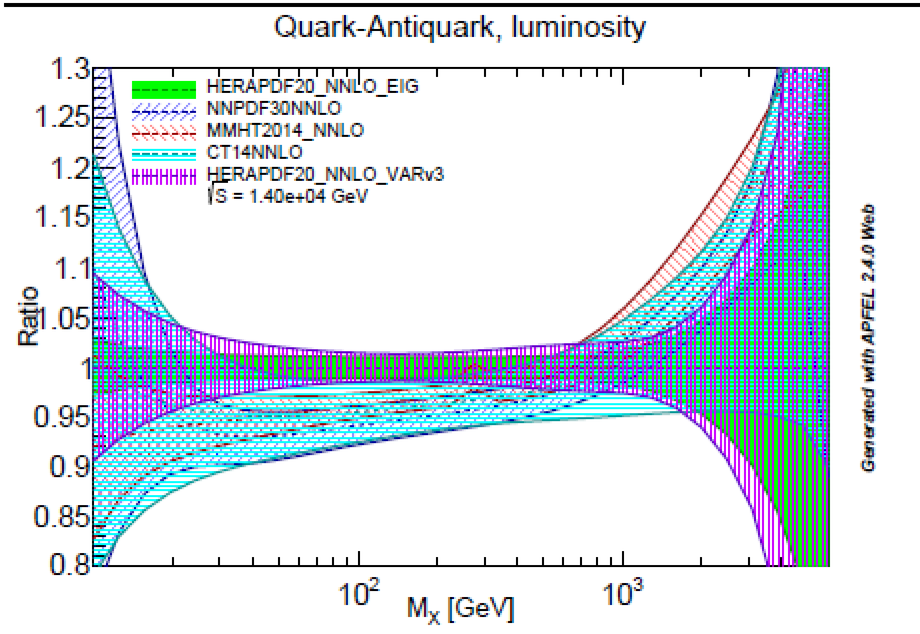


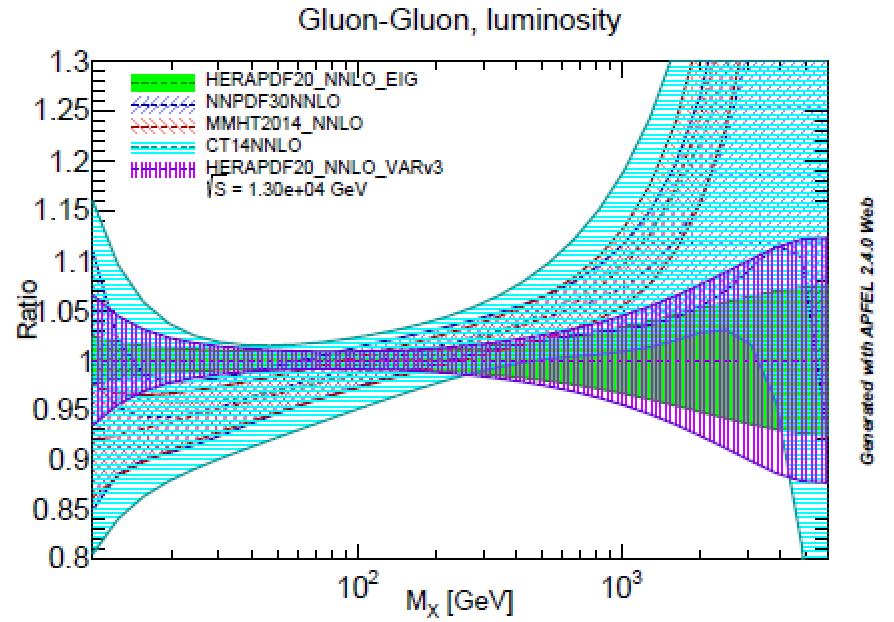
Reduction in gluon uncertainty both at low-x and high-x.
A lot of this reduction is because the model variation due to variation of Q^2 cut is not as dramatic now that we have more data.

The HERAPDF1.5 gluon was not soft compared to global PDFs. However it had a large error band.
This uncertainty on the gluon decreases and the central value moves to the lower end of its previous error band

Compare HERAPDF2.0 to other PDFs at NNLO

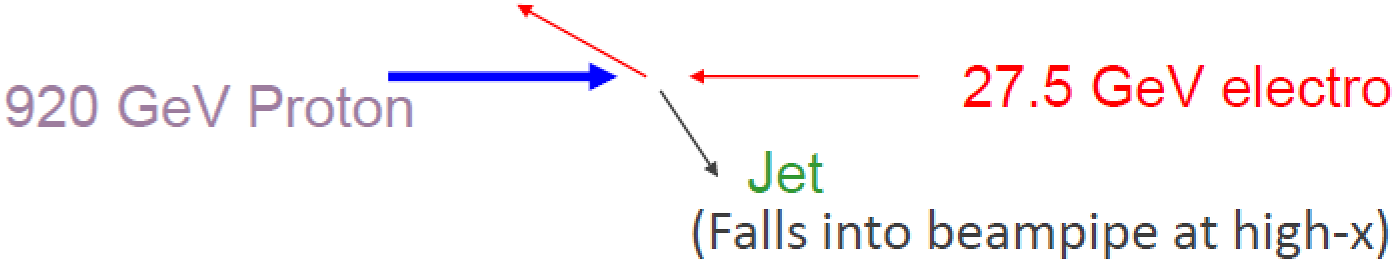






High-x valence shapes somewhat different – new high- x data and use of proton target only. Gluon and Sea are both broadly compatible with other PDFs
Comparison of q-qbar and glu-glu luminosity at 13 TeV show consequences for LHC

You will note that we do not know PDFs well at high-x. One reason that the HERA kinematic region did not allow us to measure well at high-x is that jets fall into the bea



920 GeV Proton

27.5 GeV electron

Jet

(Falls into beampipe at high-x)

Can EIC do better than HERA at high-x?

There are several advantages:

- Much higher luminosity (2 to 3 orders of magnitude)
- Run deuterons (measure neutrons)—get d_valence
- Access to lower angle jets (large crossing angle for the beams)
- Better flavor tagging.

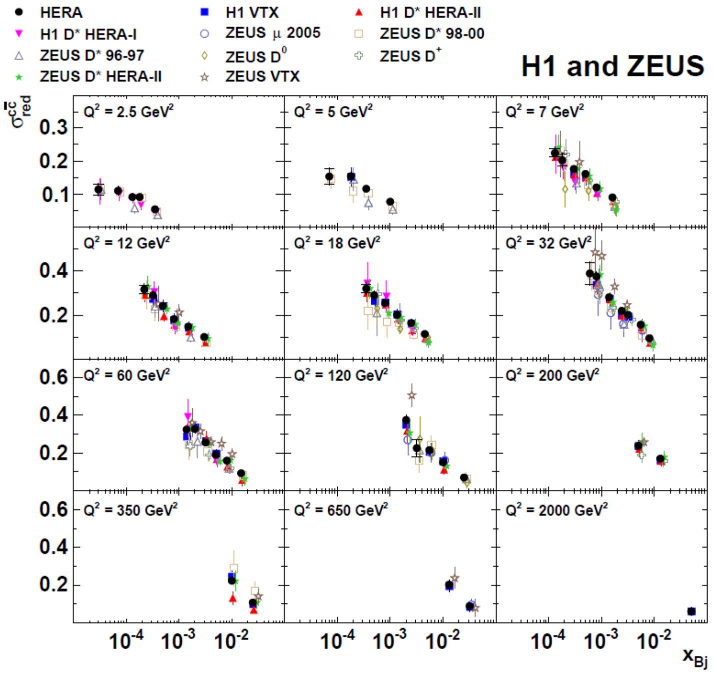
Also at least one disadvantage:

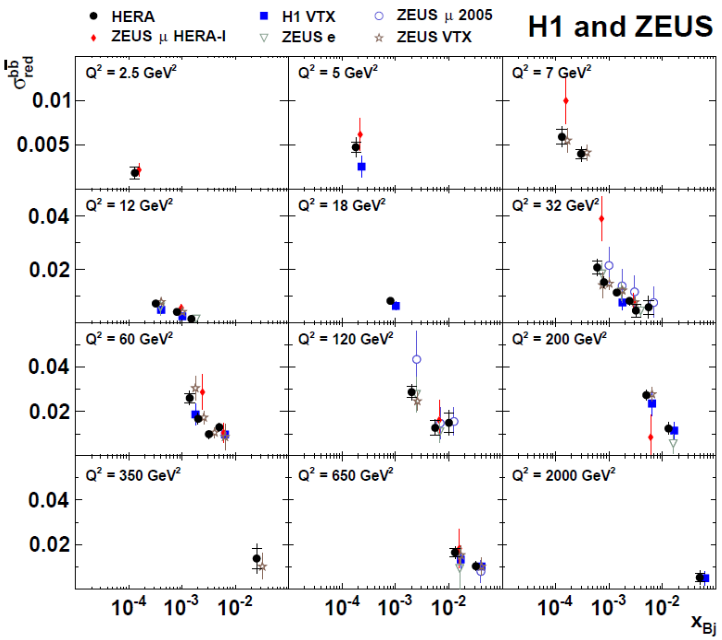
- Lower energies mean lower energy jets—worse calorimetric resolution.

(at high-x, Q^2~10 GeV^2: essentially x is measured by jet energy)

Jets could be important for improving the gluon PDF and measuring $\alpha_S(M_Z)$

Adding more data to HERAPDF2.0: heavy flavour data



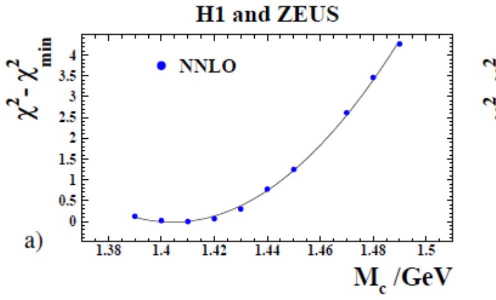


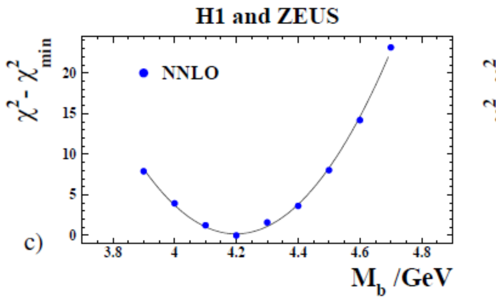
There is a new heavy flavour data combination from HERA arXIv:2018.01019.

This has been used for the latest HERAPDf2.0JetsNNLO fits arXiv:2112.01120

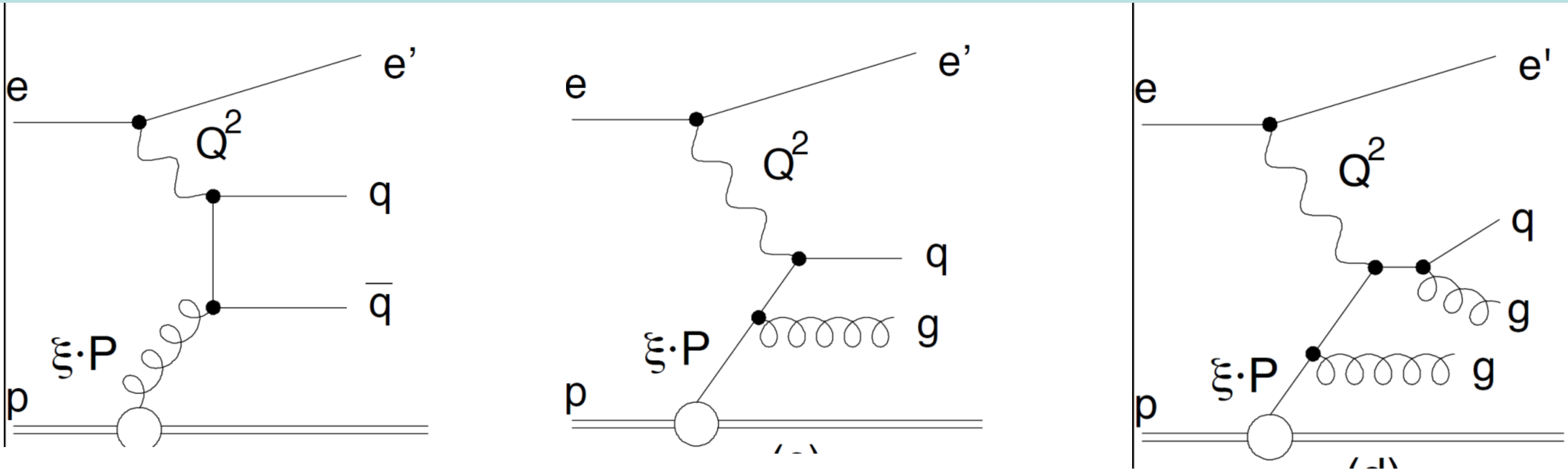
The PDFs do not change significantly due to input of heavy flavour data.

The main effect is to determine the optimal charm and beauty mass parameters and their variation variation as already done with an earlier heavy flavour combination in the standard HERAPDF2.0.





Adding more data to HERAPDF2.0: jet data (EPJC75(2015)2)



It is well known that jet data give a direct handle on the gluon PDF and can be used to measure $\alpha_S(M_Z)$

Seven data sets on inclusive jet, dijet, trijet production at low and high Q^2, from ZEUS and H1 were added to the HERAPDF2.0 fit at NLO

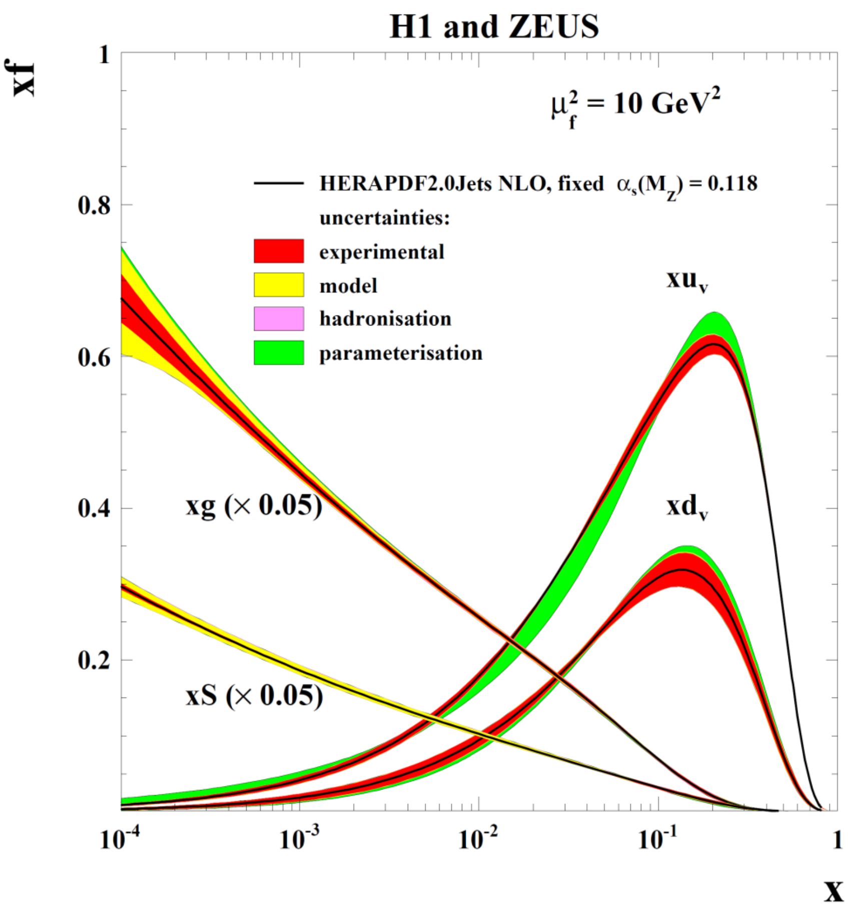
NNLO predictions became available only much more recently. This is why arXiv:1506.06042 is now supplemented by arXiv: 2112.01120

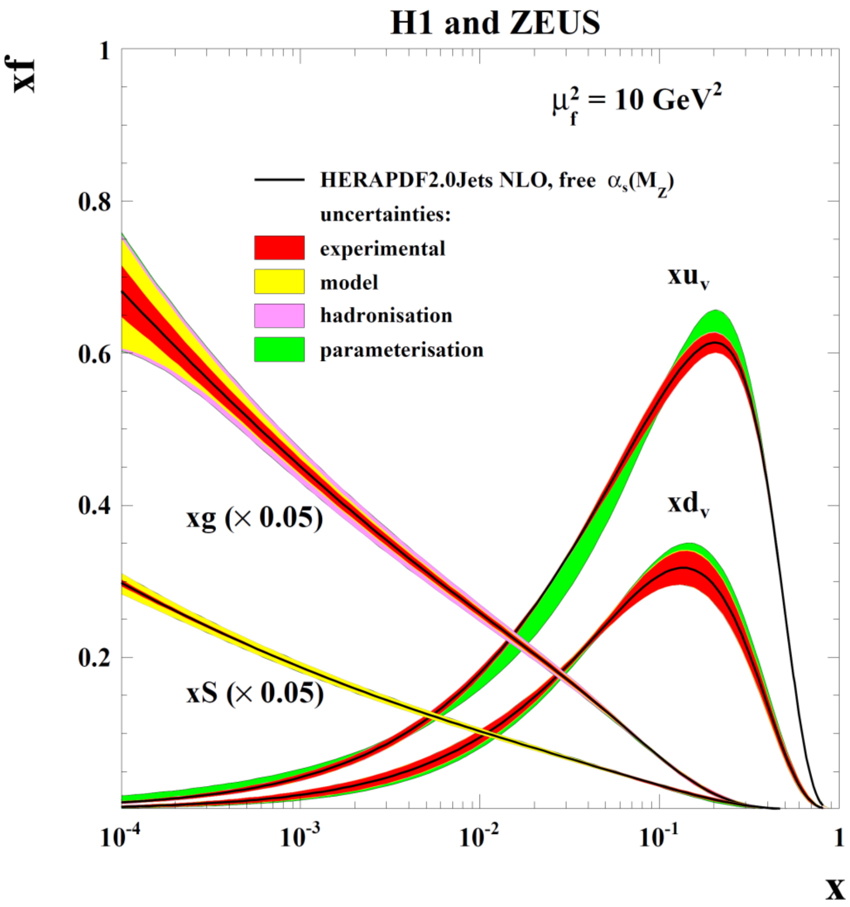
At NNLO somewhat different jet data sets were added since trijet predictions are not availale at NNLO and a new set of measurements of inclusive and dijet production became available from H1.

HERAPDF2.0Jets NLO s based on inclusive + charm + jet data

Fits are made with fixed and free $\alpha_S(M_Z)$

These PDFs are very similar since the fitted value is in agreement with the chosen fixed value. The uncertainties of gluon are not much larger when $\alpha_S(M_Z)$ is free since $\alpha_S(M_Z)$ is well determined. Scale uncertainties are not illustrated on the PDFs

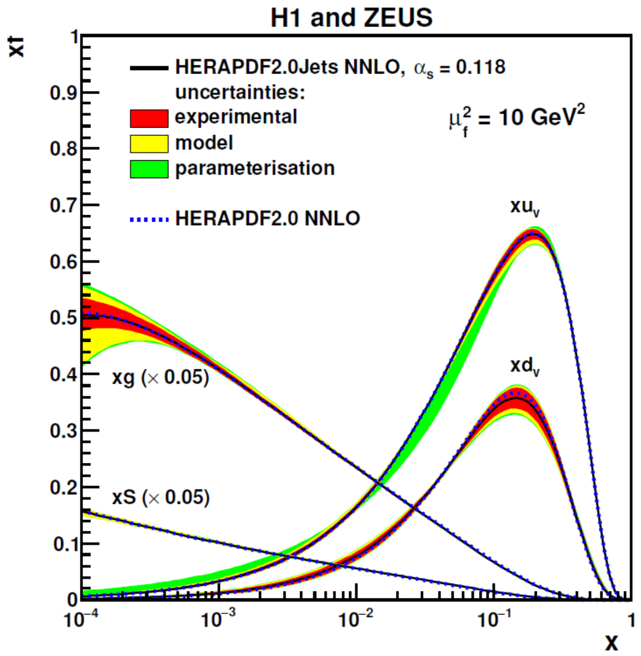




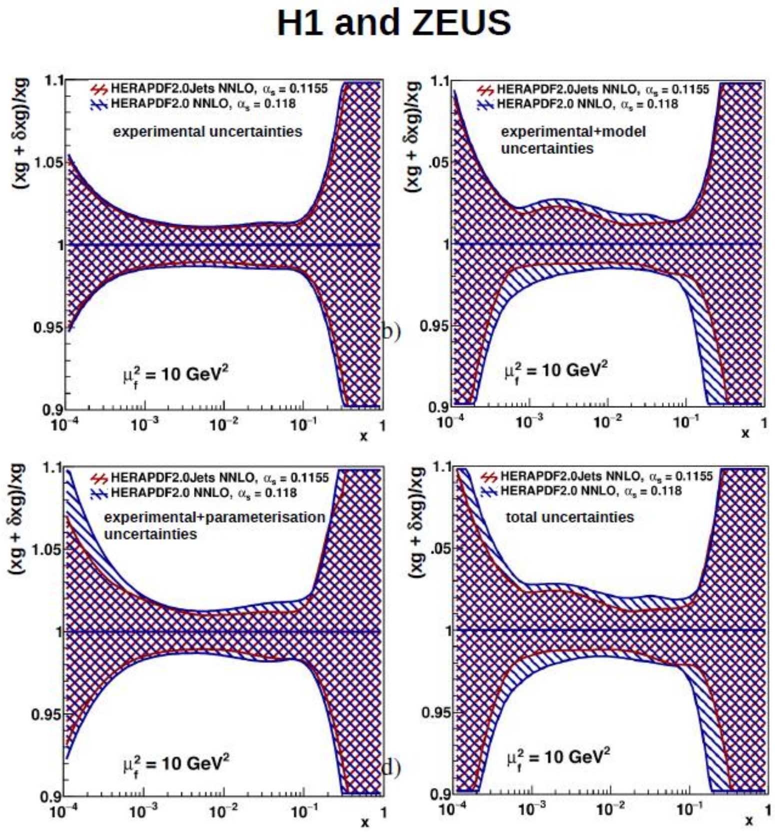
$$\alpha_S(M_Z) = 0.1183 \pm 0.0009_{(exp)} \pm 0.0005_{(model/param)} \pm 0.0012_{(had)} \, {}^{+0.0037}_{-0.0030}(\text{scale})$$

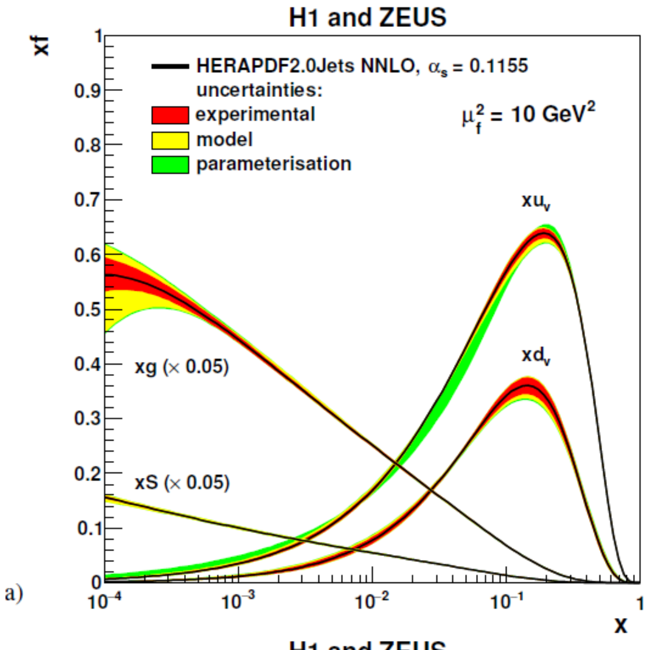
Scale variations assumed ½ correlated and ½ uncorrelated

HERAPDF2.0Jets NNLO is based on inclusive + jet data



Gluon PDF uncertainties are reduced





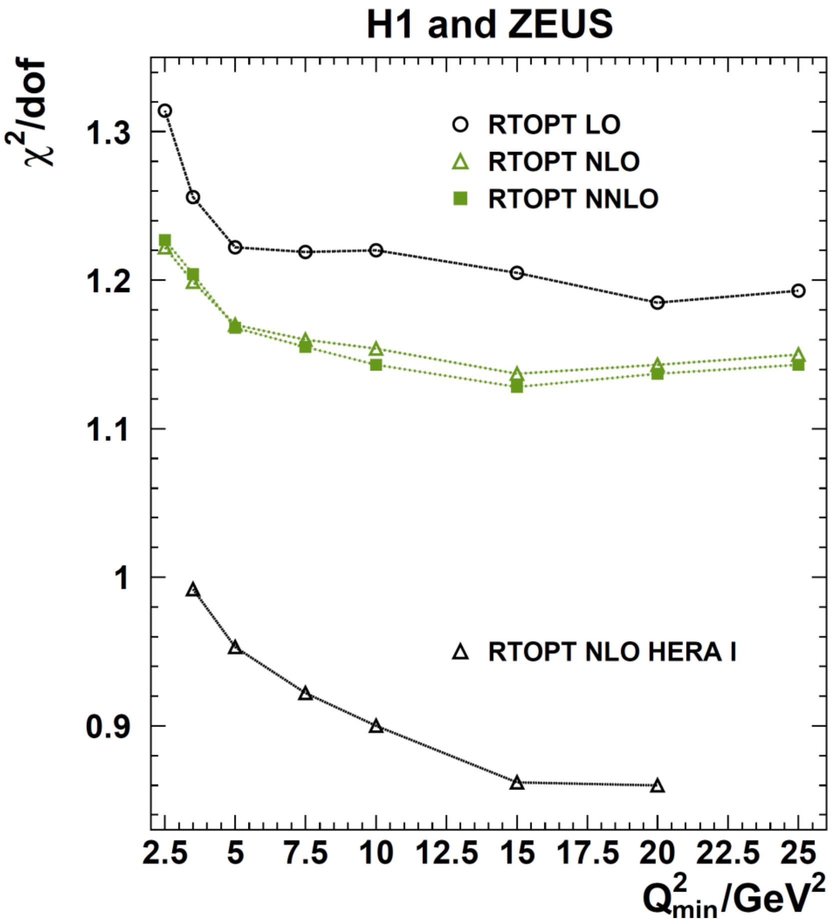
The PDFs HERAPDf2.0NNLO and JetsNNLO agree very well if the same value of $\alpha_S(M_Z)$ is used. However at NNLO the data prefer a lower value

$$\alpha_s(M_Z^2) = 0.1156 \pm 0.0011\,(\mathrm{exp})\ {}^{+0.0001}_{-0.0002}\,(\mathrm{model} + \mathrm{parameterisation})$$

$$\pm 0.0022 \text{ scale}$$

Scale variations assumed ½ correlated and ½ uncorrelated

HERAPDF specifications: minimum value of Q^2



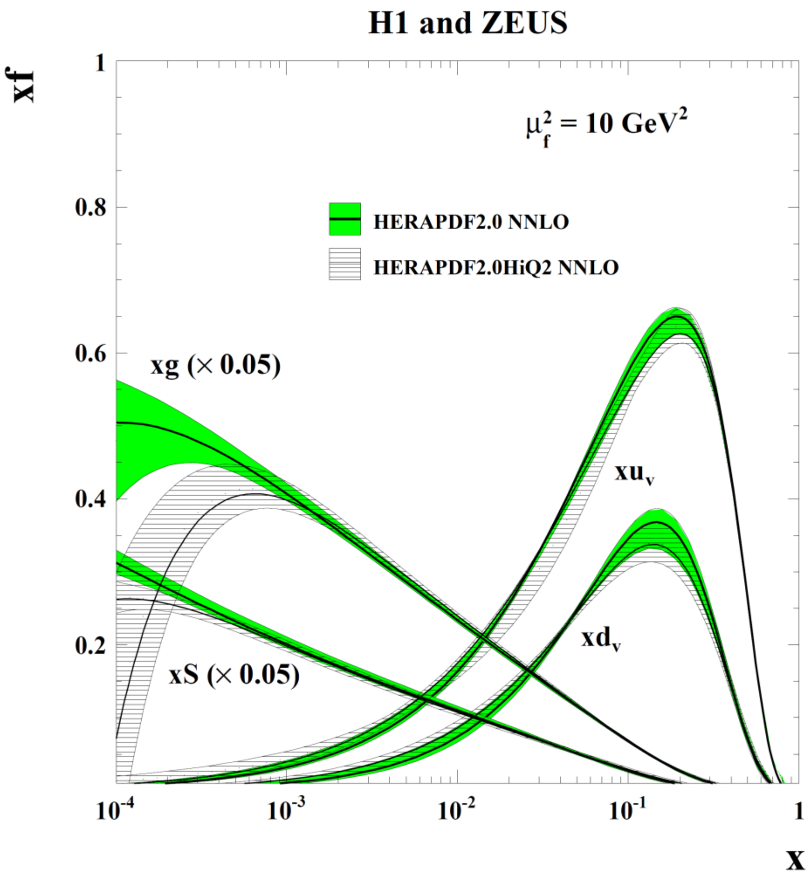
A minimum value of Q^2 for data allowed in the fit is imposed to ensure that pQCD is applicable. For HERAPDF the usual value is $Q^2 > 3.5$ GeV^2 but consider the variation of χ2 with this cut

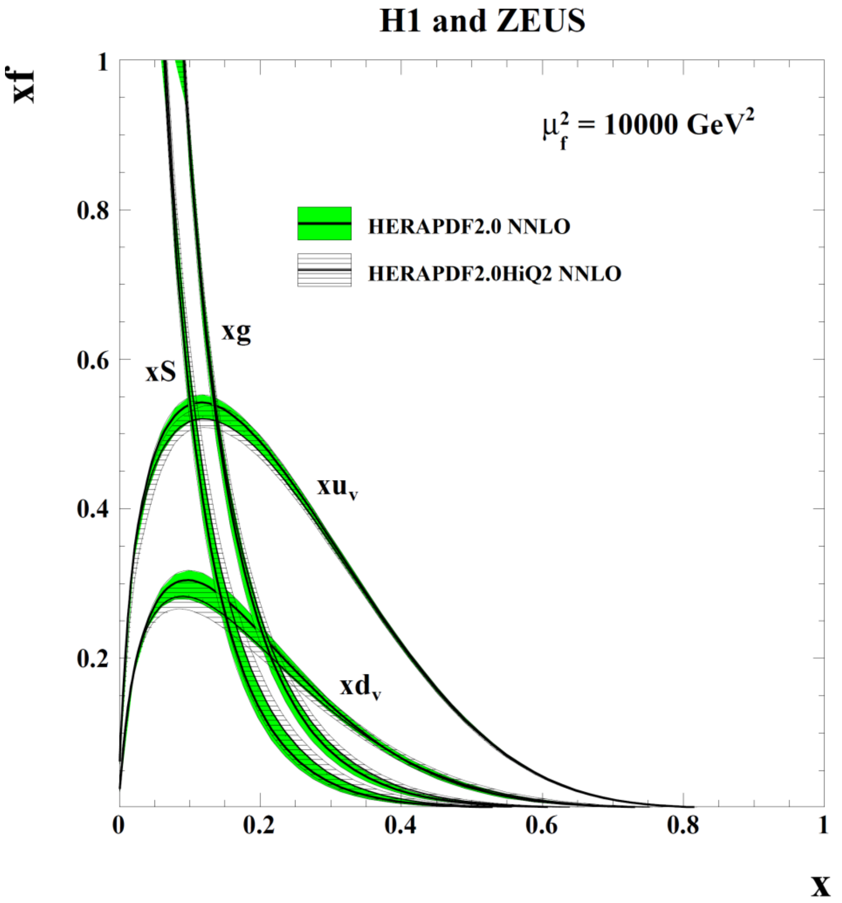
- The χ2 decreases with increase of Q^2 minimum until $Q^2_{min} \sim 10$ -15 GeV^2
- The same effect was observed in HERA-1 data
- This is independent of heavy flavour scheme
- NLO is obviously better than LO but NNLO is not significantly better than NLO, for RT

Fits for two Q^2 cuts were presented: HERAPDF2.0: $Q^2 > 3.5$ and
HERAPDF2.0HiQ2: $Q^2 >10$ GeV^2

HERA kinematics is such that cutting out low Q^2 also cuts the lowest x values, thus HERAPDF2.0HiQ2 is used to assess possible bias in HERAPDF2.0 from including a kinematic region which might require treatment of: non-perturbative effects; ln(1/x) resummation; saturation etc.

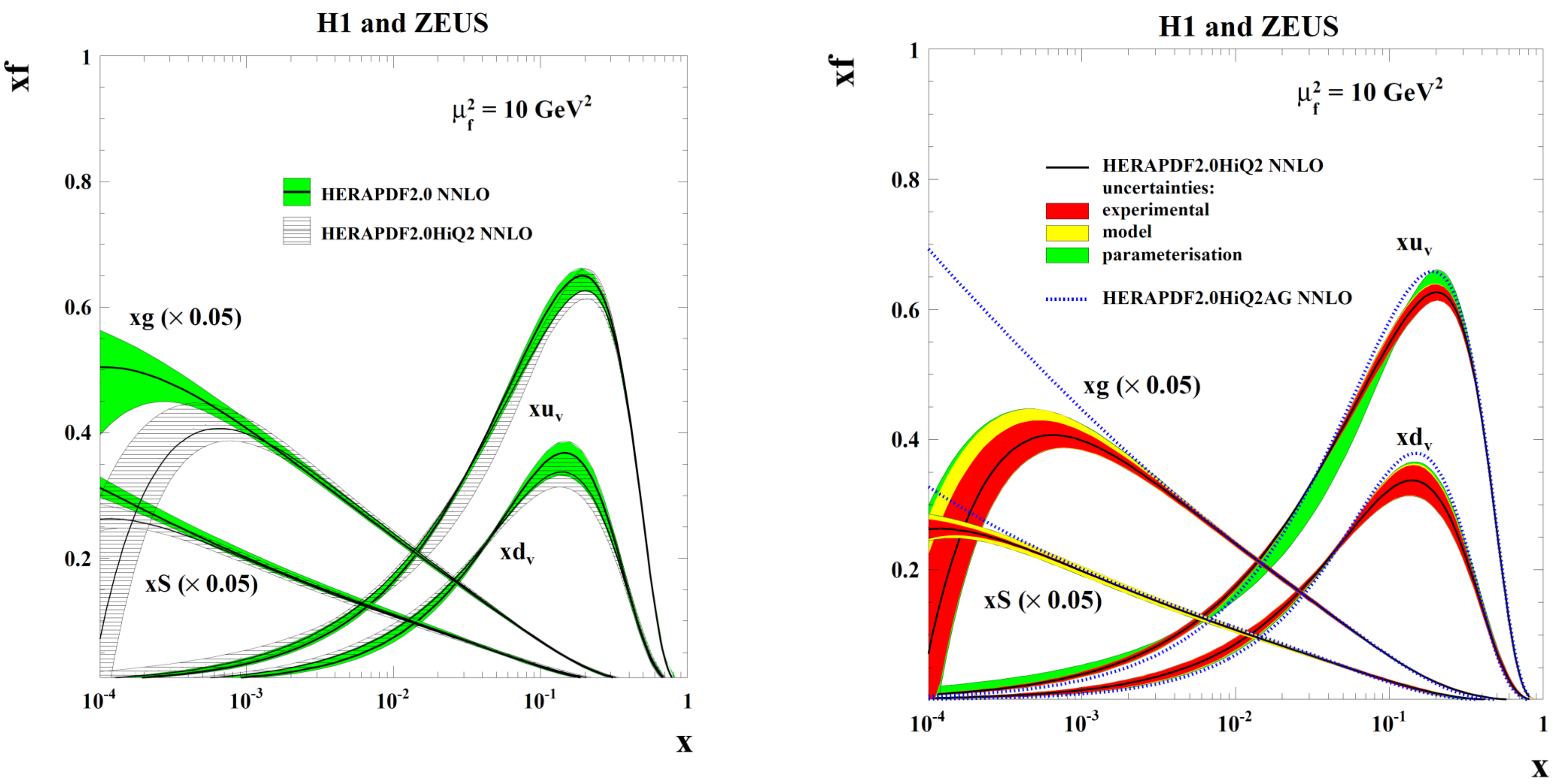
Compare HERAPDF2.0 with Q^2>10GeV^2 to the standard fit at NNLO





Fits are VERY compatible at high-x,Q^2 ---
So there is no bias from including the lower Q^2, lower x data in the fits if we move to LHC scales ---for the ATLAS,CMS kinematic regimes.
BUT the difference in shape for low-x gluon becomes pronounced- fits are no longer compatible
However at very low-x and moderate Q^2 --as in LHCb --the NNLOfit for Q^2_{min}=10 cannot be used--- the gluon becomes negative and so does the longitudinal cross section

Compare HERAPDF2.0 with Q2>10GeV^2 to the standard fit at NNLO



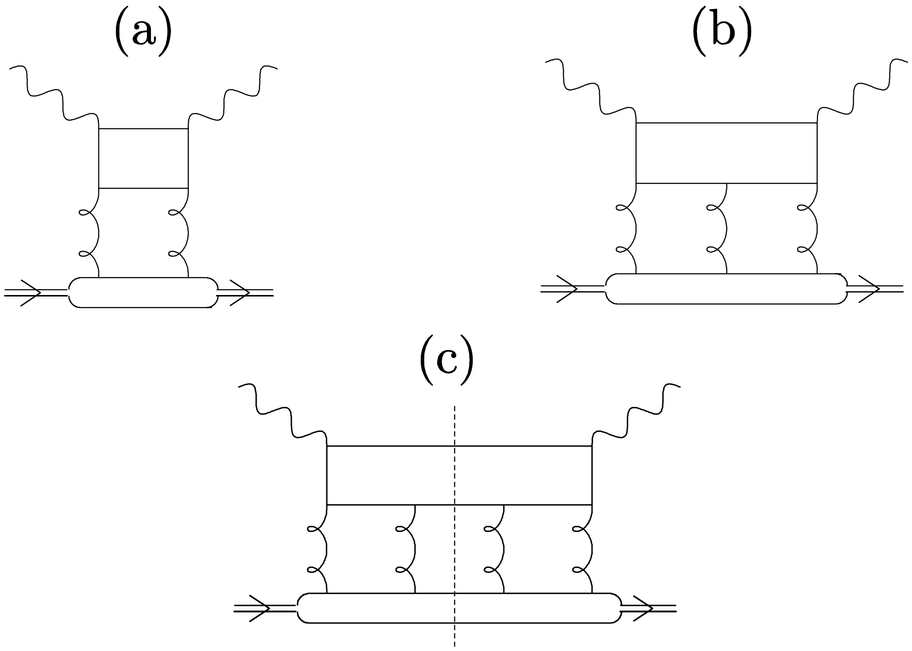
The difference in shape for low-x Sea and gluon– has now become pronounced.

At very low-x and moderate Q^2 --as in LHCb --the NNLOfit for Q^2_{min}=10 gives a negative gluon and a negative longitudinal cross section, and thus is not fit for purpose.

Can use the HERAPDF2.0HiQ2AG– alternative gluon shape—$xg(x) = A_g\, x^{Bg}\, (1-x)^{Cg}\, (1+D_g x)$, which cannot be negative at any x for $Q^2 > Q^2_0$, but fit χ2 is larger by Δχ2~+30

Does this indicate a breakdown of DGLAP at low x?

One approach: (arXiv:1604.02299) consider adding higher twist terms at low-x



Their origin COULD be connected with recombination of gluon ladders- a non-linear evolution effect.

Bartels, Golec-Biernat, Kowalski suggest that such higher twist terms would cancel between σ_L and σ_T in F_2, but remain strong in F_L

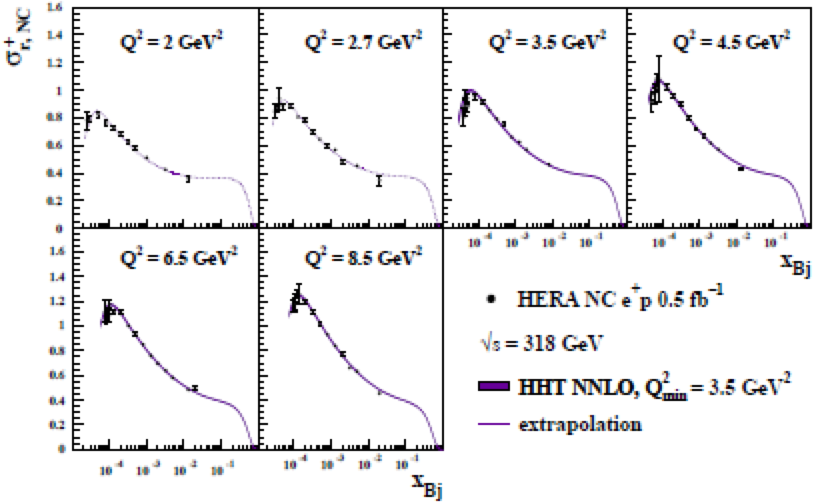
Try the simplest of possible modification to the structure functions

F_2 and F_L as calculated from HERAPDF2.0 formalism

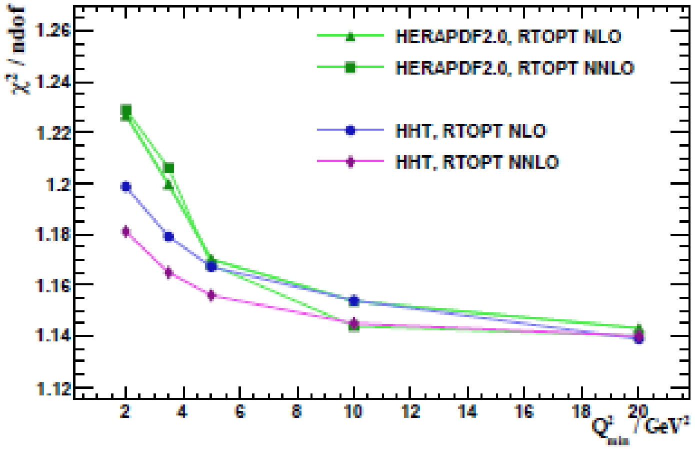
$$F_{2,L} = F_{2,L}\,(1 + A_{2,L}^{HT}/Q^2)$$

Such a modification of F_L is favoured, whereas for F_2 it is not.

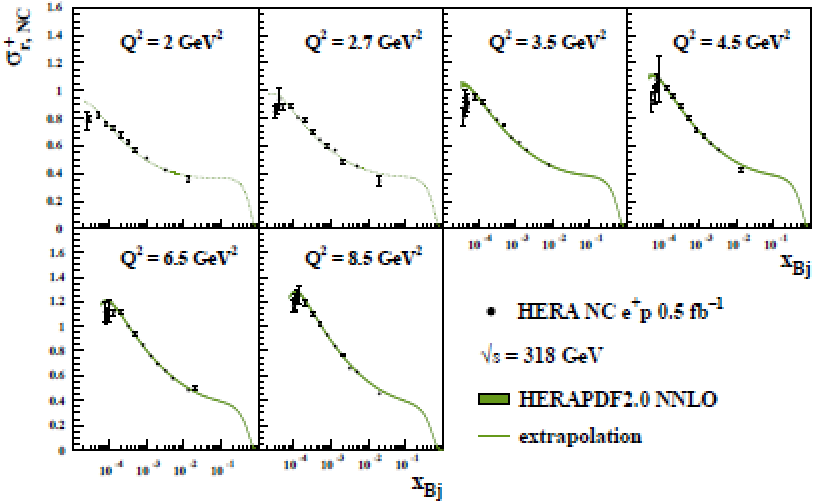
If A_L^{HT} is added $A_L^{HT} = 5.5 \pm 0.6$ GeV^2 and $\Delta\chi2 = -47$



$$\sigma^{\pm}_{r,NC} = F_2 - \frac{y^2}{Y_+} F_L \qquad F_L^{HT} = F_L^{DGLAP}\left(1 + A_L^{HT}/Q^2\right)$$

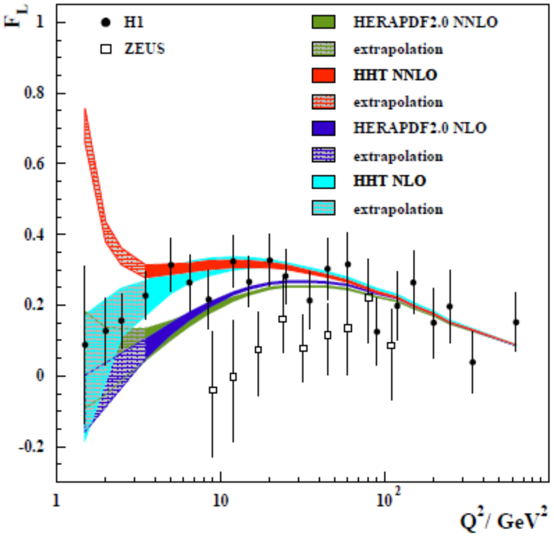


NNLO is now better than NLO

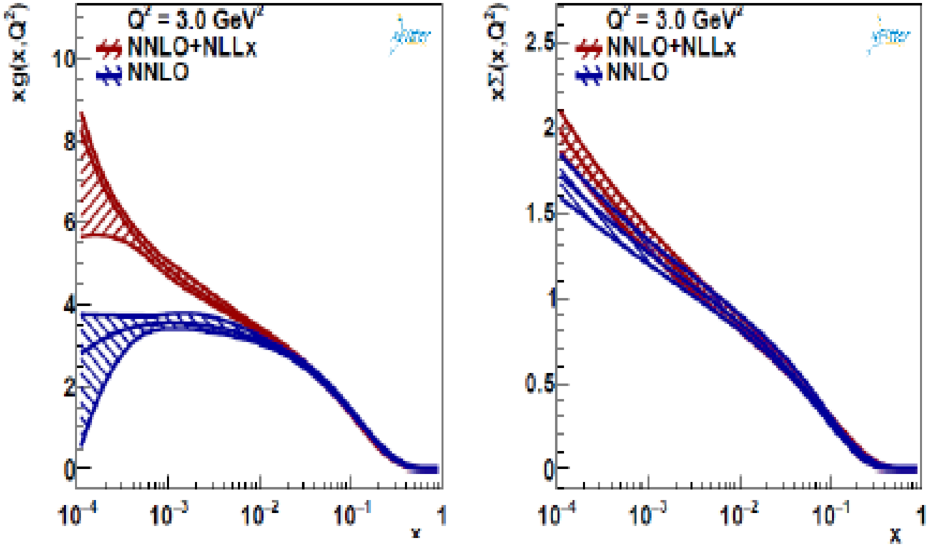


PDFs from these fits are similar at LHC scales

Data can be fitted down to $Q^2 = 2\text{GeV}^2$ - but lower Q^2 cannot be described in such a simple picture



An alternative picture: ln(1/x) resummation at low-x to NLLx (arXiV:1804.00064)

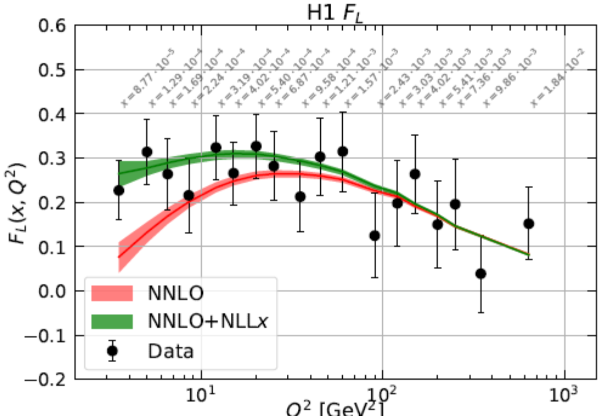


A decrease in χ2 of 74 using ln(1/x) resummation

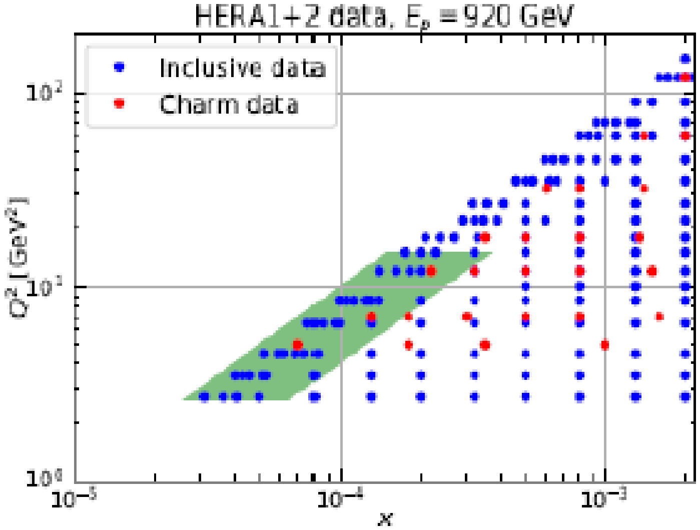
Largely due to the NC e+p 920 data

But also less need for shifts of systematic uncertainties

Gives a steep gluon at low-x, instead of the valence like gluon of DGLAP (and also of the +HT analysis)



More sensible behaviour of F_L at low Q^2



Scans of χ2 vs $Q^2_{\min}$, $x_{\min}$ and $y_{\max}$ were made to delineate the region where fits become poor: $Q^2 < 15$ GeV^2, $x < 5\ 10^{-4}$, $y > 0.4$

These low-x, low Q2 studies are suggestive but perhaps not definitive in suggesting that physics beyond DGLAP, is needed.

EIC could help here **especially with heavy ion data**
But even with proton data perhaps an interesting measurement of F_L could be made

- High luminosity at all proton beam energies– HERA failed to do that
- Well spread energies- maximize range in y^2– You can do better than HERA
- Ability to measure LOW energy electrons (sub-GeV if possible)
- High resolution electron calorimetry
- Control the background
- mostly photo-production- taggers down the rear be
- distinguish right and wrong sign electron candidates even at low angles and low energies
- needs excellent tracking and minimum inactive material

Summary

Lessons from HERA for the EIC

What can EIC do that HERA could not– even using just protons?

Measurement of PDFs up to high-x

Measurement of F_L at lower x, Q^2

Maybe consider combining HERA and EIC data for PDF fits

Do not spend your time chasing ‘new physics’

Concentrate on ‘new QCD’