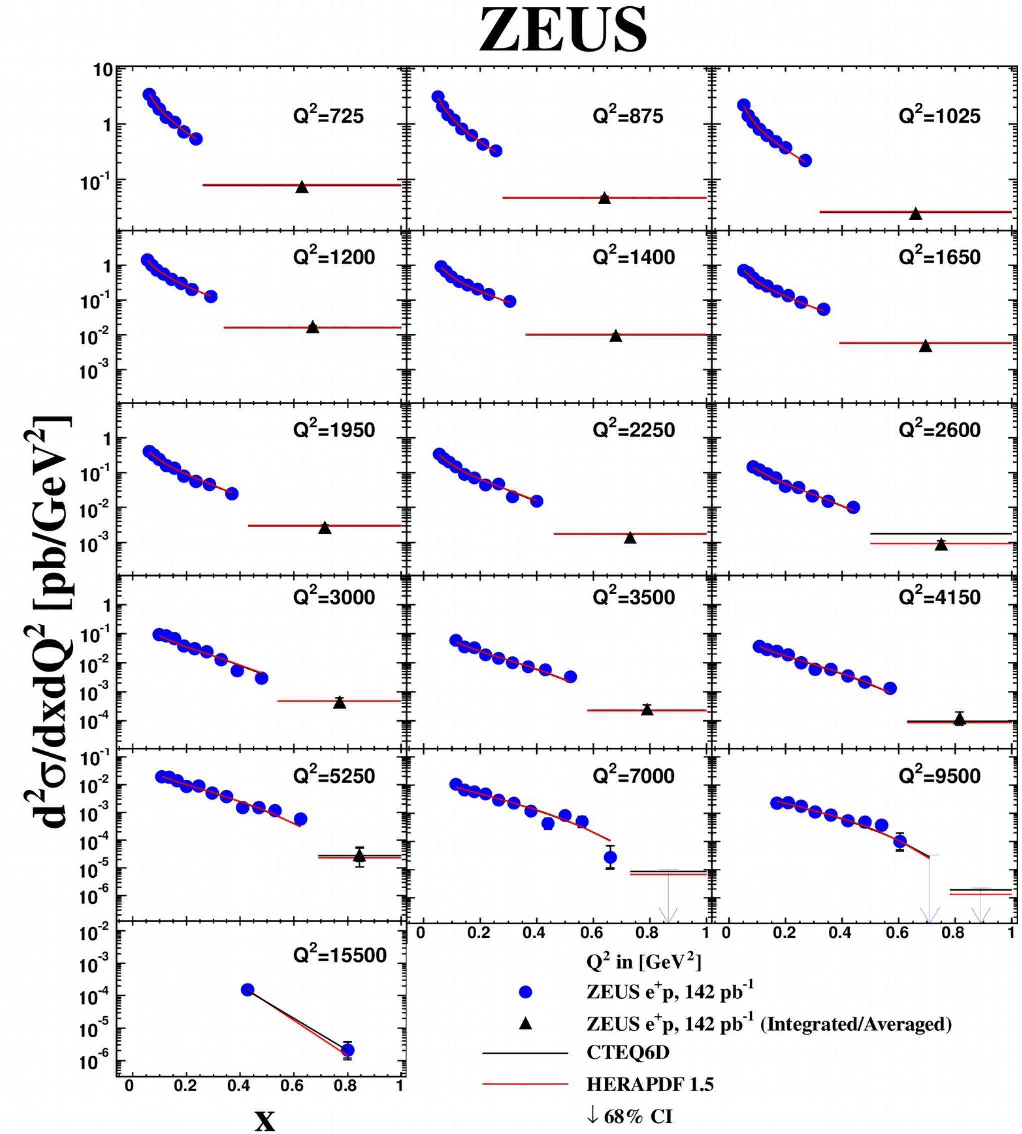
Study of proton parton distribution functions at high-x
(An update post EB2)

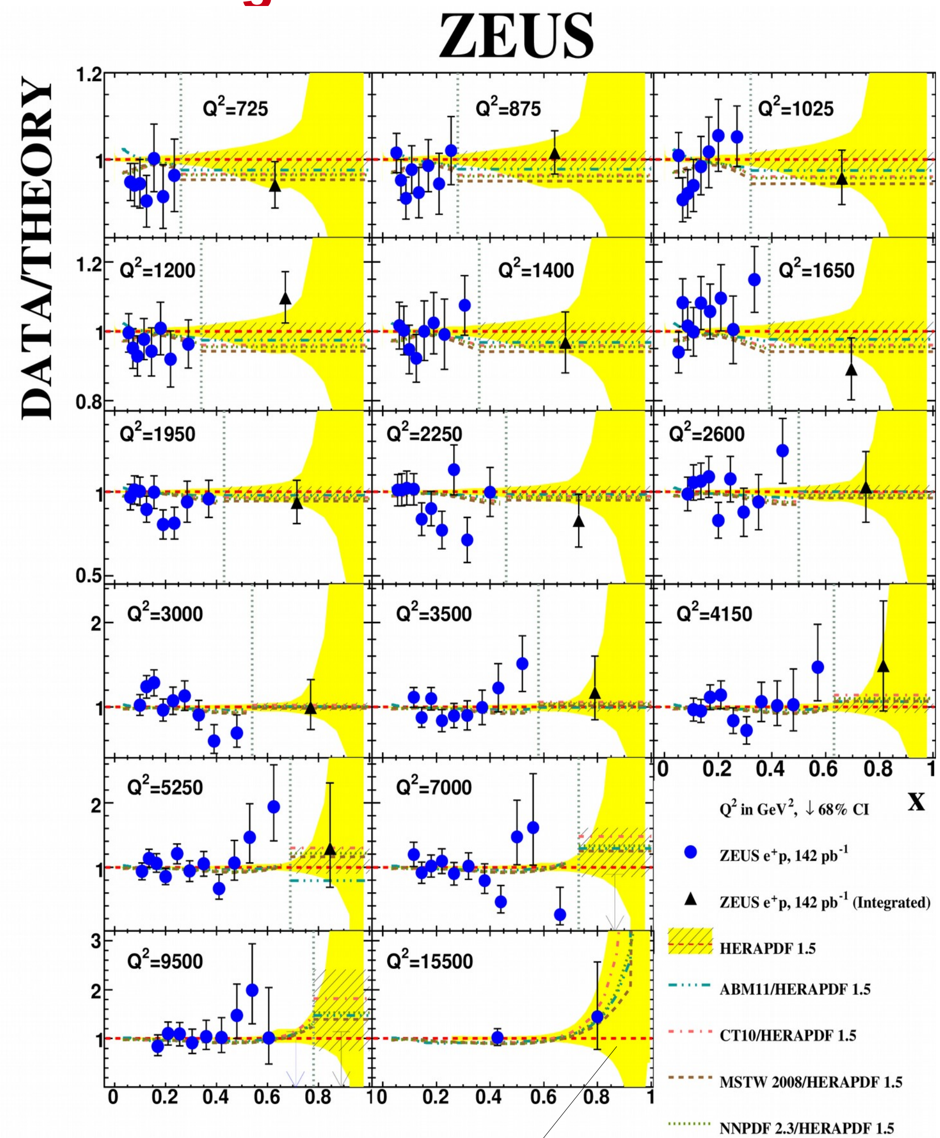
Ritu Aggarwal, Allen Caldwell

Overview

- Motivation
- Transfer Matrix for ZEUS detector for high-x data
- Using xfitter (instead of DISPRED) to calculate CTEQ5D cross sections for reweighting.
- Using ZMVFNS for CTEQ5D and corresponding ‘internally used schemes” for other Modern PDFs for reweighting (as listed in backup).
- Tables and Plots from the paper.
- Comparison of analysis (vector ν) with event by event reweighting and by using integrated born cross sections : NEW

Motivation of studying published high-x data



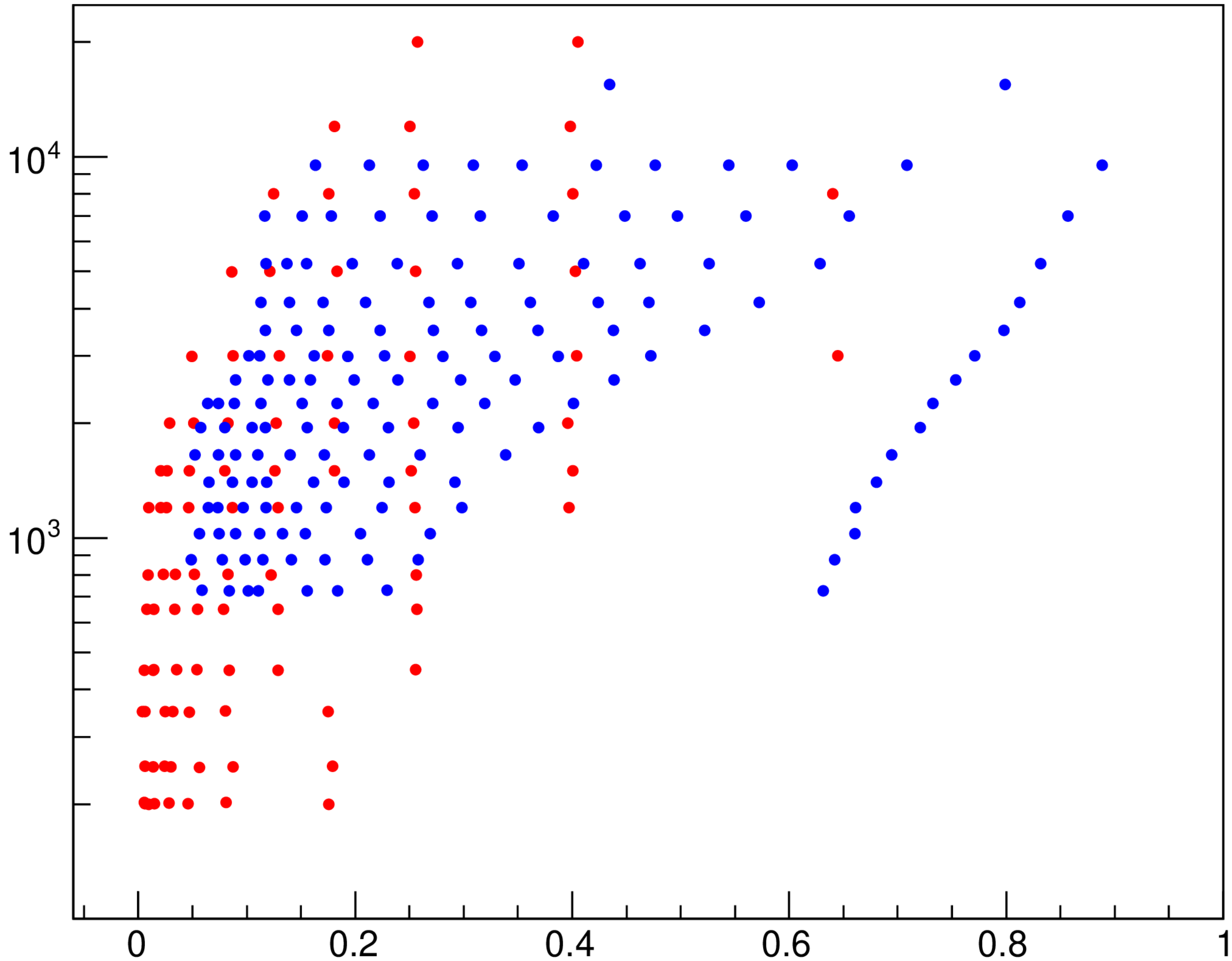


At present x upto 0.65 ZEUS data is included in PDF fits

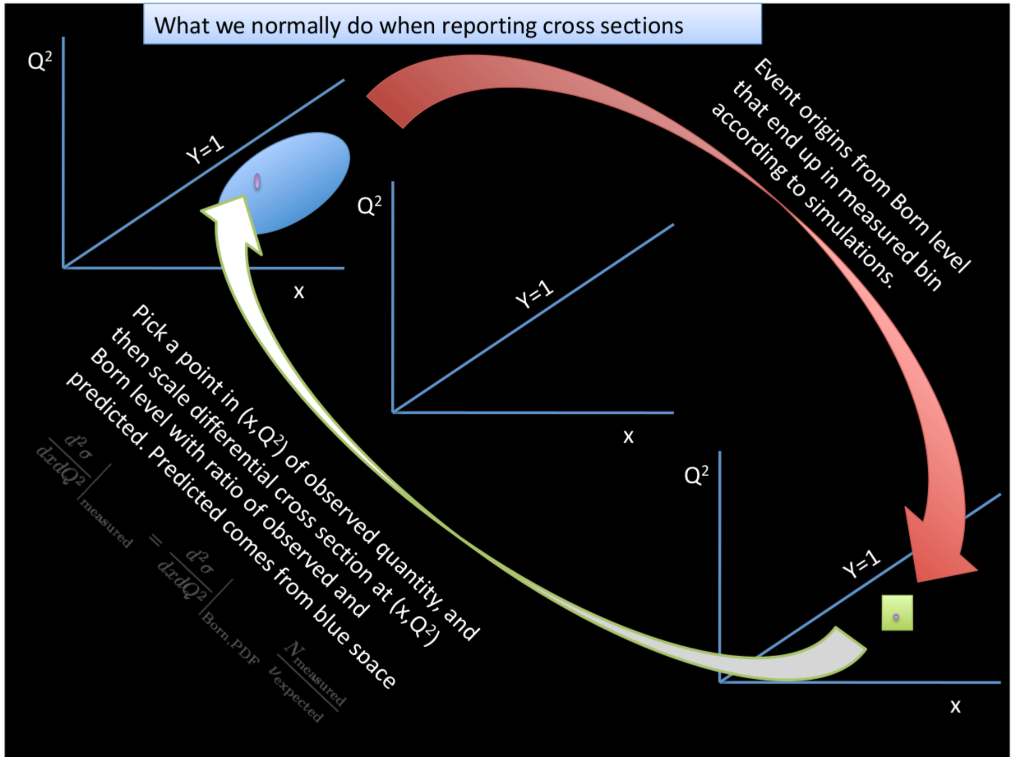
Note the uncertainity bands above x ~ 0.65, can high-x data impact here

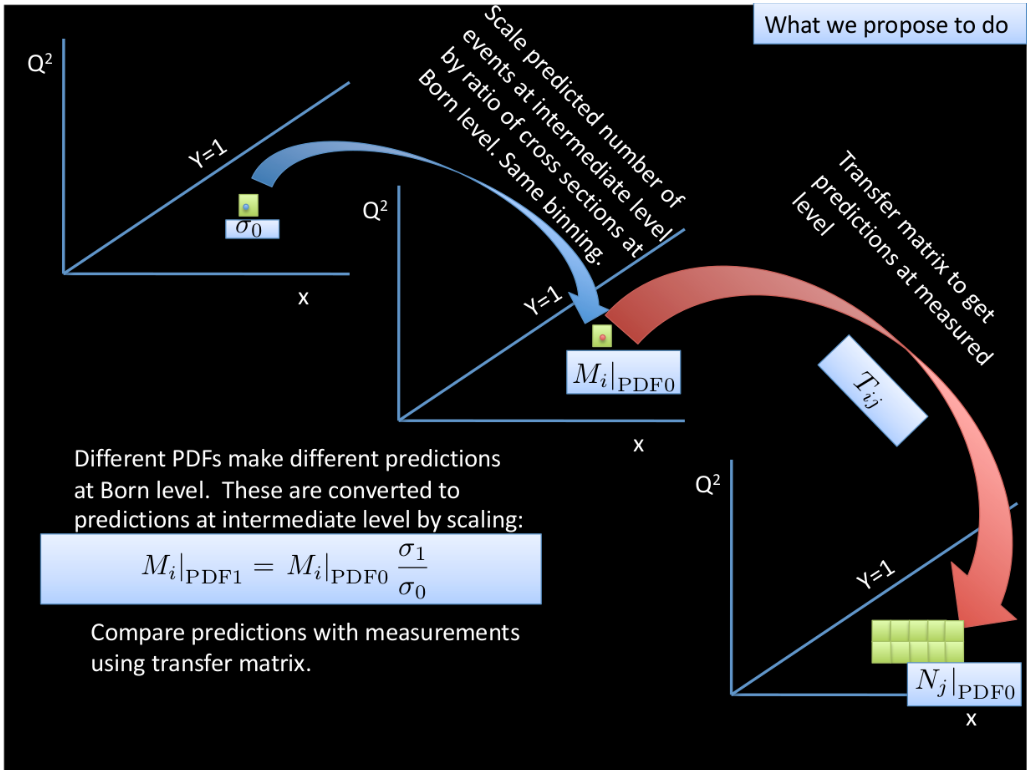
ZEUS Collaboration; H. Abramowicz et al. Measurement of Neutral Current e ± p Cross-Sections at High Bjorken x with the ZEUS Detector Phys. Rev. D 89 (2014) 072007

High-x cross sections Vs High-Q2 cross sections



From Born level to reconstruncted level





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.

Matrix for the detector response 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.

$$\mu_k = R\,\lambda_k$$

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

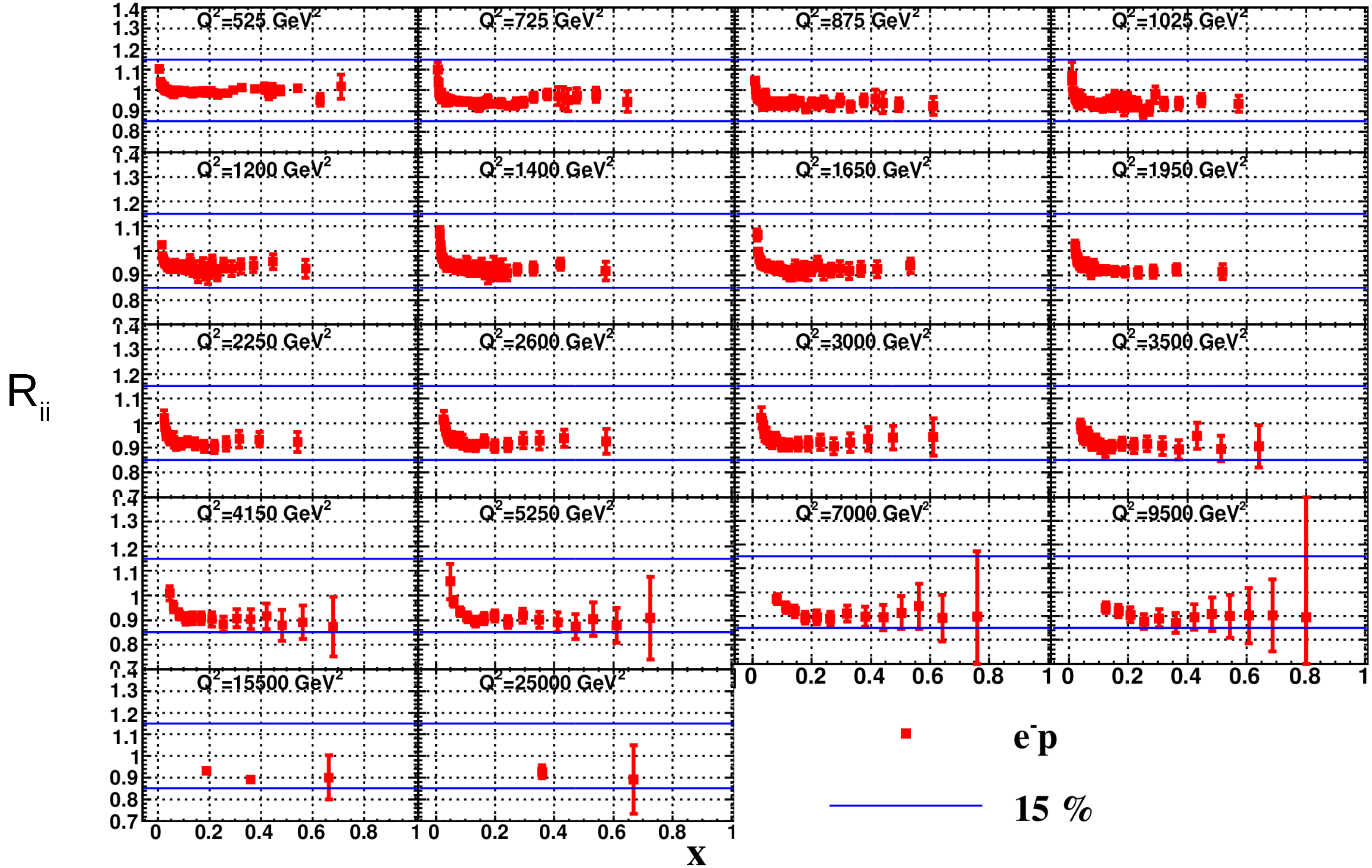
$$\nu_k = T\,\mu_k$$

R : Radiative corrections (calculated using HERACLES)
$\lambda_{i,k}$: born level cross sections in i^{th} bin for k^{th} PDF
T : has all detector and analysis effects
(probability of an event reconstructed in j^{th} bin to come from i^{th} true bin)

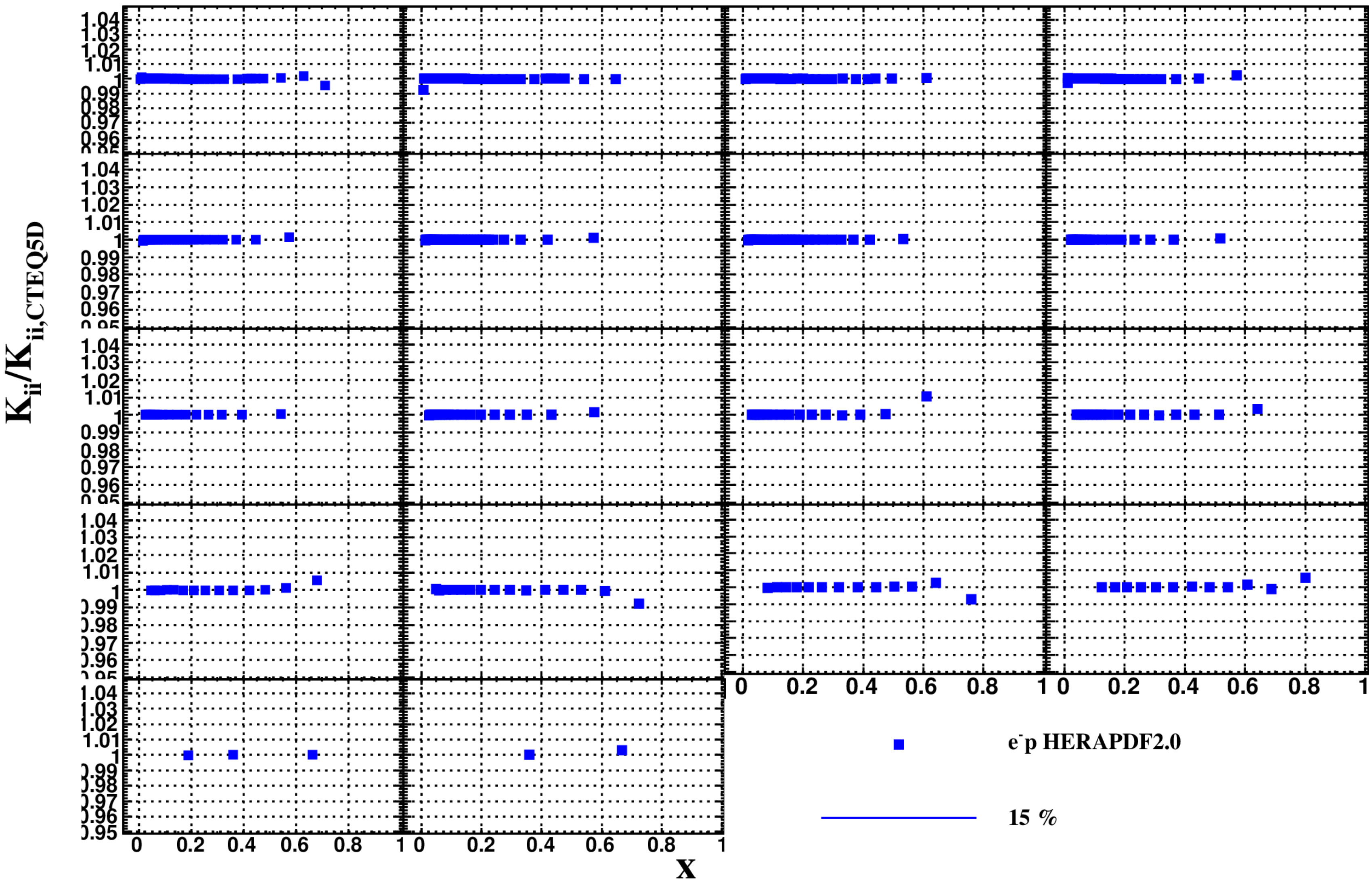
Radiative Corrections

- **Ratio of μ (high-x, with Radiative Corrections) and L*σ (Integrated cross section calculated from xfitter – master version).**
- **μ has running alpha (used in MC event generation).**
- **Integrated cross sections (σ) calculated with fixed alpha.**
- **Rii calculated for CTEQ5D and HERAPDF2.0 NNLO**
- **Comparison done**
- **A cross check is done by Andrii using MC samples generated with RAPGAP with and without radiative corrections.**

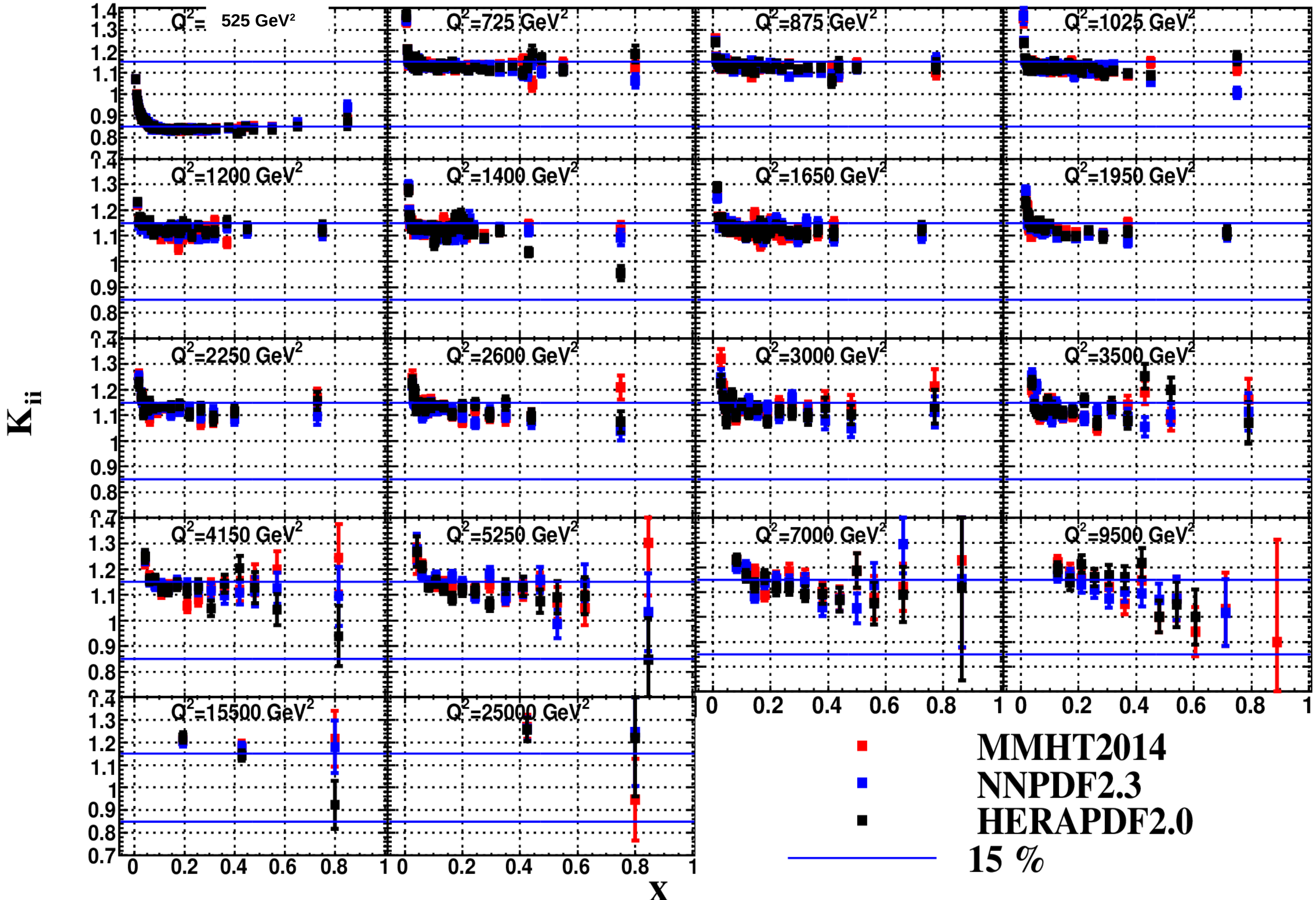
Ratio of μ_k (high-x, with Radiative Corrections) and L*σ_k (xfitter : without radiative corrections, alpha = 1/137.)



Ratio of Rii HERAPDF2.0/ Rii CTEQ5D

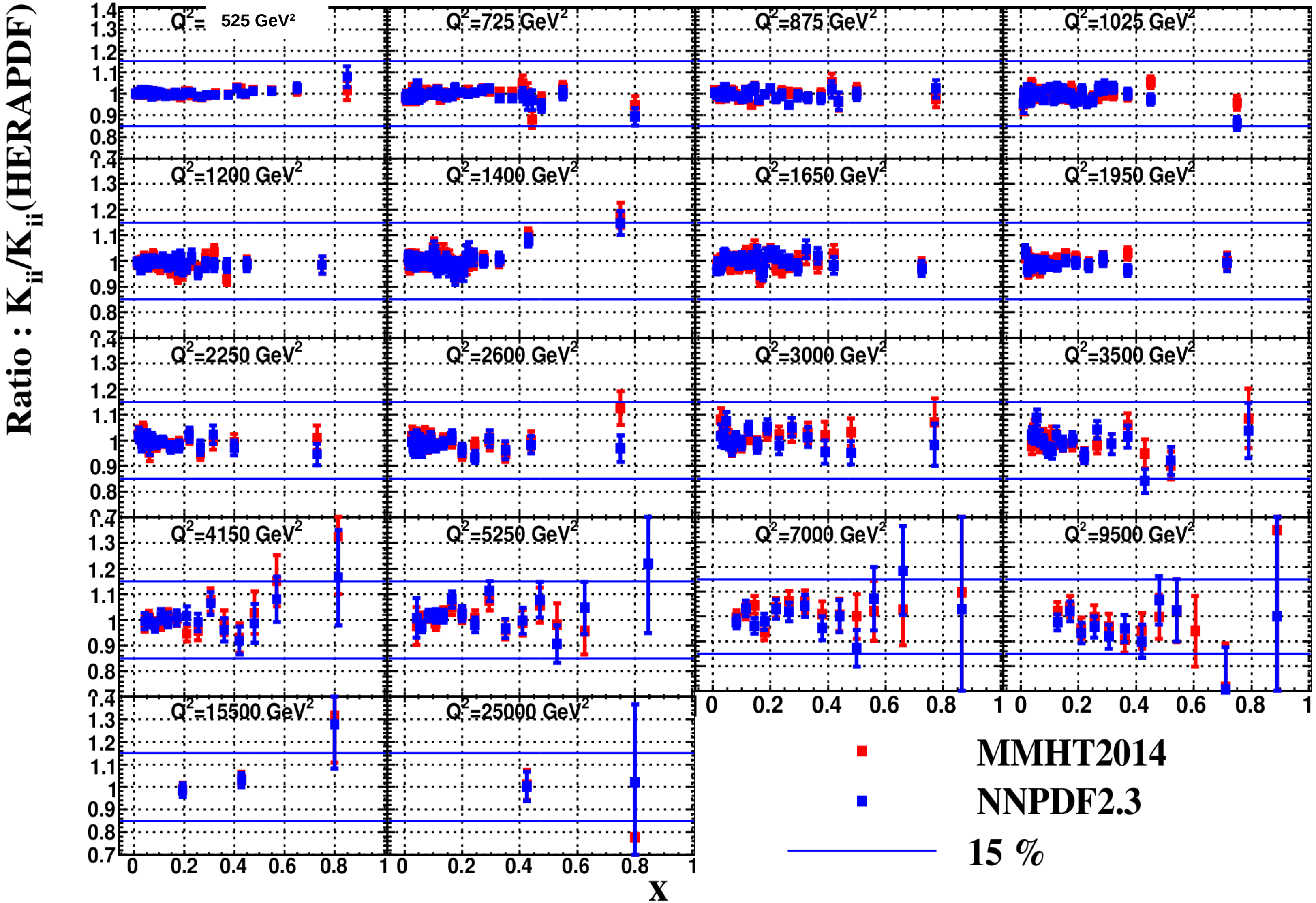


Ratio of generated events with and without radiative corrections for different PDFs using RAPGAP



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 Rii_k to $Rii_{,HERAPDF2.0}$



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

Transfer Matrix : Probability of an event reconstructed in j^{th} bin to come from i^{th} true bin

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

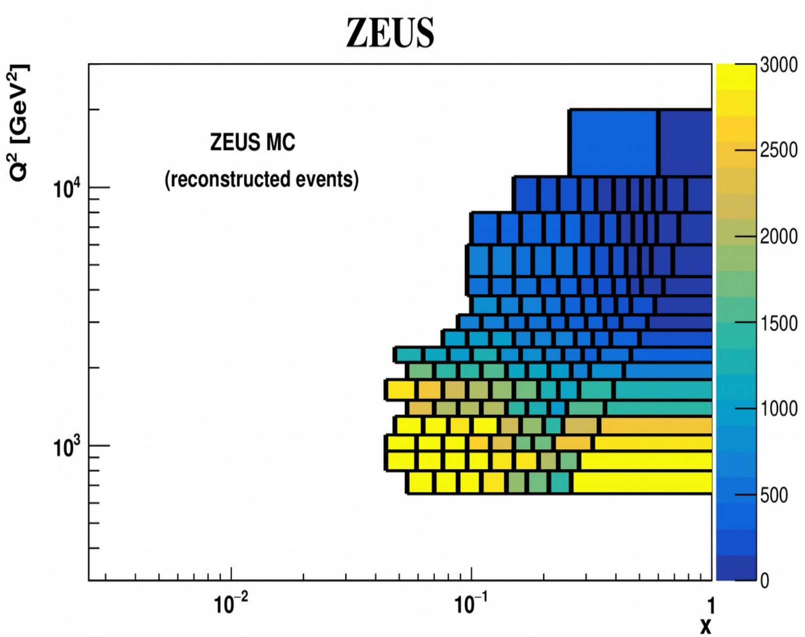
$$T_{ji} = \frac{\sum_{n=1}^{M_i} \omega_n I(n \in j)}{\sum_{n=1}^{M_i} \omega_n^{MC}}$$

T_{ij} = probability of an event reconstructed in j^{th} bin to come from i^{th} bin

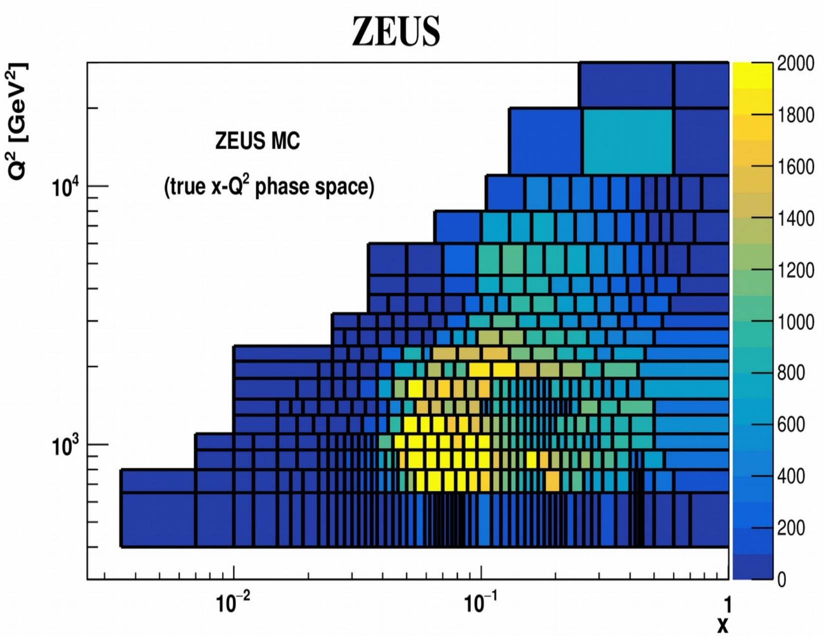
ω_m = MC weights given to m^{th} event in bin i

I = 1 if m^{th} event is reconstructed in bin j, else = 0

M_i = total events generated in i^{th} bin



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



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

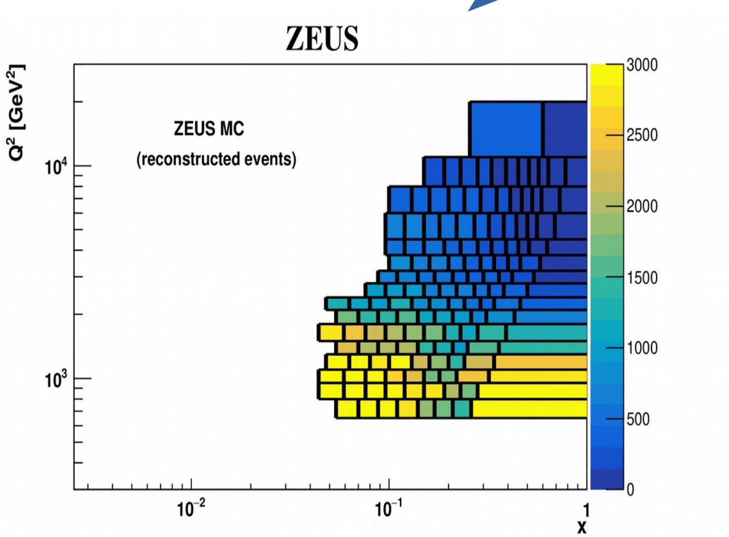
Note : MC samples used as in high-x paper.

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

$$\nu = T\, \mu$$

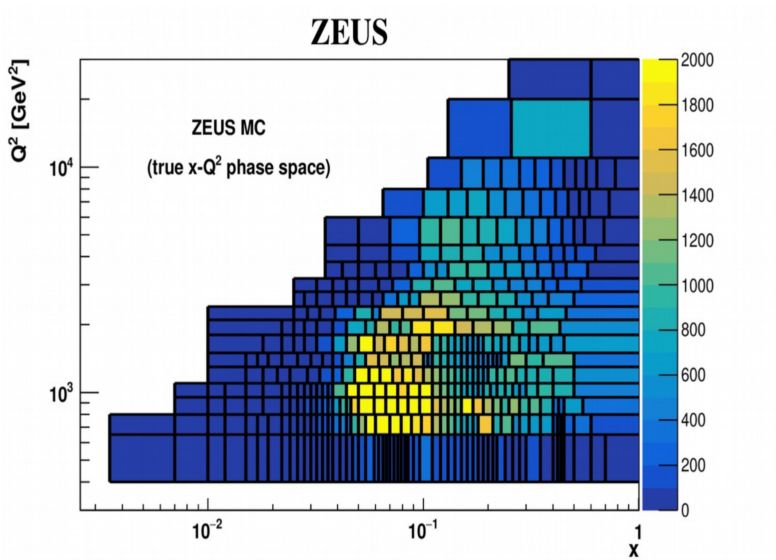
Transfer Matrix

(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)

Heavy Flavor schemes used internally for different PDFs

CT14nnlo	ACOT	Ref (https://arxiv.org/pdf/1506.07443.pdf)
ABMP16	Fixed flavor nf =3 taken	Ref(https://arxiv.org/pdf/1701.05838.pdf)
ABM 11	Fixed flavor nf =3	
MMHT2014	RT-OPT	Ref (https://arxiv.org/pdf/1412.3989.pdf)
HERAPDF2.0	RT-opt	Ref (Mandy)
NNPDF3.1	FONLL-C	Ref(https://arxiv.org/pdf/1706.00428.pdf)
NNPDF2.3	GMVFNS based on FONLL	Ref(https://arxiv.org/pdf/1207.1303.pdf)

Xfitter can work with
Nf=3 for FFN

Comparison of reconstructed events from HERAPDF2.0 with event by event reweighting and by using integrated born level cross section (e+p)

First few points

Q2	x	nu(from integrated)	nu(from reweigthing)	ratio
725	0.06	594.37	594.34	1.00005
725	0.08	440.19	440.17	1.00005
725	0.10	337.19	337.19	1
725	0.12	302.31	302.31	1
725	0.16	209.11	209.12	0.999952
725	0.19	195.19	195.20	0.999949
725	0.23	160.99	161.00	0.999938
725	0.63	415.41	415.41	1
875	0.05	590.14	590.10	1.00007
875	0.07	517.73	517.70	1.00006
875	0.09	440.96	440.95	1.00002
875	0.11	357.36	357.36	1
875	0.14	313.23	313.23	1
875	0.17	297.57	297.58	0.999966
875	0.21	215.15	215.16	0.999954
875	0.26	186.74	186.75	0.999946
875	0.64	500.88	500.90	0.99996
1025	0.05	429.06	429.02	1.00009
1025	0.07	374.53	374.50	1.00008
1025	0.09	337.53	337.52	1.00003
1025	0.11	290.99	290.99	1
1025	0.14	260.11	260.12	0.999962
1025	0.16	183.21	183.21	1
1025	0.20	182.96	182.97	0.999945
1025	0.27	279.33	279.35	0.999928
2600	0.75	28.01	28.03	0.999286
3000	0.10	73.25	73.24	1.00014
3000	0.12	80.15	80.15	1
3000	0.16	62.00	62.00	1
3000	0.19	64.36	64.36	1
3000	0.23	46.32	46.32	1
3000	0.28	42.54	42.55	0.999765
3000	0.33	35.08	35.09	0.999715
3000	0.39	23.17	23.17	1

Last few points

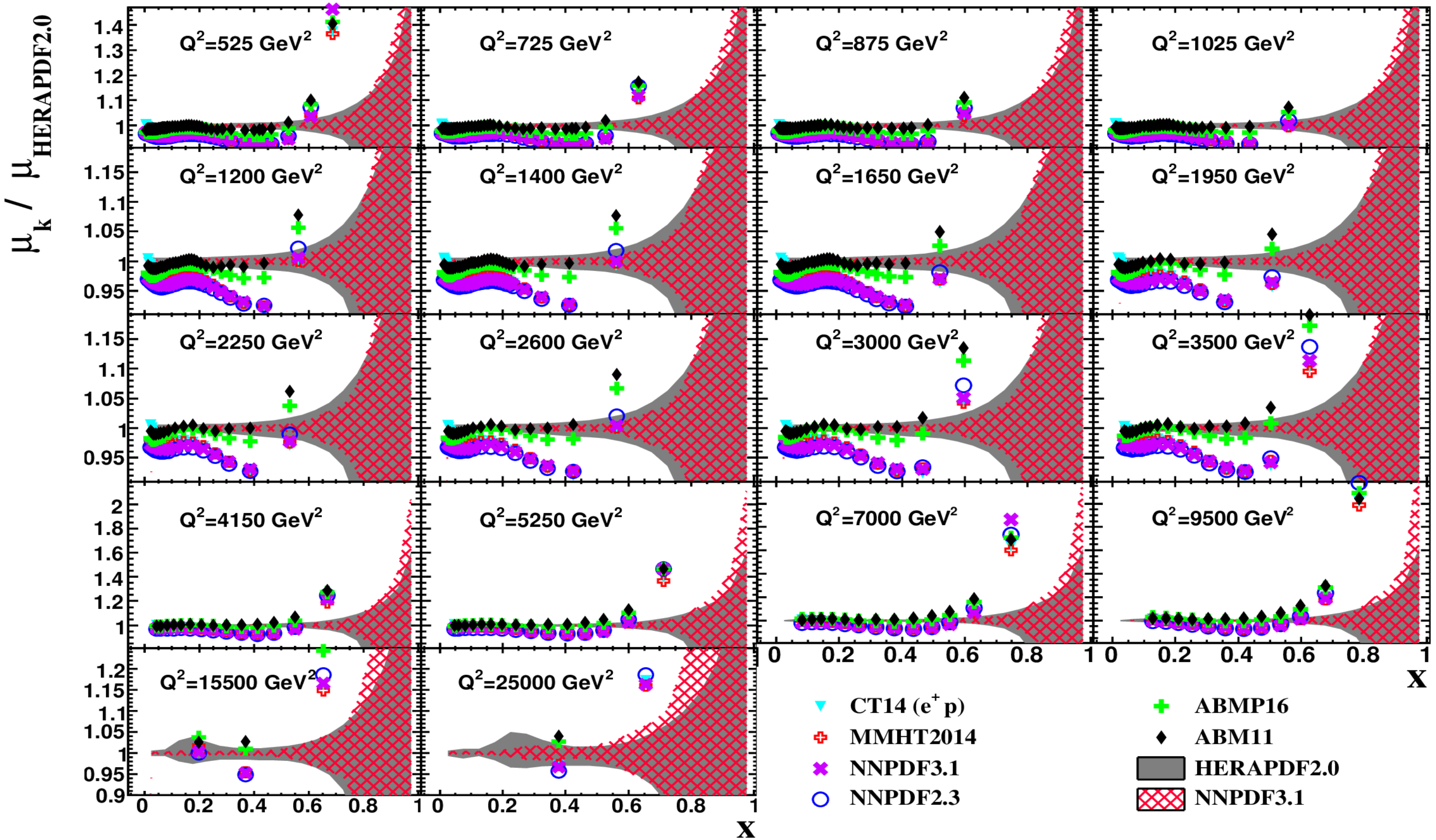
Q2	x	nu(from integrated)	nu(from reweigthing)	ratio
5250	0.47	13.49	13.49	1
5250	0.53	8.33	8.33	1
5250	0.62	7.12	7.12	1
5250	0.84	2.01	2.02	0.99505
7000	0.12	36.52	36.52	1
7000	0.15	36.85	36.85	1
7000	0.18	42.34	42.35	0.999764
7000	0.22	34.51	34.51	1
7000	0.27	33.01	33.01	1
7000	0.32	28.39	28.39	1
7000	0.38	19.58	19.58	1
7000	0.44	11.75	11.75	1
7000	0.50	7.42	7.42	1
7000	0.56	4.44	4.44	1
7000	0.66	3.57	3.57	1
7000	0.86	0.59	0.59	1
9500	0.17	22.52	22.52	1
9500	0.21	23.07	23.07	1
9500	0.26	22.40	22.40	1
9500	0.31	17.15	17.15	1
9500	0.36	15.77	15.77	1
9500	0.42	10.10	10.10	1
9500	0.48	6.44	6.44	1
9500	0.54	3.64	3.64	1
9500	0.60	2.29	2.29	1
9500	0.71	1.37	1.37	1
9500	0.89	0.15	0.15	1
15500	0.43	42.55	42.55	1
15500	0.80	2.54	2.53	1.00395

Difference Negligible!

As on Sep 10 : Integration with xfitter-master had bugs for other schemes. Bugs were reported! This check was done for HERAPDF2.0

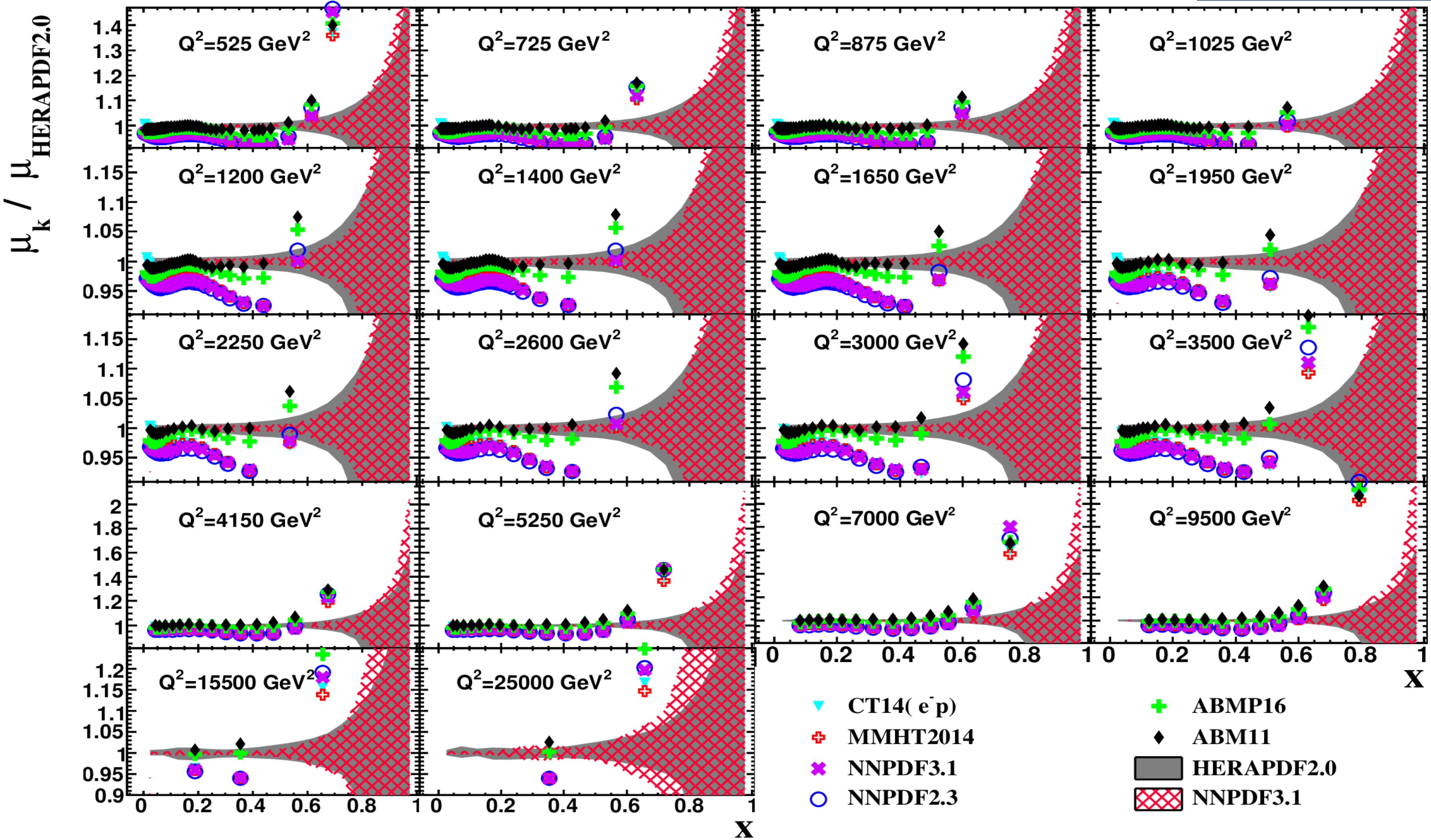
Average ratio of Born level cross sections in different PDFs to HERAPDF2.0NNLO for M bins (e+p)

Figure 2



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

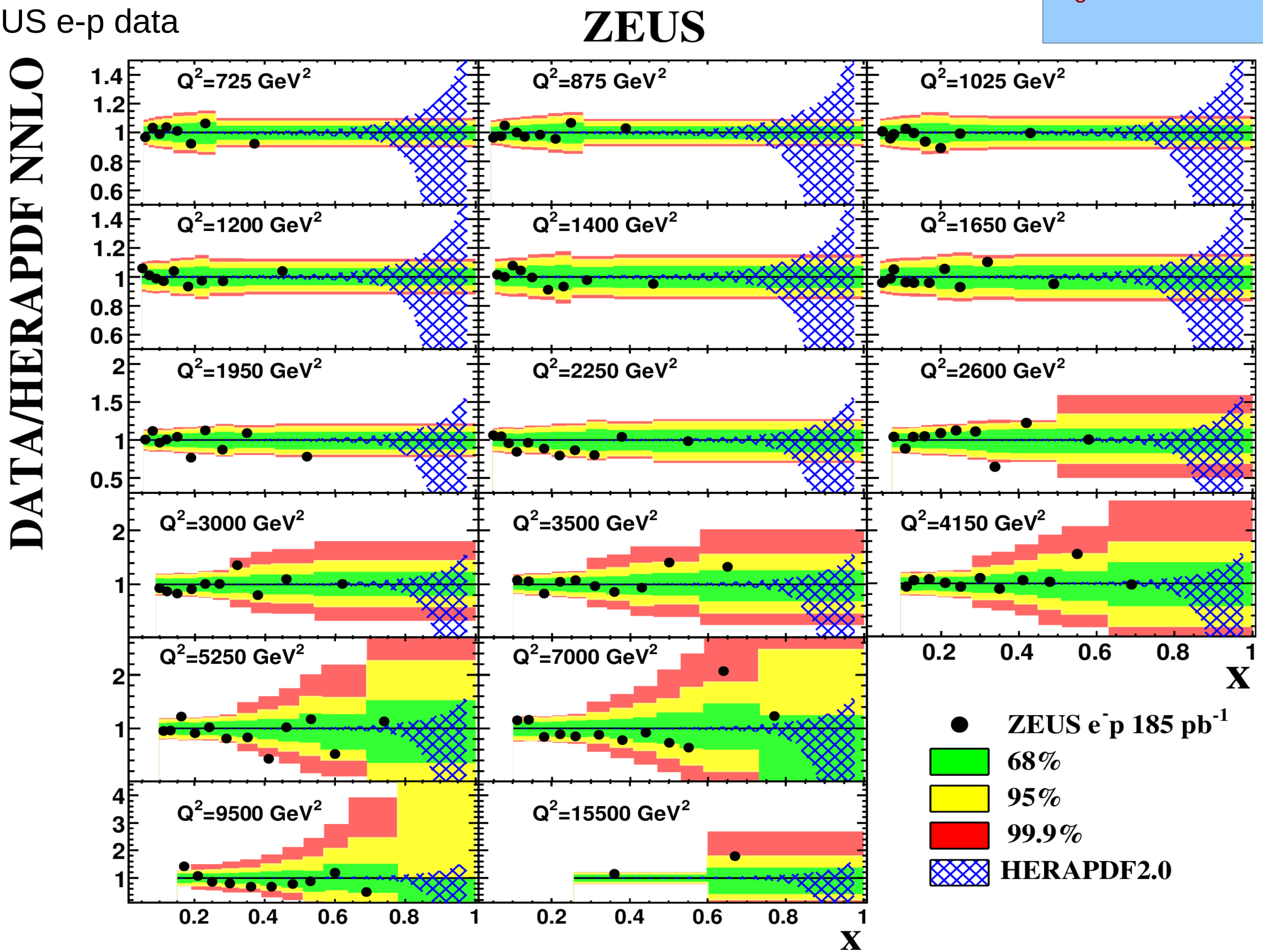
Figure 3



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

Figure 5

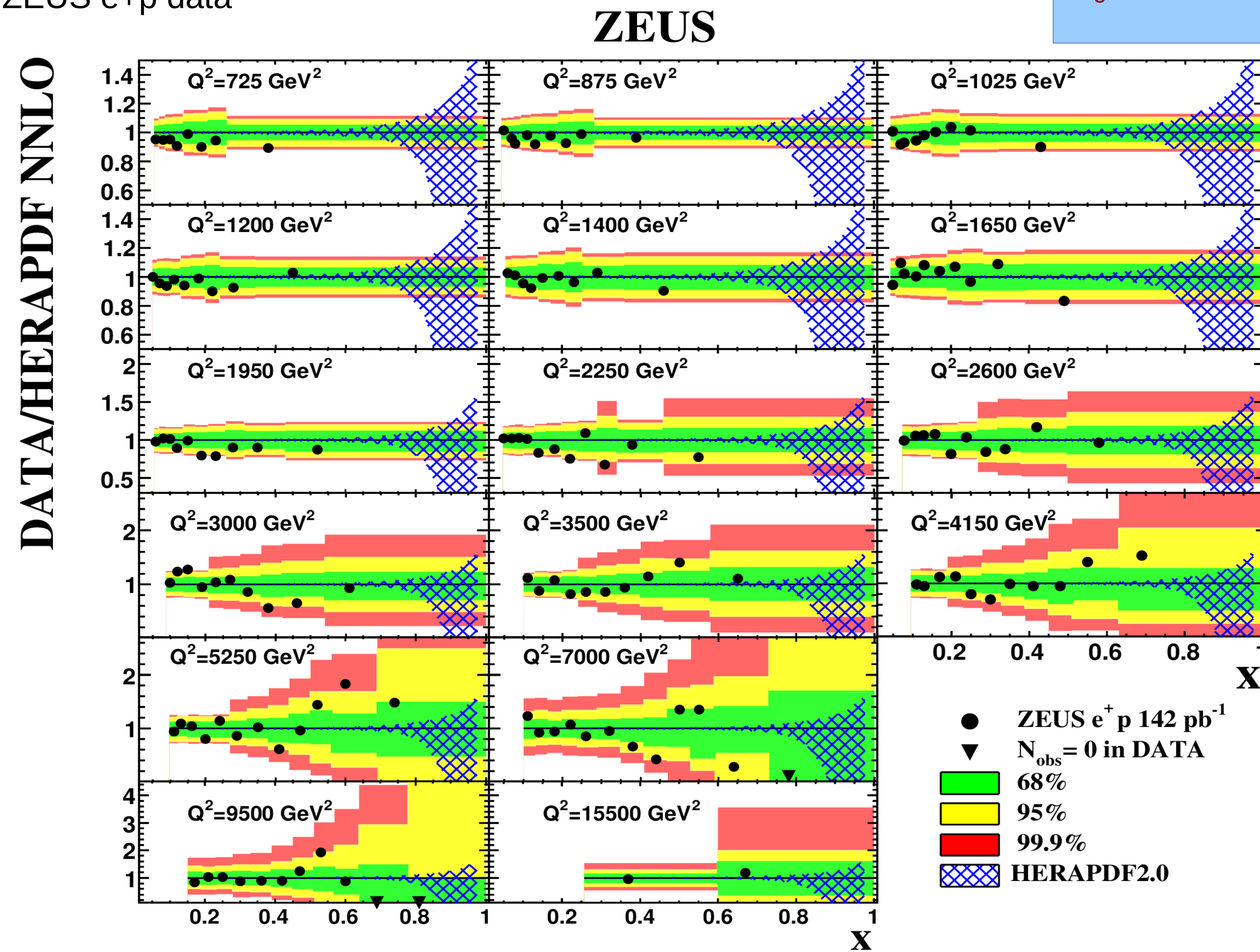
ZEUS e-p data



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

ZEUS e+p data

Figure 4



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	$\Delta\chi^2$
$HERAPDF2.0$	1.0	$2.8*10^{-02}$	0.00	1.0	0.35	0.00
$CT14$	$7.6*10^{-03}$	$3.2*10^{-03}$	9.8	$5.9*10^{+05}$	0.82	-27
$MMHT2014$	$2.1*10^{-03}$	$2.3*10^{-03}$	12	$4.7*10^{+05}$	0.82	-26
$NNPDF3.1$	$3.2*10^{-06}$	$3.9*10^{-04}$	25	$9.0*10^{+04}$	0.73	-23
$NNPDF2.3$	$2.3*10^{-07}$	$1.3*10^{-04}$	31	$4.2*10^{+04}$	0.70	-21
$ABMP16$	$9.0*10^{-01}$	$2.6*10^{-02}$	0.21	$6.1*10^{+02}$	0.64	-13
$ABM11$	$7.2*10^{-01}$	$3.3*10^{-02}$	0.67	2.8	0.45	-2.1

Table 1: *The Bayes Factor, p-values and $\Delta\chi^2$ from comparisons of predictions using different NNLO PDF sets to the observed numbers of events. 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

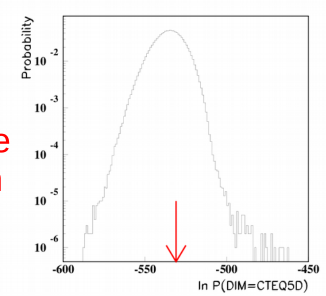


FIGURE 4. Distribution of expected values for $\ln P(D|M = CTEQ)$ for the e^+p data set. The arrow shows the value found in the data.

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			
	lower x		higher x		lower x		higher x	
PDF	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$
$HERAPDF2.0$	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0
$CT14$	$6.0 * 10^{-01}$	1.0	$1.3 * 10^{-02}$	8.7	$1.6 * 10^{+03}$	-15	$3.6 * 10^{+02}$	-12
$MMHT2014$	$7.1 * 10^{-02}$	5.3	$2.9 * 10^{-02}$	7.1	$1.3 * 10^{+03}$	-14	$3.7 * 10^{+02}$	-12
$NNPDF3.1$	$9.1 * 10^{-05}$	19	$3.5 * 10^{-02}$	6.7	$2.5 * 10^{+02}$	-11	$3.6 * 10^{+02}$	-12
$NNPDF2.3$	$8.0 * 10^{-06}$	23	$2.9 * 10^{-02}$	7.1	$1.2 * 10^{+02}$	-9.5	$3.7 * 10^{+02}$	-12
$ABMP16$	$2.3 * 10^{-01}$	3.0	$4.0 * 10^{+00}$	-2.7	$4.8 * 10^{+01}$	-7.8	$1.3 * 10^{+01}$	-5.1
$ABM11$	$2.3 * 10^{-01}$	3.0	$3.2 * 10^{+00}$	-2.3	$4.2 * 10^{+00}$	-2.9	$6.7 * 10^{-01}$	0.8

Table 2: *The Bayes Factor and $\Delta\chi^2$ (calculated relative to HERAPDF2.0) from comparisons of predictions using different NNLO PDF sets to the observed numbers of events are shown for two different x ranges. The 'higher x' region is defined as the highest three x bins in each Q^2 range. The remaining x range is labeled 'lower x'*

A MMHT, CT, NNPDF ABM better for e+P data, HERAPDF2.0 better for e-p data

Statistical and systematic uncertainties

Type of Systematic Uncertainties :

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

Type I :

1) Luminosity uncertainty scaling μ 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.

Internal Document Table 1

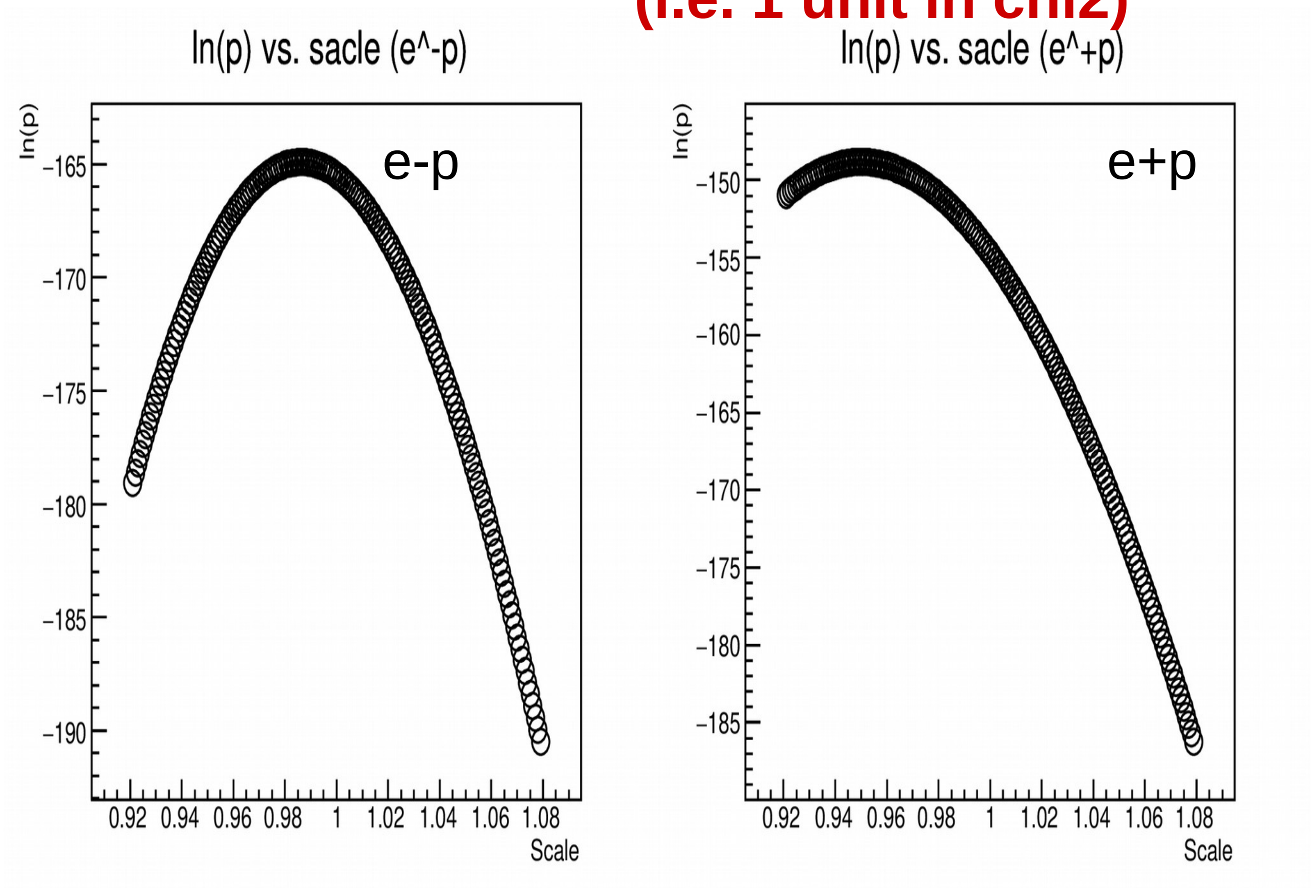
	e^-p				e^+p			
PDF	lower x		higher x		lower x		higher x	
	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$
$+1.8$ %								
$HERAPDF2.0$	$1.7 * 10^{-03}$	12.77	0.08	5.04	$8.65 * 10^{-05}$	18.71	0.01	9.79
$CT14$	$2.8 * 10^{+01}$	-6.67	15.99	-5.54	$4.94 * 10^{-02}$	6.01	0.37	2.01
$MMHT2014$	$1.1 * 10^{+02}$	-9.49	12.92	-5.12	$1.13 * 10^{-01}$	4.36	0.31	2.33
$NNPDF3.1$	$4.9 * 10^{+03}$	-16.98	15.26	-5.45	$2.16 * 10^{+00}$	-1.54	0.36	2.05
$NNPDF2.3$	$1.8 * 10^{+04}$	-19.59	18.28	-5.81	$5.95 * 10^{+00}$	-3.57	0.41	1.76
$ABMP16$	$3.7 * 10^{-01}$	2.01	0.30	2.39	$2.60 * 10^{-03}$	11.90	0.02	7.85
$ABM11$	$9.6 * 10^{-03}$	9.29	0.07	5.28	$2.95 * 10^{-04}$	16.25	0.01	9.95
-1.8 %								
$HERAPDF2.0$	$6.9 * 10^{-01}$	0.74	1.70	-1.06	$9.27 * 10^{+01}$	-9.06	30.39	-6.83
$CT14$	$4.9 * 10^{-05}$	19.85	0.01	9.35	$1.81 * 10^{-01}$	3.42	0.66	0.83
$MMHT2014$	$1.2 * 10^{-05}$	22.62	0.01	8.93	$8.01 * 10^{-02}$	5.05	0.77	0.51
$NNPDF3.1$	$3.1 * 10^{-07}$	29.99	0.01	9.26	$4.39 * 10^{-03}$	10.86	0.67	0.79
$NNPDF2.3$	$8.5 * 10^{-08}$	32.56	0.01	9.61	$1.62 * 10^{-03}$	12.85	0.59	1.07
$ABMP16$	$3.5 * 10^{-03}$	11.32	0.46	1.55	$3.26 * 10^{+00}$	-2.36	11.69	-4.92
$ABM11$	$1.2 * 10^{-01}$	4.16	1.91	-1.29	$2.77 * 10^{+01}$	-6.64	32.92	-6.99

Conclusions :

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

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

Numbers updated in the paper



Only the high-x data taken

	scale	LnP
e-p :	0.986	-164.893
	0.999	-165.394
	0.974	-165.383

	Scale	LnP
e+p	:0.95	-148.823
	0.965	-149.361
	0.936	-149.318

Data sensitivity ~ +-1.2% +-1.4%

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

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

Where Mi are the total number of events Generated in MC

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

Internal Document Table 2

	e^-p		e^+p	
Systematic	$P1/P2$	$\Delta\chi^2$	$P1/P2$	$\Delta\chi^2$
Electron energy scale	0.504 0.099	1.37 4.62	0.103 0.850	4.54 0.33
Electron energy resolution	0.224 1.309	2.99 −0.54	1.904 0.217	−1.29 3.06
Hadronic energy scale	0.770 1.146	0.52 −0.27	0.696 1.112	0.72 −0.21
MEPS/Ariadne reweighting	3.673 0.110	−2.60 4.41	1.068 0.605	−0.13 1.00
F-BCal Crack cut	0.218 4.651	3.05 −3.07	1.080 0.868	−0.15 0.28
Electron isolation cut	1.369 0.094	−0.63 4.74	0.253 11.246	2.75 −4.84
FCAL alignment (x)	1.399 0.939	−0.67 0.13	1.014 0.974	−0.03 0.05
FCAL alignment (y)	1.251 0.887	−0.45 0.24	1.608 0.623	−0.95 0.95

Table 2: (FOR ZEUS INTERNAL USE) The ratio of probabilities and the corresponding $\Delta\chi^2$ for various systematic checks performed for the e^-p and e^+p data sets.

Normalization is the main uncertainty

Systematic Errors : Considering various vectors for HERAPDF2.0

Internal Document Table 3 and 4

	e^-p				e^+p			
	lower x		higher x		lower x		higher x	
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	0.98	0.04	0.99	0.02	0.84	0.34	1.00	−0.00
2	0.99	0.02	1.00	0.00	1.12	−0.23	0.96	0.08
3	1.08	−0.15	1.05	−0.10	1.48	−0.78	1.09	−0.17
4	0.89	0.24	0.94	0.12	0.63	0.92	0.88	0.25
5	1.49	−0.79	1.19	−0.35	1.55	−0.88	2.09	−1.47
6	0.64	0.90	0.78	0.50	0.60	1.02	0.44	1.64
7	1.31	−0.54	1.22	−0.40	2.91	−2.14	1.55	−0.87
8	0.65	0.86	0.80	0.46	0.30	2.43	0.61	0.98
9	0.59	1.04	0.75	0.57	0.43	1.70	0.40	1.84
10	1.55	−0.88	1.22	−0.40	2.12	−1.51	2.28	−1.65
11	0.90	0.20	0.94	0.13	0.80	0.45	0.71	0.68
12	1.07	−0.14	1.04	−0.09	1.19	−0.35	1.34	−0.58
13	0.85	0.31	1.12	−0.23	1.65	−1.00	0.56	1.16
14	1.07	−0.14	0.81	0.42	0.55	1.18	1.62	−0.96
15	1.06	−0.12	1.10	−0.19	1.31	−0.55	1.00	−0.01
16	0.93	0.14	0.91	0.19	0.84	0.35	0.97	0.06
17	0.91	0.19	0.92	0.16	1.15	−0.28	0.95	0.10
18	1.04	−0.09	1.06	−0.11	0.80	0.44	0.98	0.04
19	0.92	0.17	1.08	−0.15	0.83	0.38	0.91	0.19
20	1.05	−0.09	0.92	0.16	1.14	−0.26	1.04	−0.07
21	1.57	−0.91	1.31	−0.54	4.10	−2.82	2.35	−1.71
22	0.42	1.73	0.62	0.96	0.15	3.77	0.30	2.39
23	1.26	−0.46	1.19	−0.35	1.17	−0.31	1.69	−1.04
24	0.88	0.26	0.81	0.43	0.72	0.67	0.82	0.39
25	0.83	0.38	0.88	0.26	0.71	0.69	0.76	0.54
26	1.30	−0.53	1.12	−0.23	1.37	−0.63	1.69	−1.05
27	1.29	−0.51	1.18	−0.34	1.78	−1.15	1.35	−0.60
28	1.15	−0.28	1.18	−0.33	1.81	−1.19	1.26	−0.46

Table 3: (FOR ZEUS INTERNAL USE) 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.

	e^-p				e^+p			
	lower x		higher x		lower x		higher x	
Variance 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	0.94	0.13	0.98	0.05	0.86	0.31	0.92	0.16
2	0.95	0.11	1.15	−0.27	1.52	−0.83	1.32	−0.56
3	0.88	0.25	0.95	0.10	0.83	0.38	0.85	0.33
4	1.05	−0.10	1.02	−0.04	1.12	−0.23	1.07	−0.14
5	1.17	−0.31	1.12	−0.23	1.06	−0.11	1.50	−0.81
6	1.13	−0.25	1.15	−0.28	1.74	−1.11	1.25	−0.44
7	0.97	0.05	0.98	0.03	0.96	0.09	0.96	0.08
8	1.00	0.01	1.01	−0.02	0.94	0.12	1.03	−0.05
9	1.31	−0.54	1.17	−0.31	1.89	−1.27	1.51	−0.82
10	0.90	0.21	0.99	0.03	0.74	0.61	0.93	0.14
11	1.42	−0.69	1.22	−0.40	2.35	−1.71	1.77	−1.14
12	0.68	0.77	0.81	0.43	0.48	1.48	0.61	0.99
13	1.22	−0.40	1.19	−0.36	2.15	−1.53	2.36	−1.72

Table 4: (FOR ZEUS INTERNAL USE) 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 vector (labeled 0) is the nominal prediction.

Summary of Changes (Post EB2)

- Using xfitter (instead of DISPRED) to calculate CTEQ5D cross sections for reweighting.
- Using ZMVFNS for CTEQ5D and RT-OPT for all other Modern PDFs for reweighting.
- All the tables and Plots updated.
- Figures containing the comparison of reconstructed events with data with different PDFs icnluded in a separate document for external sharing.
- Tables for the systematical error studies included in s separate document for internal use only.
- Check done for HERAPDF for the analysis with event by event reweighting and by born cross sections, differences found to be negligible.
- Tables with details of events reconstructed from different PDFs in the appedix.
- Changes in text in the Introduction and other sections done for better understanding.

Thanks

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.

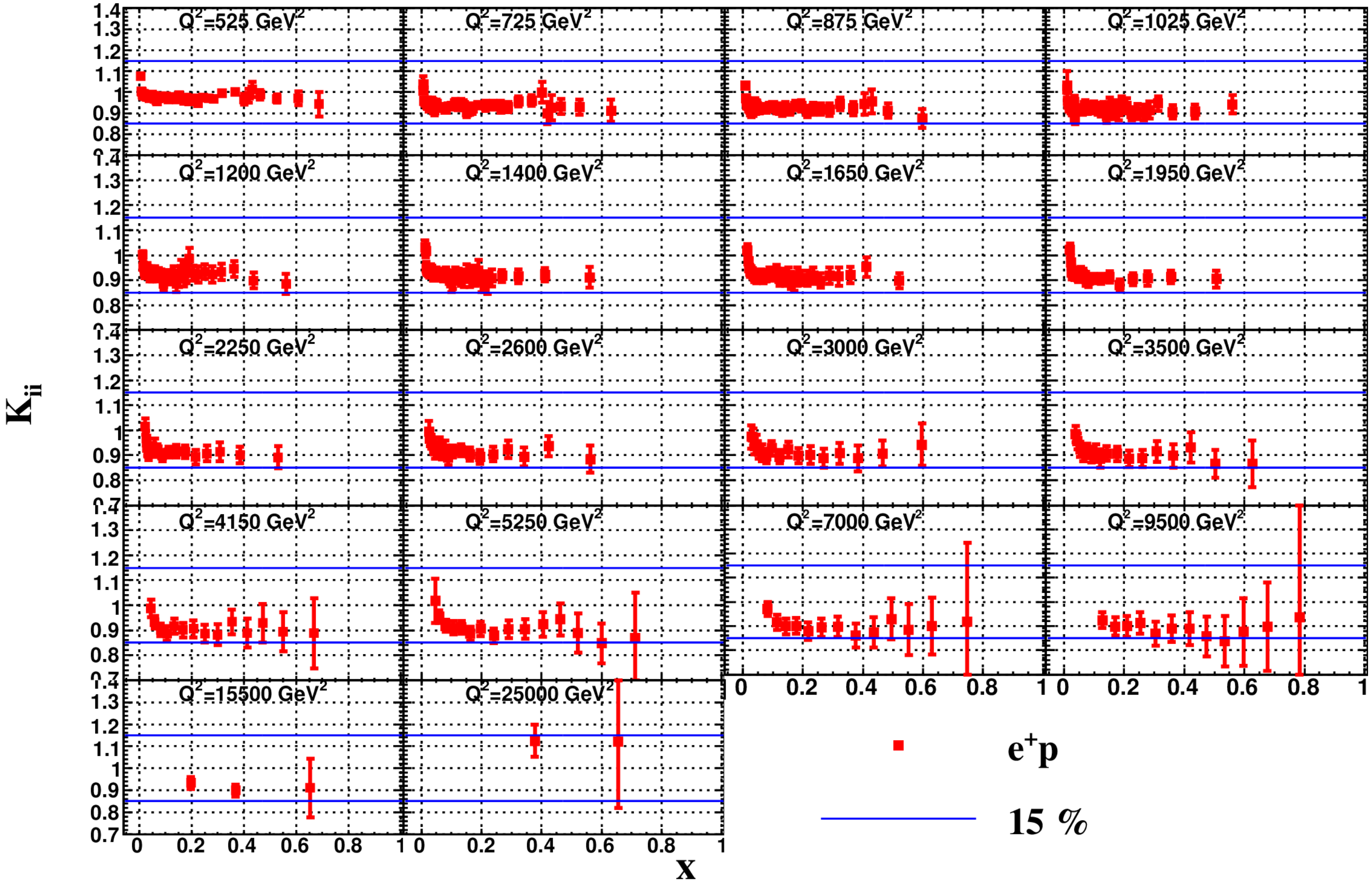
Comparing Total Probability for different Pdfs in different x range (x <= 0.6 & x > 0.6)

	e^-p				e^+p			
	lower x		higher x		lower x		higher x	
PDF	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$	P1/P2	$\Delta\chi^2$
$CT14$	$2.64 * 10^{-03}$	11.87	0.48	1.46	4722.06	−16.92	112.06	−9.44
$HERAPDF2.0$	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
$MMHT2014$	$3.79 * 10^{-03}$	11.15	0.53	1.26	4192.28	−16.68	111.37	−9.43
$NNPDF2.3$	$7.79 * 10^{-07}$	28.13	0.68	0.77	488.82	−12.38	96.02	−9.13
$NNPDF3.1$	$1.01 * 10^{-05}$	23.01	0.74	0.61	953.37	−13.72	100.33	−9.22
$ABMP16$	$4.96 * 10^{-01}$	1.40	1.48	−0.78	2305.38	−15.49	53.32	−7.95
$abm11$	$1.17 * 10^{-05}$	22.71	1.42	−0.70	61.68	−8.24	0.66	0.83

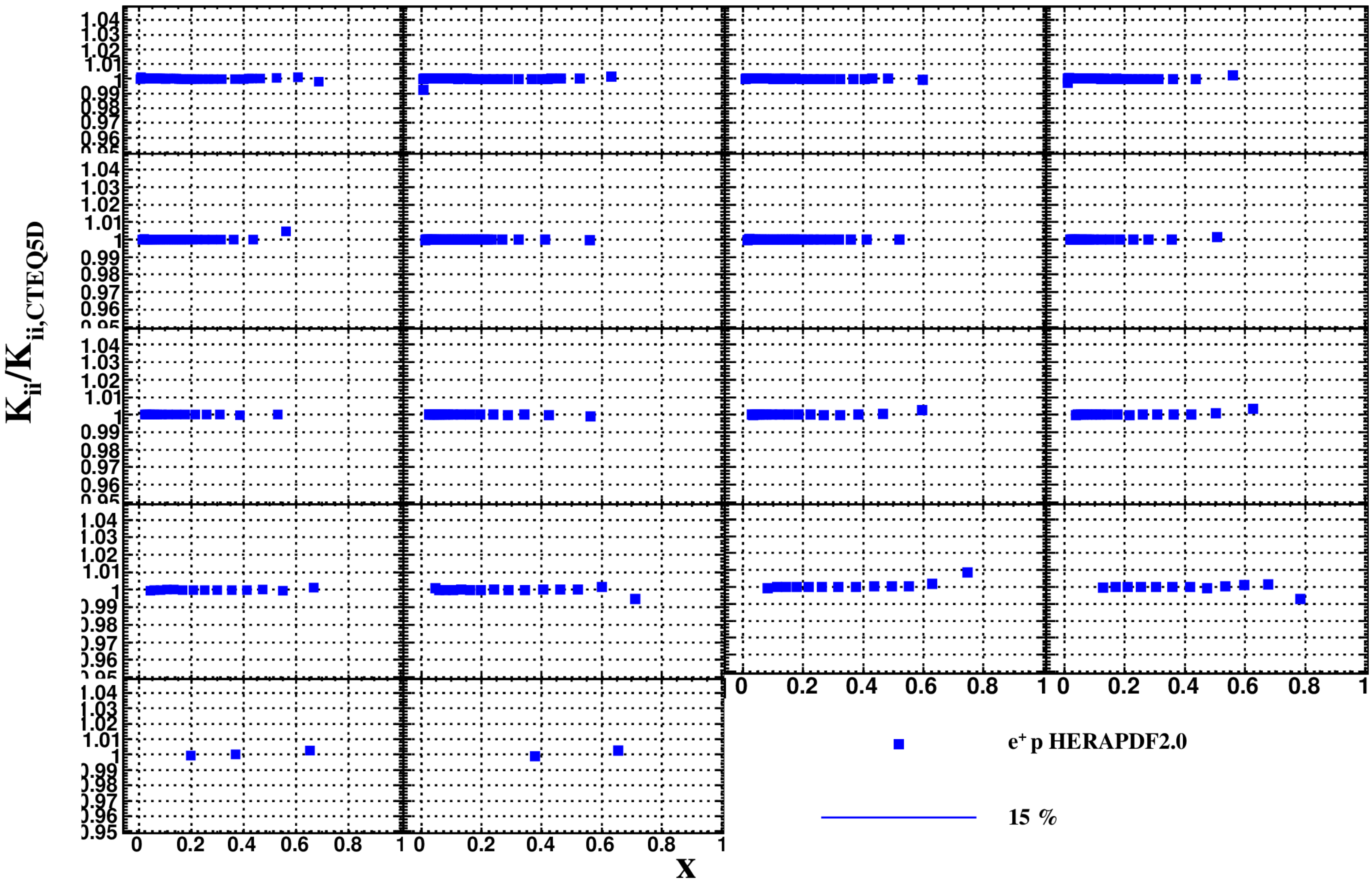
Table A.7: *(OLD TABLE)The Bayes Factor and $\Delta\chi^2$ (calculated relative to HERAPDF2.0) from comparisons of predictions (at NNLO) using different PDF sets to the observed numbers of events are shown for two different x ranges for the e^+p and e^-p data sets.*

At high x MMHT, CT, NNPDF ABM better for e+P data. Disagreement primarily from low x!

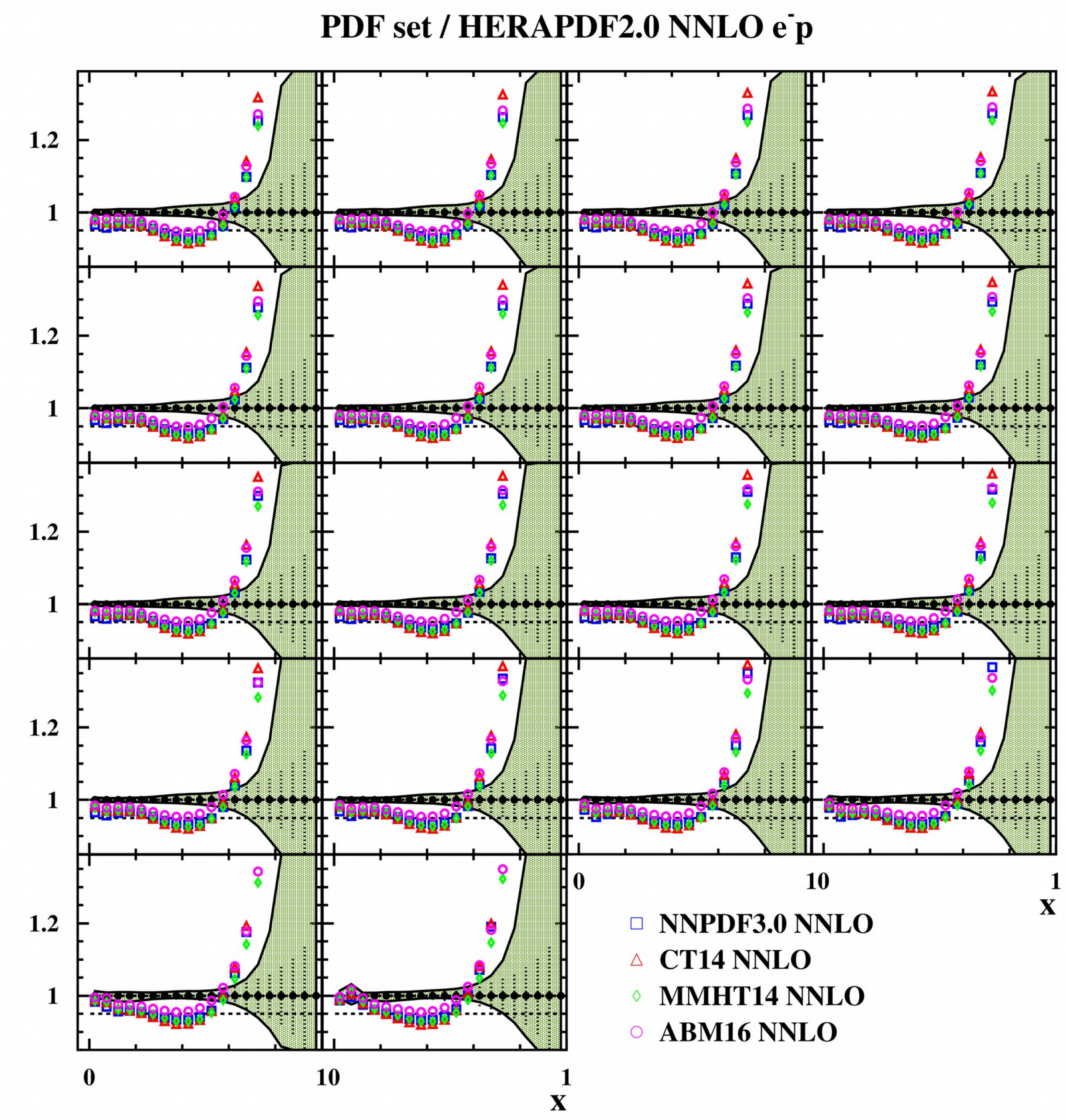
Ratio of m_k (high-x, with Radiative Corrections) and $L^*\sigma_k$ (xfitter : without radiative corrections, alpha = 1/137.)



Ratio of Kii HERAPDF2.0/ Kii CTEQ5D



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 !!!)

Comments from PP-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 at PP2

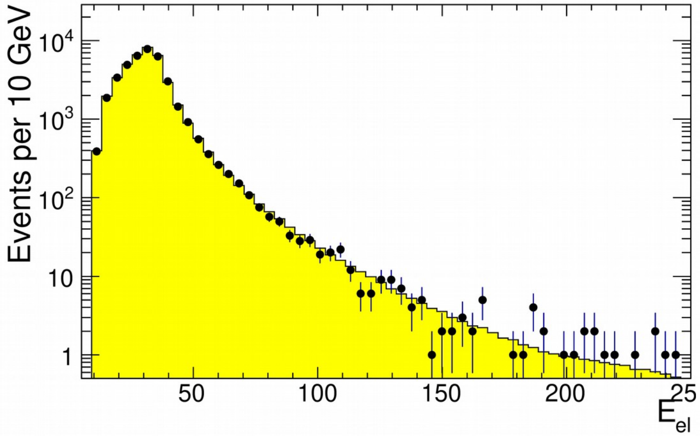
Major Changes in text :

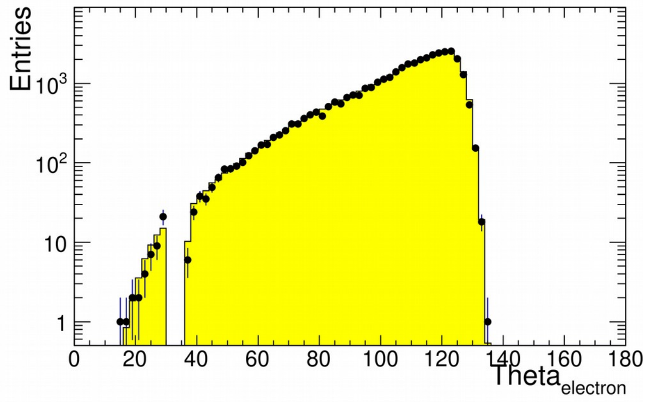
inclusion of a section discussing how the transfer matrix approach can be used in future PDF extractions

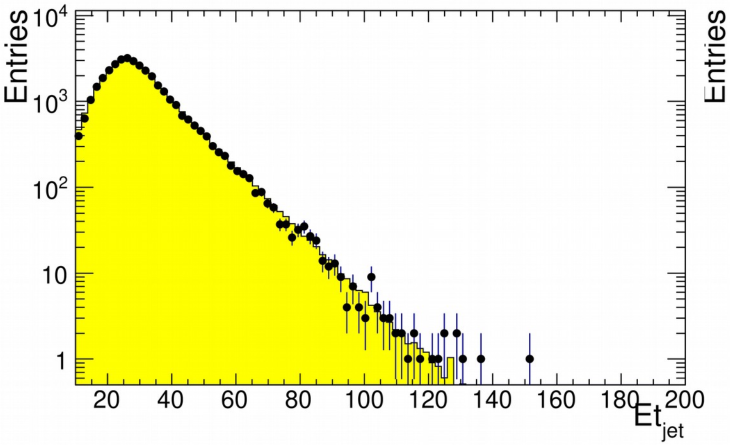
as well as a discussion of the effective power of the high-x data points along the lines suggested by Ola

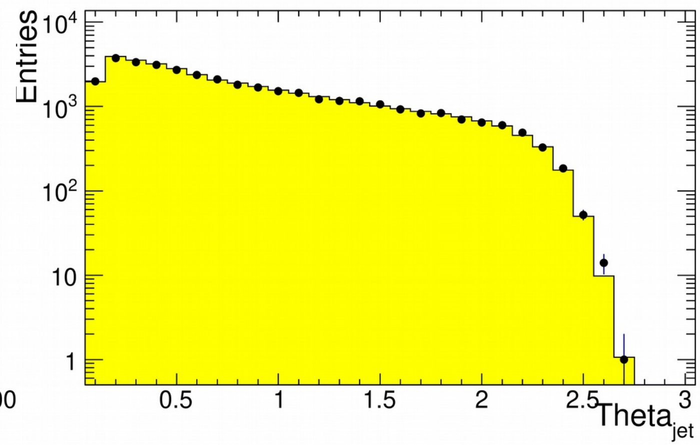
Changes expected :

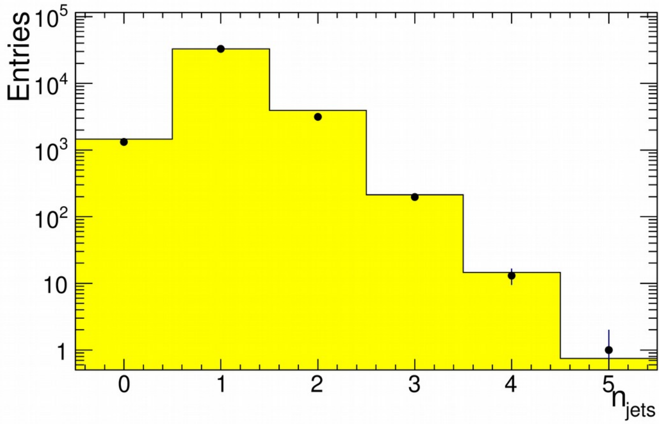
Update in the tables and plots with the given improvements in the reweighting proceedure

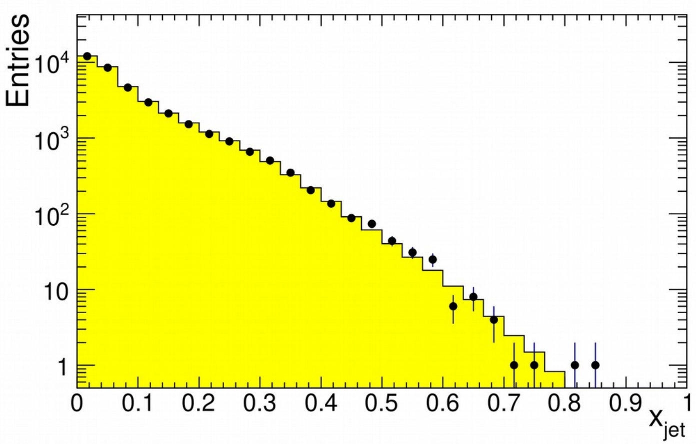

Events per 10 GeV
E_{el}

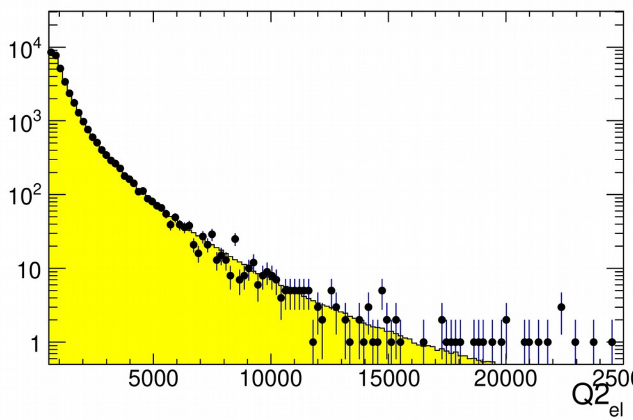

Entries
$Theta_{electron}$


Entries
Et_{jet}


Entries
$Theta_{jet}$


Entries
n_{jets}


Entries
x_{jet}


$Q2_{el}$

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: $\mathcal{L}(\theta,\lambda) = P(n_j|\nu_j(\theta,\lambda))P(\lambda)$

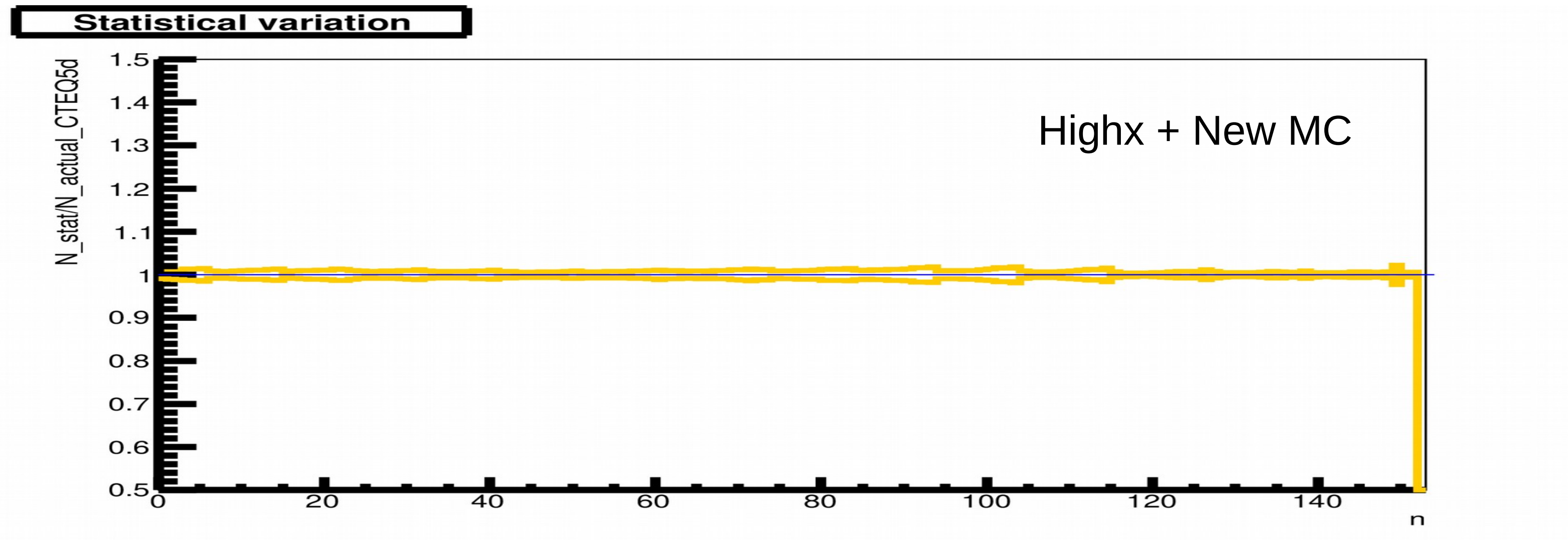
where the $P(\lambda)$ is a product of Gauss distributions:

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

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

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

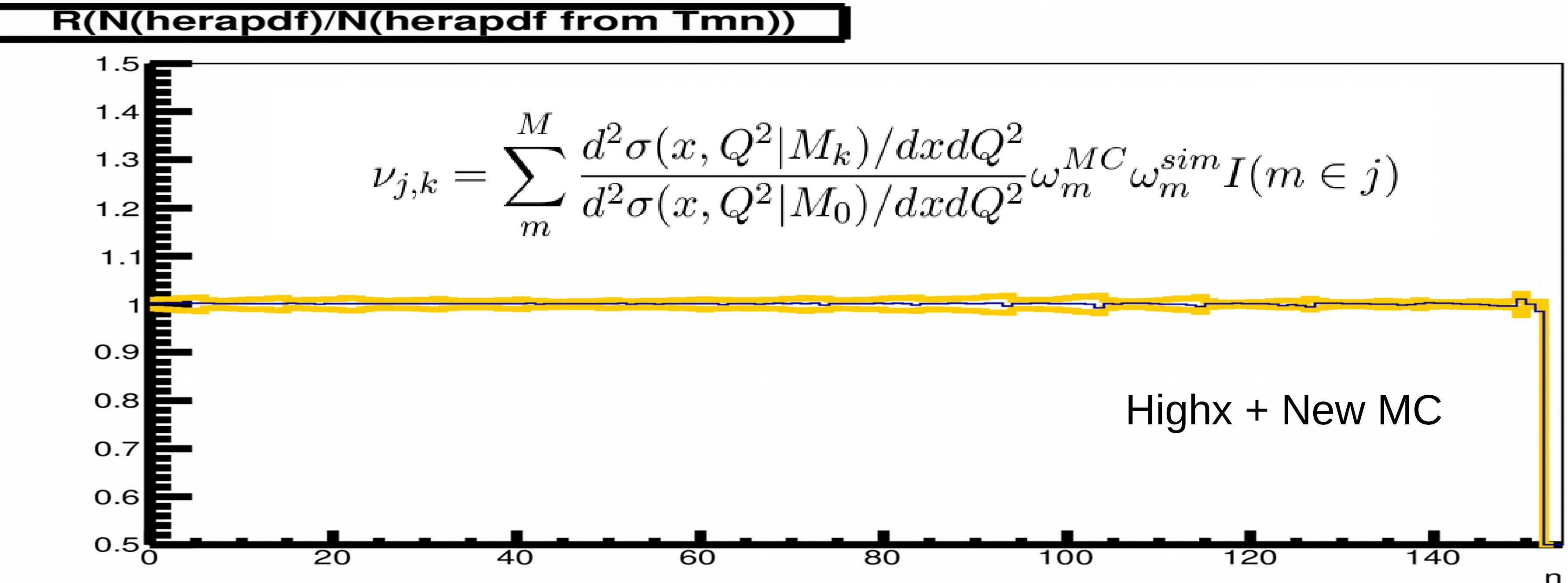
Statistical Error in MC in various Xsec Bins (with in 1%)

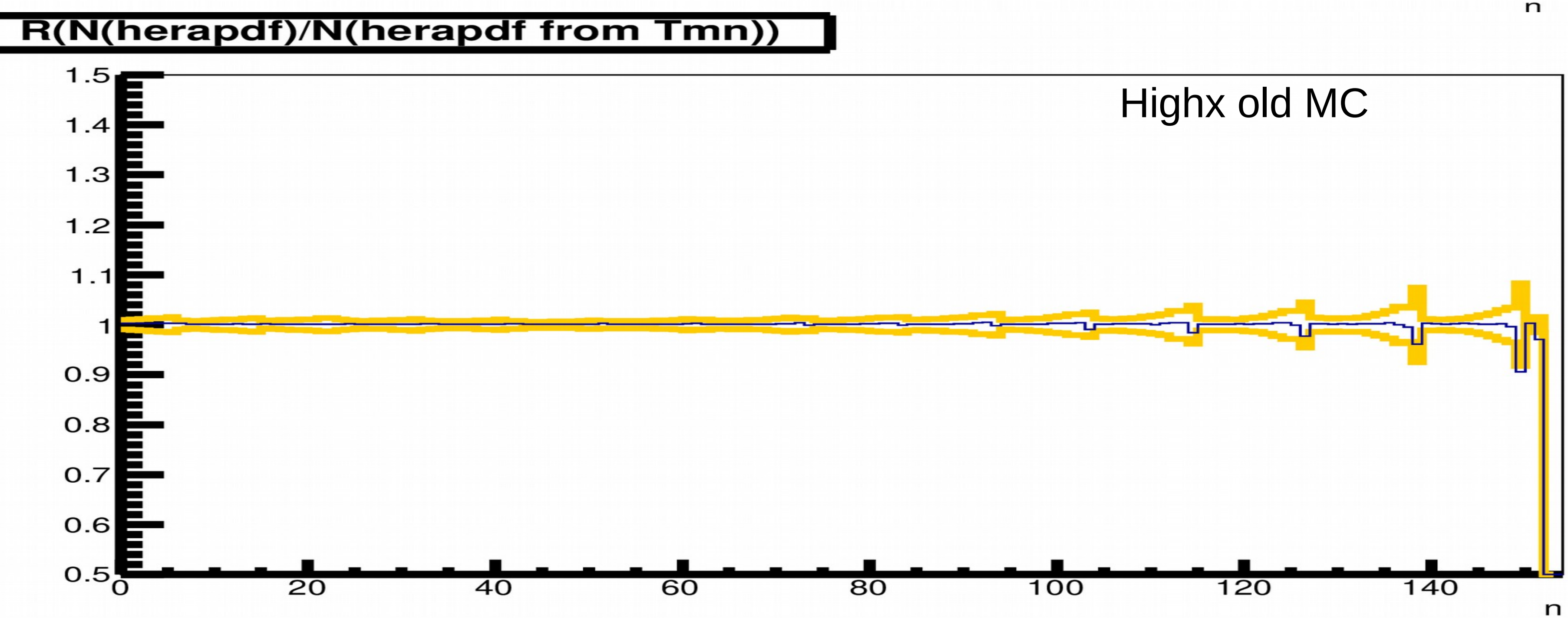


$$\delta a_{ij}^{\mathrm{stat}} = \sqrt{\frac{a_{ij}(1-a_{ij})}{M_i}}$$

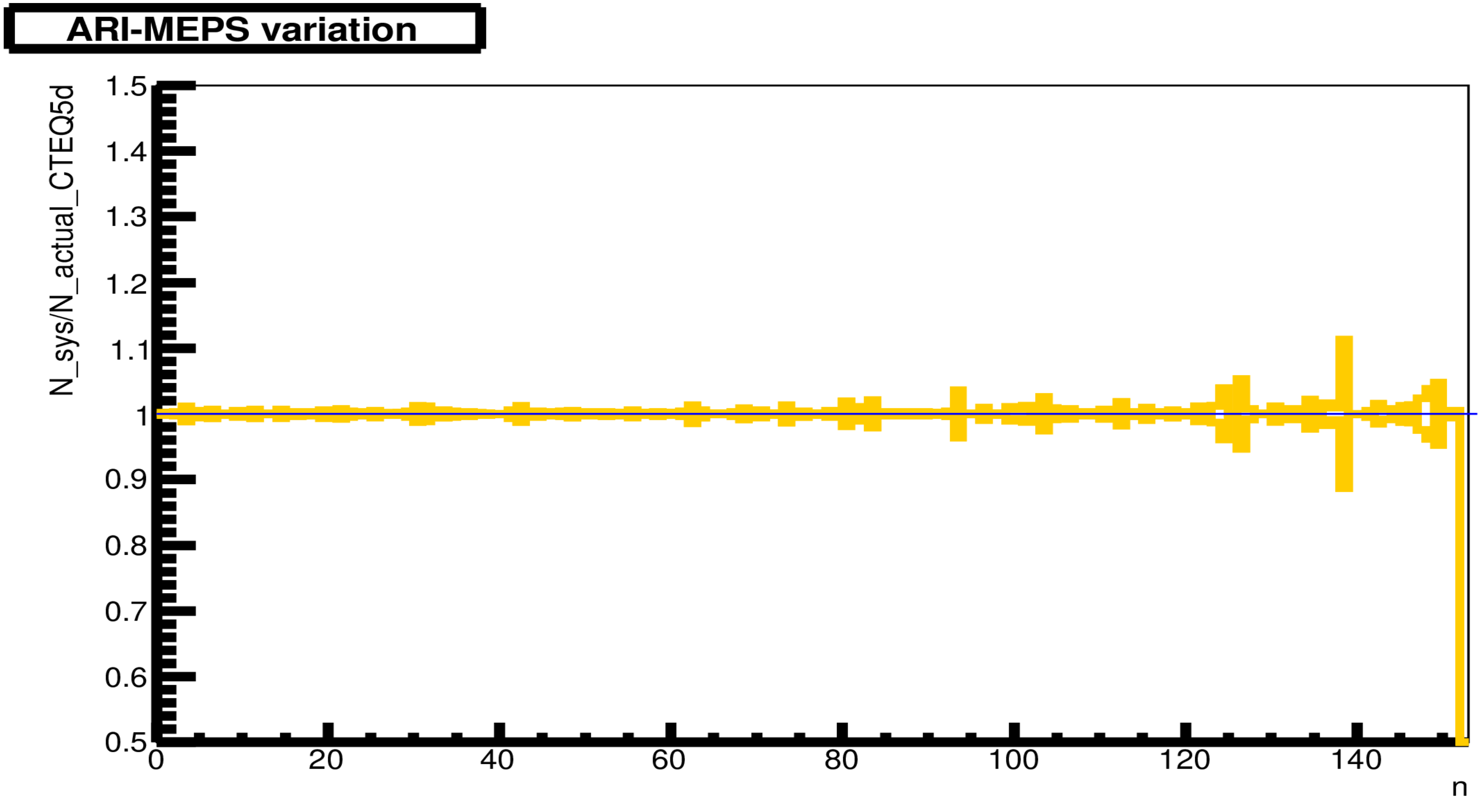
Where Mi are the total number of events Generated in MC

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



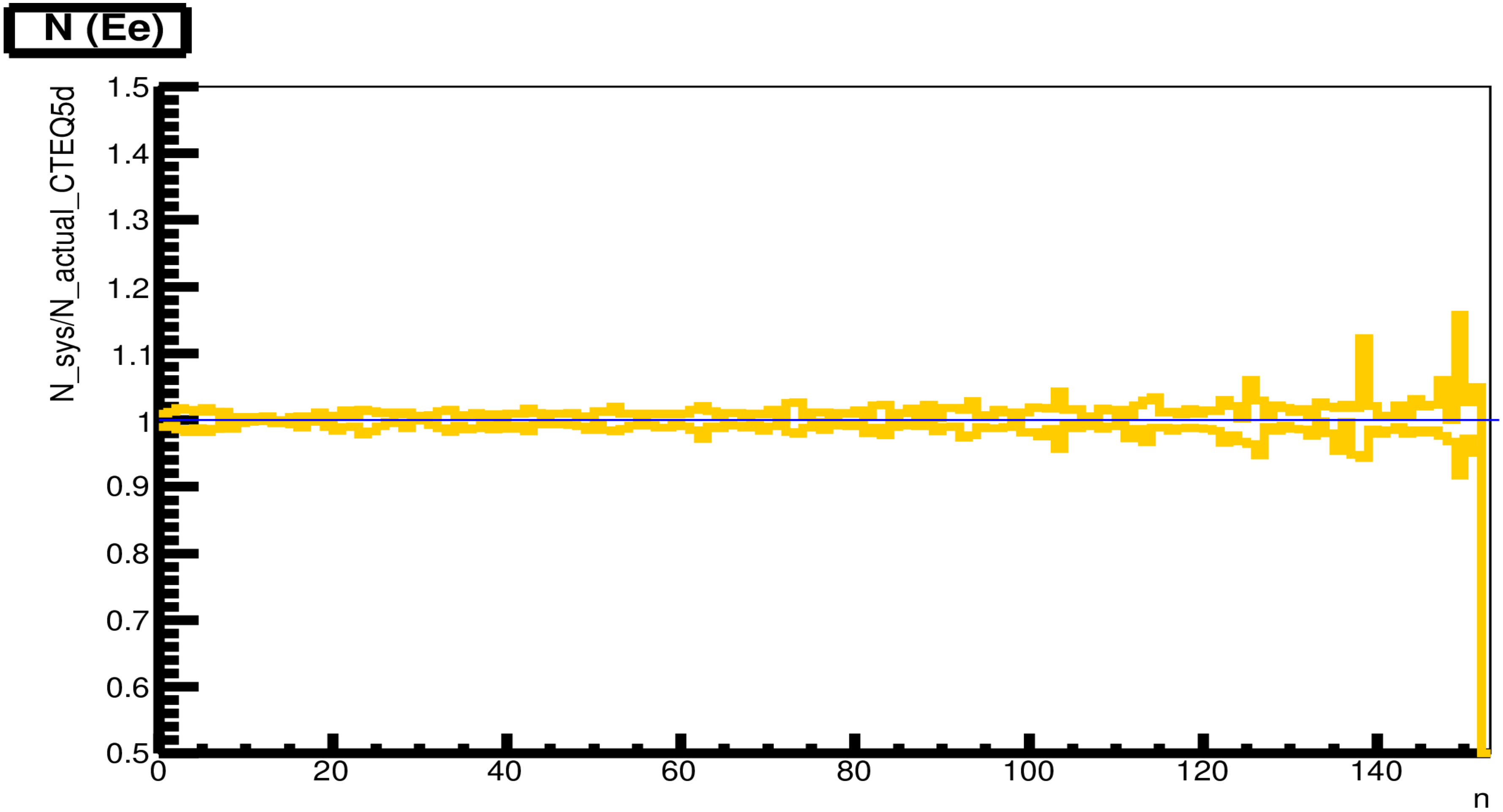


Ariadne-MEPS variation: The ARI-MEPS combination is varied in construction of Transfer Matrix.



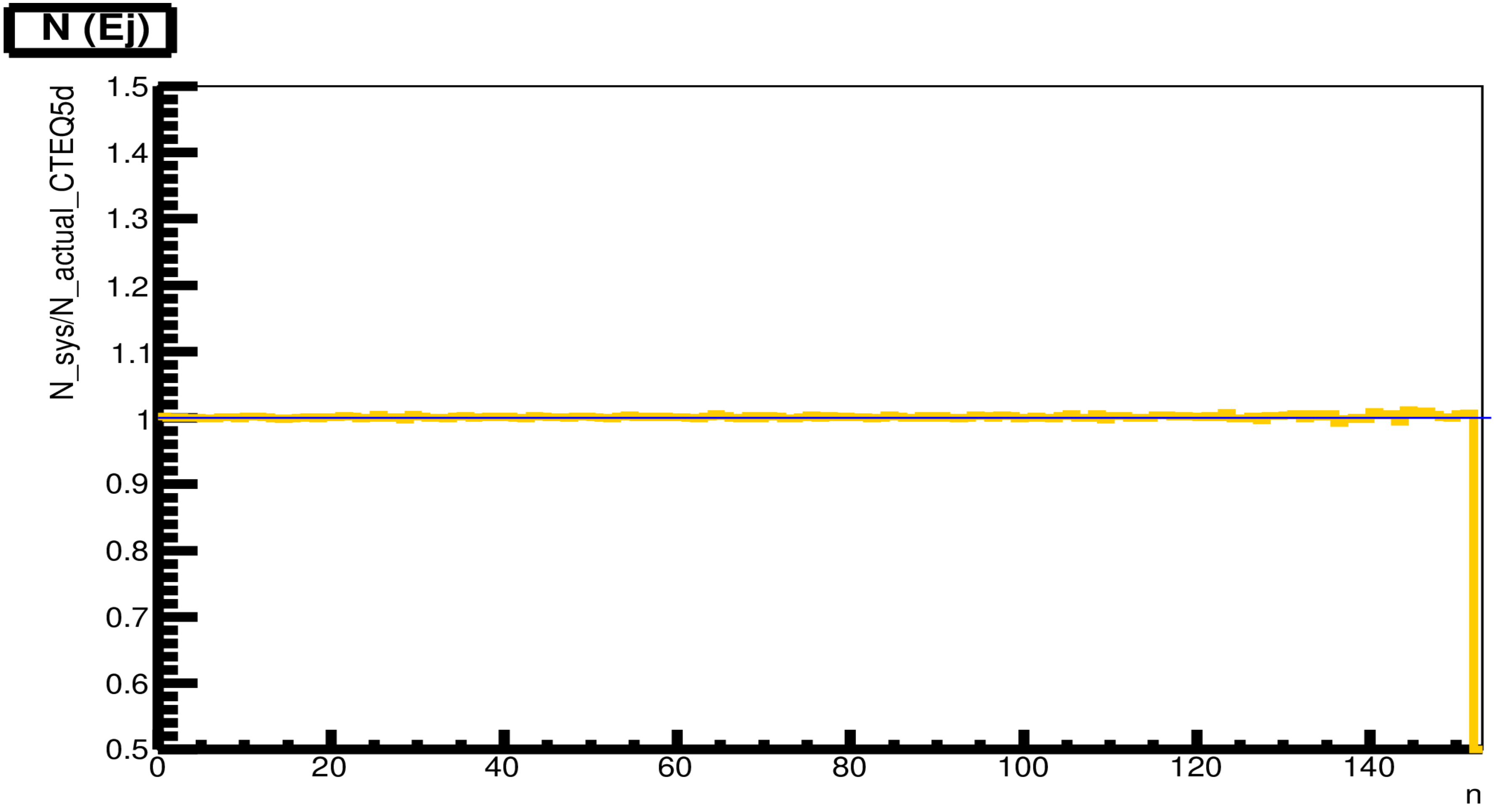
For most of the bins with in 1%, increases to 2-10% in the highest x-bins at high Q2.

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



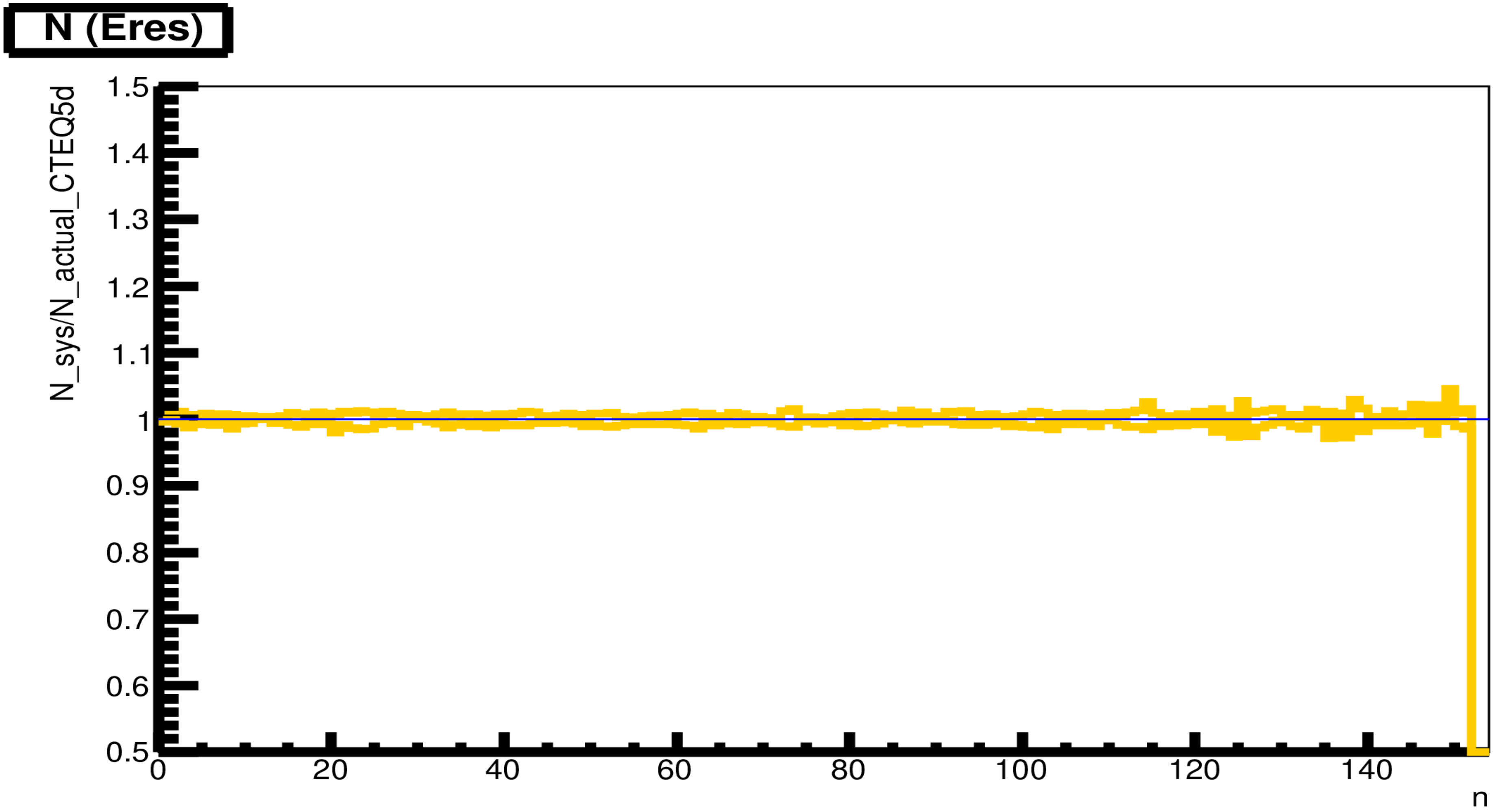
For most of the bins with in 1%, increases to 2-12% in the highest x-bins at high Q2.

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



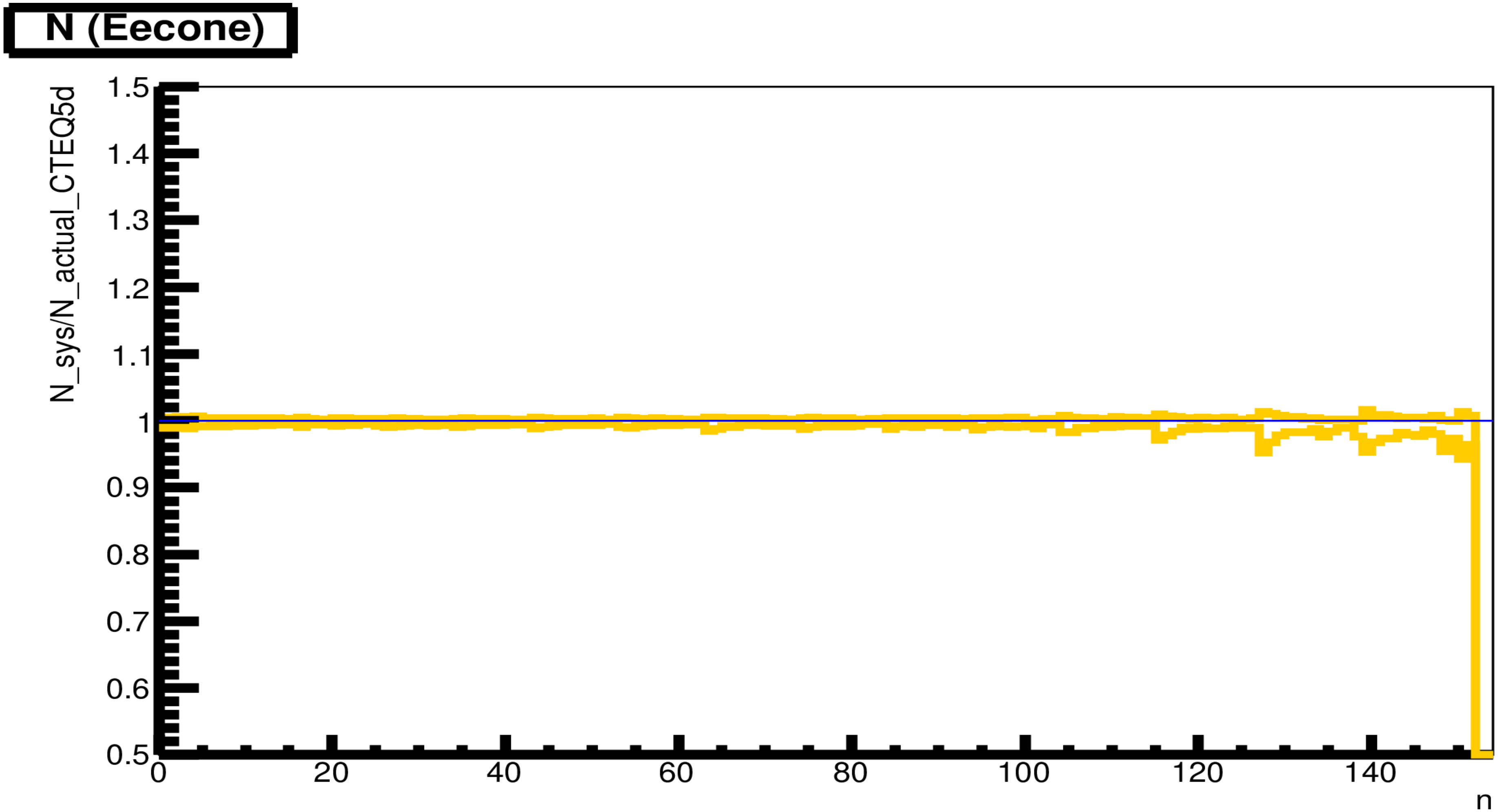
Negligible

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

$$\mathcal{L} \propto e^{-\chi^2/2}$$

and

$$\ln \mathcal{L}_1 - \ln \mathcal{L}_2 = \frac{1}{2}(\chi_2^2 - \chi_1^2)$$

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

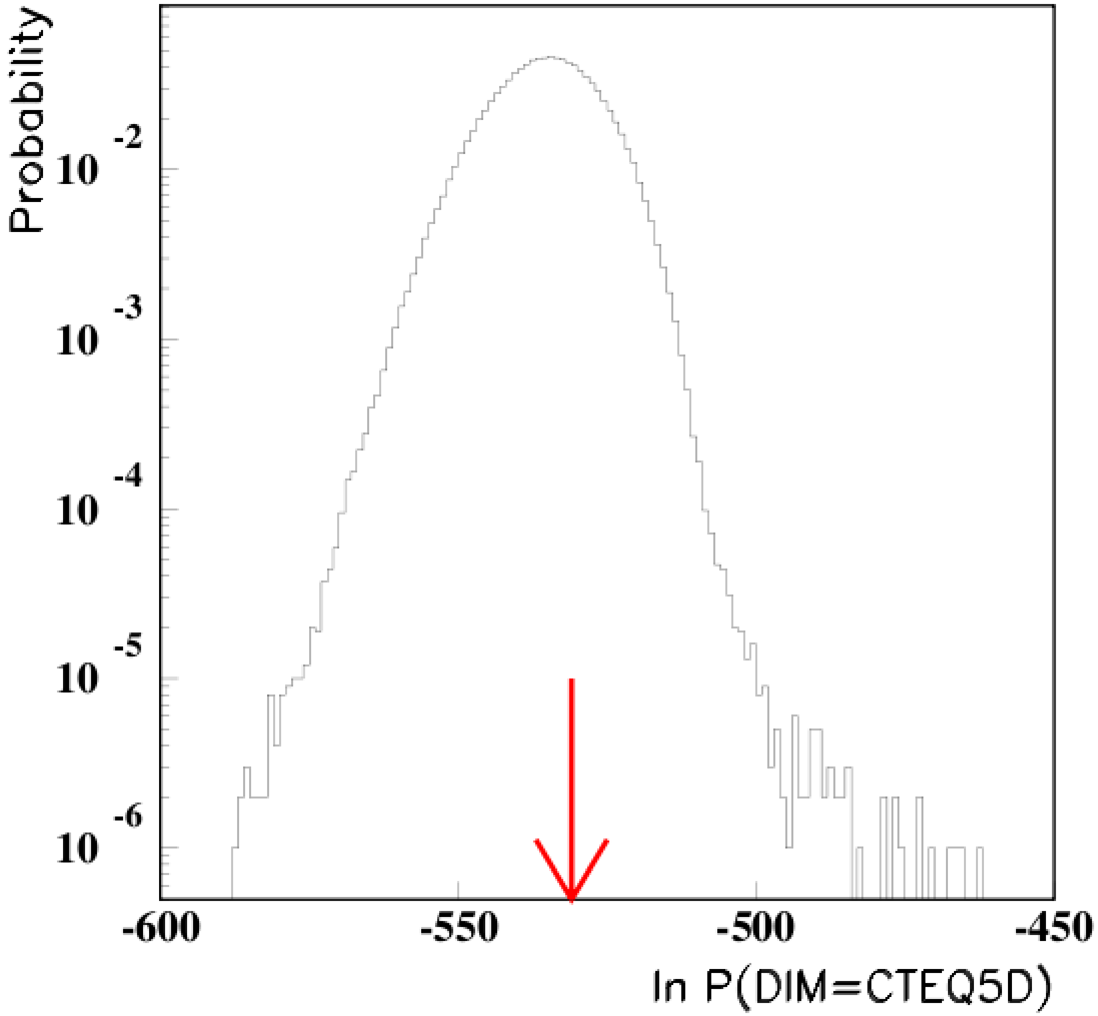


FIGURE 4. Distribution of expected values for $\ln P(D|M = CTEQ)$ for the e^+p data set. The arrow shows the value found in the data.

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,