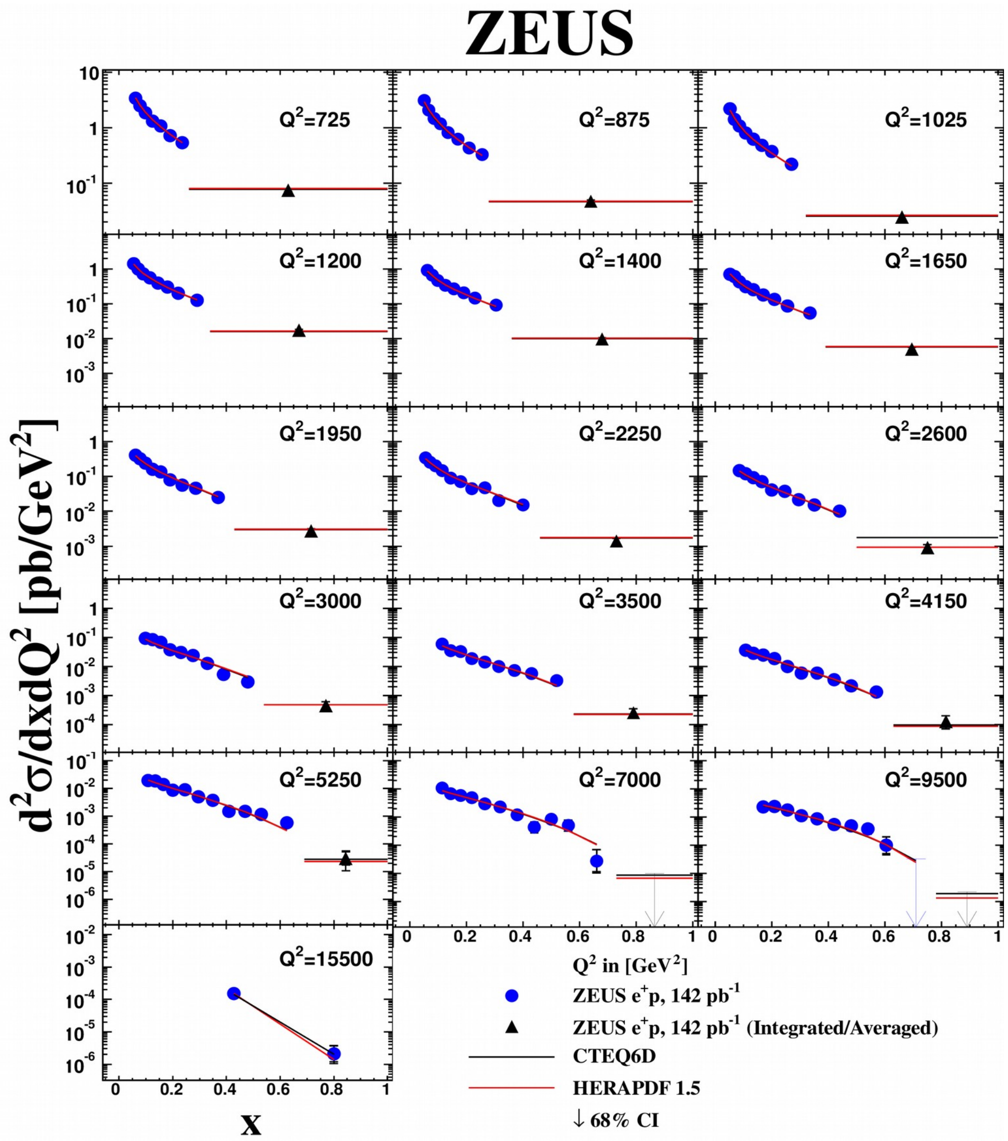
Proton parton densities at high-x

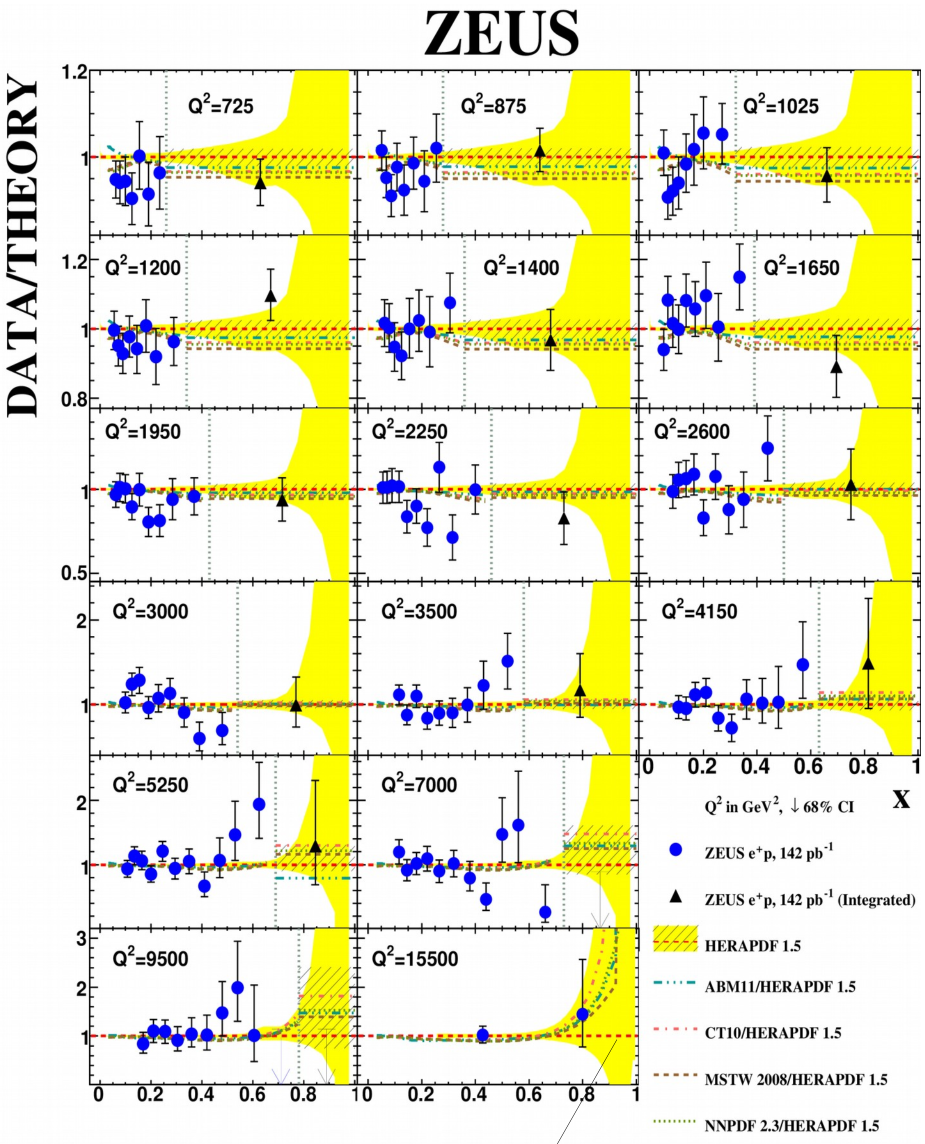
Ritu Aggarwal, Allen Caldwell

Overview

- Motivation
- Transfer Matrix for ZEUS detector for high-x data
- Using xfitter (instead of DISPRED) to calculate CTEQ5D cross sections for reweighting
- The differences
- Tables from the paper.
- Comparisons done by Katarzyna.
- Study on Radiative Corrections : Radiative corrections from CTEQ5D and HERAPDF2.0
- Checks on the integration process (using sample mean from random numbers)
- Data Sensitivity : change in normalization corresponding to 1 delta chi2.
- Status of paper

Motivation of studying published high-x data





At present x upto 0.65 ZEUS data is included in PDF fits

Note the uncertainity bands above x ~ 0.65, can high-x data impact here

ZEUS Collaboration; H. Abramowicz et al. Measurement of Neutral Current e $^{\pm}$p Cross-Sections at High Bjorken x with the ZEUS Detector Phys. Rev. D 89 (2014) 072007

Current Analysis : Extension of ZEUS high-x paper

Data & MC samples (same as high-x paper)
04-06 e-p data (187 pb -1) & 06/07 e+p data (142 pb^{-1})
DJANGOH 1.6, Ariadne 4.12, CTEQ-5D MCs

Using a combination of Ariadne and MEPS MC to get best representation of data.
(same as high-x paper)

Also included high-x specific samples

Generated and preserved by Katarzyna, funnelled and reprocessed by Andrii
(Q2 > 4000, 10000, 20000 with x > 0.1, > 0.5)

Selection Cuts :
Please refer backup for details (same as in high-x paper)

Other Inputs to MC :
(termed as simulation weights in further presentation : w^{SM}_{MC})

- Calibrations
- Track Matching Efficiency
- Track Veto inefficiency
- Zvtx Reweighting

(same as in high-x paper)

High-x data is still not used..

1) Some of the bins have low number of events / few have zero, so poisson errors are quoted.

2) Ofcourse it has a subset of data (high-Q2 ZEUS data) already included in fits, but high-x data has more to say.

Transfer Matrix for the detector is developed using which number of events reconstructed in data can be predicted from any PDF as below.

- Get a prediction for the generator/hadron level number of events, which is luminosity x radiative corrections x Born cross section.

i.e. $$\nu_{i,k} = \mathcal{L} K_{ii} \sigma_{i,k}$$

- Apply transfer matrix a_{ij} to get the number of events in a bin j.

$$\nu_{j,k} \approx \sum_i a_{ij} \nu_{i,k}$$

L : data luminosity
K_{ii} : Radiative corrections (calculated using HERACLES)
$\sigma_{i,k}$: born level cross sections in i^{th} bin for k^{th} PDF
$\boldsymbol{a_{ij}}$ has all detector and analysis effects
(probability of an event reconstructed in j^{th} bin to come from i^{th} true bin)

Transfer Matrix : Probability of an event reconstructed in jth bin to come from ith true bin

Tracing back the path of MC reconstructed events in the generated x-Q^2 phase space

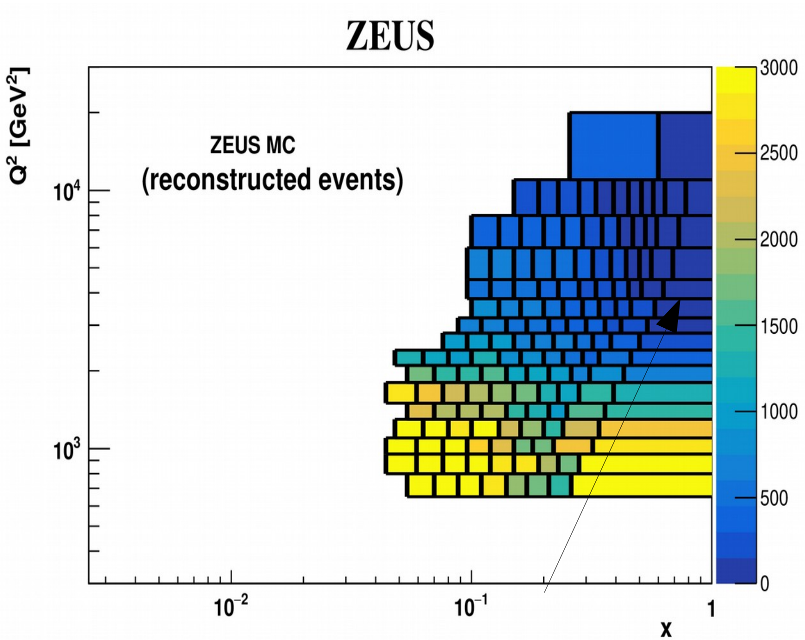
$$a_{ij} = \frac{\sum_{m=1}^{M_i} \omega_m I(m \in j)}{\sum_{m=1}^{M_i} \omega_m^{MC}}$$

a_{ij} = probability of an event reconstructed in jth bin to come from ith bin

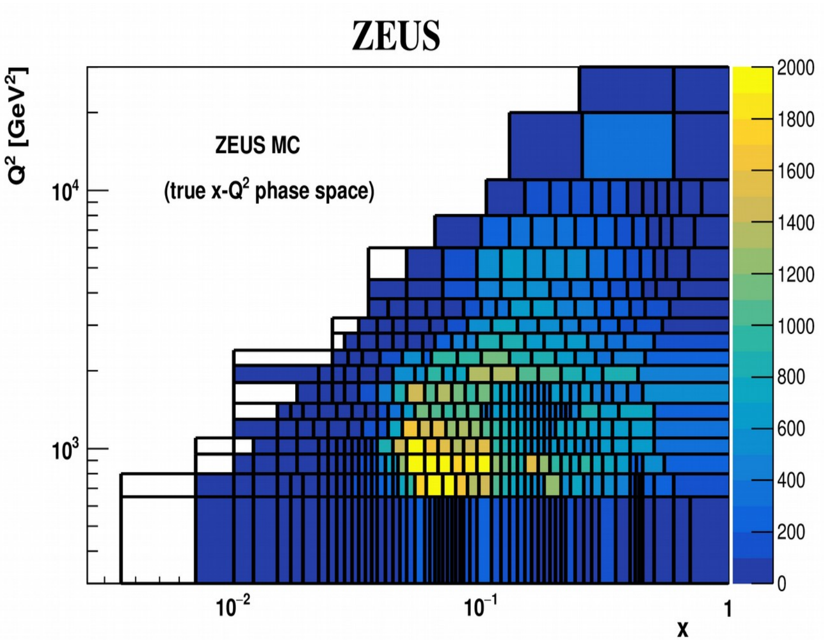
ω_m = MC weights given to mth event in bin i

I = 1 if mth event is reconstructed in bin j, else = 0

M_i = total events generated in ith bin



Reconstructed MC events in xsection binning ‘N’ (total 153 bins)



Generated distribution of these events in extended binning ‘M’ (total 429 bins)

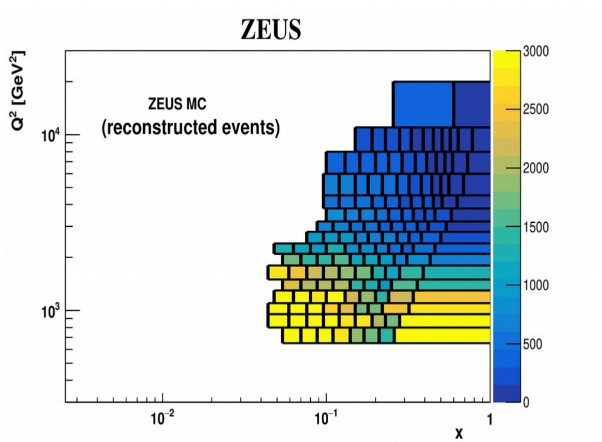
Note : MC samples used as in high-x paper.

Using Transfer matrix to predict no. of events reconstructed in a given cross section bin

N = T M

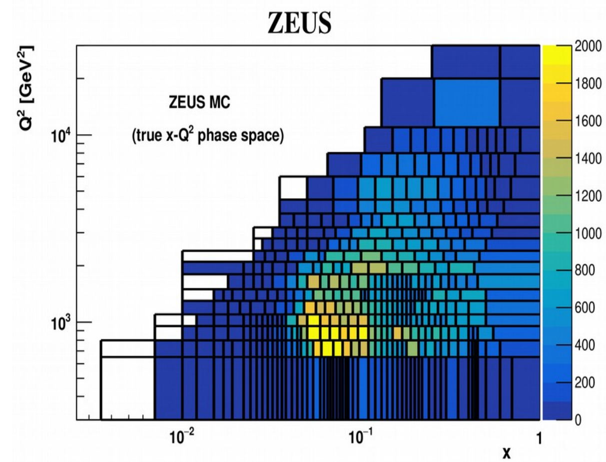
Transfer Matrix

(153 X 429 elements)



Predicted x-Q2 events in Cross section binning

(153 elements in N Vector = number of cross section bins)



Generated x-Q2 events in Extended binning

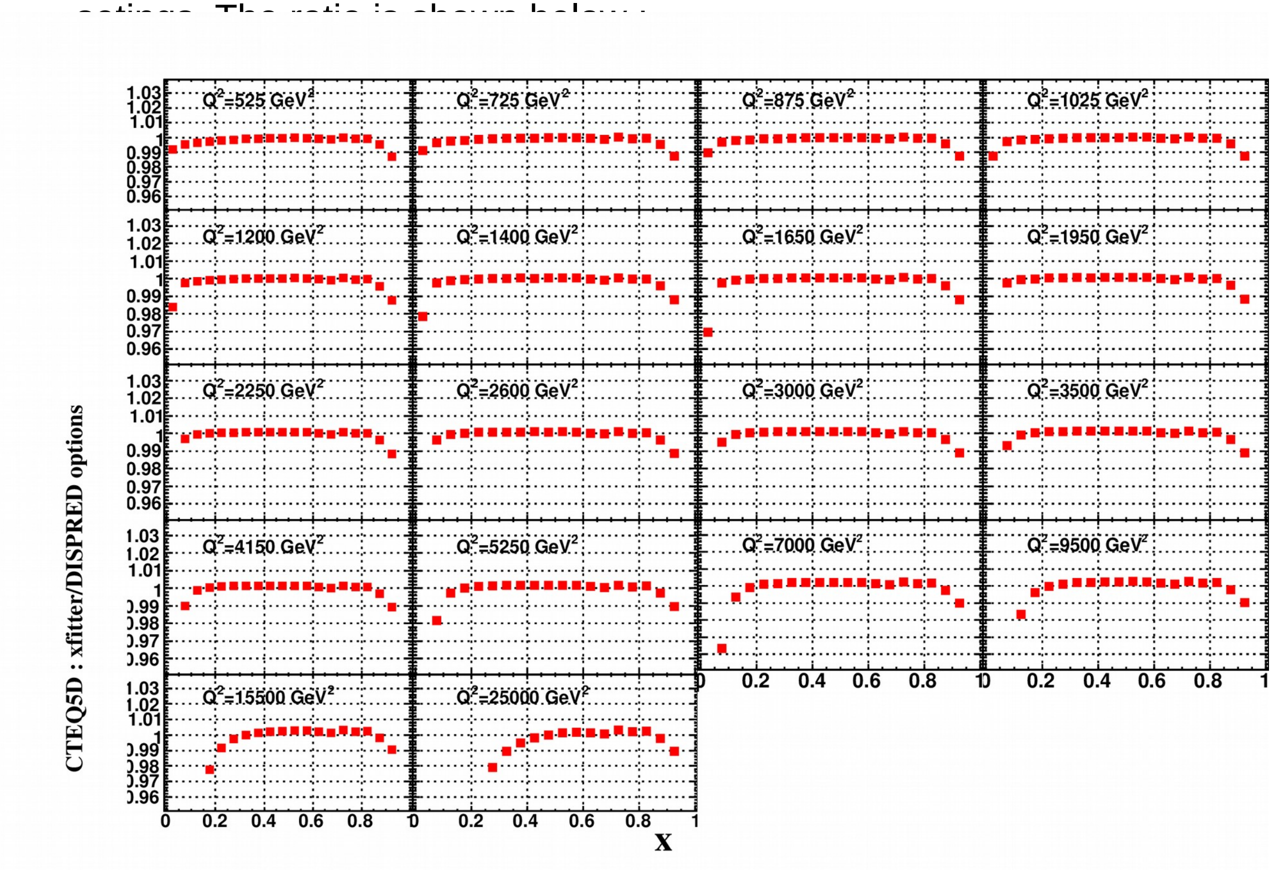
(429 elements in M Vector = number of generated bins)

M from Different PDFs

M from different PDFs can be obtained by reweighting the events to the new PDF by taking ratio of cross section from the new PDF to that of from CTEQ5D
--> CTEQ5D cross sections were obatined from DISPred (it can use LHAPDF5's CTEQ5D)
--> Other PDFs weights were calculated using xfitter (New PDFs are defined in LHAPDF6)

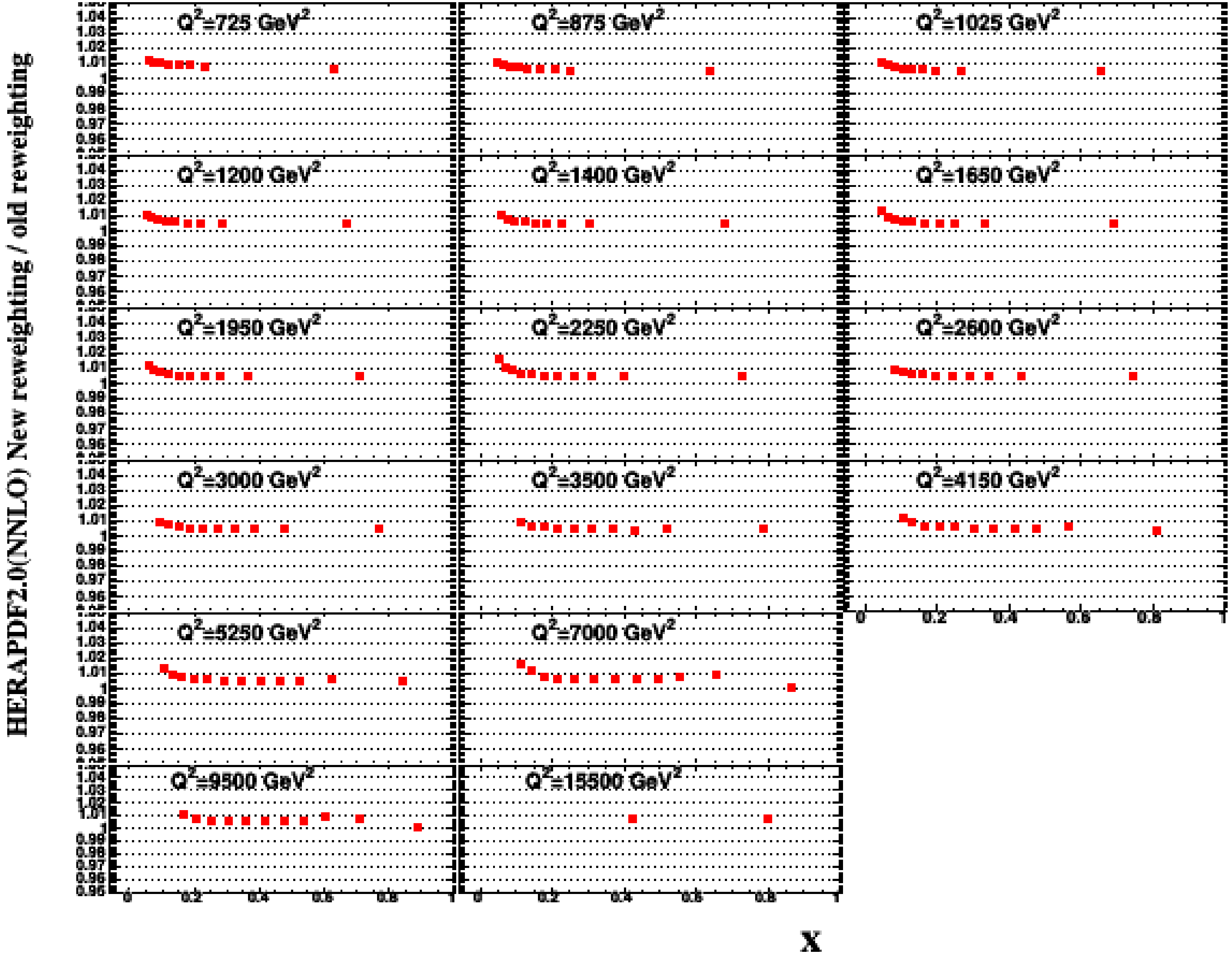
Problem : Dispred had different settings than used in xfitter, Like : ZVMFNS in DISpred and RT-OPT in xfitter (definition of variables)

Check done: CTEQ5D cross sections were calculated in using xfitter, with DISPRED settings and usual xfitter settings. The ratio is shown below :



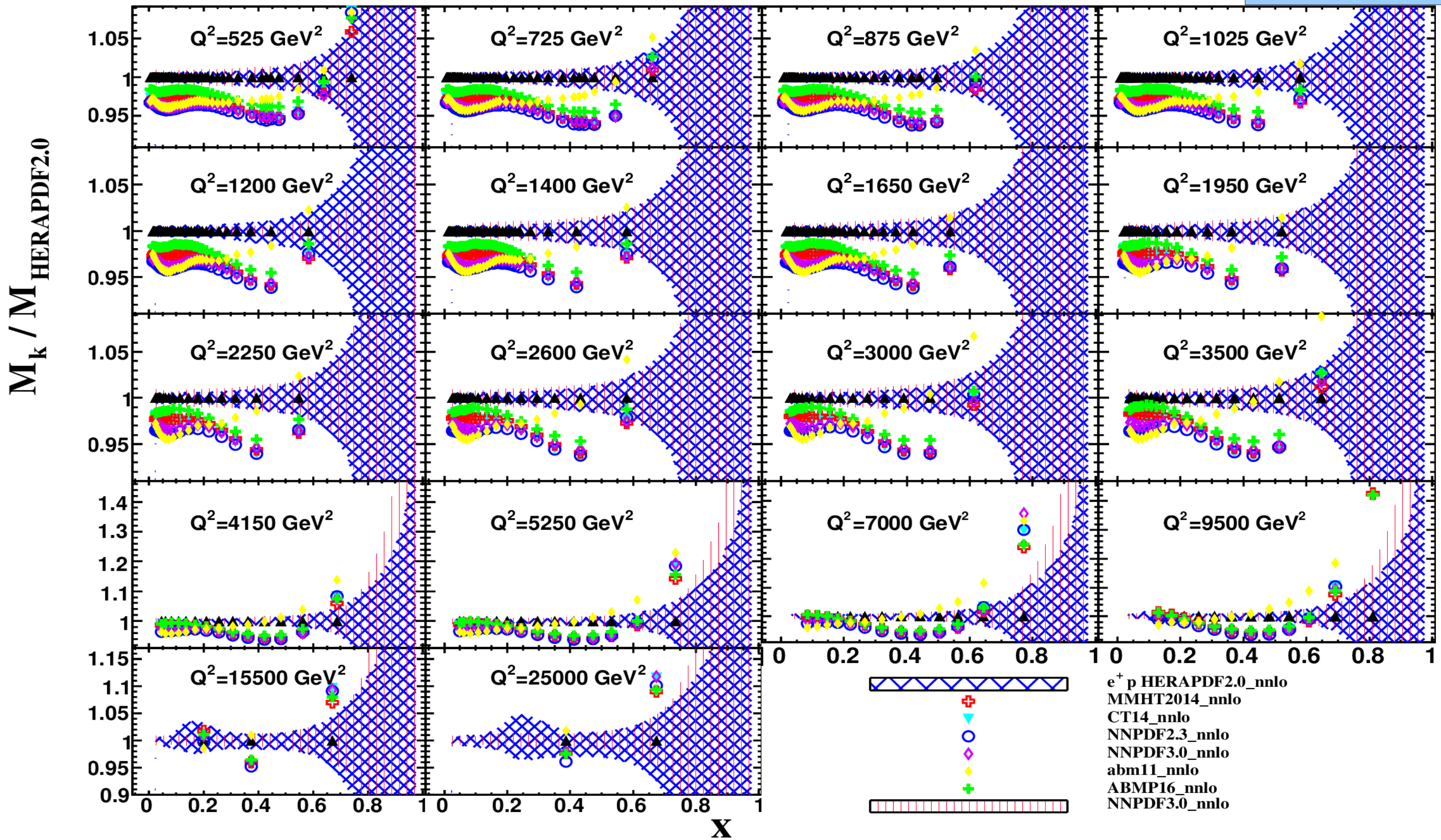
Weights are calculated again, analysis redone and all the plots and the numbers will be updated.

N (HERAPDF2.0 NNLO) from new reweighting vs. the old reweighting



Average ratio of Born level cross sections in different PDFs to HERAPDF2.0NNLO for M bins (e+p)

New reweighting



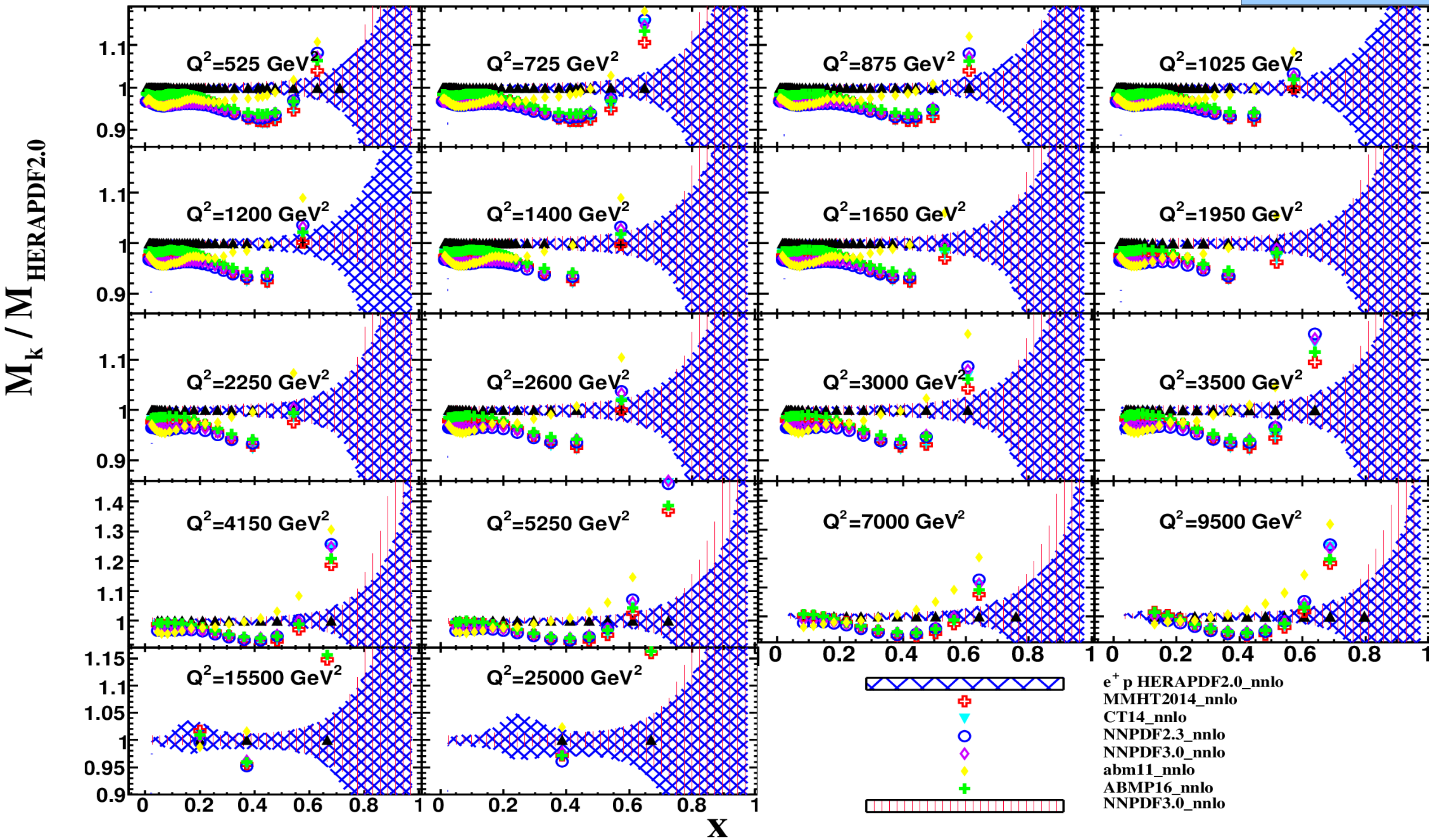
Another Issue : bin boundaries are defined by the true variables from exchanged boson information

PDF weights are given wrt the true variables from lepton information (with QED radiation, these could give wrong results).

Using the exchanged boson information for PDF reweighting (updated plots in page 12,13)

Average ratio of Born level cross sections in different PDFs to HERAPDF2.0NNLO for M bins (e+p)

Exhanged Boson Info

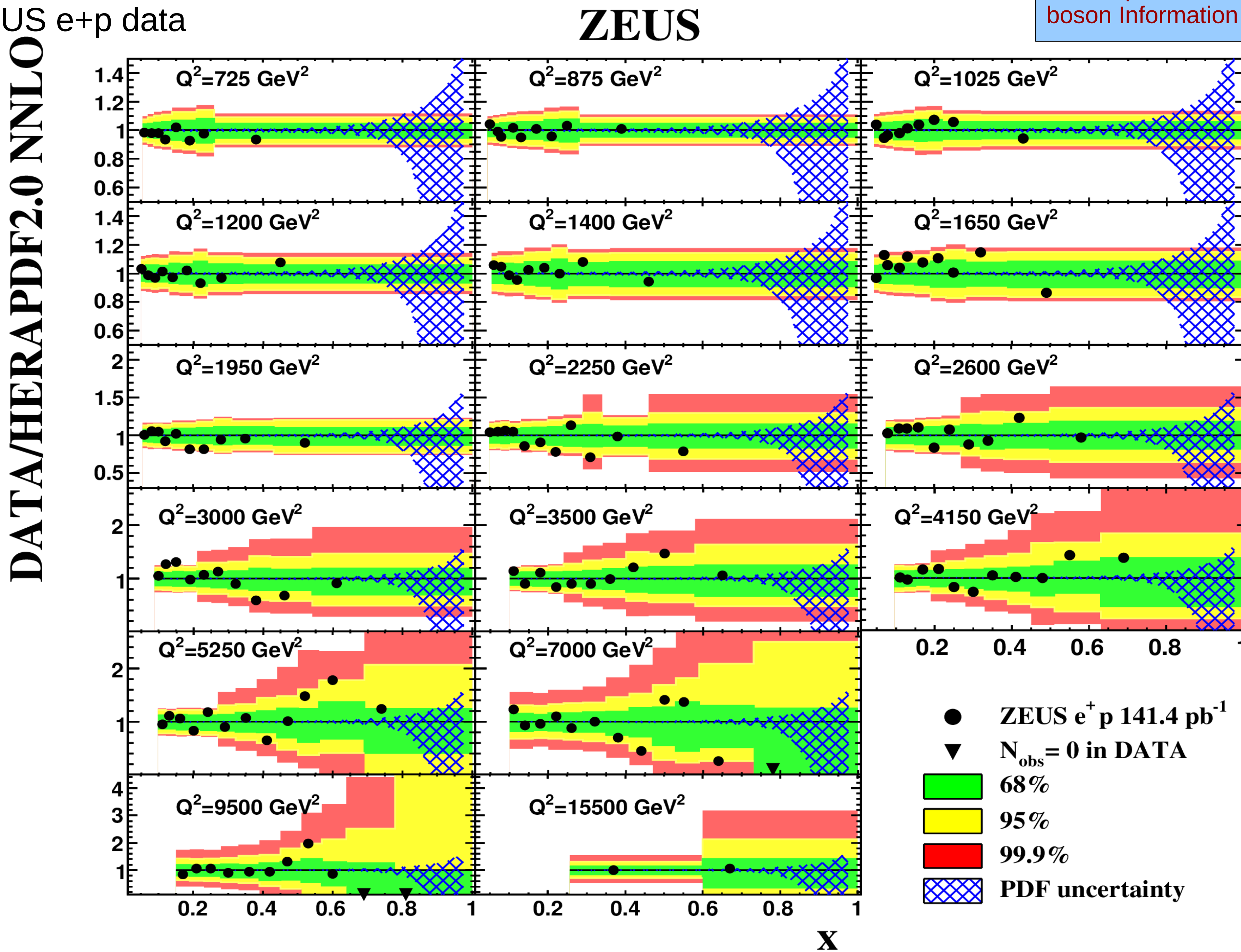


Using the exchanged boson information for PDF reweighting
(similar plot will be produced for e-p, not shown here)

Ratio of No. of events in data to HERAPDF2.0 NLO and 1,2,3 sigma bands from Poisson Statistics

To be updated Exchanged boson Information

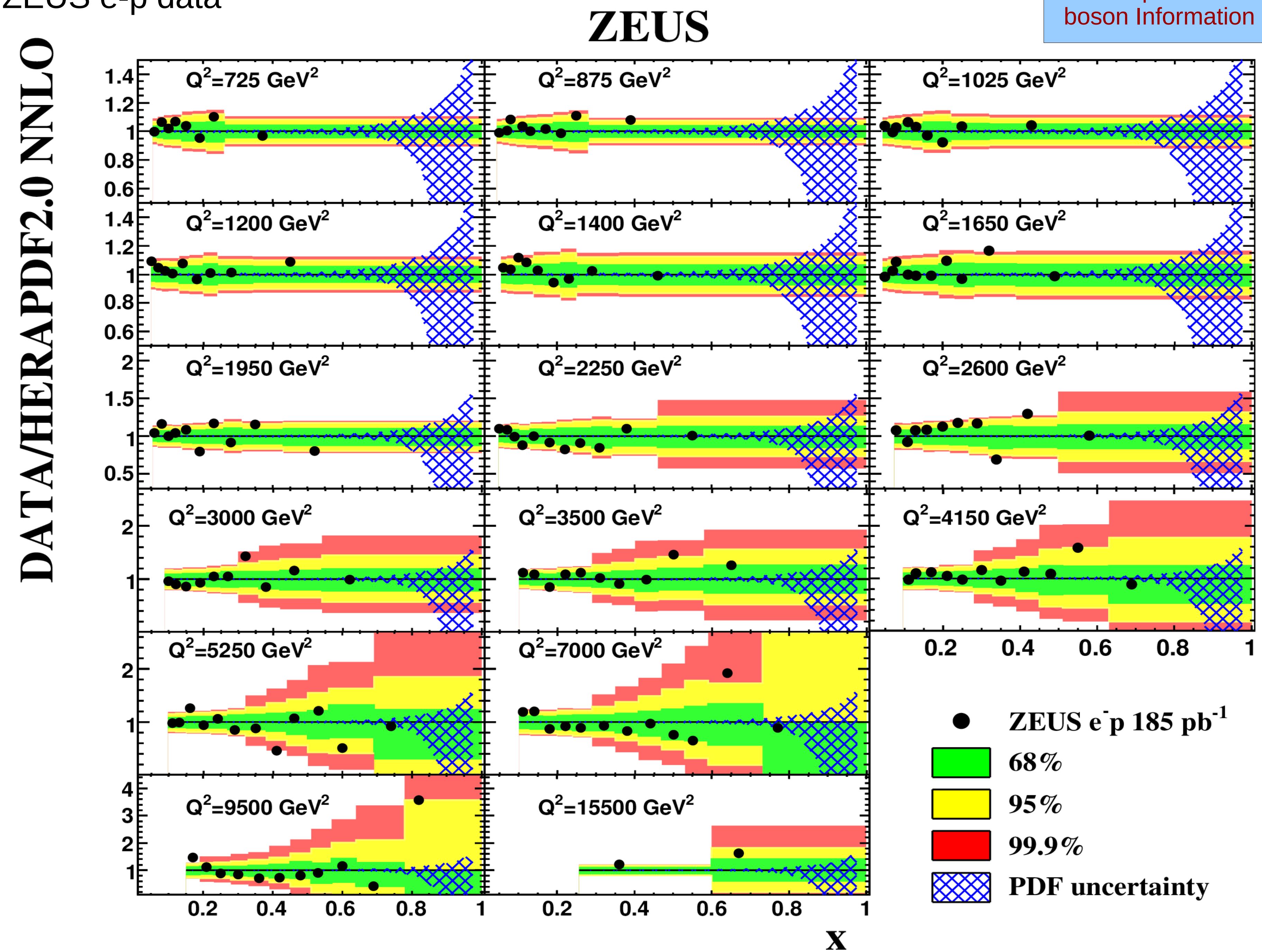
ZEUS e+p data



Ratio of No. of events in data to HERAPDF2.0 NLO and 1,2,3 sigma bands from Poisson Statistics

ZEUS e-p data

To be updated Exchanged boson Information



Probability for explaining data from different PDFs

Total probability for each PDF :

$$P(D|M_k) = \prod_j \frac{e^{-\nu_{j,k}}\nu_{j,k}^{n_j}}{n_j!}$$

n_j = events in data in j^{th} bin
k : k^{th} PDF index

Calculating the **relative Probablity wrt. HERAPDF**

	e^-p			e^+p		
PDF	P1/P2	p-value	$\Delta\chi^2$	P1/P2	p-value	$\Delta\chi^2$
$CT14$	$2.0*10^{-04}$	0.001589	17.01	31761.18	0.79	-20.73
$HERAPDF2.0$	1.0	0.041332	0.00	1.00	0.45	0.00
$MMHT2014$	$2.2*10^{-04}$	0.001943	16.88	31132.26	0.80	-20.69
$NNPDF2.3$	$9.8*10^{-09}$	0.000076	36.88	2440.60	0.67	-15.60
$NNPDF3.0$	$1.7*10^{-07}$	0.000236	31.22	5558.04	0.71	-17.25
$ABMP16$	$7.1*10^{-02}$	0.013981	5.29	11350.31	0.78	-18.67
$abm11$	$7.5*10^{-07}$	0.000756	28.21	43.08	0.57	-7.53

Table 1: *The Bayes Factor, p-values and $\Delta\chi^2$ from comparisons of predictions at NNLO using different PDF sets. The Bayes Factor is calculated relative to HERAPDF2.0, as is $\Delta\chi^2$. The results are shown separately for the e^-p and e^+p data sets.*

Eg. of P-value determination

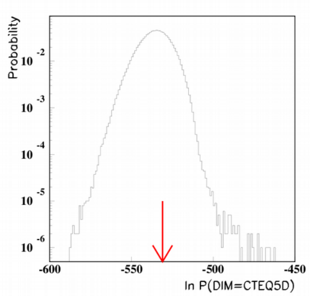


FIGURE 4. Distribution of expected values for $\ln P(D|M = CTEQ)$ for the e^+p data set. The arrow shows the value found in the data.

P-value is calculated by integrating out the probability from the left edge till red for the given PDF

Equivalent Delta chi2 determination

$$\Delta\chi^2_{k,l} = -2\ln\frac{P(D|M_k)}{P(D|M_l)} = -2\left(\sum_j \nu_{j,l} - \nu_{j,k} + n_j \cdot \ln\frac{\nu_{j,k}}{\nu_{j,l}}\right)$$

MMHT2014, CT14nlo, NNPDF2.3, ABM better than HERAPDF2.0 for e^+P, much worse for e^-P.

Comparing Total Probability for different Pdfs in different x range (integrated bins +2 preceding x bins in each Q2)

	e^-p				e^+p			
	low x		high x		low x		high x	
PDF	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$
$CT14$	$1.16 * 10^{-02}$	8.92	0.02	8.09	247.15	-11.02	128.64	-9.71
$HERAPDF2.0$	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
$MMHT2014$	$7.28 * 10^{-03}$	9.84	0.03	7.03	240.33	-10.96	129.67	-9.73
$NNPDF2.3$	$7.46 * 10^{-07}$	28.22	0.01	8.66	17.64	-5.74	138.52	-9.86
$NNPDF3.0$	$5.48 * 10^{-06}$	24.23	0.03	6.99	42.95	-7.52	129.54	-9.73
$ABMP16$	$4.67 * 10^{-01}$	1.52	0.15	3.76	166.50	-10.23	68.24	-8.45
$abm11$	$3.47 * 10^{-07}$	29.75	2.16	-1.54	2.18	-1.56	19.79	-5.97

Table 10: *The Bayes Factor, p-values and $\Delta\chi^2$ from comparisons of predictions (at NNLO) using different PDF sets to the observed numbers of events are shown. The p-values are given for two different x ranges for the e^+p and e^-p data sets.*

At high x MMHT, CT, NNPDF ABM better for e+P data. disagreement comes primarily from lower x in e-p

Statistical and systematic uncertainties

Type of Systematic Uncertainties :

1) Affecting the predictions at generator level (M values)
2) Affecting the Transfer Matrix T

Type I :
1) Luminosity uncertainty scaling M values

Type II :
1) MC statitical fluctuations (uncorrelated uncertainty)
2) All correlated and uncorrelated systematic uncertainties as in high-x paper
3) Choice of PDF for building T

Nomalization Error : Vary M by 1.8 % up and down and calculate ln P.

	e^-p				e^+p			
PDF	$x < 0.6$		$x \geq 0.6$		$x < 0.6$		$x \geq 0.6$	
	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$
$HERAPDF2.0$	0.003	11.76	0.32	2.26	$3.41 * 10^{-05}$	20.57	0.06	5.62
$CT14$	0.542	1.23	1.64	−0.98	$9.39 * 10^{+01}$	−9.08	22.04	−6.19
$MMHT2014$	0.487	1.44	1.68	−1.03	$7.65 * 10^{+01}$	−8.67	19.69	−5.96
$NNPDF2.3$	0.006	10.24	1.50	−0.81	$3.66 * 10^{+02}$	−11.80	25.74	−6.50
$NNPDF3.0$	0.020	7.86	1.77	−1.14	$2.74 * 10^{+02}$	−11.22	18.80	−5.87
$ABMP16$	1.326	−0.56	1.91	−1.30	$3.72 * 10^{+00}$	−2.63	6.17	−3.64
$ABM11$	0.013	8.75	0.88	0.26	$3.72 * 10^{+01}$	−7.23	0.47	1.51
−1.8 %								
$HERAPDF2.0$	$1.35 * 10^{-01}$	4.00	1.35	−0.60	102.21	−9.25	8.84	−4.36
$CT14$	$1.06 * 10^{-10}$	45.94	0.09	4.75	8.36	−4.25	129.09	−9.72
$MMHT2014$	$1.10 * 10^{-10}$	45.87	0.11	4.40	10.83	−4.76	130.87	−9.75
$NNPDF2.3$	$3.42 * 10^{-17}$	75.83	0.07	5.23	0.01	8.94	130.80	−9.75
$NNPDF3.0$	$1.57 * 10^{-15}$	68.17	0.12	4.32	0.11	4.37	123.38	−9.63
$ABMP16$	$1.10 * 10^{-06}$	27.44	0.36	2.06	131.37	−9.76	89.28	−8.98
$ABM11$	$1.22 * 10^{-14}$	64.07	0.90	0.21	0.01	9.09	22.74	−6.25

Table 5: *The results from comparisons of predictions using different PDF sets increased by 1.8 % (top) and decreased by 1.8 % (bottom) to the observed numbers of events. The p-values are given for two different x ranges for the e^-p and e^+p data sets.*

Conclusions :

- **p-values from different PDFs change differently while moving up or down by 1.8%**

Major Systematic Errors : New a_ij according to systematic variation up and down.

- **Electron Energy Scale : varied by 0.5%**
- **Electron energy resolution : varying the smearing factor by 10%**
- **Jet Energy : varied by 1%**
- **Jet X-Projection on FCAL varied by 5 mm**
- **Jet Y-Projection on FCAL varied by 5 mm**
- **Isolation cut varied by 2 GeV**
- **Ariadne-MEPS combination varied (0.3+-0.3)**
- **The FCAL-BCAL Crack cut on electron angle varied by 0.015rad**

Including the systematic uncertainty :

1) Re-evaluate the Transfer Matrix with the given systematic check
2) Calculate the new predicttion to the data
3) Calculate the new probability from the prediction
4) Evaluate Bayes factor and chi square wrt to the nominal MC

Statistical Uncertainty calculated using bionomial errors and found to be very small (with in 1%) when the high-x Specific MC is included

$$\delta a_{ij}^{\text{stat}} = \sqrt{\frac{a_{ij}(1-a_{ij})}{M_i}}$$

Where Mi are the total number of events Generated in MC

Major Systematic Errors : New a_ij according to systematic variation up and down.

	e^-p		e^+p	
Systematic	$P1/P2$	$\Delta\chi^2$	$P1/P2$	$\Delta\chi^2$
Electron energy scale	0.469 0.094	1.51 4.72	0.100 0.908	4.61 0.19
Electron energy resolution	0.230 1.283	2.94 −0.50	2.059 0.190	−1.44 3.32
Hadronic energy scale	0.739 1.226	0.61 −0.41	0.685 1.146	0.76 −0.27
MEPS/Ariadne reweighting	3.861 0.108	−2.70 4.45	1.025 0.621	−0.05 0.95
F-BCal Crack cut	0.223 4.637	3.00 −3.07	1.089 0.860	−0.17 0.30
Electron isolation cut	1.190 0.133	−0.35 4.04	0.222 14.426	3.01 −5.34
FCAL alignment (x)	1.425 0.947	−0.71 0.11	1.003 0.960	−0.01 0.08
FCAL alignment (y)	1.281 0.866	−0.50 0.29	1.584 0.616	−0.92 0.97

Normalization is the main uncertainty

Systematic Errors : Considering various vectors for HERAPDF2.0

	e^-p				e^+p			
	$x < 0.6$		$x \geq 0.6$		$x < 0.6$		$x \geq 0.6$	
Eigen Vector	P1/P2	$-\Delta\chi^2$	P1/P2	$-\Delta\chi^2$	P1/P2	$-\Delta\chi^2$	P1/P2	$-\Delta\chi^2$
0	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
1	0.97	−0.06	1.00	−0.00	0.75	−0.59	1.00	−0.00
2	1.01	0.03	1.00	−0.00	1.33	0.57	1.00	0.00
3	1.11	0.20	1.09	0.18	1.95	1.33	1.24	0.43
4	0.85	−0.32	0.91	−0.19	0.49	−1.41	0.80	−0.44
5	1.22	0.39	1.23	0.42	1.47	0.77	1.66	1.01
6	0.79	−0.47	0.79	−0.47	0.66	−0.82	0.59	−1.05
7	0.99	−0.02	0.91	−0.19	1.94	1.33	0.81	−0.42
8	0.92	−0.17	1.09	0.18	0.48	−1.45	1.23	0.42
9	0.73	−0.63	0.75	−0.58	0.33	−2.20	0.53	−1.27
10	1.21	0.39	1.28	0.50	2.78	2.04	1.83	1.20
11	0.78	−0.49	0.87	−0.29	0.47	−1.49	0.71	−0.68
12	1.21	0.38	1.14	0.26	1.97	1.36	1.39	0.66
13	1.05	0.10	0.67	−0.80	1.26	0.46	0.40	−1.82
14	0.92	−0.18	1.36	0.62	0.78	−0.49	2.32	1.68
15	1.18	0.33	1.06	0.12	1.51	0.82	1.15	0.28
16	0.83	−0.37	0.94	−0.13	0.70	−0.70	0.86	−0.30
17	1.12	0.23	1.03	0.06	1.23	0.41	1.05	0.11
18	0.88	−0.26	0.96	−0.08	0.76	−0.54	0.92	−0.18
19	1.18	0.32	1.11	0.22	1.99	1.38	1.22	0.41
20	0.81	−0.43	0.89	−0.23	0.52	−1.31	0.79	−0.46
21	0.67	−0.80	0.83	−0.37	0.34	−2.17	0.60	−1.01
22	1.21	0.38	1.10	0.19	2.06	1.44	1.37	0.63
23	0.89	−0.24	1.06	0.13	1.29	0.50	1.07	0.13
24	0.89	−0.23	0.80	−0.44	0.57	−1.14	0.65	−0.86
25	1.41	0.69	1.15	0.29	1.32	0.56	1.36	0.62
26	0.77	−0.53	1.10	0.19	0.89	−0.22	1.30	0.52
27	1.23	0.41	1.31	0.54	4.55	3.03	1.86	1.24
28	0.87	−0.27	0.79	−0.47	0.53	−1.25	0.62	−0.97

Table 3: *The results from the different Eigen Vector variants in HERAPDF2.0. The Bayes Factor and respective $\Delta\chi^2$ values are given for the e^-p and e^+p data sets. The first Eigenvector (labeled 0) is the nominal prediction.*

Variance vectors

	e^-p				e^+p			
	$x < 0.6$		$x \geq 0.6$		$x < 0.6$		$x \geq 0.6$	
Eigen Vector	P1/P2	$-\Delta\chi^2$	P1/P2	$-\Delta\chi^2$	P1/P2	$-\Delta\chi^2$	P1/P2	$-\Delta\chi^2$
0	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
1	1.02	0.04	1.04	0.08	1.02	0.04	1.10	0.19
2	0.97	−0.05	0.96	−0.08	0.97	−0.06	0.91	−0.18
3	1.02	0.04	1.05	0.10	1.03	0.06	1.13	0.24
4	0.98	−0.04	0.97	−0.06	0.97	−0.05	0.93	−0.14
5	0.94	−0.12	0.94	−0.13	0.63	−0.91	0.85	−0.32
6	1.14	0.26	1.13	0.24	1.89	1.28	1.34	0.58
7	0.99	−0.02	0.98	−0.03	0.93	−0.16	0.96	−0.07
8	1.01	0.02	1.01	0.01	1.05	0.11	1.02	0.03
9	1.04	0.07	1.02	0.03	1.10	0.20	1.04	0.08
10	0.96	−0.07	0.98	−0.05	0.88	−0.24	0.95	−0.11
11	0.93	−0.14	1.04	0.08	0.98	−0.05	1.07	0.13
12	1.03	0.06	0.96	−0.09	0.94	−0.13	0.93	−0.15
13	1.09	0.18	1.95	1.34	1.61	0.95	3.00	2.20

Table 4: *The results from the different Variance Vector variants in HERAPDF2.0 The Bayes factor and respective $\Delta\chi^2$ and respective p-values are given for the e^-p and e^+p data sets. The first Eigenvector (labeled 0) is the nominal prediction.*

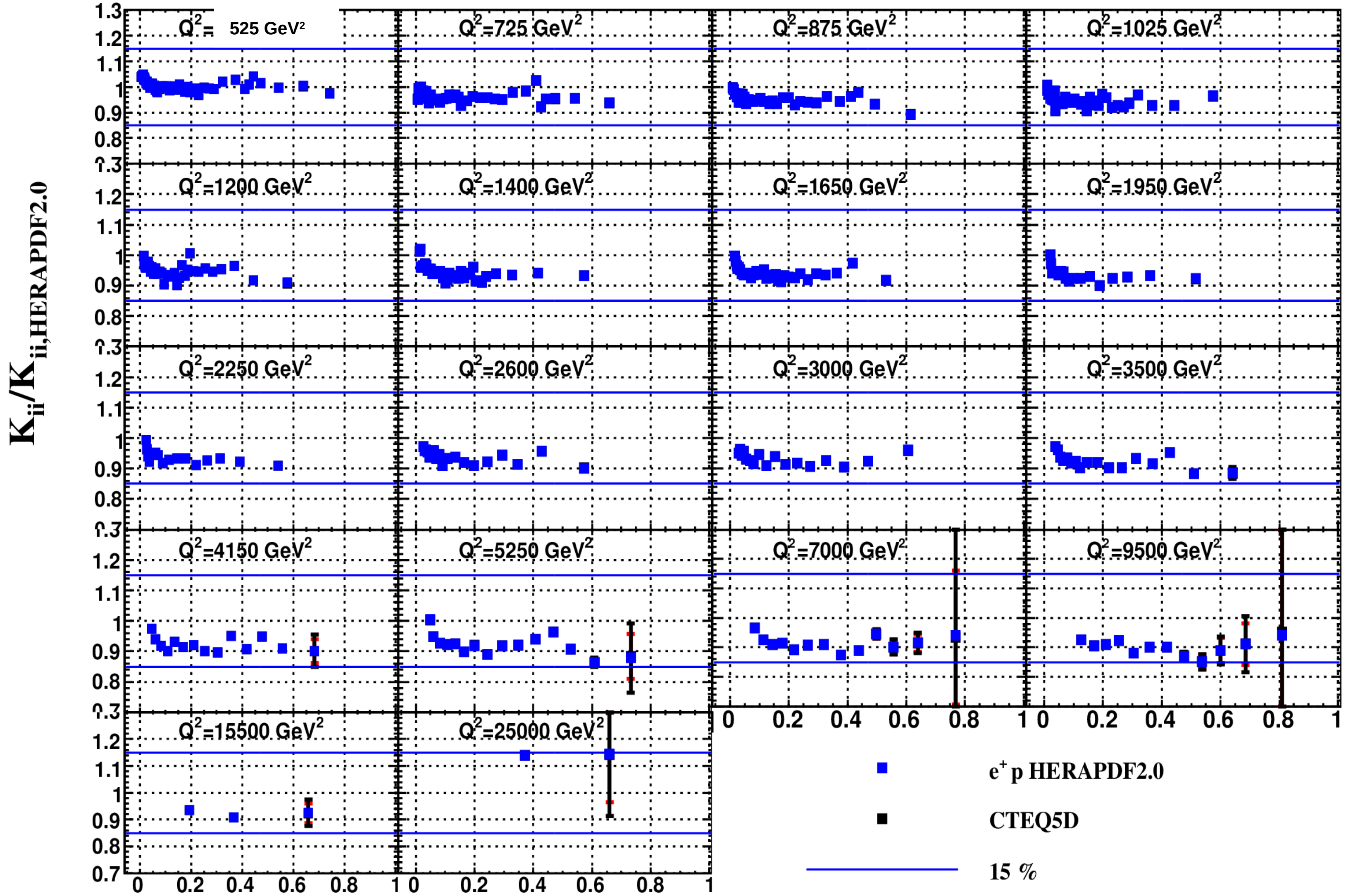
Radiative Corrections (2 studies done)

- **Ratio of M (high-x, with Radiative Corrections) and L*σ (Mandy : without radiative corrections)**

  (here running alpha_em was used by Mandy)

- **Ratio of events in x,Q2 bins generated using RAPGAP with and without radiative corrections for different PDFs (smaples produced by Andrii)**

Ratio of M_k (high-x, with Radiative Corrections) and $L*\sigma_k$ (xfitter : without radiative corrections, alpha = 1/137.)

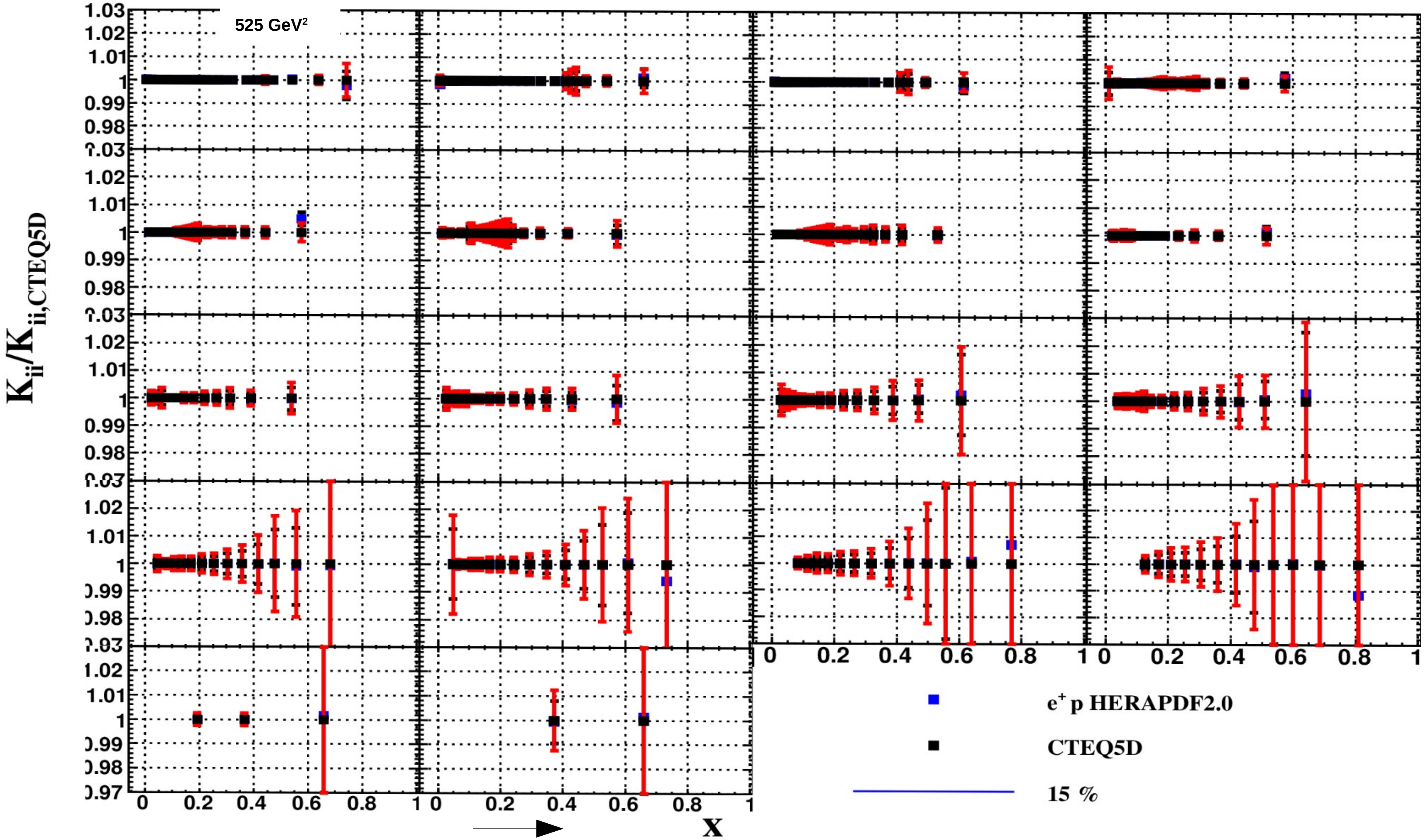


For each PDF, sigma is calculated as :

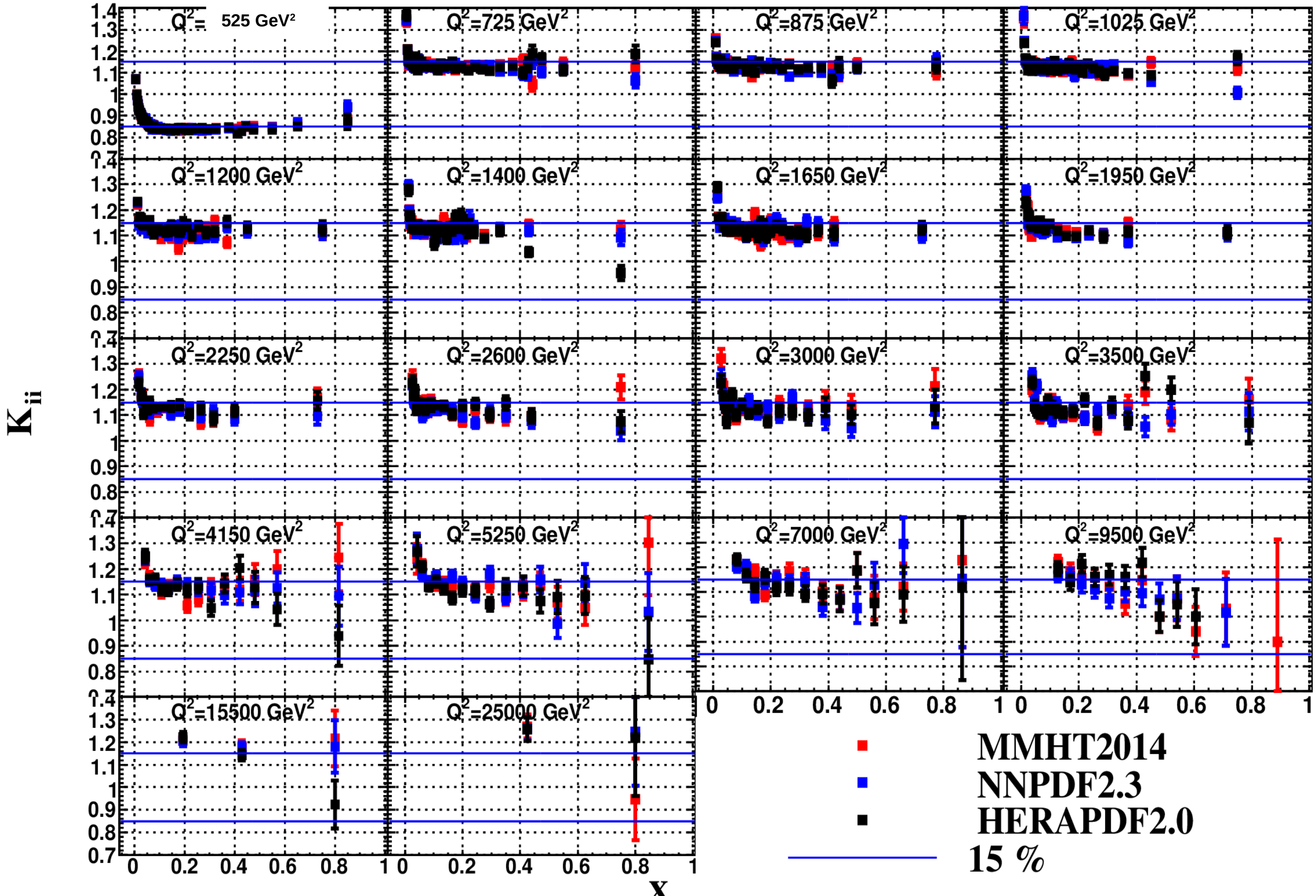
-- dividing each x,Q2 bin into 200x200 small bins. Integrating the double differential cross section in these bins

.-- applying conversion factors : reduced to double differential (where alpha =1/137.), delta x * delta Q2, natural units to pb

Ratio of Kii HERAPDF2.0/ Kii CTEQ5D

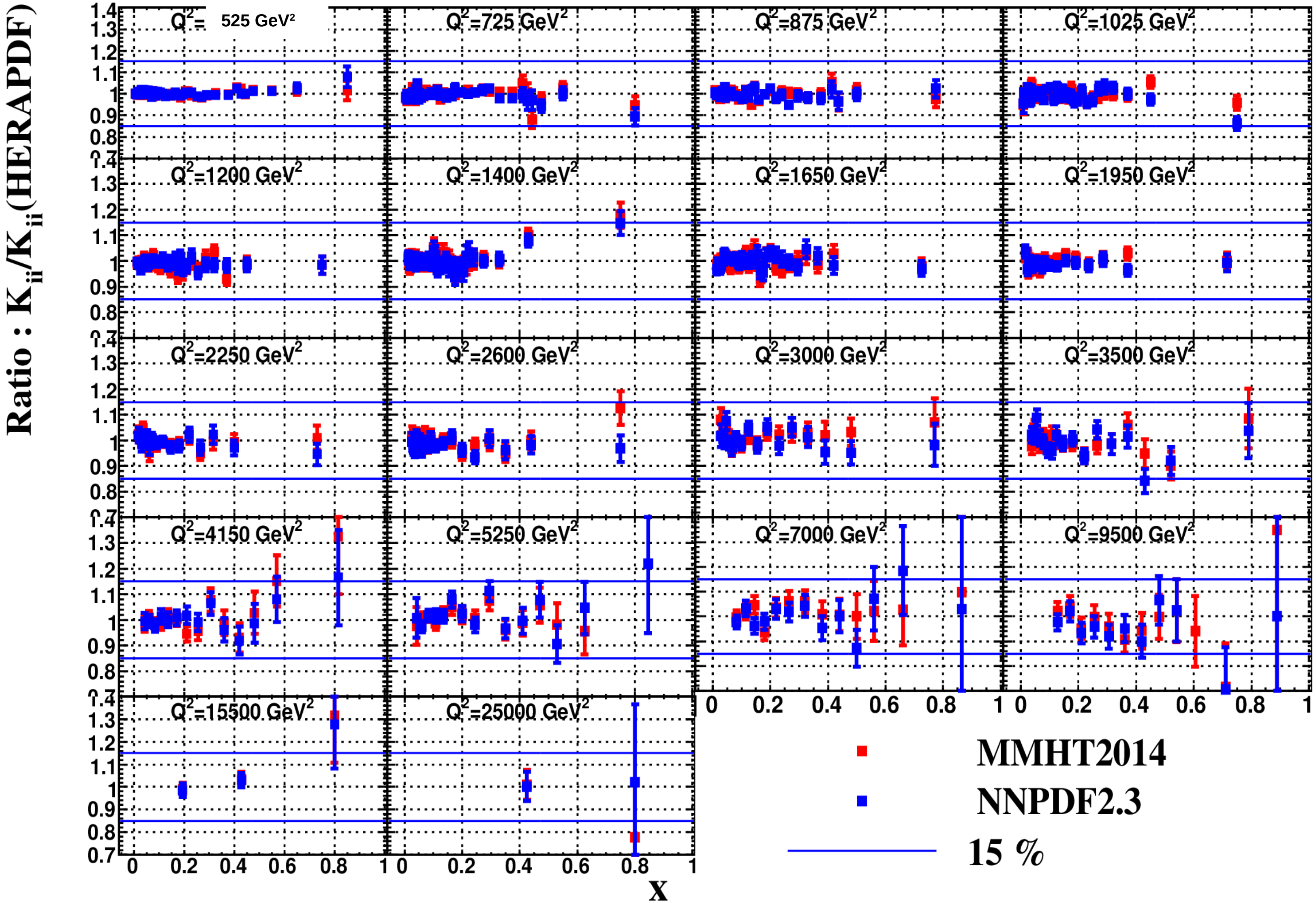


Ratio of generated events with and without radiative corrections for different PDFs using RAPGAP



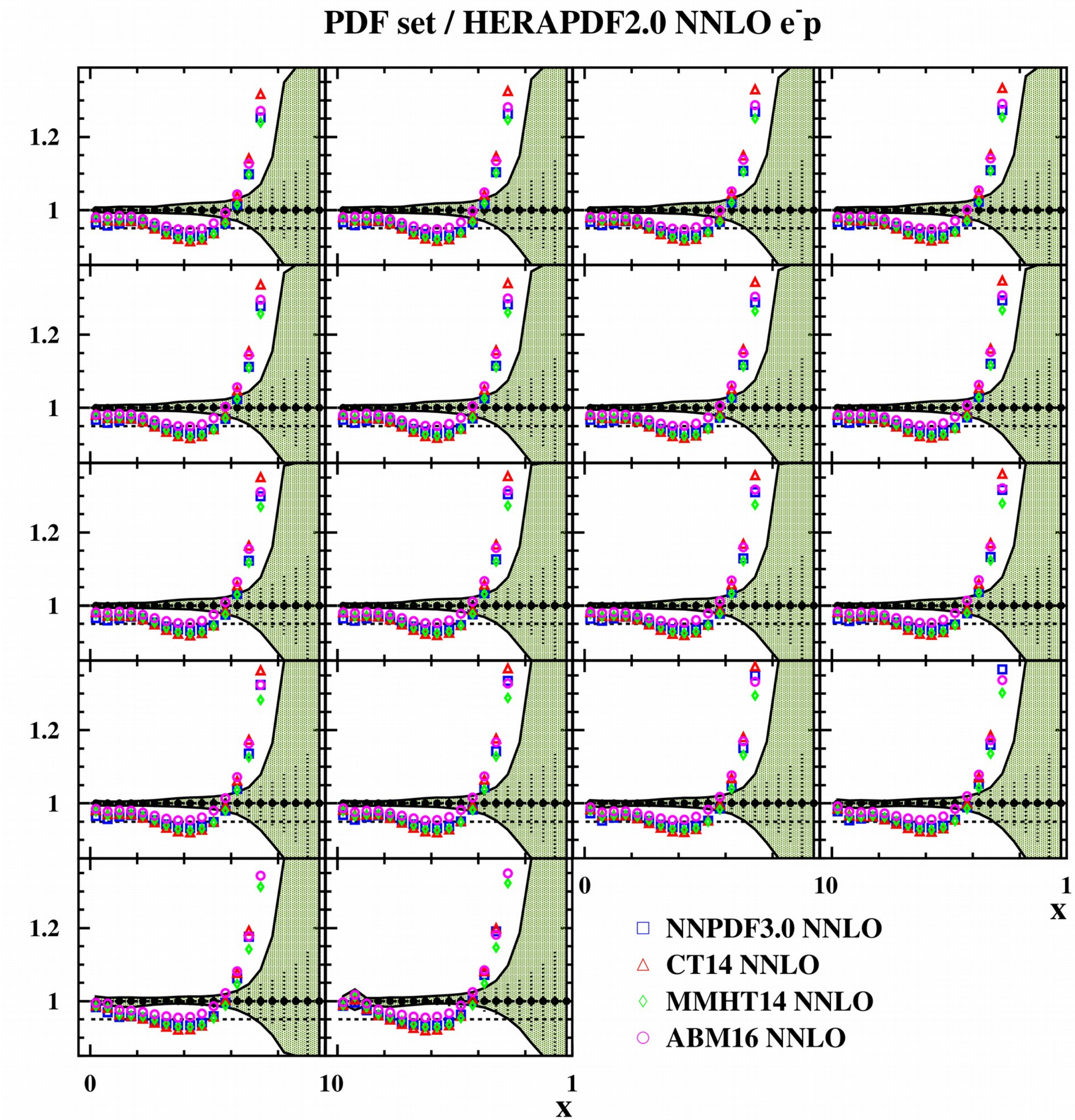
The shape at lowest Q2 and at low x at each Q2 bin can be reproduced using just Radiative corrections (samples from Andrii).

Ratio of Kii_k to $Kii,_{HERAPDF2.0}$



Also as the samples are generated using RAPGAP,
the ratio is ~1 even at the high-x bins

Cross Checks by Katarzyna



Recalculated the PDF uncertainty as parametrization is to be taken as an envelope and not to be added in quadrature

xfitter environment is good enough to be used for the analysis

actual numbers that are used for getting number of events on figures 2-4 were cross checked, (random samples)

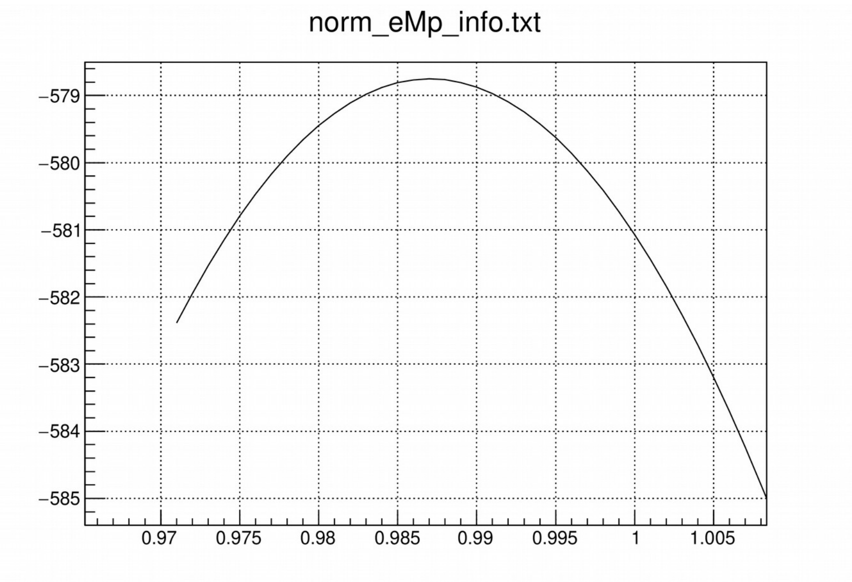
different PDFs differ more that PDF uncertainty allows

(As demsonstrated in this plot prouced by Katarzyna,

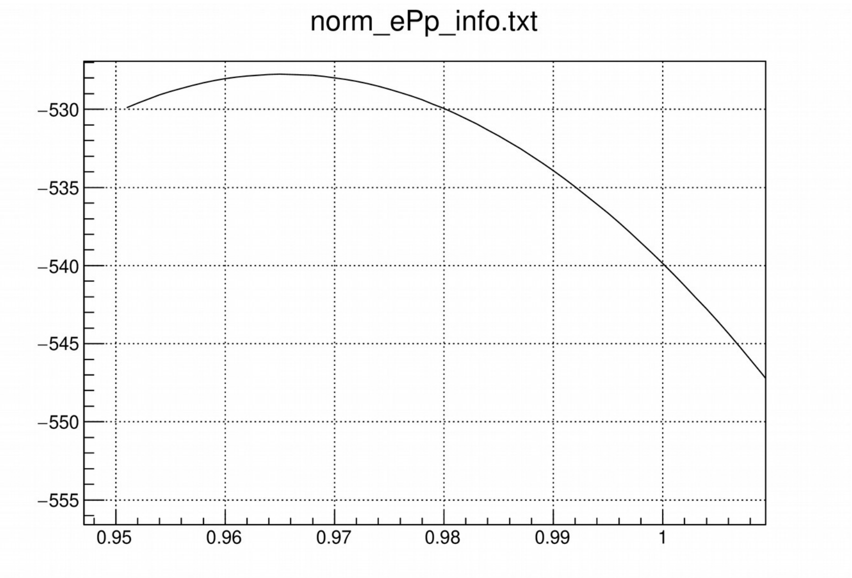
“second analysis” on this remark !!!)

Data sensitivity : change in scale corresponding to 0.5 change in LogP (i.e. 1 unit in chi2)

e-p



e+p



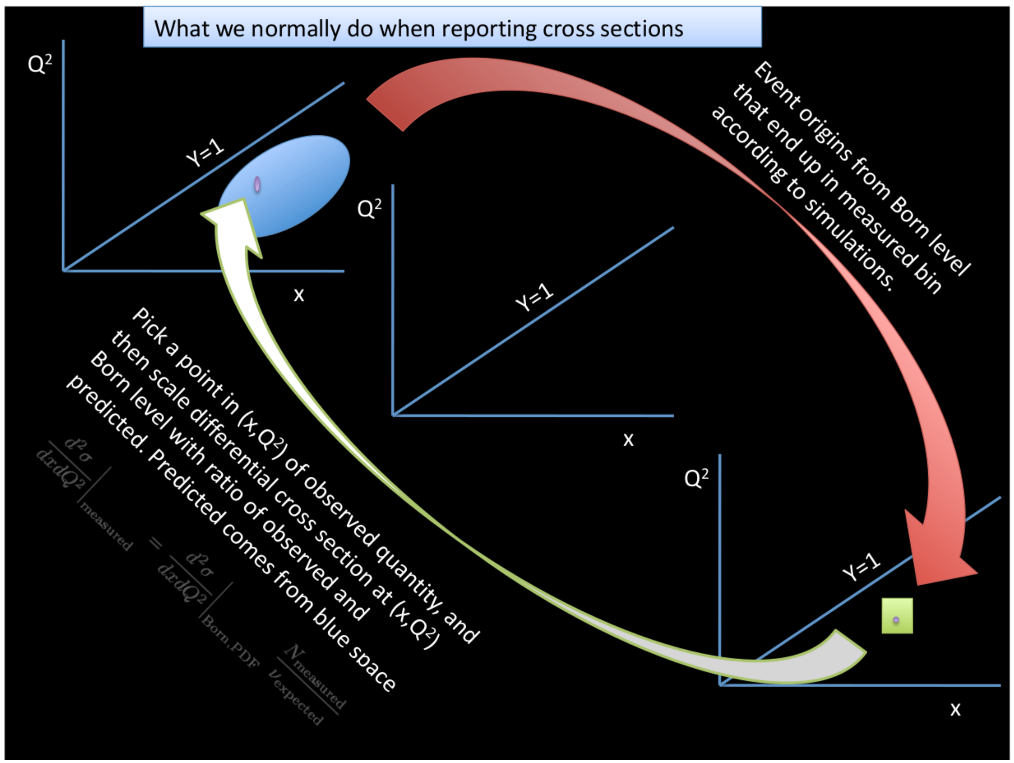
e-p	e+p
scale 0.987 logp0 = -578.751	scale 0.965 logp0 = -527.756
scale 0.981 logp0 = -579.265	scale 0.958 logp0 = -528.31
scale 0.993 logp0 = -579.244	scale 0.972 logp0 = -528.213

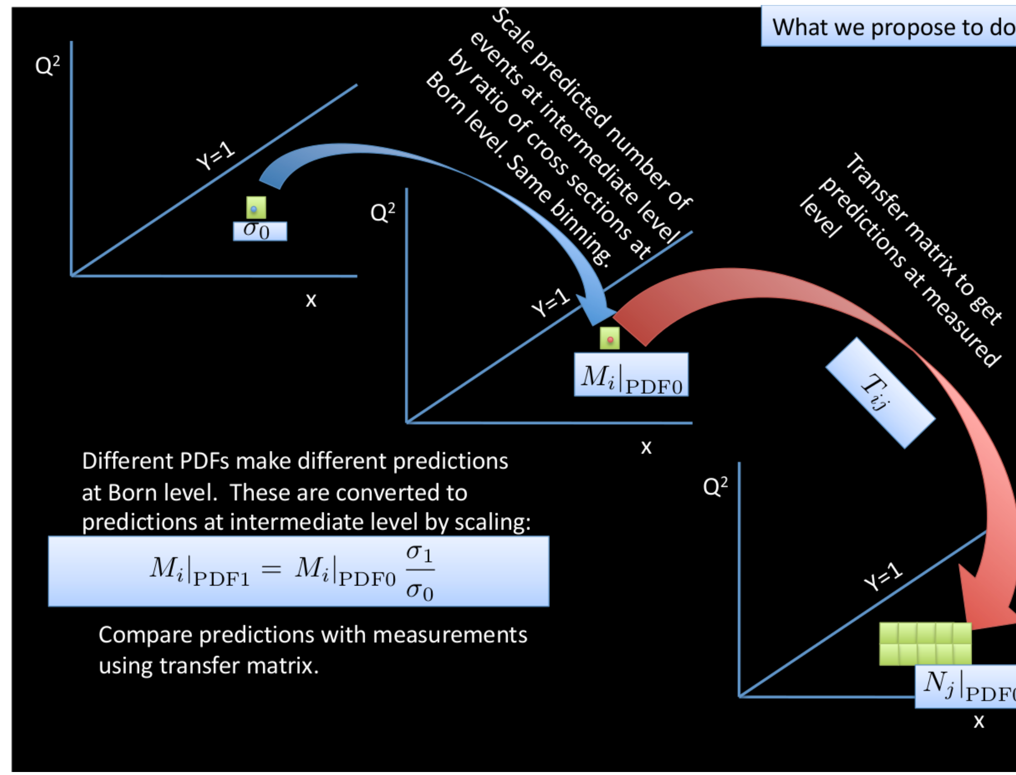
Data sensitivity +-0.6% +-0.7%

Comments from EB-1

The method has to be explained more clearly in the text. It is hard to understand for non-experts

proposed to add explanation plots or summarize them in a couple of sentences





Comments from EB-1

The discrepancies are not entirely new. They are now more visible, because a linear x-scale is used. This should be somehow reflected in the text.

perhaps even point at the figure in the e+p NC ZEUS paper

EB should make suggestion.

Slang should be avoided.

agreed

The nomalisation issue should be discussed in a clear way. Each PDF fit ends up with a different luminosity used. This is on the several % level.

EB should make text suggestions

Olaf and Allen will discuss whether there is an "overall number" that reflects the strength of the data

Done

All plots showing HERAPDF and its uncertainties have to show

NNLO plus full uncertainties [with parameterisation as an envelope]

done

The comparison of cross section predictions was shown for, x values different than in the paper and no integration was performed.

That is okay for a check at this level, BUT it leaves the integration

agreed in last discussion that this check OK. Integration is still being worked on.

Integration was compared to Mandy's old computation on radiative corrections. There were differences on the % level.

Uncertainties on the integration will be reduced to well below the % level by better integration methods.

agreed, but calculations are slow because xFitter is slow and this takes some time

2 bins checked which had difference of ~1.2% with Mandy’s numbers, now with 2 million random number smapling difference is % level.

X bin	Mandy	Random	Integration old
0.6-0.7	0.8255	0.8210	0.8150
0.7-1.0	0.1912	0.1907	0.1894

Changes and Status of paper

Major Changes in text :

inclusion of a section discussing how the transfer matrix approach can be used in future PDF extractions

as well as a discussion of the effective power of the high-x data points along the lines suggested by Ola

Changes expected :

Update in the tables and plots with the given improvements in the reweighting proceedure

Back Up

(some Old slides)

Data & MC sample:

04-06 e-p data (185 pb -1) & 06/07 e+p data (141.44 pb^{-1})
DJANGOH 1.6, Ariadne 4.12, CTEQ-5D MCs (Standard Orange)

Selection:

Vertex:
Valid vertex && |Zvtx| < 50. cm

Electron:
EM finder
e- candidate with Ee>15GeV
EmProb >0.001 (θ_e>0.3) else EmProb > 0.01
Econe (w/o e+) < 4.0 GeV
QEDC rejection

Fiducial volume cuts:
BCAL+FCAL e-s
no cracks, no RCAL
|DME| > 1.4 cm && | DCE| > 0.6 cm

In CTD Acceptance
DCA < 10 cm
Superlayers > 4
TrkP > 5. GeV

Not in Acc. Of CTD
Pt elec > 30. GeV

Trigger selection:
DST 14

Kinematics:
40<Empz<65
Pt/SqrtEt < 5 GeV
y_el < 0.80

Jets

1,2,3(<4) jet events
Box cut (40.40 cm^2)
Et (all jets) > 10 GeV

0 jet events (including events rejected in box cut & Et cut) to be assigned to highest x-bin.

Differences : from old reweighting to the new reweighting

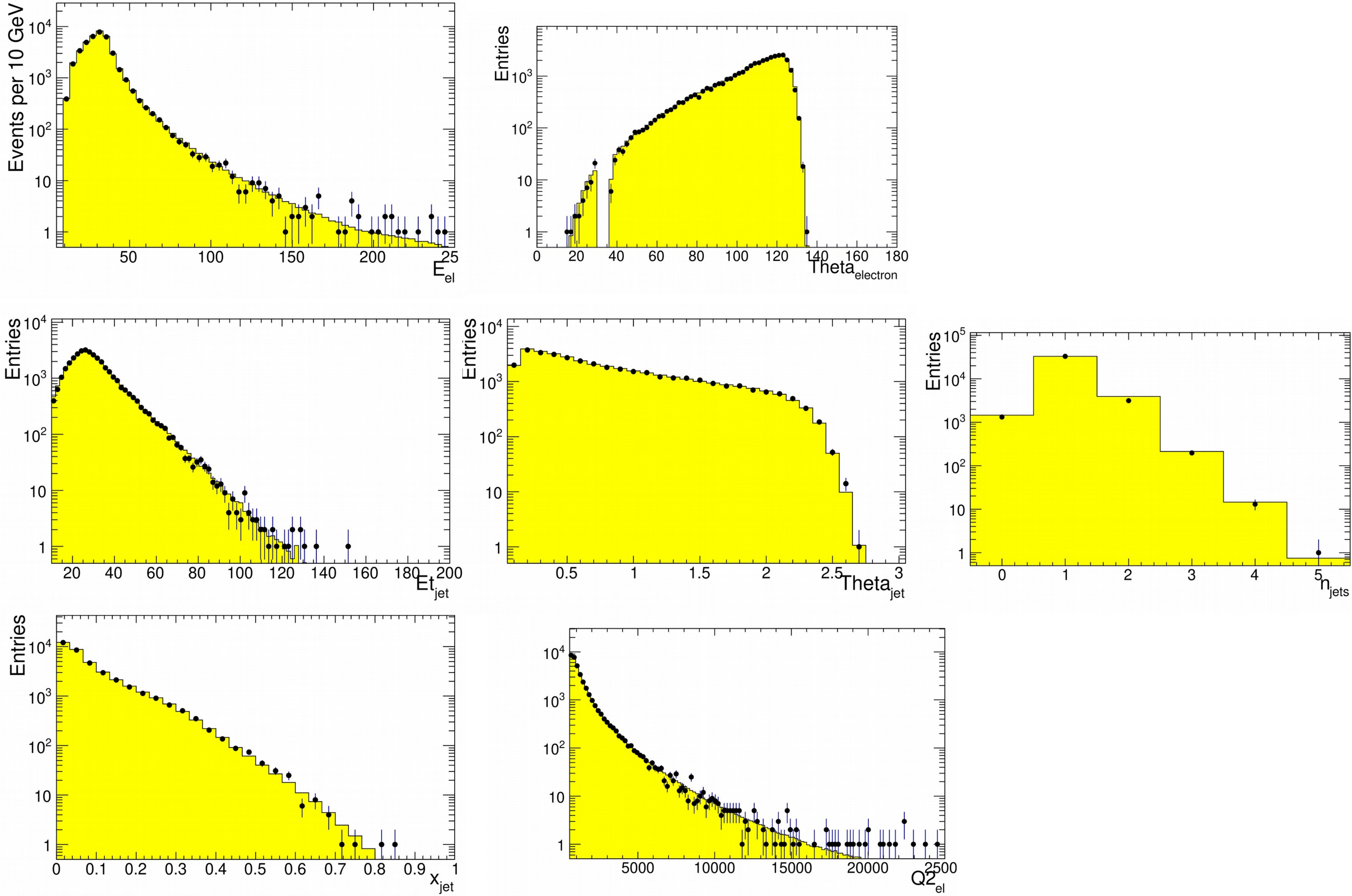
	e+p		e-p	
	New	old	New	old
CTEQ5D	-531.545	-531.545	-577.814	-577.814
CT14	-525.866	-525.687	-583.267	-588.318
HERAPDF2.0	-539.865	-536.053	-581.073	-579.814
MMHT2014	-525.941	-525.707	-583.205	-588.253
NNPDF2.3	-526.687	-528.253	-590.693	-598.253
NNPDF3.0	-526.416	-527.43	-588.465	-595.423
ABMP16	-528.124	-526.716	-579.363	-582.457
abm11	-531.232	-532.29	-587.842	-593.92

	xfitter	DISPred
HF_SCHEME =	'RT OPT'	ZMVFN

This input was given by me, however there are some other internal variables inside the two program which are defined slightly different,

which could also generate a difference (few as below).

	DISPred	xfitter
WBosonMass =	80.398	80.385d0
mbt =	4.2	4.5d0
mtp =	171.2	173d0
Vcb =	0.00393	0.04156
D0Vub =	0.0412	0.00358d0
Sin2ThetaW =	0.22308	0.23127d


Events per 10 GeV
E_{el}
Entries
$Theta_{electron}$
Entries
Et_{jet}
Entries
$Theta_{jet}$
Entries
n_{jets}
Entries
x_{jet}
$Q2_{el}$

Comparing Total Probability for different Pdfs in different x range (x cut which mainly allows integrated bins)

	e^-p				e^+p			
	$x < 0.6$		$x \geq 0.6$		$x < 0.6$		$x \geq 0.6$	
PDF	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$
$CT14$	$3.50 * 10^{-04}$	15.92	0.58	1.09	440.98	-12.18	72.07	-8.56
$HERAPDF2.0$	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
$MMHT2014$	$3.37 * 10^{-04}$	15.99	0.64	0.89	453.96	-12.24	68.61	-8.46
$NNPDF2.3$	$2.00 * 10^{-08}$	35.46	0.49	1.42	31.16	-6.88	78.35	-8.72
$NNPDF3.0$	$2.48 * 10^{-07}$	30.42	0.67	0.80	85.37	-8.89	65.10	-8.35
$ABMP16$	$5.77 * 10^{-02}$	5.71	1.24	-0.42	357.09	-11.76	31.82	-6.92
$abm11$	$5.60 * 10^{-07}$	28.79	1.34	-0.58	9.67	-4.54	4.46	-2.99

Table 2: *The Bayes Factor, p-values and $\Delta\chi^2$ from comparisons of predictions (at NLO) using different PDF sets to the observed numbers of events are shown. The p-values are given for two different x ranges for the e^+p and e^-p data sets.*

NNLO

At high x MMHT, CT, NNPDF ABM better for e+P data.

disagreement comes primarily from lower x in e-p

Comparing Total Probability for different Pdfs in different x range (integrated bins +2 preceding x bins in each Q2)

	e^-p				e^+p			
	low x		high x		low x		high x	
PDF	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$
CT14	$1.16 * 10^{-02}$	8.92	0.02	8.09	247.15	-11.02	128.64	-9.71
HERAPDF2.0	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
MMHT2014	$7.28 * 10^{-03}$	9.84	0.03	7.03	240.33	-10.96	129.67	-9.73
NNPDF2.3	$7.46 * 10^{-07}$	28.22	0.01	8.66	17.64	-5.74	138.52	-9.86
NNPDF3.0	$5.48 * 10^{-06}$	24.23	0.03	6.99	42.95	-7.52	129.54	-9.73
ABMP16	$4.67 * 10^{-01}$	1.52	0.15	3.76	166.50	-10.23	68.24	-8.45
abm11	$3.47 * 10^{-07}$	29.75	2.16	-1.54	2.18	-1.56	19.79	-5.97

Table 10: *The Bayes Factor, p-values and $\Delta\chi^2$ from comparisons of predictions (at NNLO) using different PDF sets to the observed numbers of events are shown. The p-values are given for two different x ranges for the e^+p and e^-p data sets.*

New Table

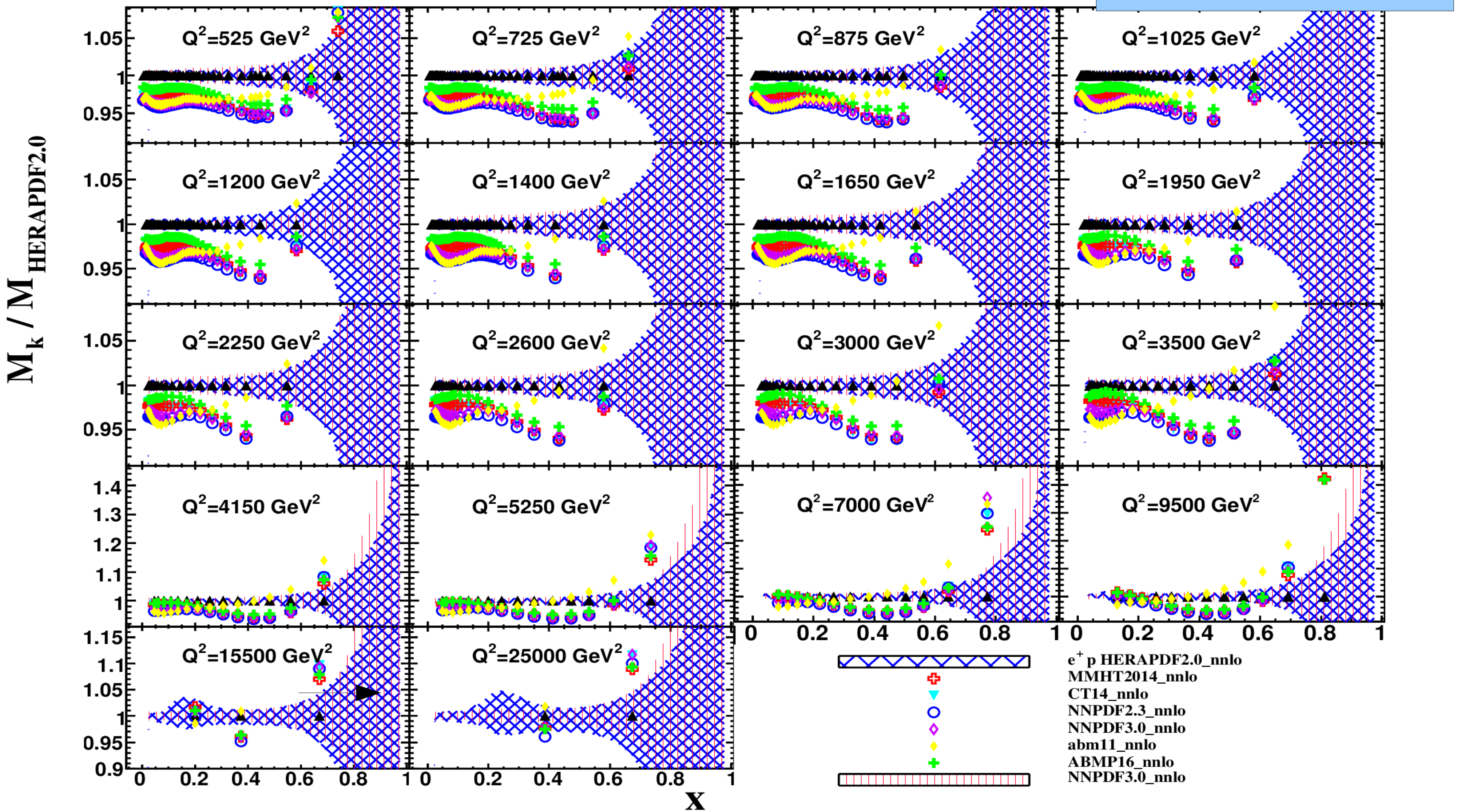
	e^-p				e^+p			
	low x		high x		low x		high x	
PDF	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$
CT14	1.71	-1.08	0.07	5.46	4092.86	-16.63	293.54	-11.36
HERAPDF2.0	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
MMHT2014	1.15	-0.29	0.10	4.55	3858.37	-16.52	288.88	-11.33
NNPDF2.3	$1.22 * 10^{-03}$	13.42	0.05	5.81	1516.26	-14.65	348.28	-11.71
NNPDF3.0	$5.41 * 10^{-03}$	10.44	0.11	4.34	2340.22	-15.52	296.19	-11.38
ABMP16	1.41	-5.30	0.39	1.87	1013.33	-13.84	123.97	-9.64
abm11	$2.75 * 10^{-04}$	16.40	4.19	-2.86	173.82	-10.32	32.30	-6.95

Table 4: *The Bayes Factor, p-values and $\Delta\chi^2$ from comparisons of predictions (at NNLO) using different PDF sets to the observed numbers of events are shown. The p-values are given for two different x ranges for the e^+p and e^-p data sets.*

At high x MMHT, CT, NNPDF ABM better for e+P data. disagreement comes primarily from lower x in e-p

Average ratio of Born level cross sections in different PDFs to HERAPDF2.0NNLO for M bins (e+p)

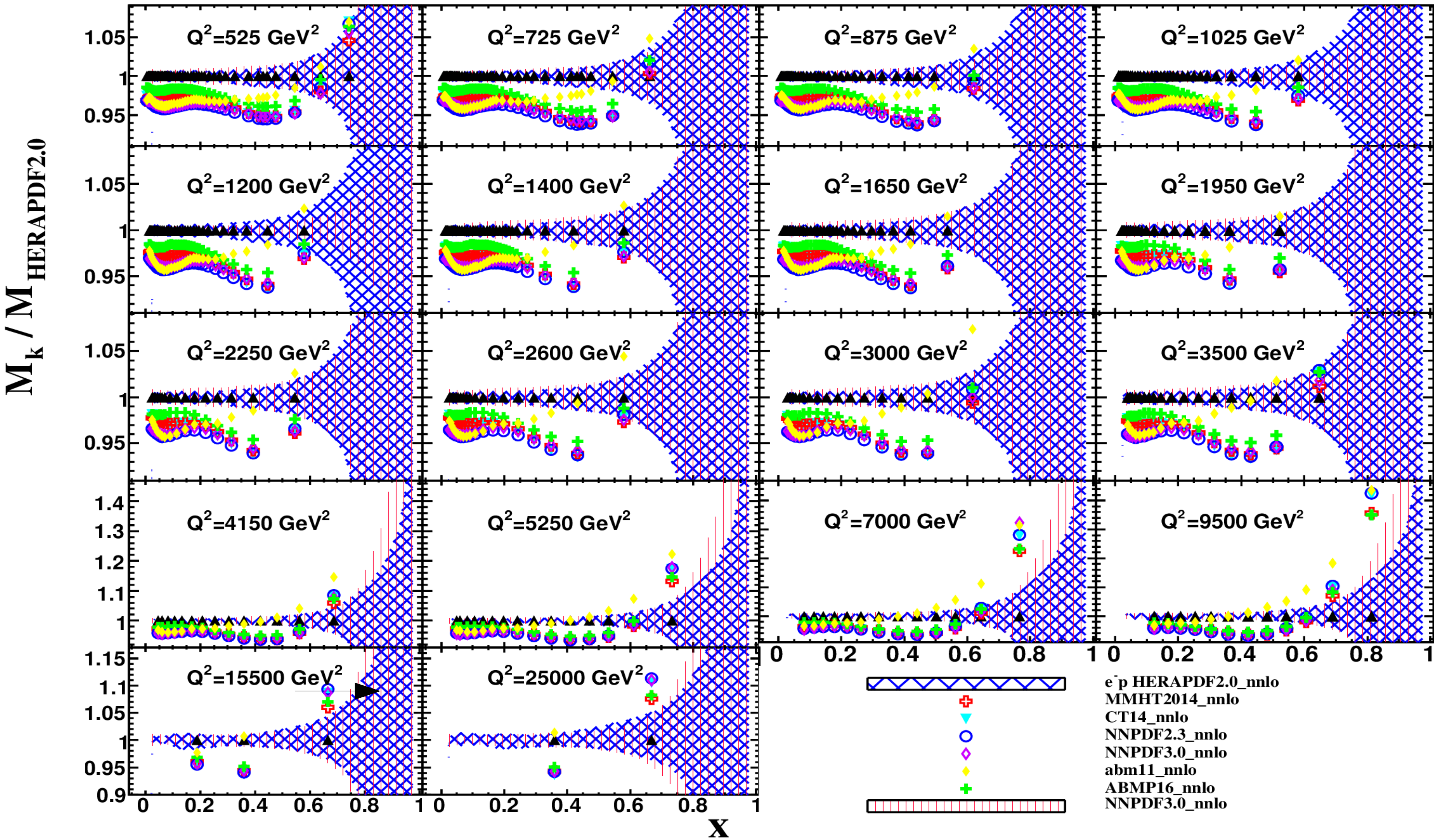
Old reweighting



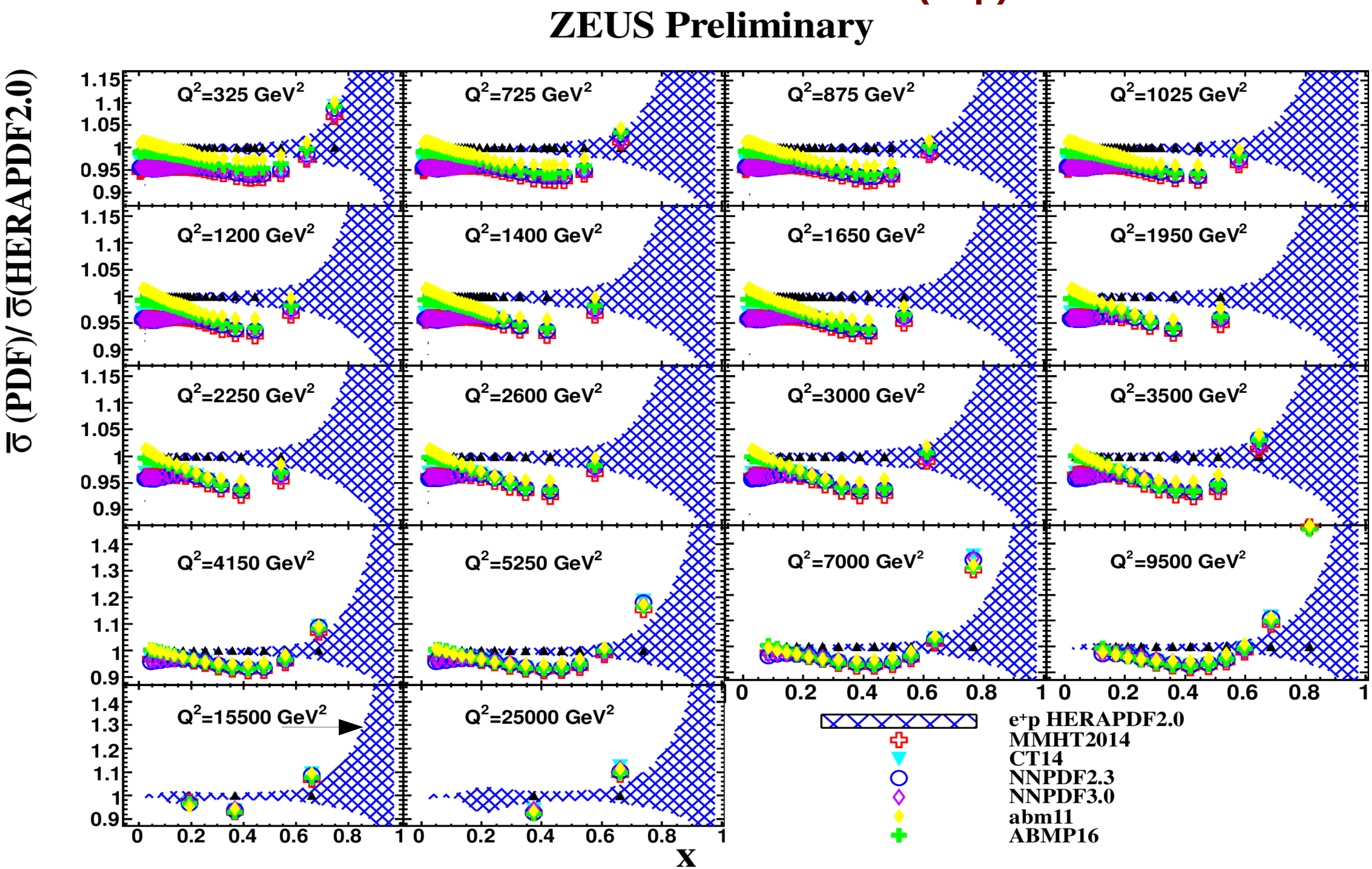
PDFs differ and the difference is not covered by PDF uncertainty !

Average ratio of Born level cross sections in different PDFs to HERAPDF2.0NNLO for M bins (e-p)

Old reweighting

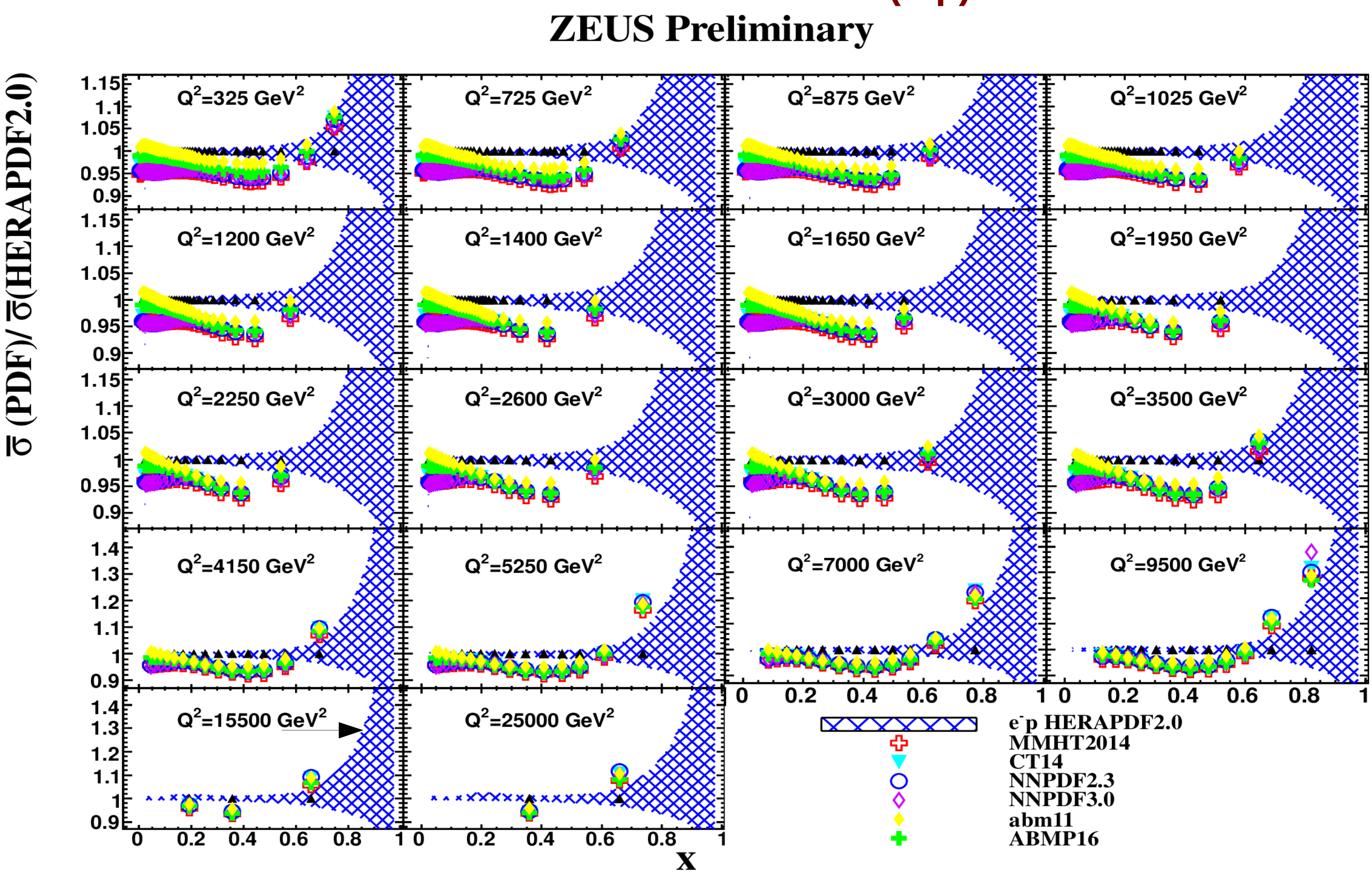


Average ratio of Born level cross sections in different PDFs to HERAPDF2.0NLO for M bins (e+p)



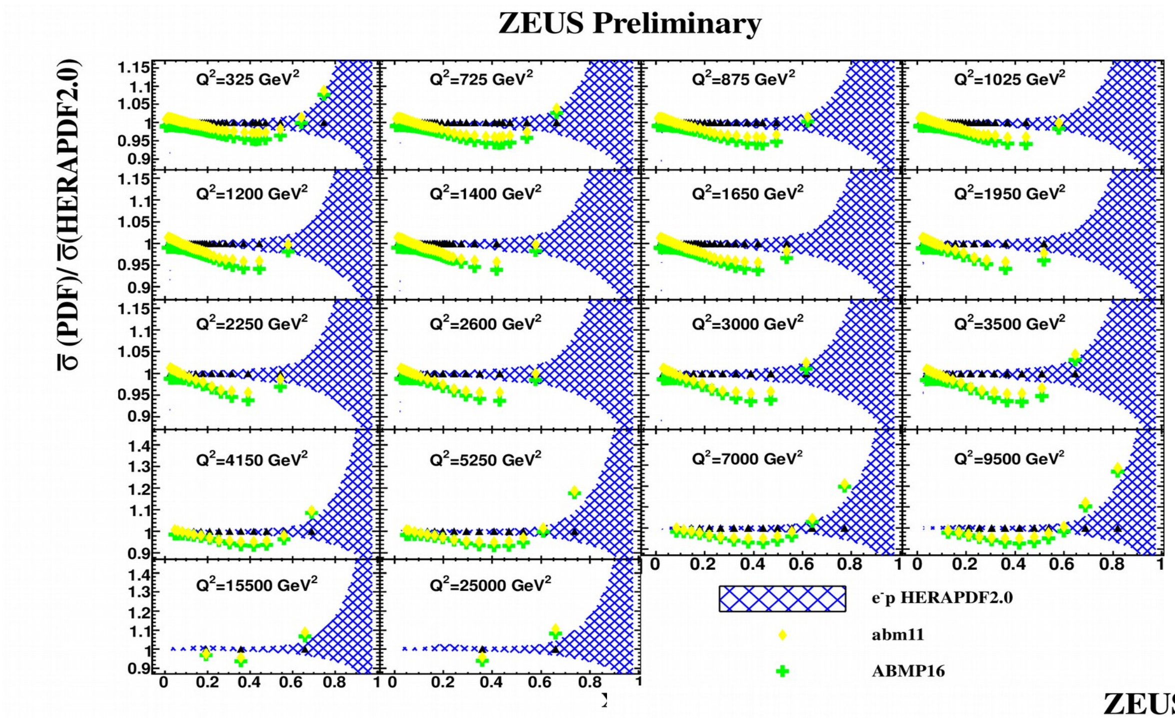
There is a shape difference between HERAPDF & other PDFs, approaches 10% at x ~ 0.4, well outside PDF uncertainties.

Average ratio of Born level cross sections in different PDFs to HERAPDF2.0NLO for M bins (e-p)

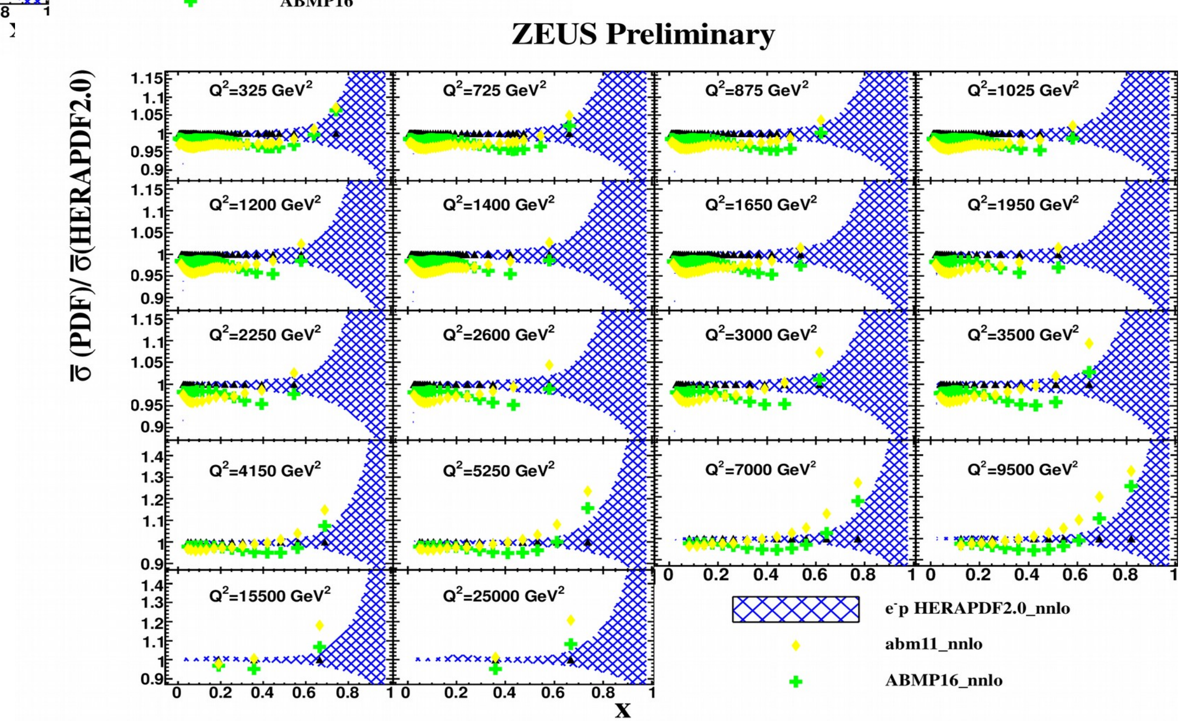


There is a shape difference between HERAPDF & other PDFs, approaches 10% at x ~ 0.4, well outside PDF uncertainties.

Average ratio of Born level cross sections in ABM PDFs to HERAPDF2.0 for M bins (e-p)

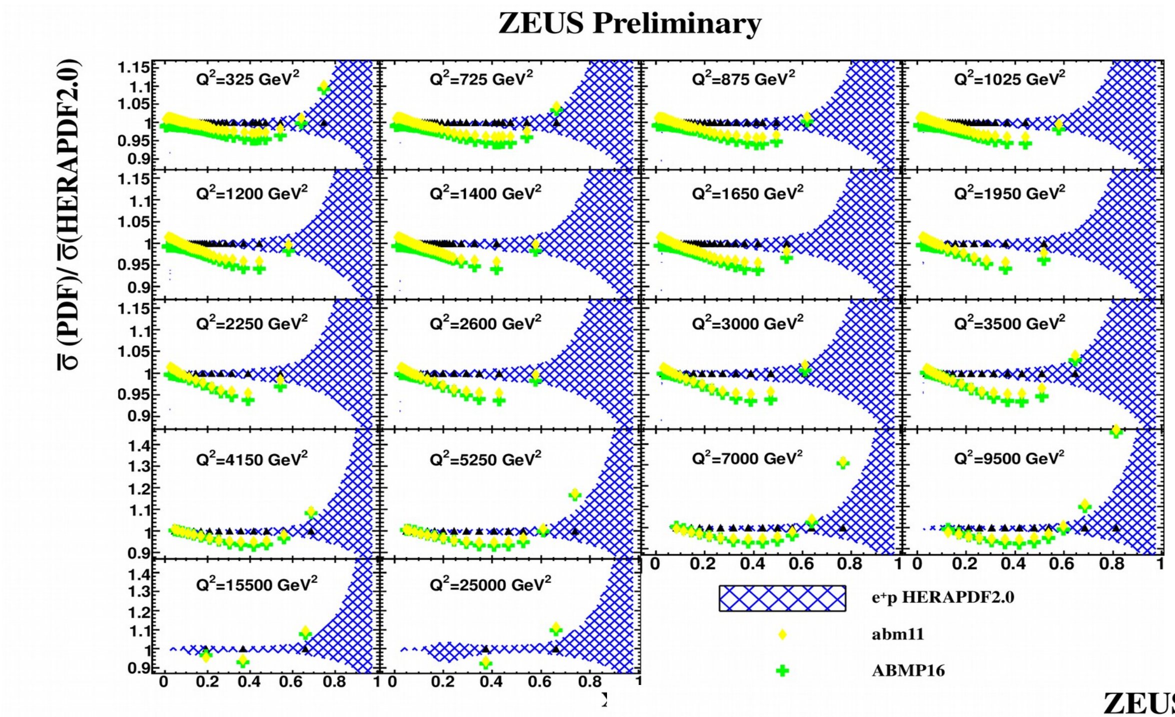


NLO

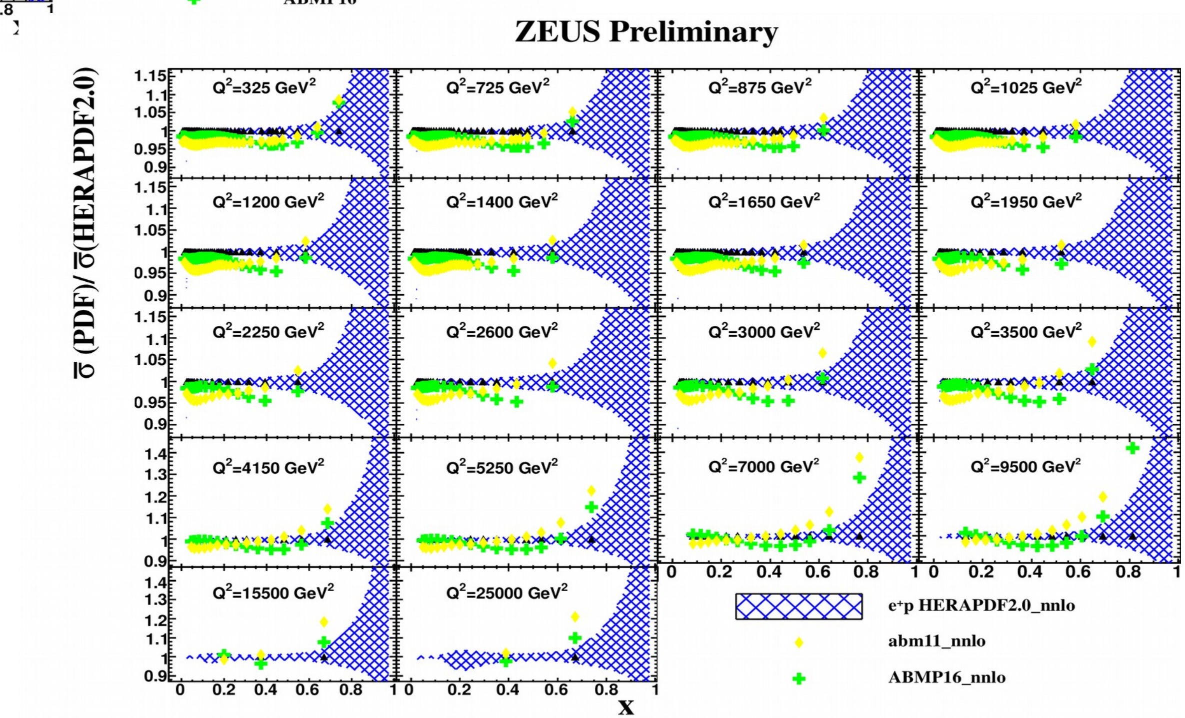


NNLO

Average ratio of Born level cross sections in ABM PDFs to HERAPDF2.0 for M bins (e+p)



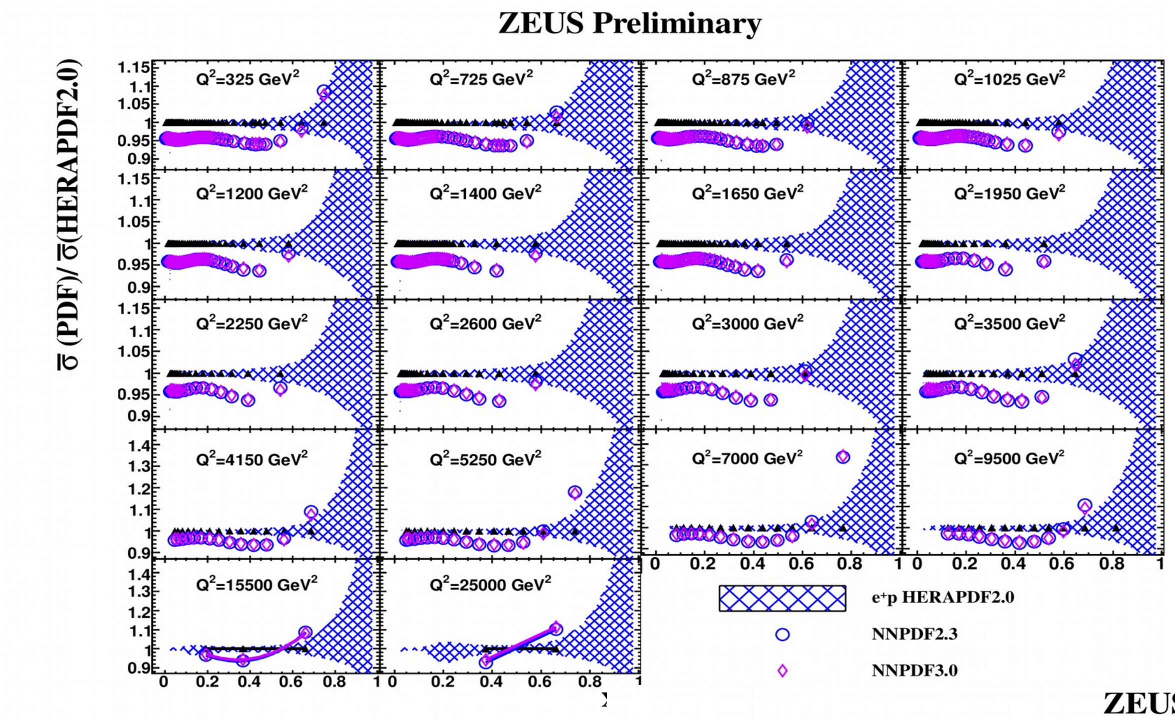
NLO



NNLO

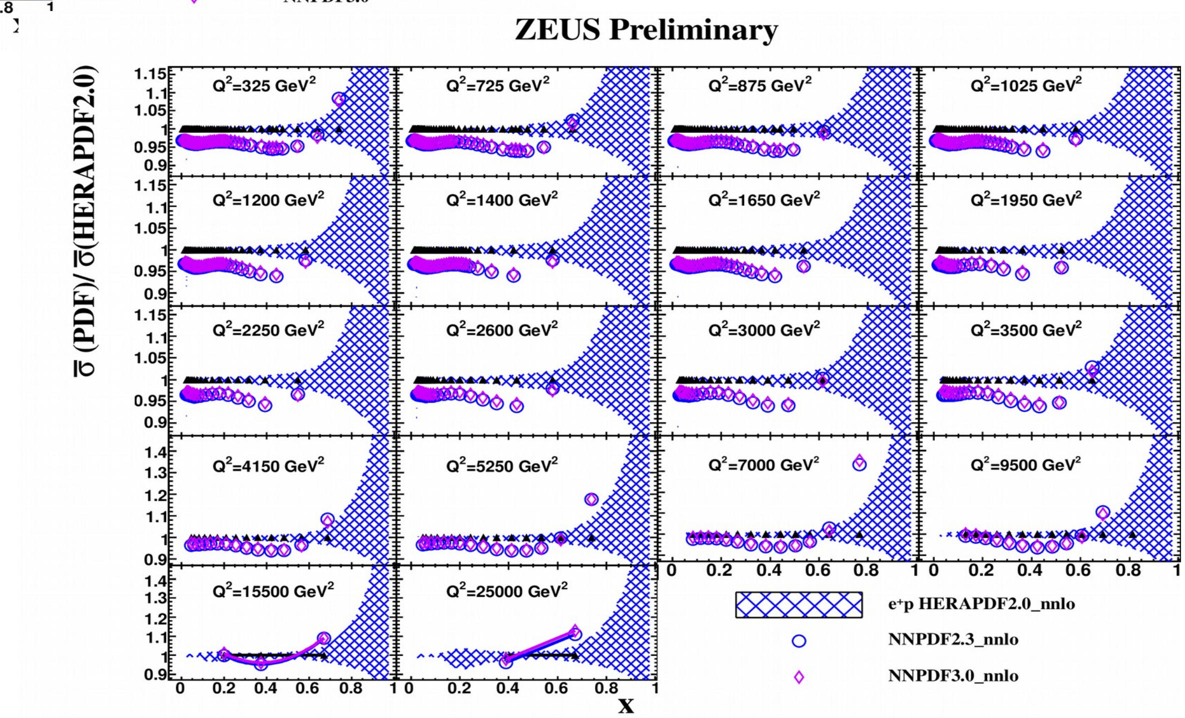
Average ratio of Born level cross sections in NNPDF to HERAPDF2.0 for M bins (e+p)

$$\nu_{i,k} = \sum_{m}^{M_i} \frac{d^2\sigma(x, Q^2|M_k)/dxdQ^2}{d^2\sigma(x, Q^2|M_0)/dxdQ^2}\omega_m^{MC}$$



NLO

NNLO



Prescription of model fitting to high-x data

Probability of observing Data with given set of PDF parameters θ and nuisance parameters λ:

$$P(D|\theta,\lambda) = \prod_j P(n_j|\nu_j(\theta,\lambda))$$

Predicted number of events ν_j is given as :

$$\nu_j = \sum_i \nu_i(1 + 0.018 \cdot \lambda_0) a_{ij} (1 + \sum_{k=1} \lambda_k \delta_{ij}^k)$$

δ's : one standard deviation due to k correlated systematic sourses
λo : modification in normalization in units of standard deviatiom
λk : shifts in the systematic errors

Where a penalty is added to the loglikelihood function: $\mathcal{L}(\theta,\lambda) = P(n_j|\nu_j(\theta,\lambda))P(\lambda)$

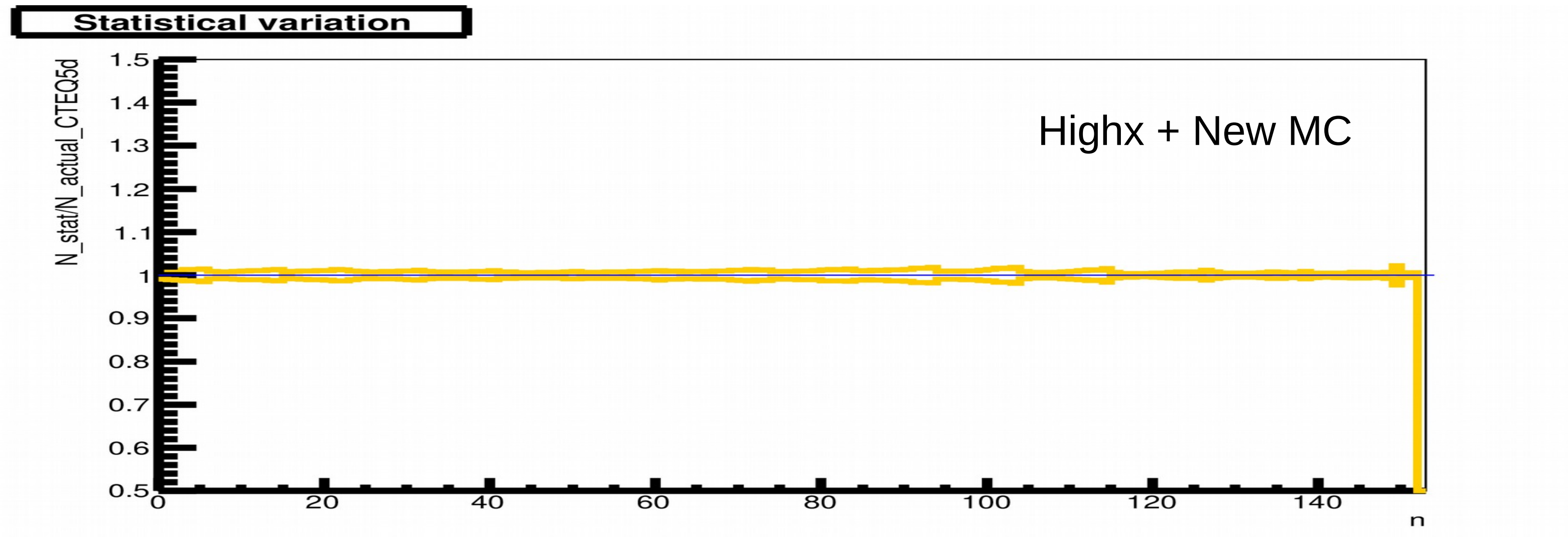
where the $P(\lambda)$ is a product of Gauss distributions:

$$P(\lambda) = \prod_{k=0} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\lambda_k^2}$$

Uncorrelated uncertainties can be taken into account by folding a Gauss distribution for them with the Poisson distribution :

$$P(n_j|\nu_j) = \int \frac{e^{-\nu_j(1+\epsilon_j)}(\nu_j(1+\epsilon_j))^{n_j}}{n_j!} \frac{1}{\sqrt{2\pi}\delta_j} e^{-\frac{1}{2}(\frac{\epsilon_j}{\delta_j})^2} d\epsilon_j$$

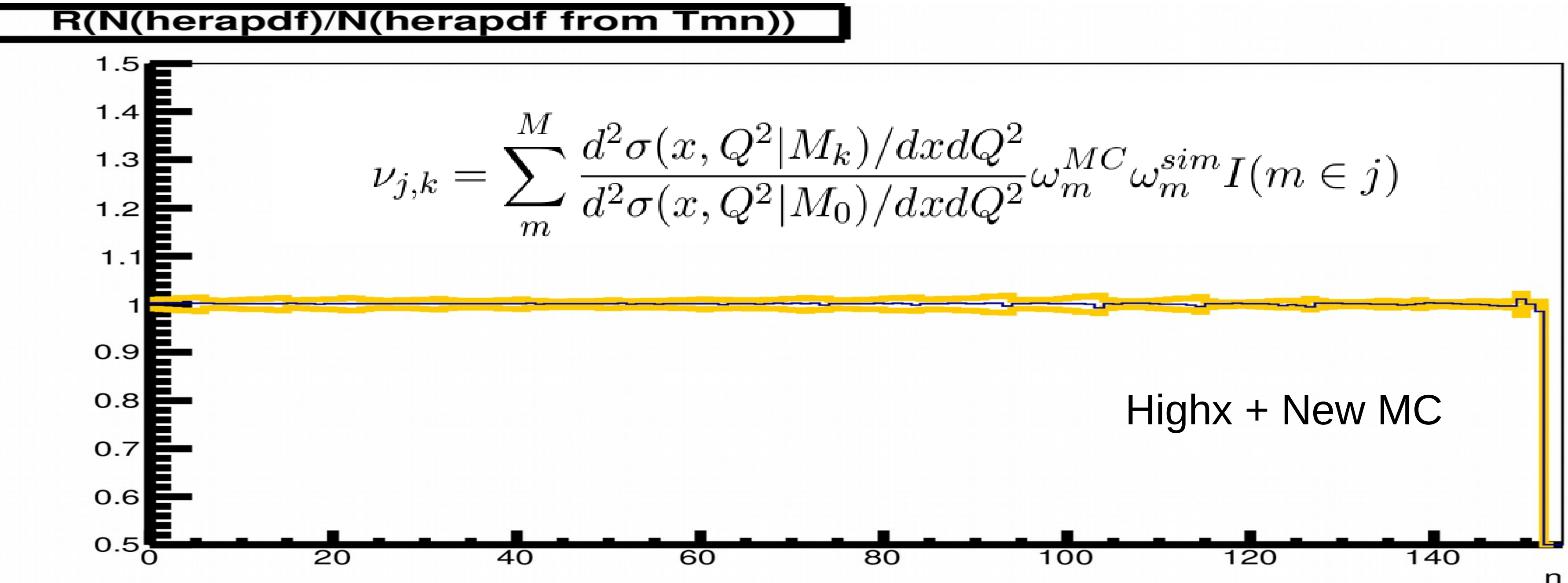
Statistical Error in MC in various Xsec Bins (with in 1%)

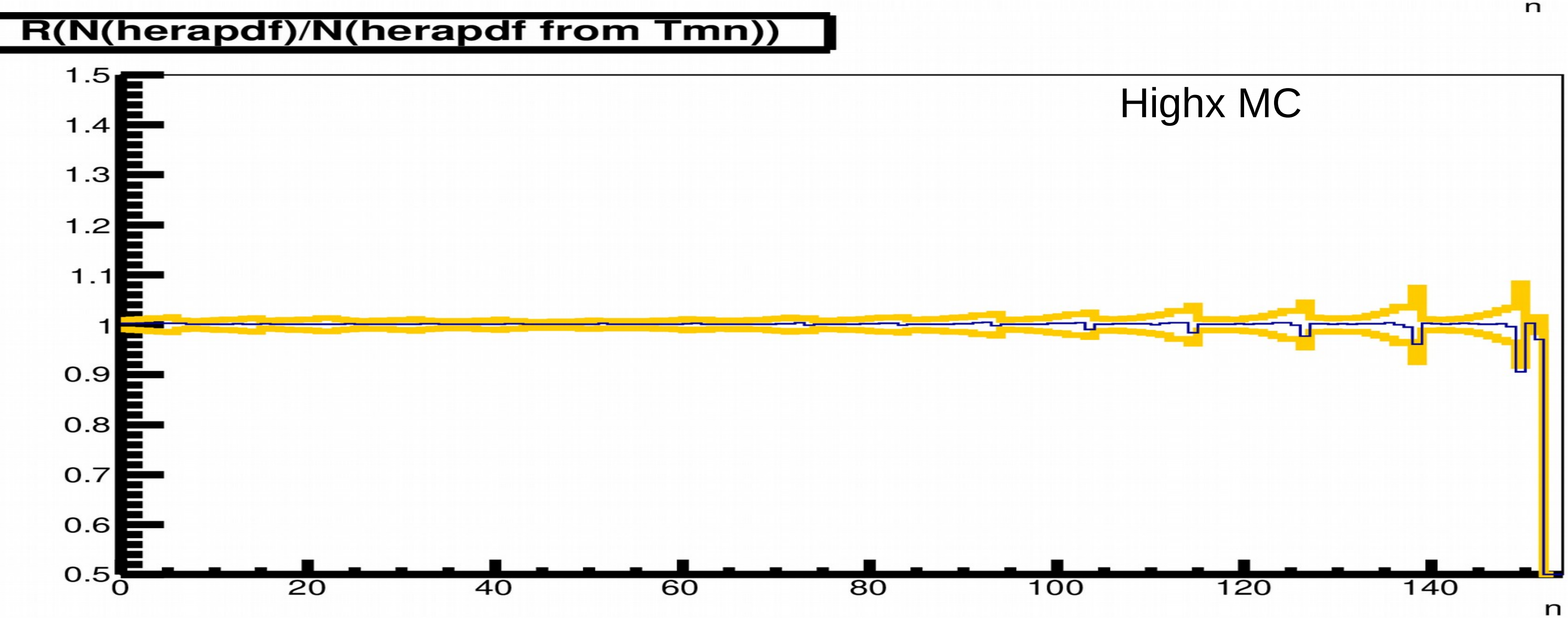


$$\delta a_{ij}^{\mathrm{stat}} = \sqrt{\frac{a_{ij}(1-a_{ij})}{M_i}}$$

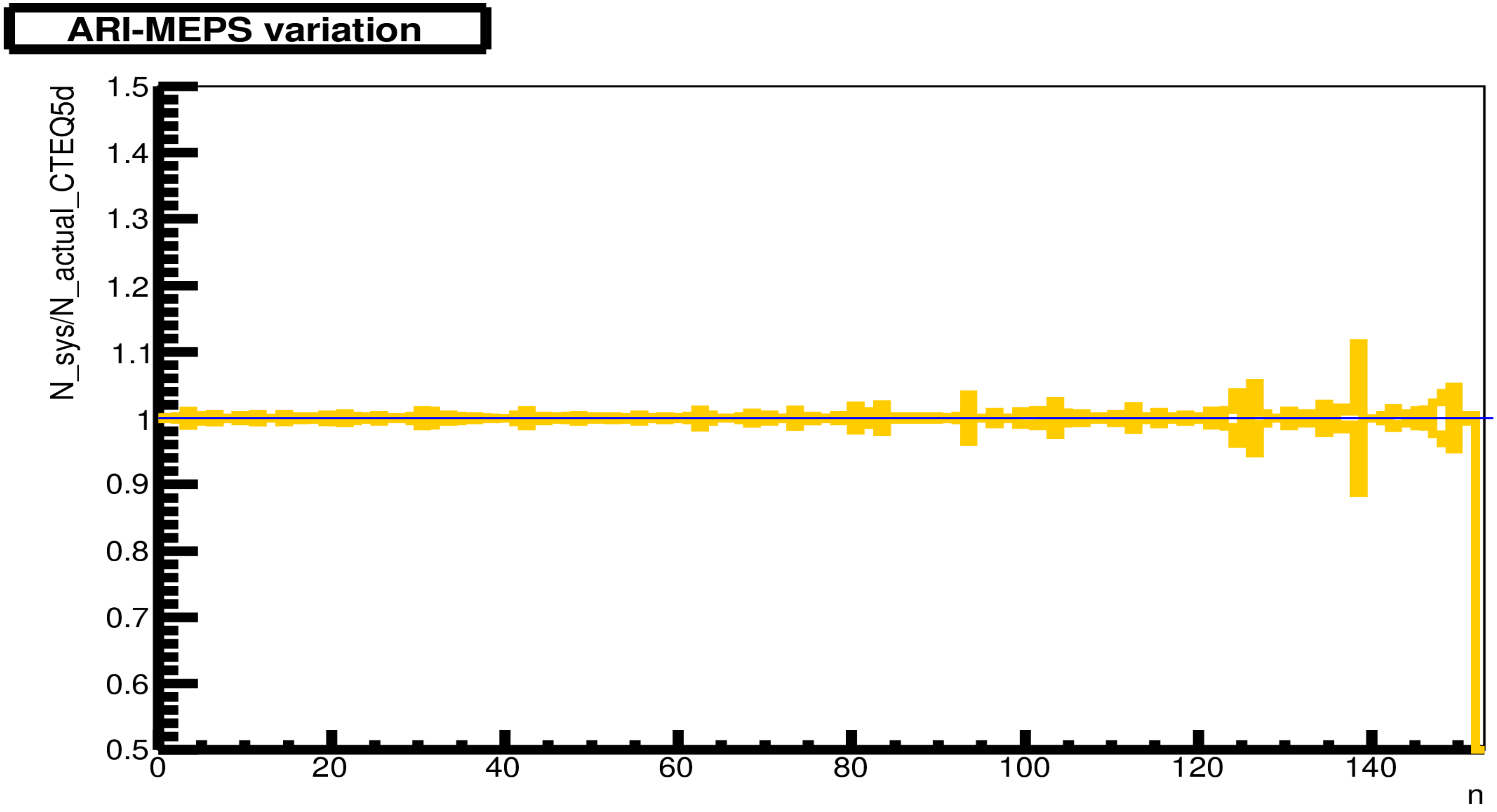
Where Mi are the total number of events Generated in MC

Ratio of N (w/o using Tmn) and N (using calculated using Tmn) for HERAPDF2.0 : An estimate of choice of PDF to build Tmn



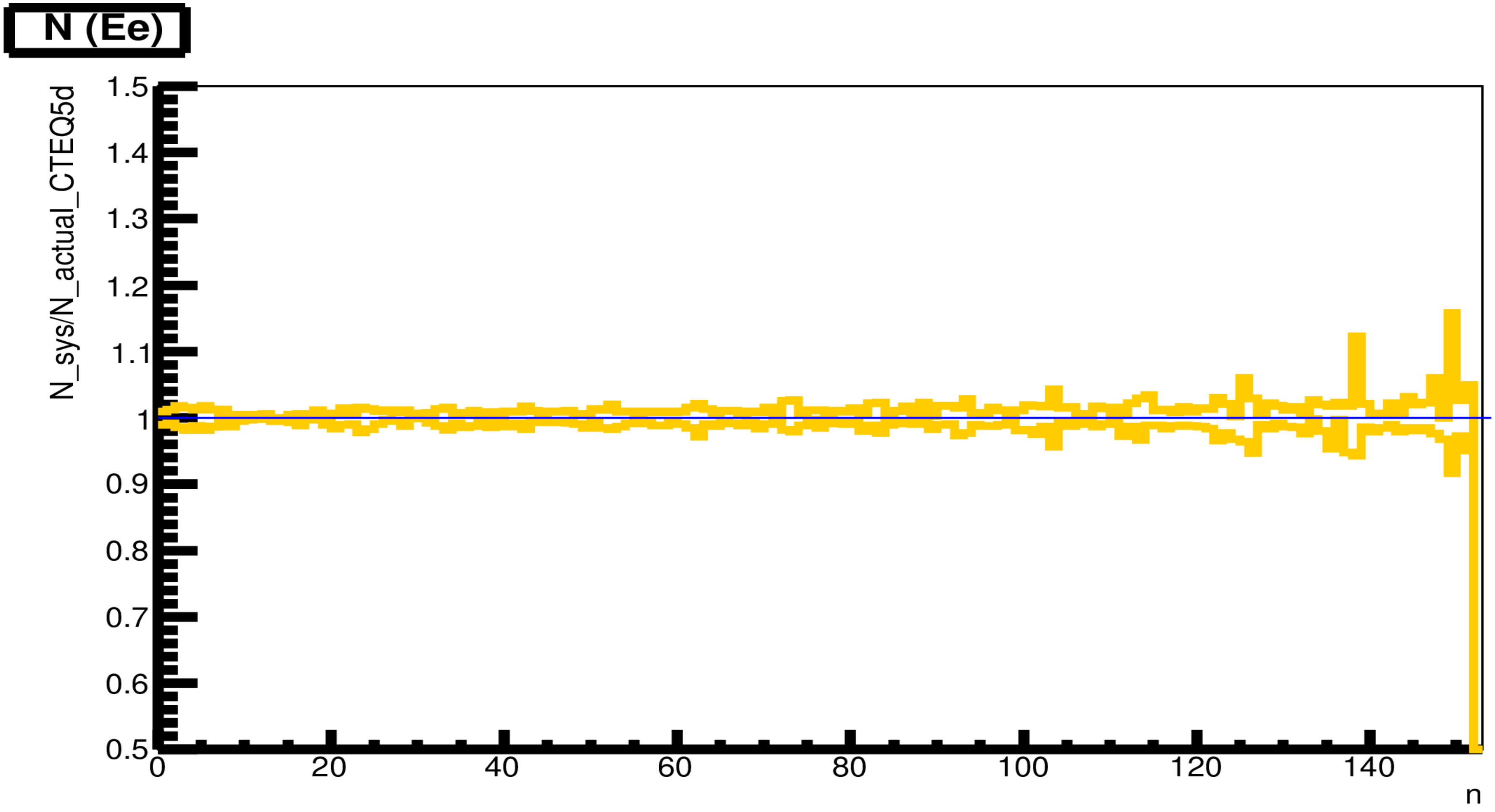


Ariadne-MEPS variation: The ARI-MEPS combination is varied in construction of Transfer Matrix.



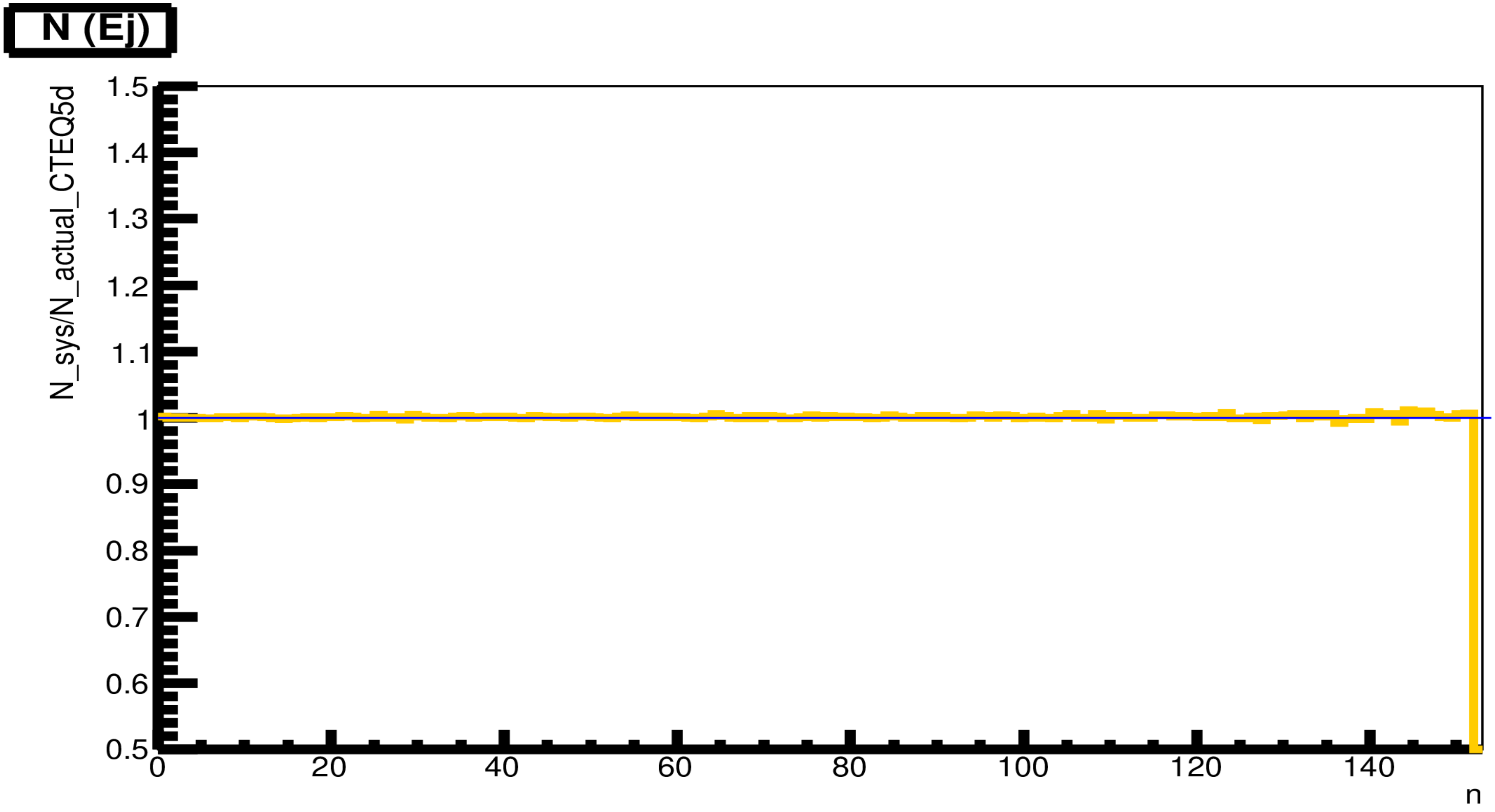
For most of the bins with in 1%, increases to 2-10% in the highest x-bins at high Q2.

Other Systematic Variation : Ee varied up and down and new Transfer Matrix constructed .



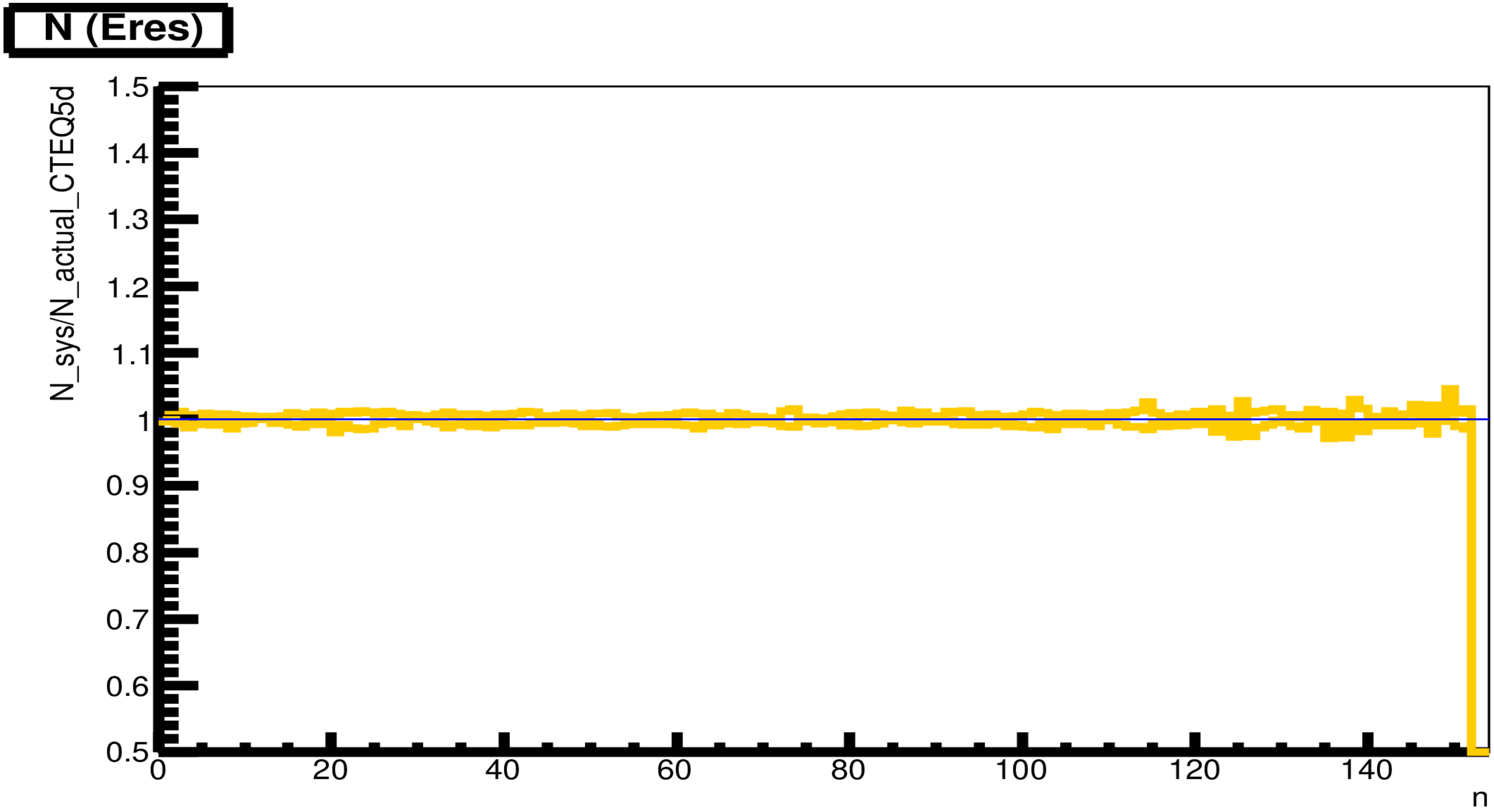
For most of the bins with in 1%, increases to 2-12% in the highest x-bins at high Q2.

Other Systematic Variation : Ejet varied up and down and new Transfer Matrix constructed .



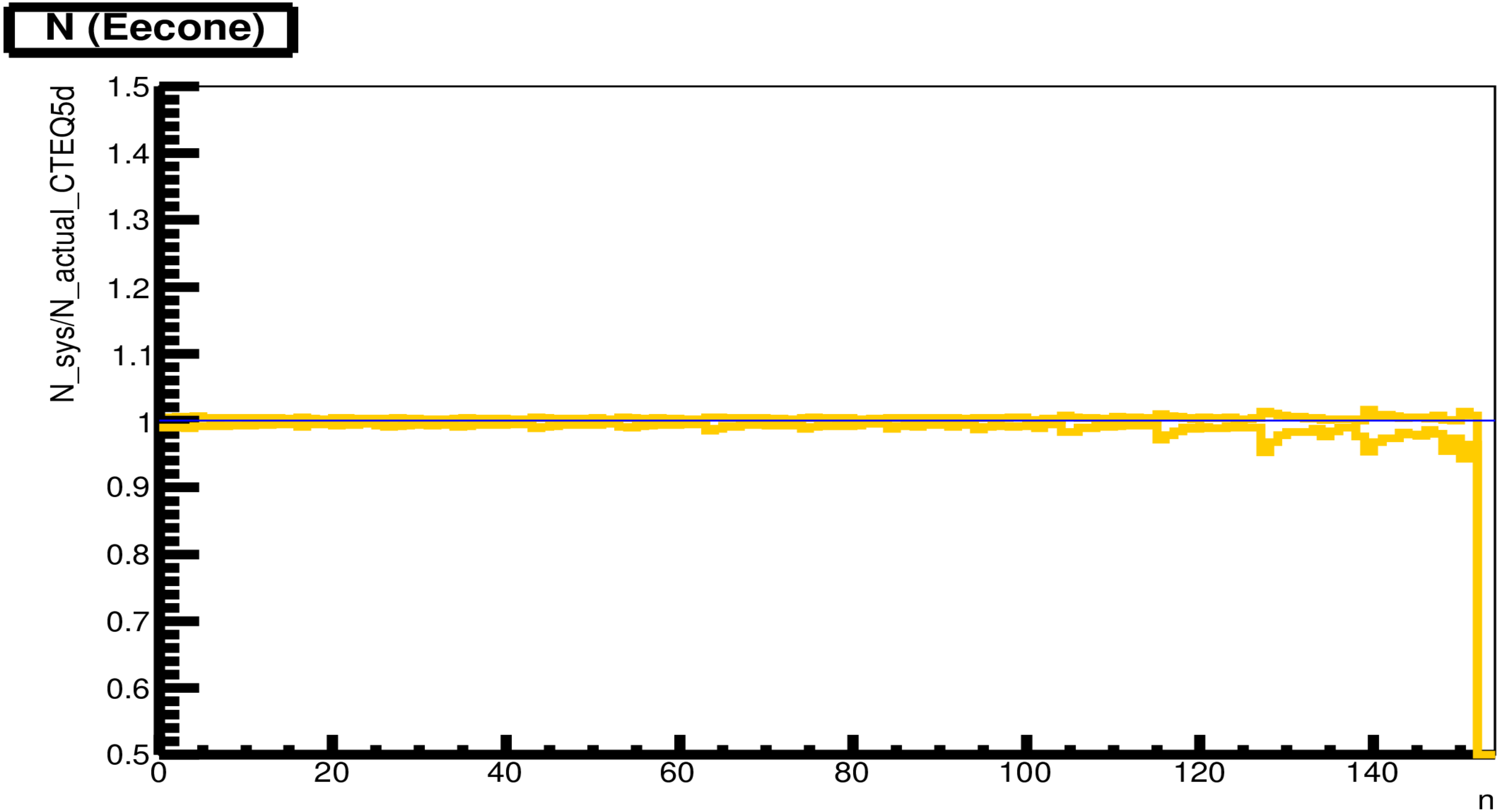
Negligible

Other Uncorr Systematic Variation : Eres varied up and down and new Transfer Matrix constructed .



For most of the bins with in 1%, increases to 2-3% in the bins at high Q2.

Other Uncorr Systematic Variation : Econe varied up and down and new Transfer Matrix constructed .



For most of the bins with in 1%, increases to 2-5% in the bins at high Q2.

Why do we study in Probability numbers

What types of probabilities do we expect ?

E.g., imagine you expect 1 event, and measure 1, then the probability is

$$P(n|\nu) = e^{-\nu}\frac{\nu^n}{n!} = e^{-1} \approx 0.37$$

E.g., imagine you expect 10 events, and measure 8, then the probability is

$$P(n|\nu) = e^{-\nu}\frac{\nu^n}{n!} = e^{-10}\frac{10^8}{8!} \approx 0.11$$

E.g., imagine you expect 100 events, and measure 90, then the probability is

$$P(n|\nu) = e^{-\nu}\frac{\nu^n}{n!} = e^{-100}\frac{100^{90}}{90!} \approx 0.02$$

If we have 150 bins with probabilities ranging from a few % to few 10 %, then

$$P(\{n\}|\{\nu\}) = \prod_{i=1}^{150} e^{-\nu_i}\frac{\nu_i^{n_i}}{n_i!} \text{ maybe } 10^{-200} \quad \ln P \approx -500$$

Slide from Allen

Why do we study in Probability numbers

If the likelihood (product of the data probabilities) is a product of Gaussian distributions, then we have

$$\mathcal{L} \propto e^{-\chi^2/2} \quad \text{and} \quad \ln \mathcal{L}_1 - \ln \mathcal{L}_2 = \frac{1}{2}(\chi_2^2 - \chi_1^2)$$

So we can translate differences in the ln of the probabilities (multiplied by -2) to equivalent chi squared differences

If we look at ratios of probabilities, and again assuming Gaussian distributions, then

$$\frac{P_1}{P_2} = e^{-(\chi_1^2 - \chi_2^2)/2}$$

so taking -2* the natural logarithm of a probability ratio is again equivalent to a chi squared difference

Slide from Allen

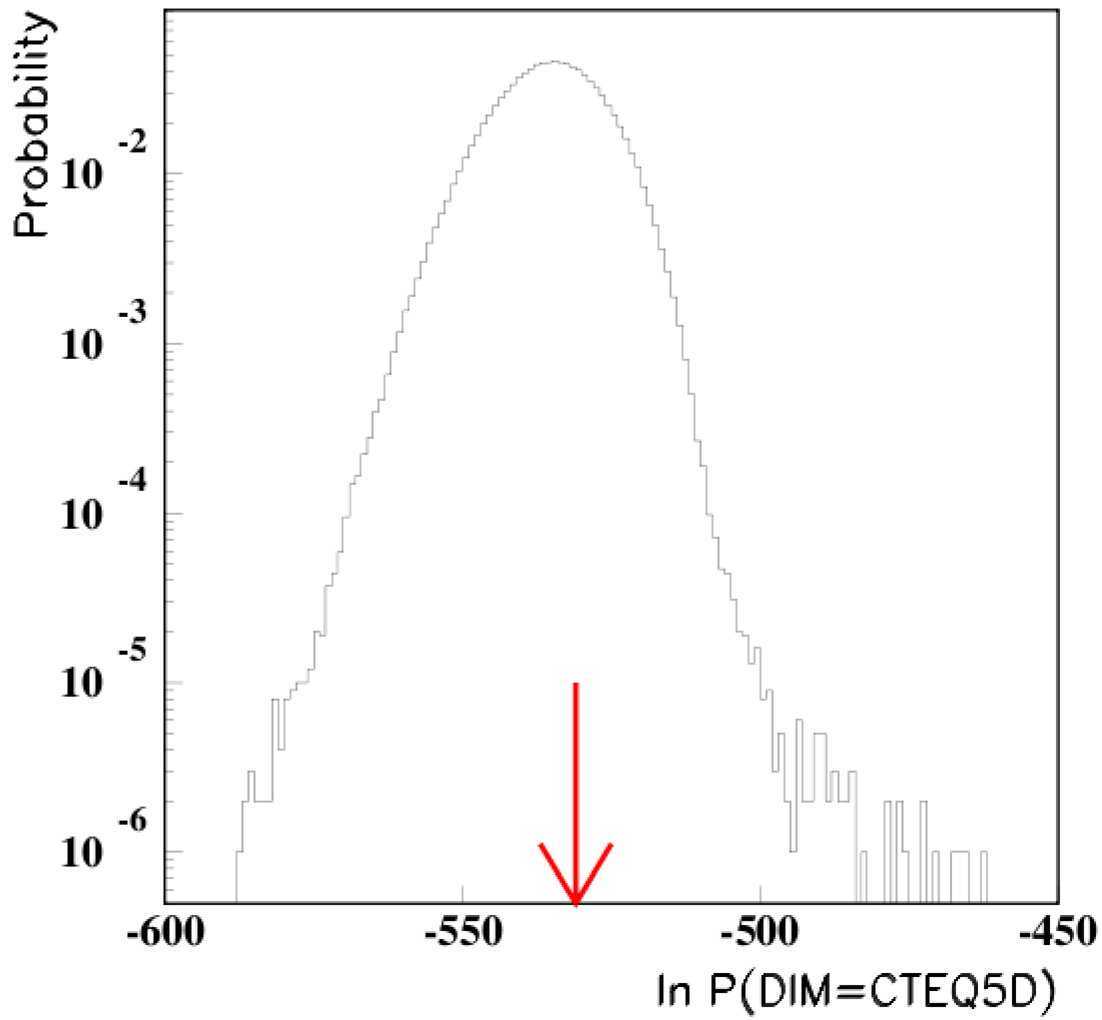


FIGURE 4. Distribution of expected values for $\ln P(D|M = CTEQ)$ for the e^+p data set. The arrow shows the value found in the data.

The high x bins with x_bin_centre > 0.6

Q2	x	N_data
650 - 800,	0.26 - 1.00,	504,
800 - 950,	0.28 - 1.00,	671,
950 - 1100,	0.32 - 1.00,	414,
1100 - 1300,	0.34 - 1.00,	368,
1300 - 1500,	0.36 - 1.00,	202,
1500 - 1800,	0.39 - 1.00,	173,
1800 - 2100,	0.43 - 1.00,	74,
2100 - 2400,	0.46 - 1.00,	51,
2400 - 2800,	0.50 - 1.00,	36,
2800 - 3200,	0.54 - 1.00,	19,
3200 - 3800,	0.58 - 1.00,	17,
3800 - 4500,	0.63 - 1.00,	5,
4500 - 6000,	0.69 - 1.00,	3,
6000 - 8000,	0.59 - 0.73,	10,
6000 - 8000,	0.73 - 1.00,	1,
8000 - 11000,	0.57 - 0.64,	4,
8000 - 11000,	0.64 - 0.78,	1,
8000 - 11000,	0.78 - 1.00,	1,
11000 - 20000,	0.60 - 1.00,	8,