Legacy of HERA for the LHC

A M Cooper-Sarkar

March 2010

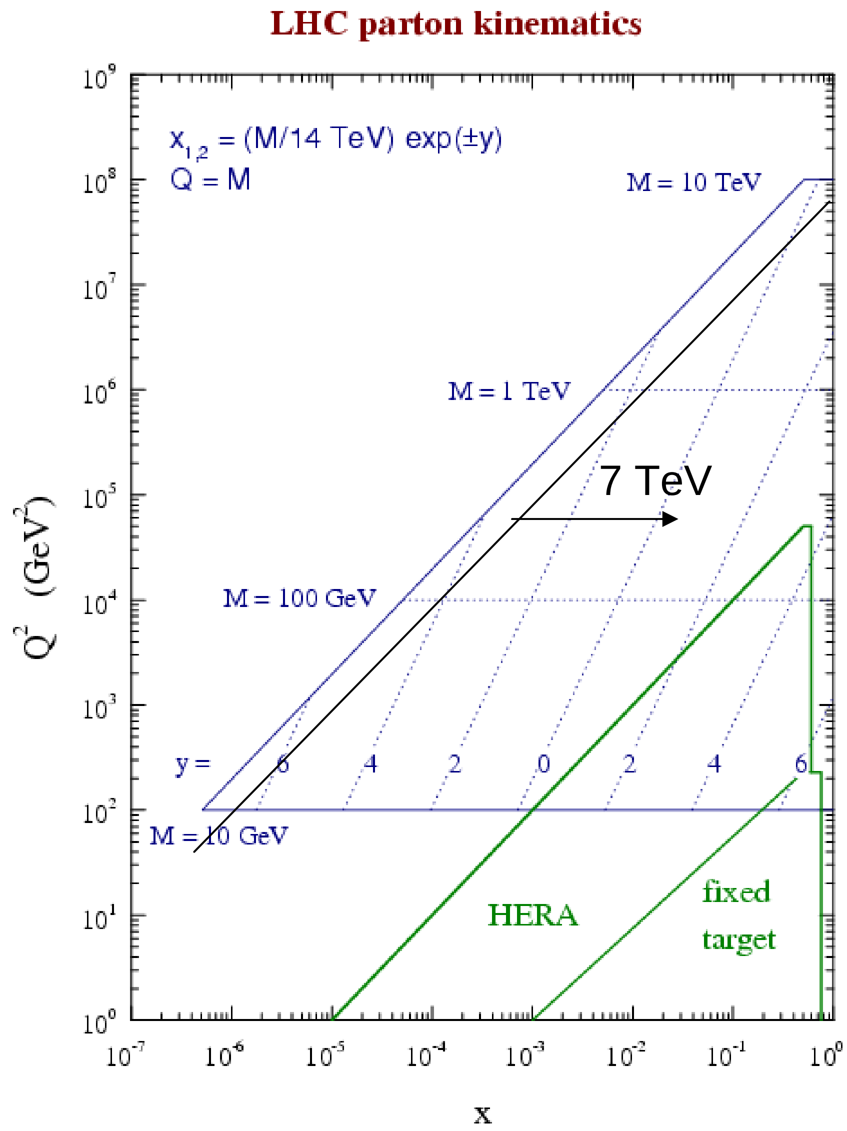
Combination of ZEUS and H1 data:

1. Data on multi-leptons at high-pt arxiv:0907:3627 -BSM constraints
2. Isolated leptons and missing pt: arxiv:0911.0858 -BSM constraints
3. Inclusive cross-sections HERA-1 (1992-2000):arxiv:0911.0884 - improved constraints at low-x
4. Heavy flavour data: F2c – preliminary -better understanding of correct heavy flavour treatment
5. Low energy runs – FL- 2007- preliminary -improved constraints at low Q2- transition to non-perturbative regime
6. Inclusive cross-sections HERA-II (2003-2007) -improved constraints at high-x, electroweak physics
   Jet cross-sections -improved constraints on gluon PDF and αs(MZ)

Why combine ZEUS and H1 data? At the LHC we collide protons – 7 TeV collisions should be happening for the first time today!

Protons are full of partons. Our knowledge of partons comes from Deep Inelastic Scattering data. HERA dominates these data and is most relevant for the kinematic region of early LHC data

We think we know how to extrapolate in Q^2 using (N)NLO QCD (using the DGLAP equations) but we don't a priori know the shapes of the parton distributions in x. The HERA data is our best guide

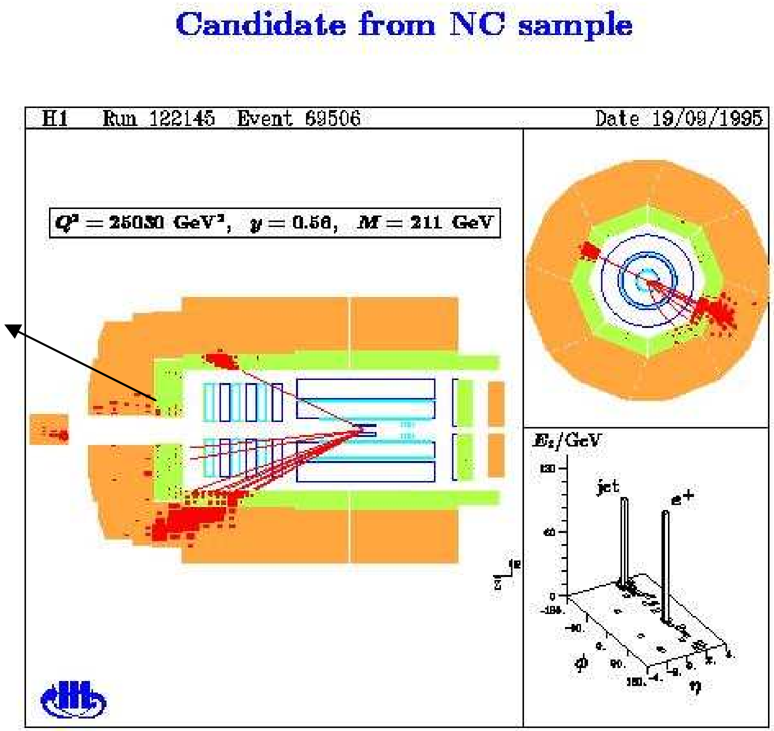


$$\frac{dq(x,Q^2)}{d\ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{dy}{y} \left[Pqq(z) q(y,Q^2) + Pqg(z) g(y,Q^2) \right]$$

$$\frac{dg(x,Q^2)}{d\ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{dy}{y} \left[\Sigma_q Pgq(z) q(y,Q^2) + Pgg(z) g(y,Q^2) \right]$$

DGLAP eqns

PDFs were first investigated in deep inelastic lepton-hadron scattering -DIS

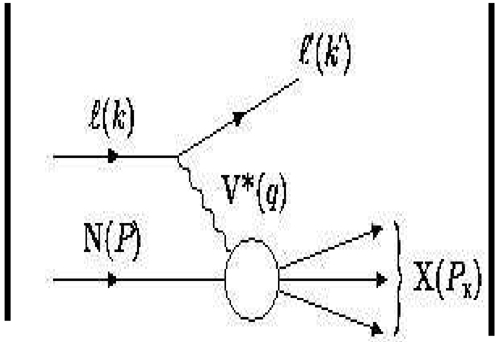


E♦

E_e

E_p

$d\sigma \sim$



2

Leptonic tensor - calculable

$L^{\mu\nu} W_{\mu\nu}$

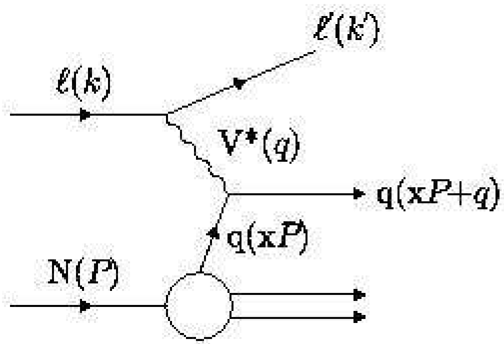
Hadronic tensor- constrained by Lorentz invariance

$q = k - k'$, $Q^2 = -q^2$

$x = Q^2 / (2p.q)$

$y = (p.q)/(p.k)$

$Q^2 = s\ x\ y$



$s = 4\ E_e\ E_p$

$Q^2 = 4\ E_e\ E'\ \sin^2\theta_e/2$

$y = (1 - E'/E_e \cos^2\theta_e/2)$

$x = Q^2/sy$

The kinematic variables are measurable and x represents The FRACTIONAL momentum of the incoming nucleon taken by the struck quark

$(xP+q)^2 = x^2p^2 + q^2 + 2xp.q \sim 0$

for massless quarks and $p^2 \sim 0$ so

$x = Q^2/(2p.q)$

Where does the information on parton distributions come from?

CC e-p

CC e+p

$$\frac{d^2\sigma(e^-p)}{dxdy} = \frac{G_F^2 M^4_W}{2\pi x(Q^2+M^2_W)^2}[x(u+c) + (1-y)^2 x(\bar{d}+\bar{s})]$$

$$\frac{d^2\sigma(e^+p)}{dxdy} = \frac{G_F^2 M^4_W}{2\pi x(Q^2+M^2_W)^2}[x(\bar{u}+\bar{c}) + (1-y)^2 x(d+s)]$$

- We can use the reduced cross-sections to learn about high-x valence PDFs

For NC e+ and e-

$$\frac{d^2\sigma(e\pm N)}{dxdy} = \frac{2\pi\alpha^2 s}{Q^4} Y_+ \left[F_2(x,Q^2) - \frac{y^2}{Y_+} F_L(x,Q^2) \pm \frac{Y\_}{Y_+} xF_3(x,Q^2)\right], \quad Y_\pm = 1 \pm (1-y)^2$$

$F_2 = F_2^{\gamma} - v_e P_Z F_2^{\gamma Z} + (v_e^2+a_e^2)P_Z^2 F_2^Z$

$xF_3 = \quad - a_e P_Z xF_3^{\gamma Z} + 2v_e a_e P_Z^2 xF_3^Z$

Where $P_Z^2 = Q^2/(Q^2 + M^2_Z)\ 1/\sin^2\theta_W$, and at LO

$[F_2, F_2^{\gamma Z}, F_2^Z] = \Sigma_i\ [e_i^2, 2e_i v_i, v_i^2+a_i^2][xq_i(x,Q^2) + x\bar{q}_i(x,Q^2)]$

$[xF_3^{\gamma Z}, xF_3^Z] = \Sigma_i\ [e_i a_i, v_i a_i] \quad [xq_i(x,Q^2) - x\bar{q}_i(x,Q^2)]$

So that $xF_3^{\gamma Z} = 2x[e_u a_u u_v + e_d a_d d_v] = x/3\ (2u_v+d_v)$

Where $xF_3^{\gamma Z}$ is the dominant term in xF_3

The difference between NC e+ and e- cross-sections gives the valence structure function xF3 due to γ/Z interference and Z exchange

Note this is obtained on a pure proton target so

- No heavy target corrections
- No assumptions on strong isospin

(Unlike xF3 determined from neutrino scattering on heavy isocalar targets)

A substantial part of the uncertainty on parton distributions comes from the need to use many different input data sets with large systematic errors and questionalble levels of consistency

- Averaging H1 and ZEUS HERA-I data provides a model independent tool to study consistency of the data and to reduce systematic uncertainties:

$\Rightarrow$ Experiments cross calibrate each other

- The combination method includes accounting for full systematic error correlations.

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

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

Data Sets

2009 average based on the complete HERA-I inclusive NC and CC DIS data:

$\Rightarrow$ Ep=820 ($\sqrt{s}$=300) and Ep=920 ($\sqrt{s}$=320) GeV

200 pb^{-1} of e+p , 30 pb^{-1} of e-p

- CC e^-p data: H1 98, ZEUS 98 (250 ≤ Q2 ≤ 15000 GeV2)

- CC e^+p data: H1 94-97, H1 99-00, ZEUS 94-97, ZEUS 99-00 ($250 \leq Q^2 \leq 15000$ GeV^2)

- NC e^-p data: H1 98, ZEUS 98 ($200 \leq Q^2 \leq 30000$ GeV^2

- NC e^+p data: ZEUS 96-97, ZEUS 99-00, H1 99-00 "high Q^2",H1 96-00 "bulk", H1 95-00 "low-Q2", ZEUSBPC/BPT, SVX95 "low-Q2" ($0.045 < Q^2 < 30{,}000$ GeV^2)

The NCe+p data sets cover 5 decades of the kinematic plane in Q2 and x. The scaling violations of these data give us our best handle on the low-x gluon which is very important for Standard Model LHC physics at the W/Z scale

The new combination supercedes all these data sets

This gives 110 correlated systematic error sources

But they are all small !

3 "procedural uncertainties" related to the averaging procedure

Averaging procedure

- Swim all points to a common x-Q^2 grid
- Moved 820 GeV data to 920 GeV p-beam energy
- Calculate average values and uncertainties

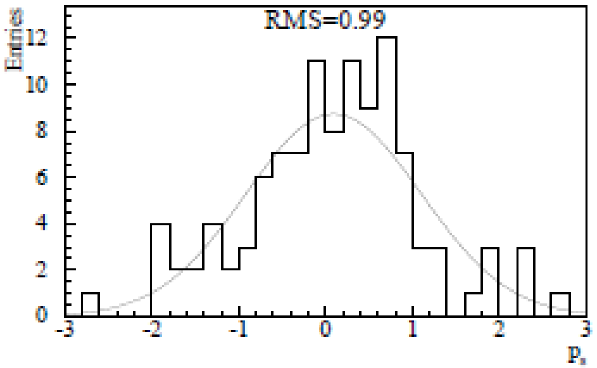
This is done by making a χ2 fit to the data points of both experiments which simply assumes that for each process (NC or CC, e+ or e-) and each x, Q^2 point (i) there is only one 'true' value of the cross-section- these are the predictions m_i whereas there can be several measurements of this value, 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 systematics of the data points- each data point can have several such uncertainties Γ, hence sum over j for each data point i, but these uncertainties are common to all data points for large sub-sets of data. The fit determines the value of the cross-sections m_i and the systematic shift parameters b_j
- Evaluate further uncertainties due to choices in combination procedure,e.g. Correlations between ZEUS and H1

1402 data points are averaged to 741 combined data points

χ2/ndf =637/656

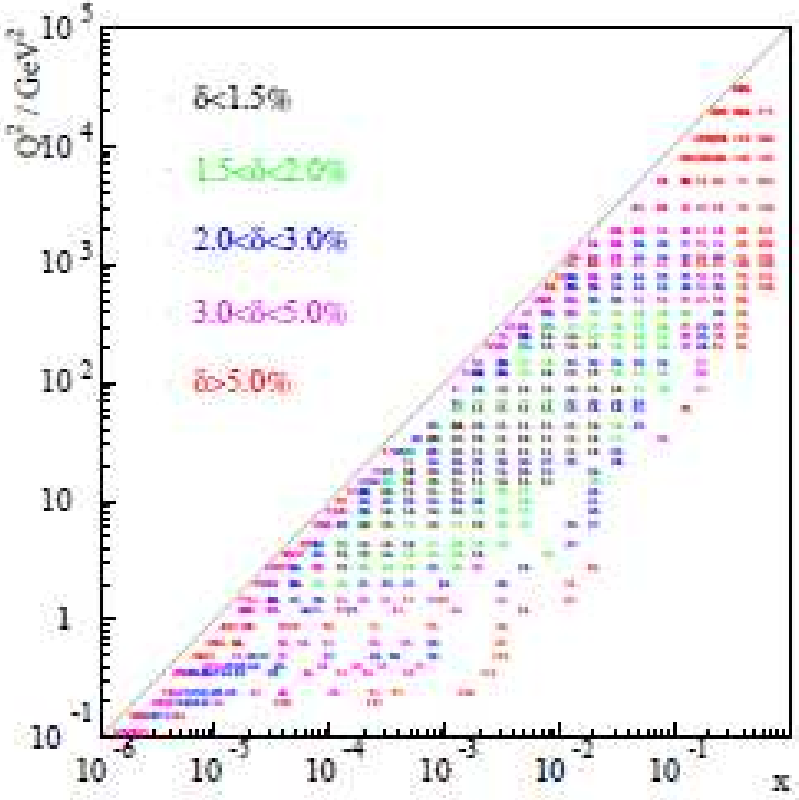


Systematic shift parameters b, shift most systematics < 1 std deviation

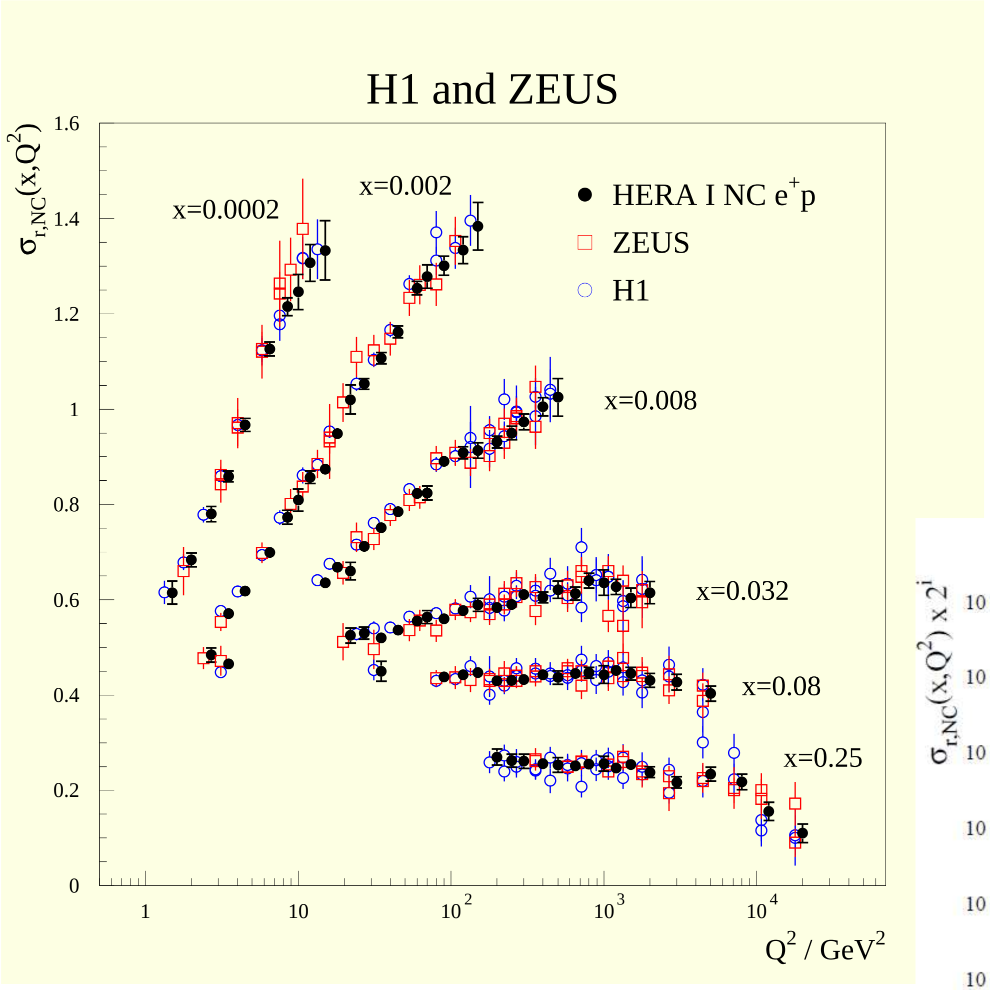
But the fit also determines uncertainties on the shift parameters Δb, some of these are much reduced e.g

ZEUS γp background uncertainty is reduced by a factor of 3

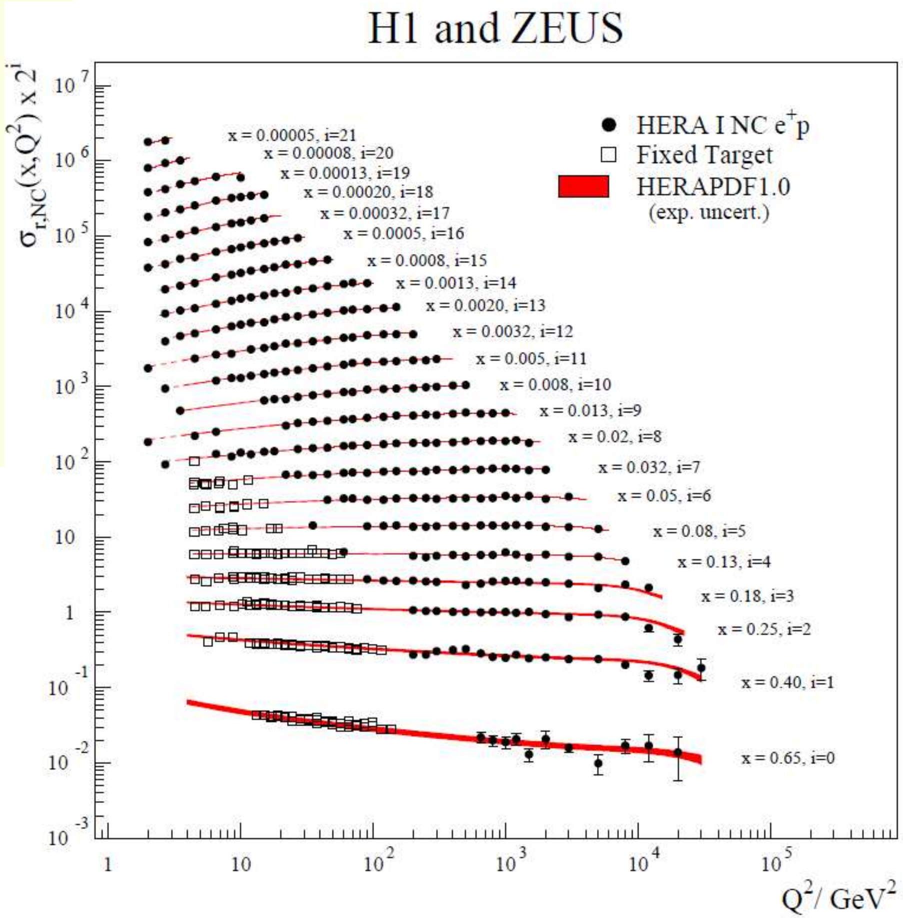
H1 LAr hadron calorimeter energy scale uncertainty is halved



Resulting total uncertainties are <2% over a large part of the kinematic plane AND the contribution of correlated systematics to this errors is now < statistical error



This page shows NC e+ combined data



Results of the combination compared to the separate data sets

Now to use these data for extracting parton distributions: HERAPDF1.0

Motivation

Some of the debates about the best way of estimating PDF uncertainties concern the use of many different data sets with varying levels of consistency.

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.

Whereas the data set does not give information on every possible PDF flavour it does:

•Give information on the low-x Sea (NCe+ data)

•Give information on the low-x Gluon via scaling violations (NCe+ data)

•Give information on high-x u (NCe+/e- and CCe-) and d (CCe+ data) valence PDFs

•Give information on u and d-valence shapes down to x~3 10^{-2} (from the difference between NCe+ and NCe-)

Furthermore, the kinematic coverage at low-x ensures that these are the most crucial data when extrapolating predictions from W, Z and Higgs cross-sections to the LHC

Correlated systematic uncertainties, $\chi 2$ and $\Delta\chi 2$

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

Even though there are 113 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 the kinematic region used in the QCD fits

This means that the method of treatment of correlated systematic uncertainties in our PDF fits is not crucial. We obtain similar results treating all systematic errors as correlated or as uncorrelated.

For our PDF fits we combine 110 sources of systematic uncertainty from the separate experiments in quadrature and OFFSET the 3 procedural systematics which derive from the method of data combination.

We set the experimental uncertainties on our PDFs at 68% CL by the conventional $\chi 2$ tolerance

$$\Delta\chi 2 = 1$$

Theoretical framework

Fits are made at NLO in the DGLAP formalism -using QCDNUM 17.04

The Thorne-Roberts massive variable flavour number scheme is used (2008 version) and compared with ACOT

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)

Parametrisation and model assumptions (all values in green are varied)

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.

B(d-valence) = B(u-valence), B(Dbar) = B(Ubar),

A(Ubar) = A(Dbar) (1-fs), where sbar = fs Dbar, so that ubar $\rightarrow$ dbar as x$\rightarrow$ 0 (fs=0.31)

Uncertainties due to model assumptions are evaluated by varying input values

Variation of heavy quark masses:

Mc=1.35 to1.65 GeV (the pole-mass)

Mb= 4.3 to 5.0 GeV

Variaion of the sea fraction

Fs=s/(d+s) = 0.23 to 0.38

Variation of the minimum Q2 cut on data entering the fit

Q^2_{min}= 2.5 to5.0 GeV^2

We also vary the value of the starting scale Q^2_0 from1.5 to 2.5 GeV^2: this is considered as a parametrisation uncertainty rather than a model uncertainty

Parametrisation uncertainties- indicative, not exhaustive

The central fit is chosen as follows: start with a 9 parameter fit with all D and E parameters = 0 and then add D and E parameters one at a time noting the χ2 improvement. Chose the fit with the lowest χ2. This has E(u-valence) ≠ 0 and χ2/ndf = 574/582.

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

This happens to be the central fit

PDF	A	B	C	D	E
xg	sum rule	FIT	FIT	-	-
xu_{val}	sum rule	FIT	FIT	-	FIT
xd_{val}	sum rule	$=B_{u_{val}}$	FIT	-	-
$x\overline{U}$	$\lim_{x\to 0} \overline{U}/\overline{D} \to 1$	FIT	FIT	-	-
$x\overline{D}$	FIT	$=B_{\overline{U}}$	FIT	-	-

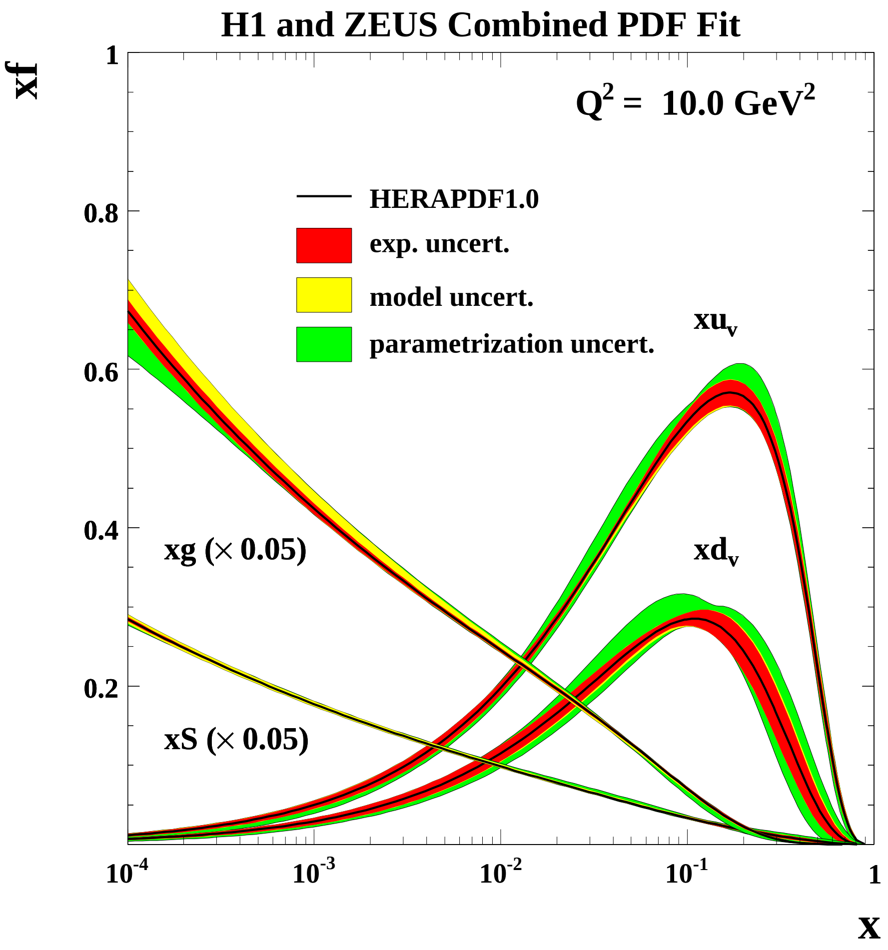
However the procedure is continued. We then start with this 10 parameter fit and add all the other D and E parameters one at a time noting the χ2 improvement. It turns out that there is no significant further improvement in χ2 for 11 parameter fits.

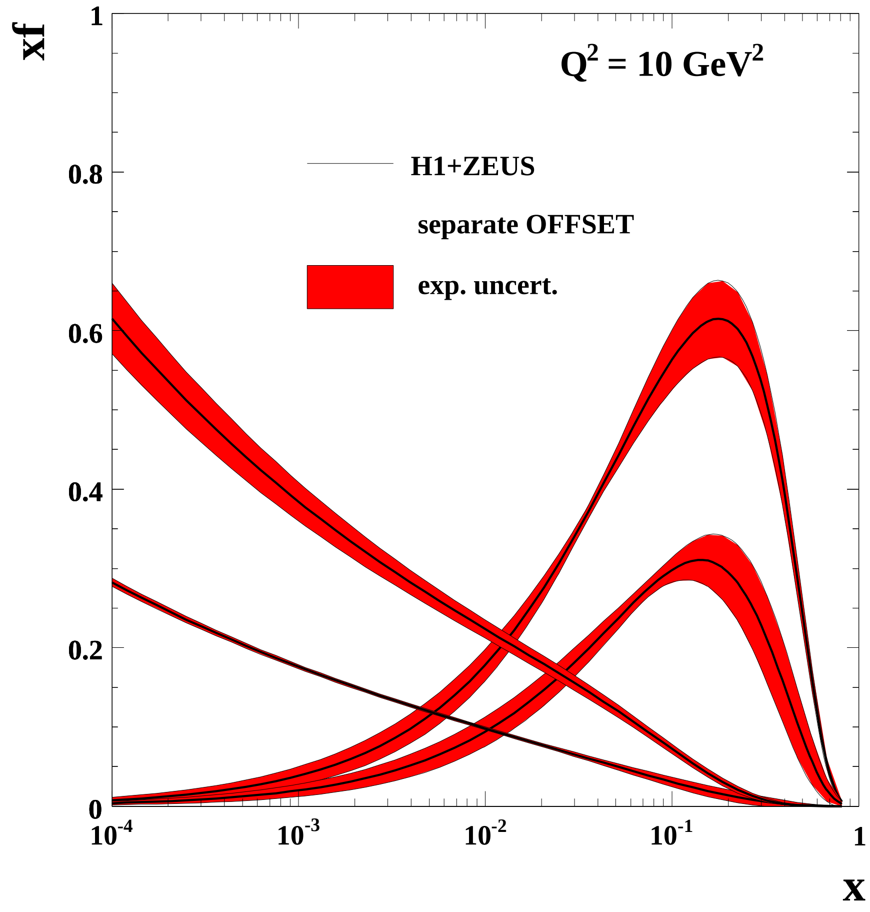
An envelope of the shapes of these 11 parameter fits is formed and used as a parametrization error. So far this addresses parametrization uncertainty at high-x.

Low-x is also addressed by considering the following additional variations:

1. Bdv free –this results in Bdv ≈ Buv
2. A negative gluon term: - A $x^B(1-x)^C$ is added to the usual gluon term, when the starting scale of the fit is lowered to Q^2_0=1.5 GeV^2 – this results in a small –ve gluon term

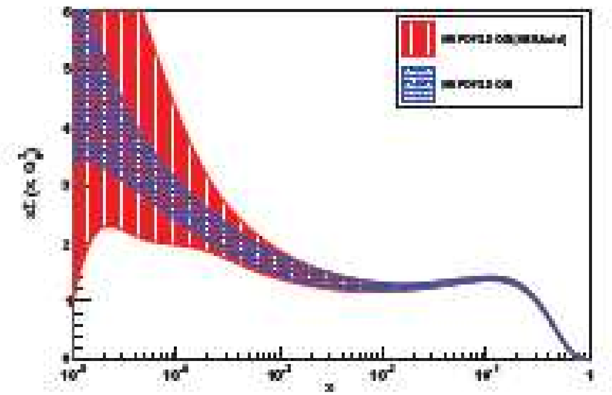
RESULTS for HERAPDF1.0 –arxiv:0911.0884



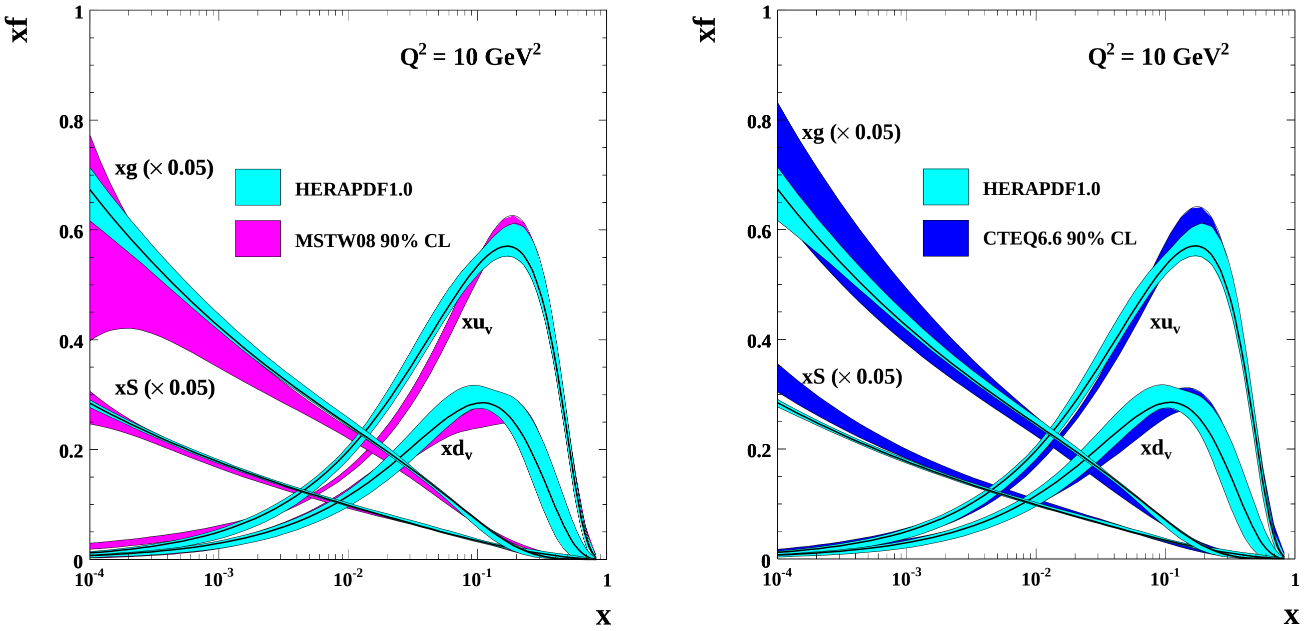


And here is a summary plot of the PDF results

To appreciate how much better this is than uncombined HERA data compare the red experimental errors to this plot which shows the experimental errors for a smilar PDf fit to uncombined data

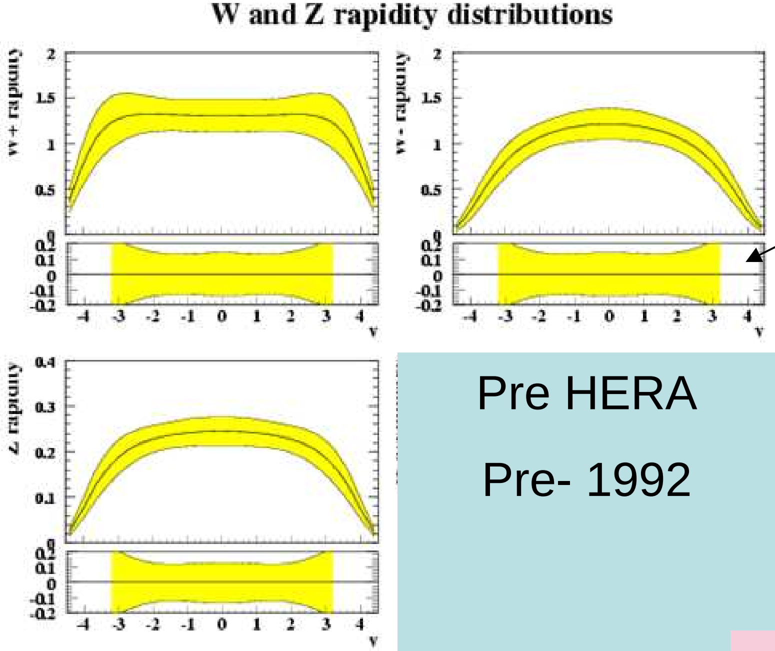


The NNPDF global PDF fitting group have already incorporated the combined HERA data into their fit and here is the improvement to the Sea PDF- with uncombined HERA data you get the red- with combined you get the blue

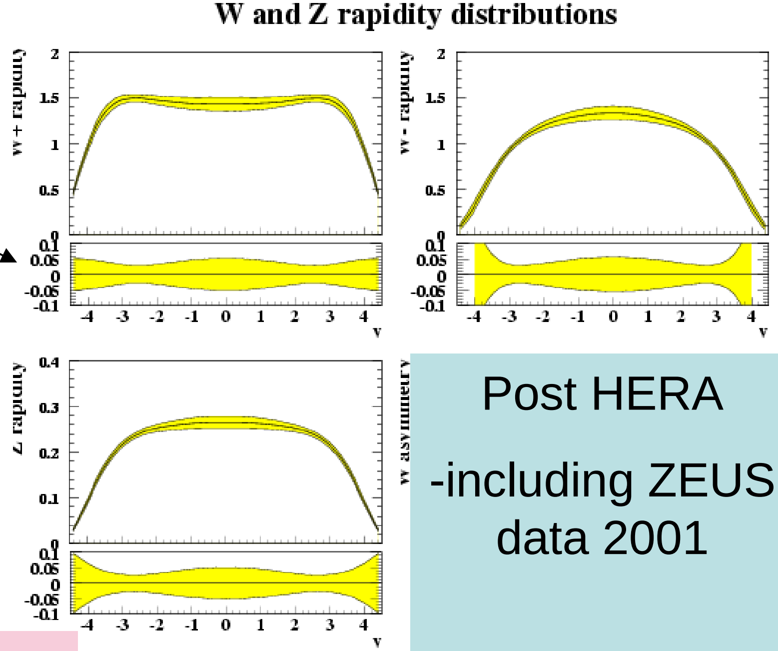


Compare HERAPDF1.0 with current CTEQ6.6 and MSTW08 PDFs input of the new data should bring some reduction in uncertainty

Consequences for W and Z production at the LHC Look at predictions for W/Z rapidity distributions: Pre- and Post-HERA



Note difference in scale for fractional errors



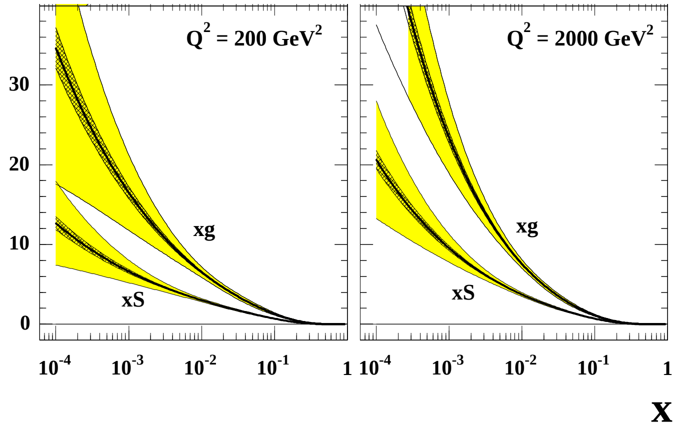
Pre HERA

Pre- 1992

Why such an improvement ?

Post HERA

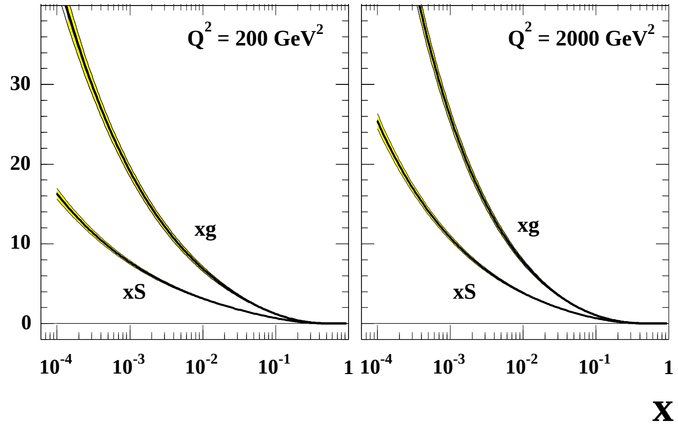
-including ZEUS data 2001



It's due to the improvement in the low-x sea and gluon

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

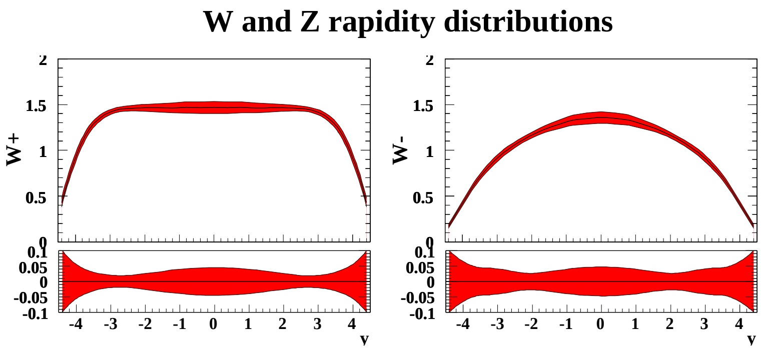
And at $Q^2 \sim M_Z^2$ **the sea is driven by the gluon**

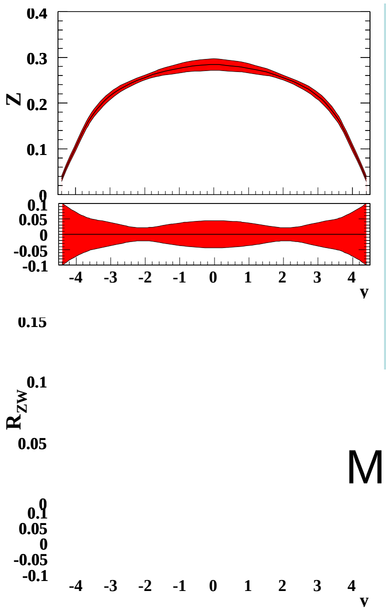


These illustrations at 14 TeV

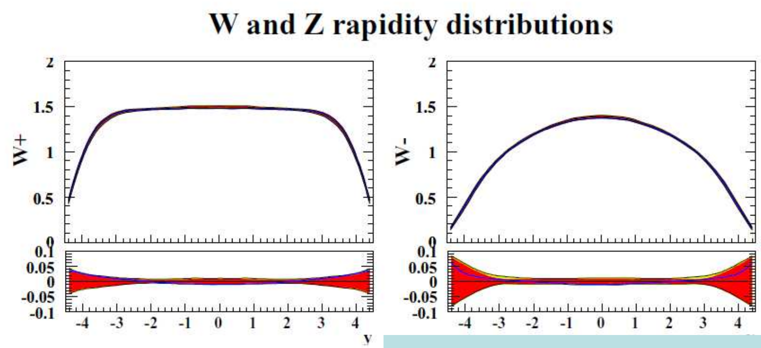
And now we have much better HERA data from the H1/ZEUS combination

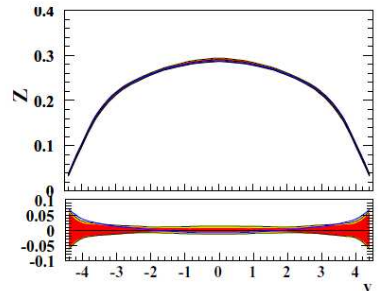
Use the HERAPDF to predict W and Z rapidity distributions at the LHC





W/Z rapidity distributions predicted from a PDFs extracted from ZEUS+ H1 separate data sets: pre-2009





W/Z rapidity distributions predicted from HERAPDF1.0 fit to HERA combined data:2009

Maybe able to use W and Z cross-sections as ‘standard candles’ for luminosity measurement

Combination of ZEUS and H1 data is ongoing

1. Data on multi-leptons at high-pt -arxiv:0907:3627 BSM constraints
2. Isolated leptons and missing pt:-arxiv:0911.0858 BSM constraints
3. Inclusive cross-sections HERA-1 (1992-2000):arxiv:0911.0884 – improved constraints at low-x
4. Heavy flavour data: F2c – better understanding of correct heavy flavour treatment
5. Low energy runs – FL- 2007- preliminary- improved constraints at low Q2- transition to non-perturbative regime
6. Inclusive cross-sections HERA-II (2003-2007) - improved constraints at high-x
7. Jet cross-sections- improved constraints on gluon PDF and $\alpha_s(M_Z)$