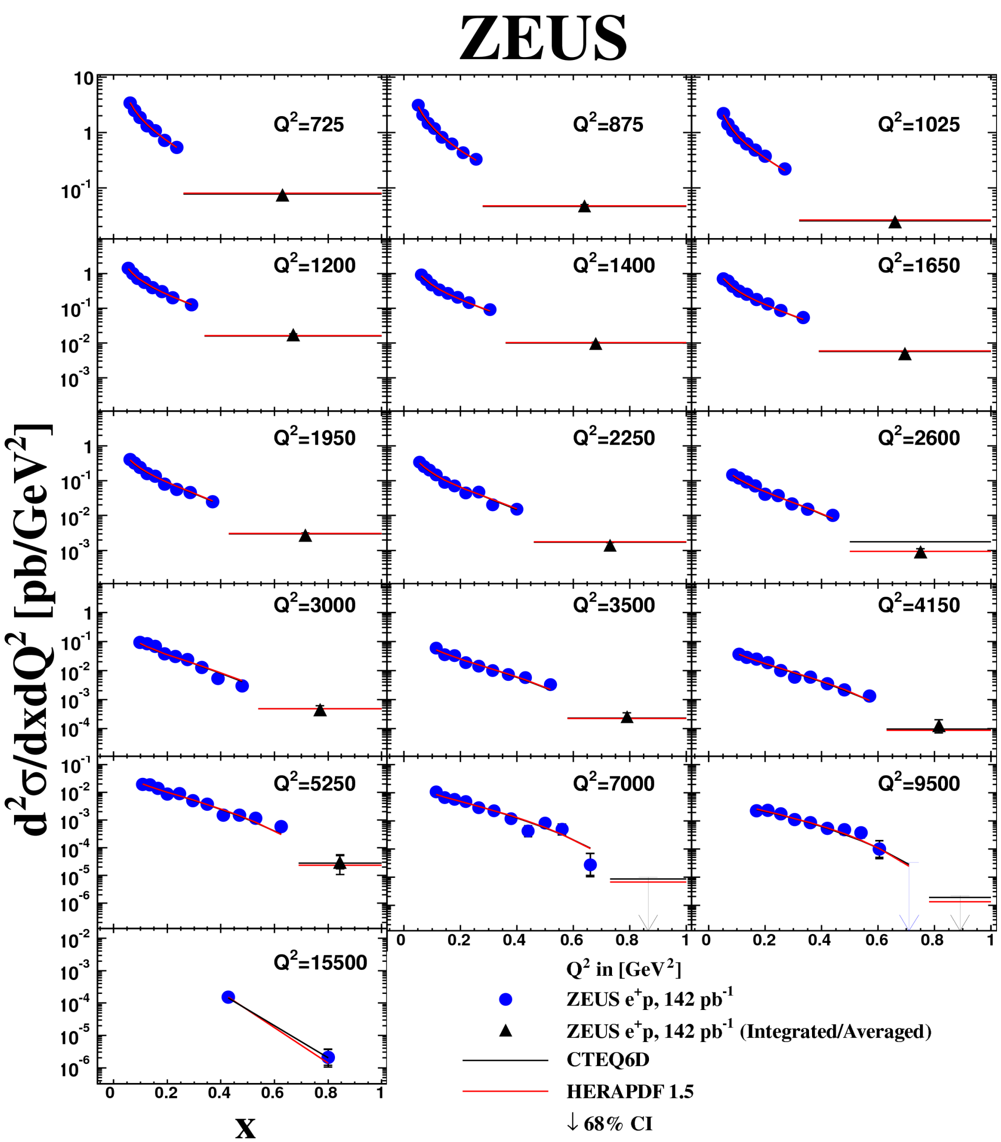
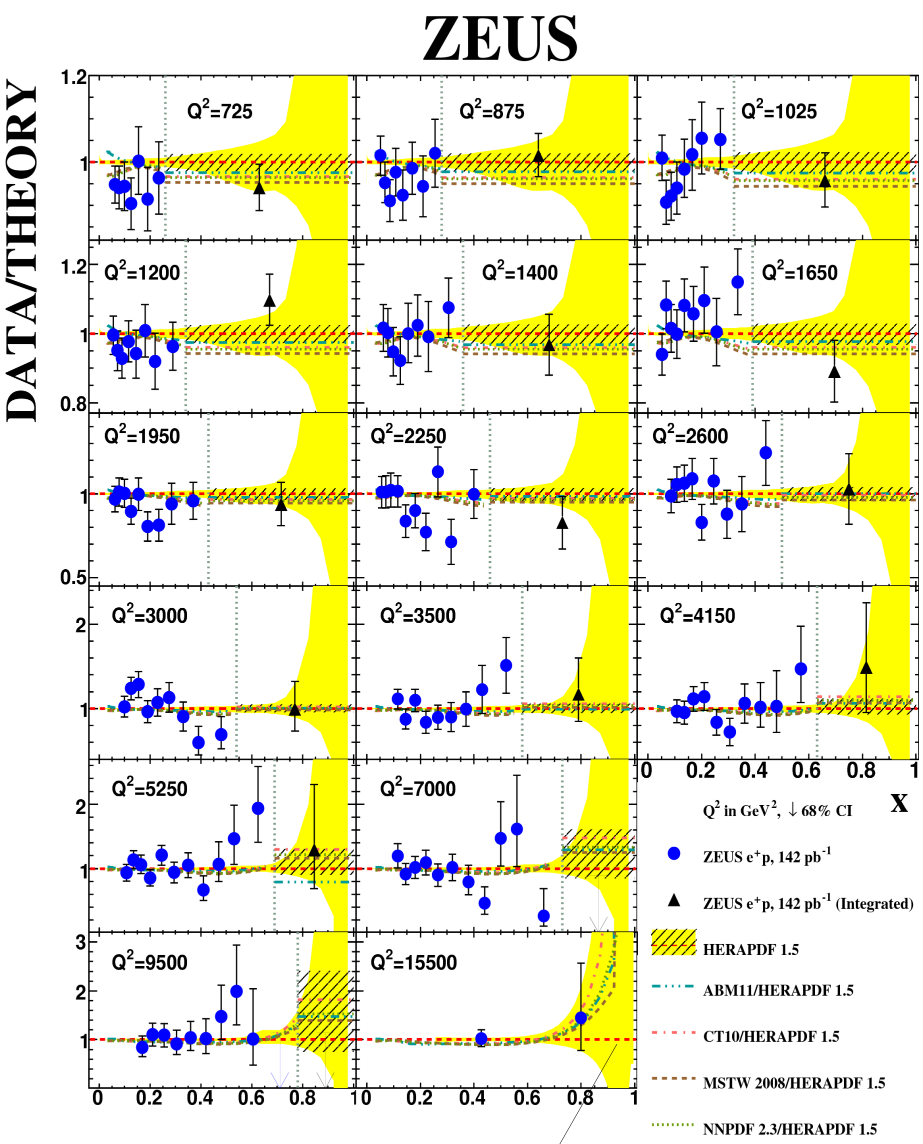
Transfer matrix for high-x MC and Study of impact of high-x data on parton distribution functions

Ritu Aggarwal, Allen Caldwell

Motivation of studying published high-x data



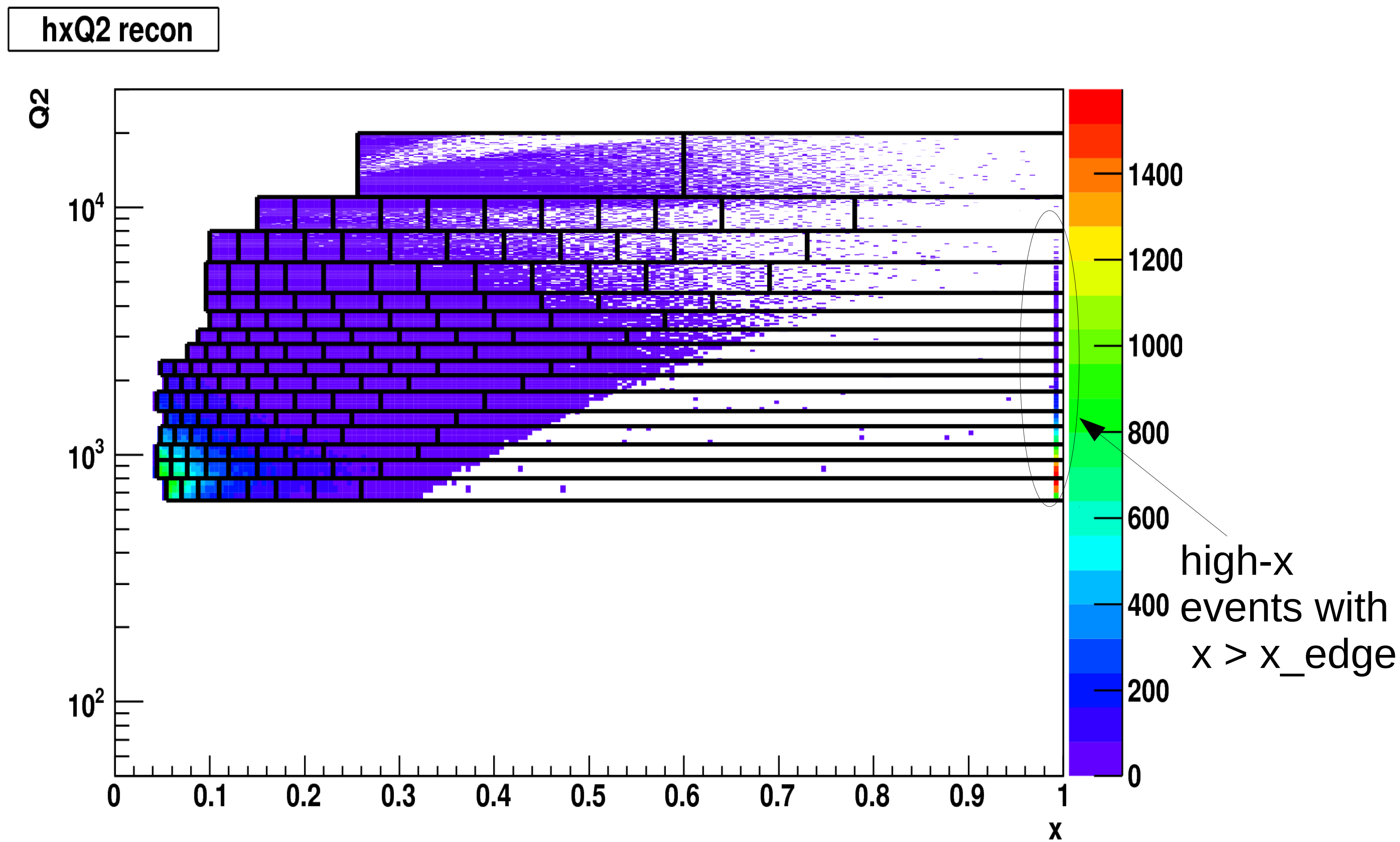


At present x upto 0.65 ZEUS data is included in PDF fits

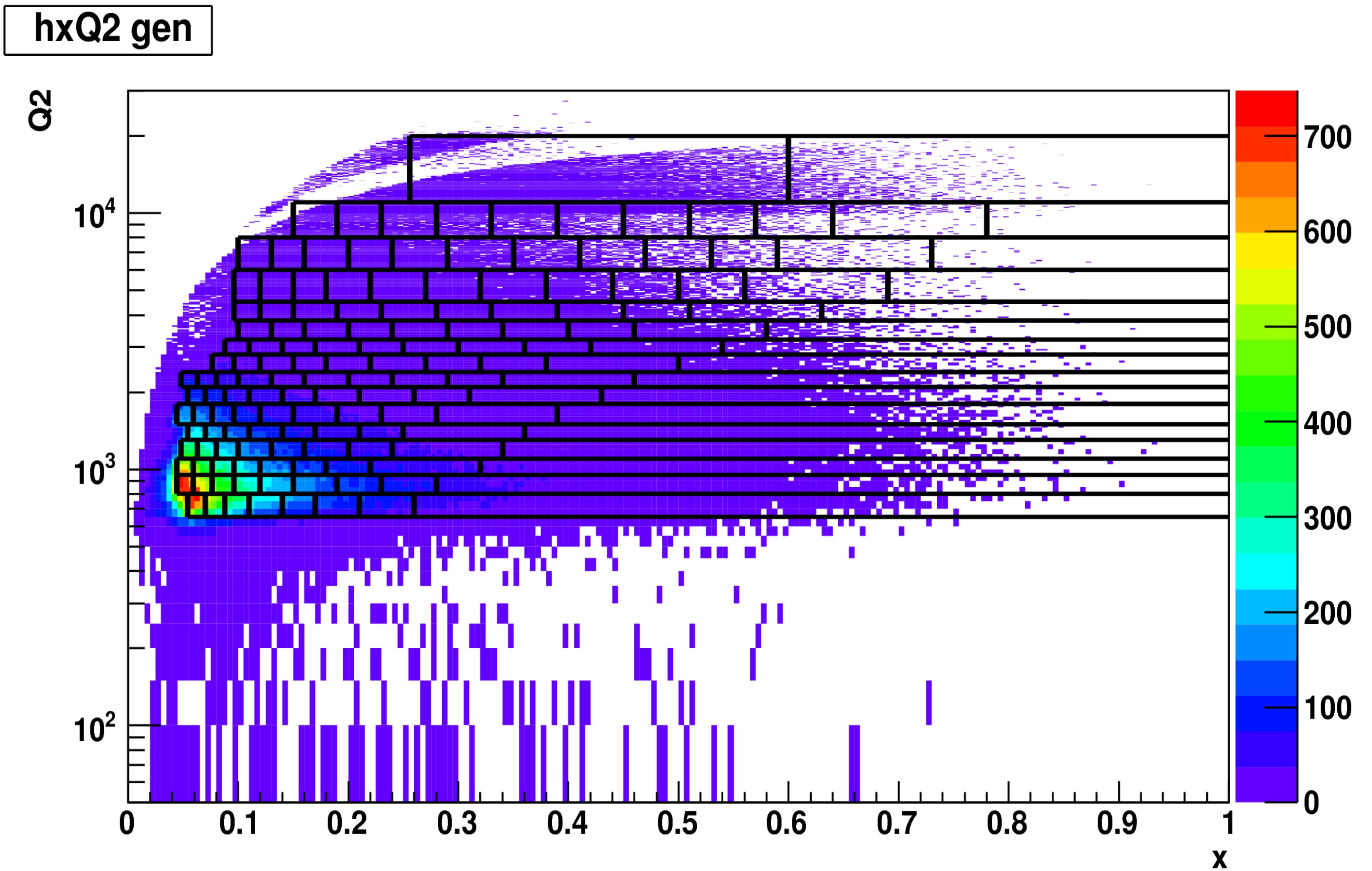
Note the uncertainity bands above x ~ 0.65, can high-x data impact here

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

Reconstructed MC events in Xsection Binning (total 153 bins)



True x-Q2 distribution of events in Xsection Binning



extended binning in M has total 429 bins

Definition of Transfer matrix

Each element in Transfer Matrix is represented as

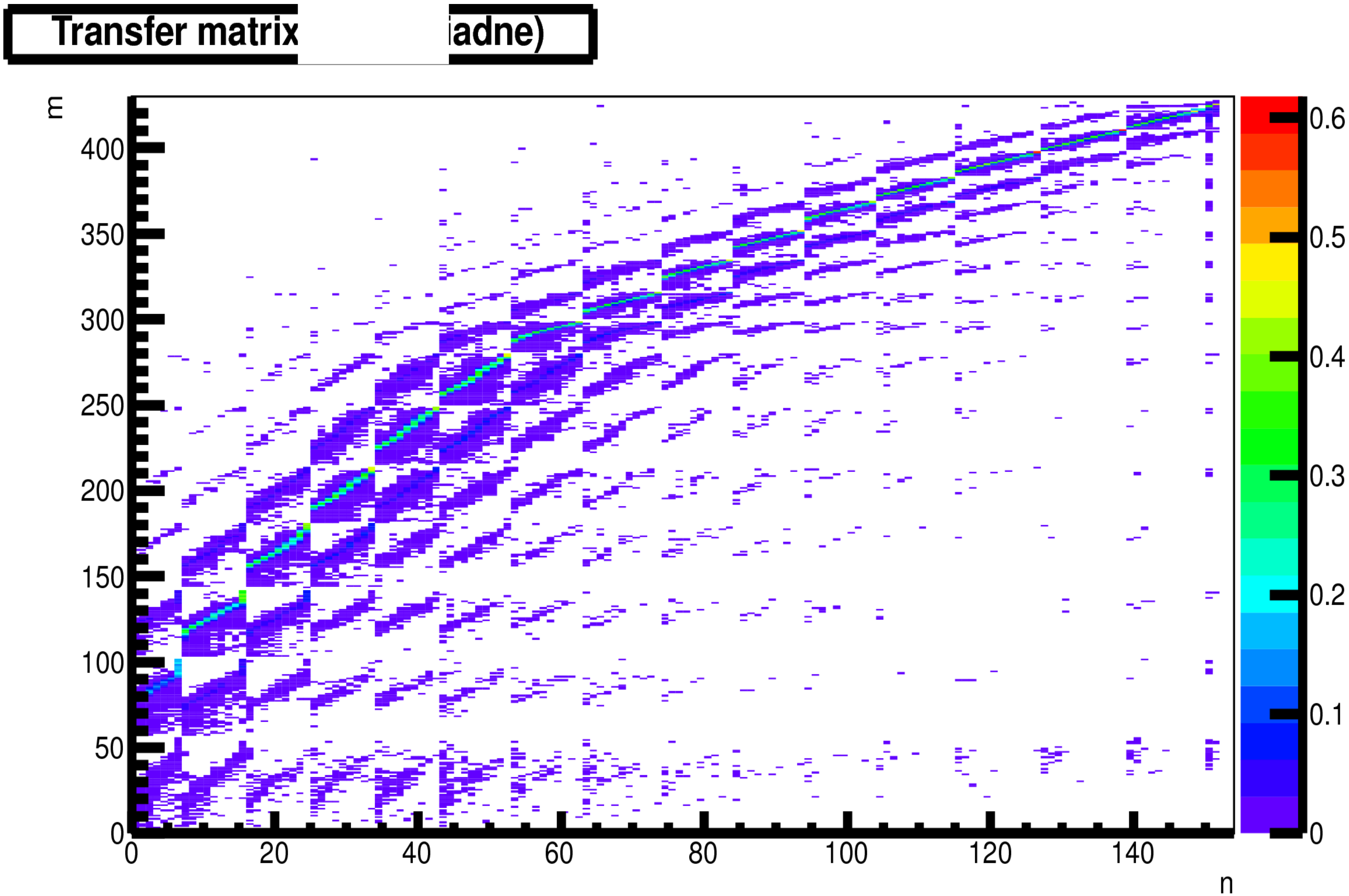
$$a_{ij} = \frac{\sum_{k=1}^{M_i} \omega_k I(k \in j)}{\sum_{k=1}^{M_i} \omega_k}$$

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

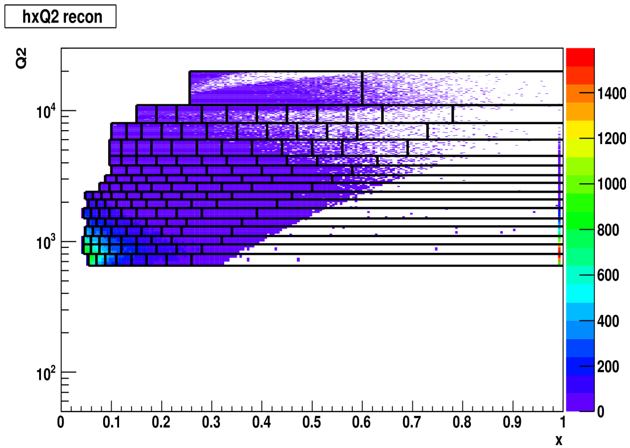
ω_k = weight given to kth event in bin i

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

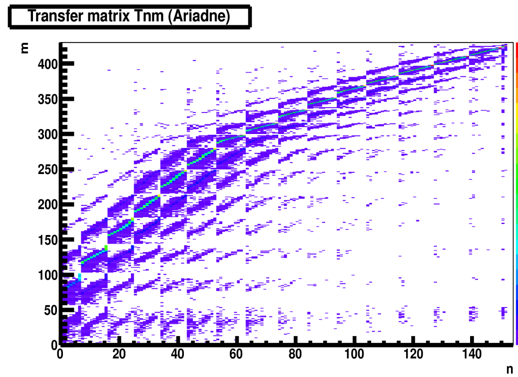
Transfer Matrix



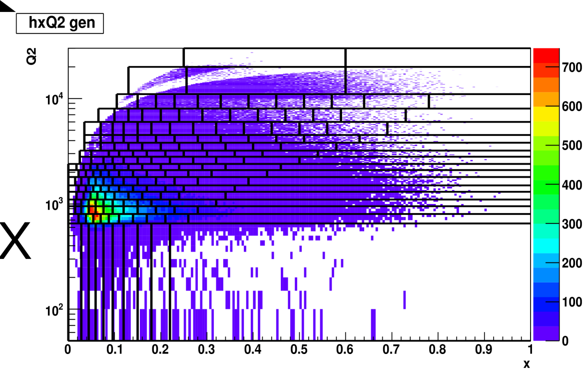
Using Transfer matrix to predict no. of events reconstructed in a given cross section bin



N = A M



X



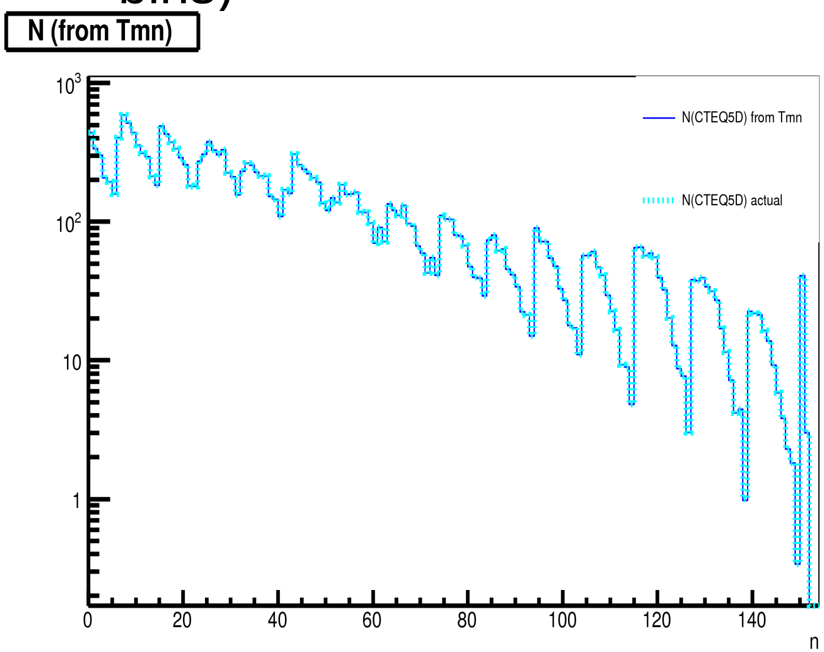
Predicted x-Q2 events in Cross section binning (153 elements in N Vector = number of cross section bins)

Transfer Matrix

(153 X 429 elements)

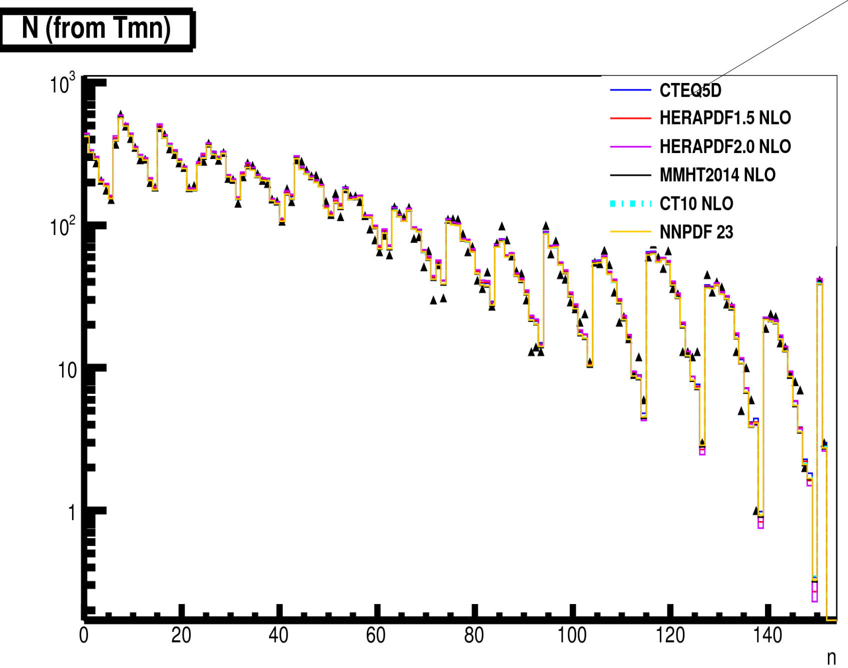
Generated x-Q2 events in Extended binning

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

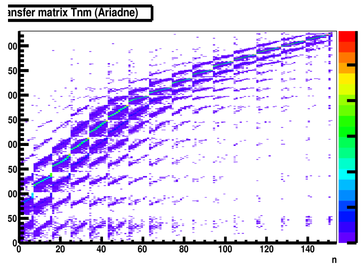


Using Transfer matrix to predict
no. of events reconstructed in a given cross section bin

N = AM

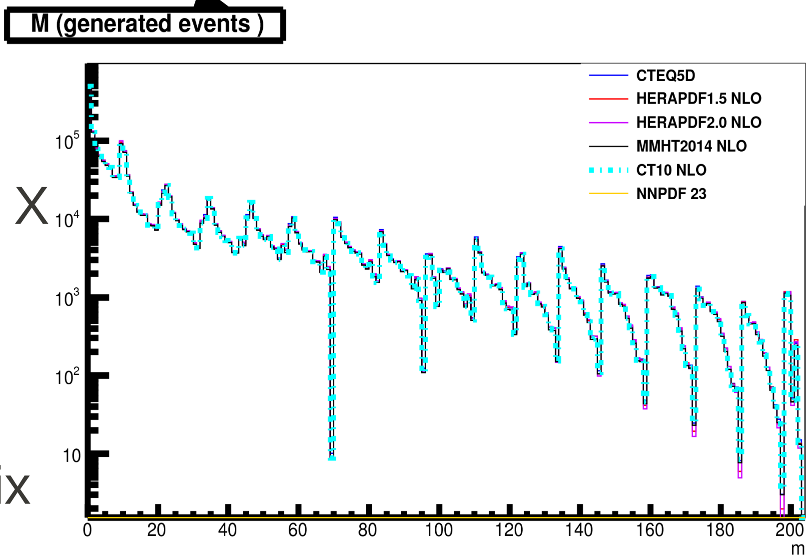


=



Transfer Matrix

X



Predicted x-Q2 events in
Cross section binning

(for different PDFs)

Generated x-Q2
events in
Extended binning

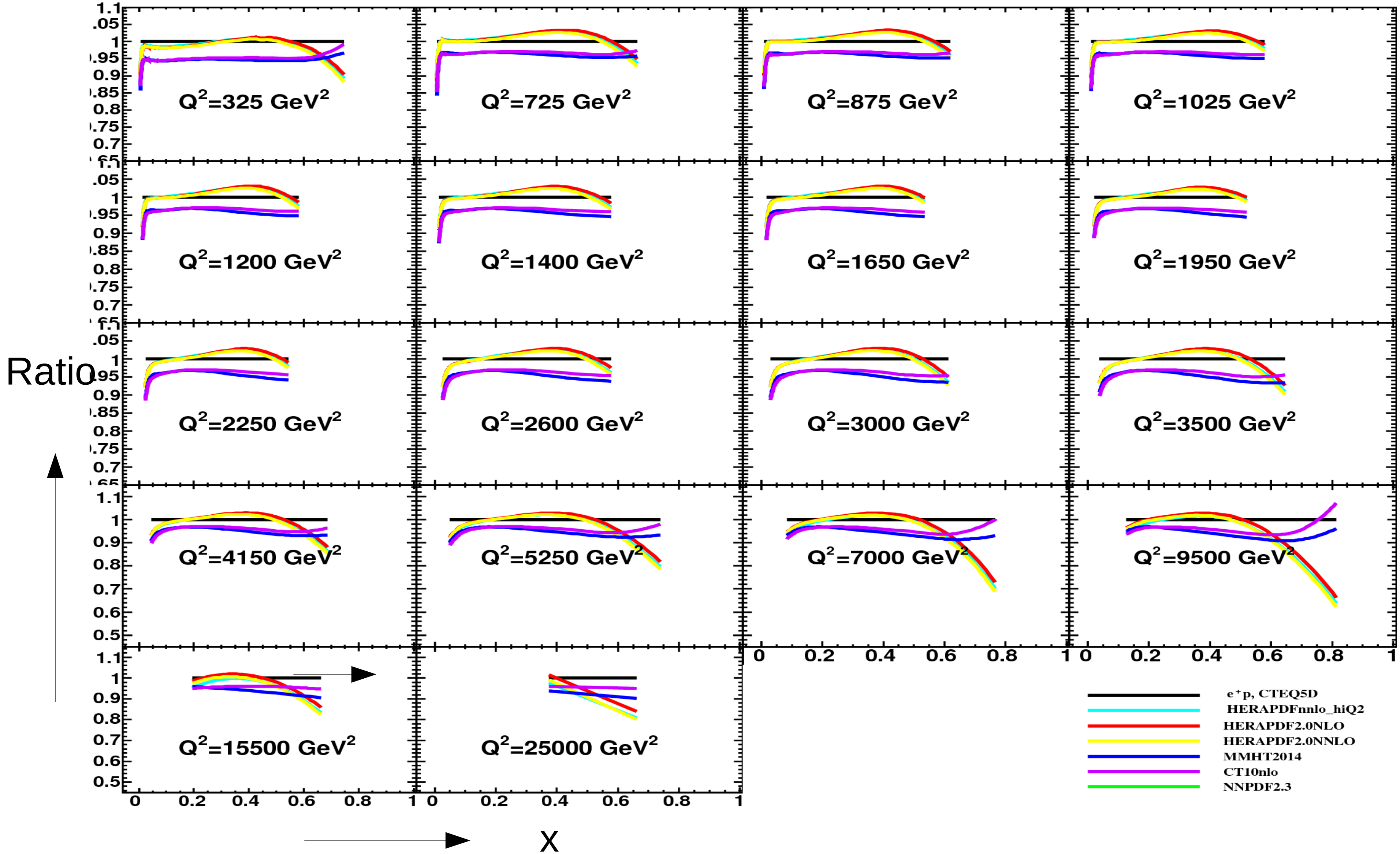
(for different PDFs)

Where for each (Q2,x) bin **M_PDF = M_CTEQ5d *<u>(RedXsec_PDF / RedXsec_CTEQ5D)</u>**

Comparison Shown on next slide

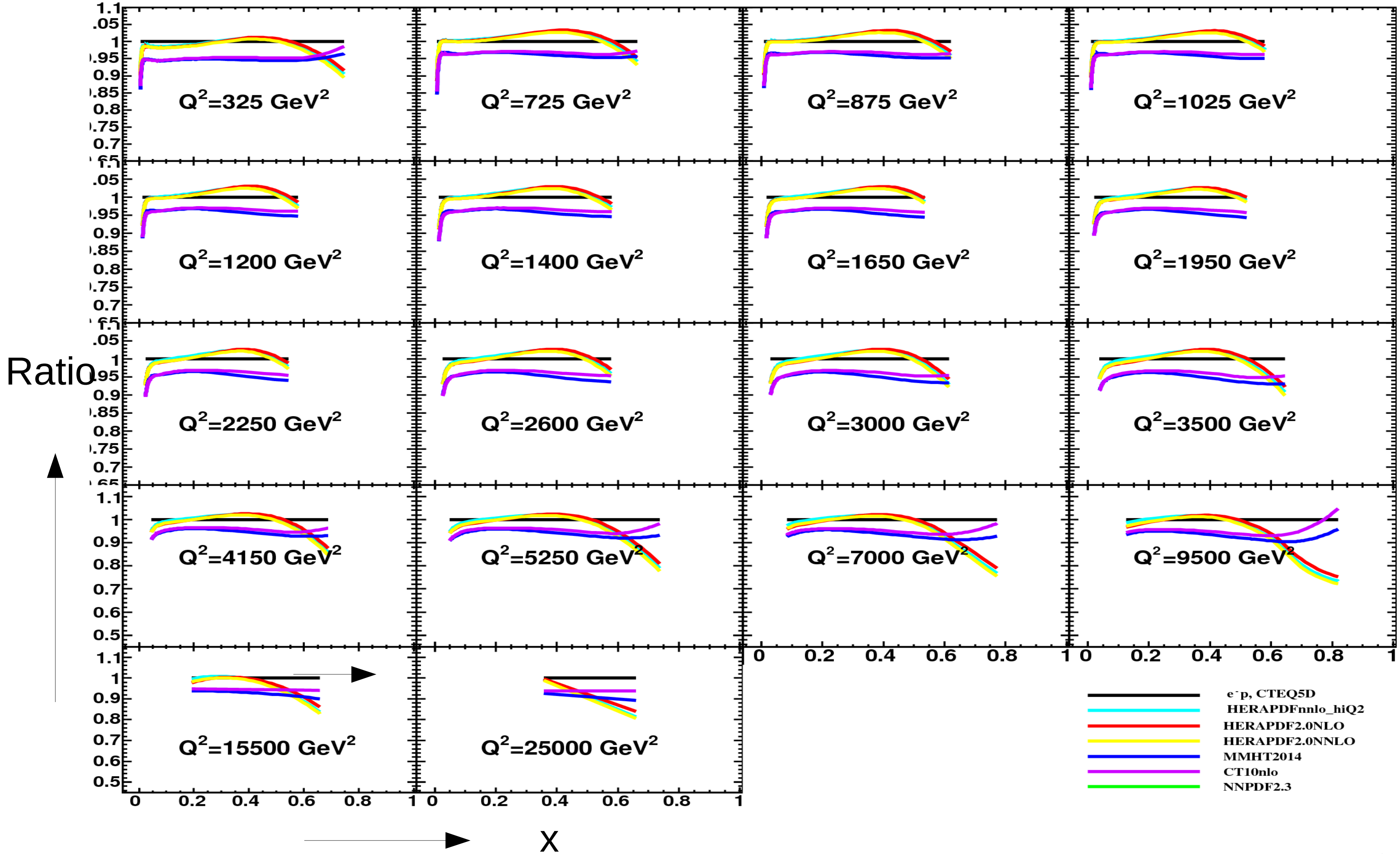
Note : Each event is reweighted according to above formula at the given Q2 and x

Ratio of Vector M for the given PDF in extended binning (e+p) to CTEQ5D



There is a normalization difference between HERAPDF & other PDFs!

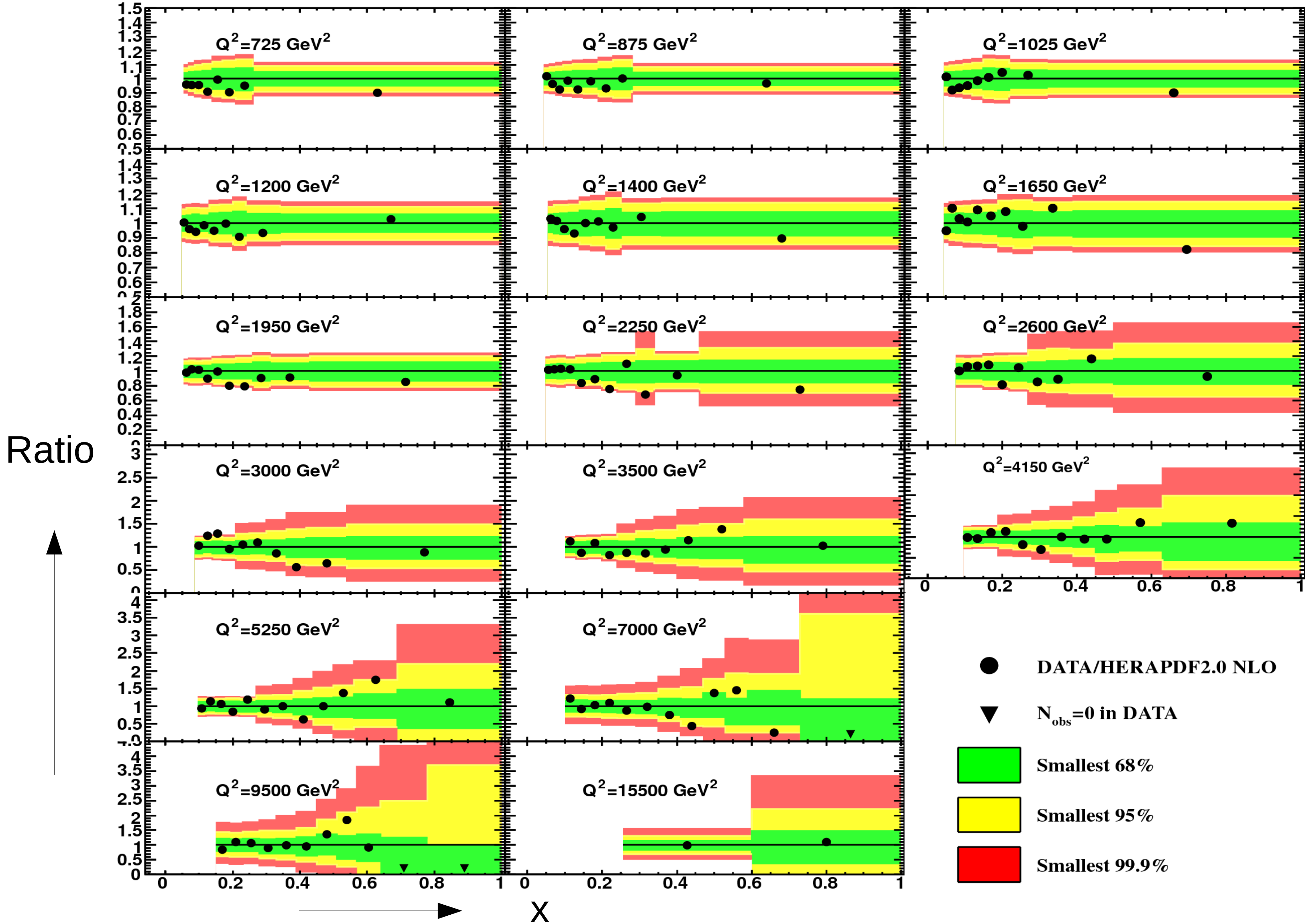
Ratio of Vector M for the given PDF in extended binning (e-p) to CTEQ5D



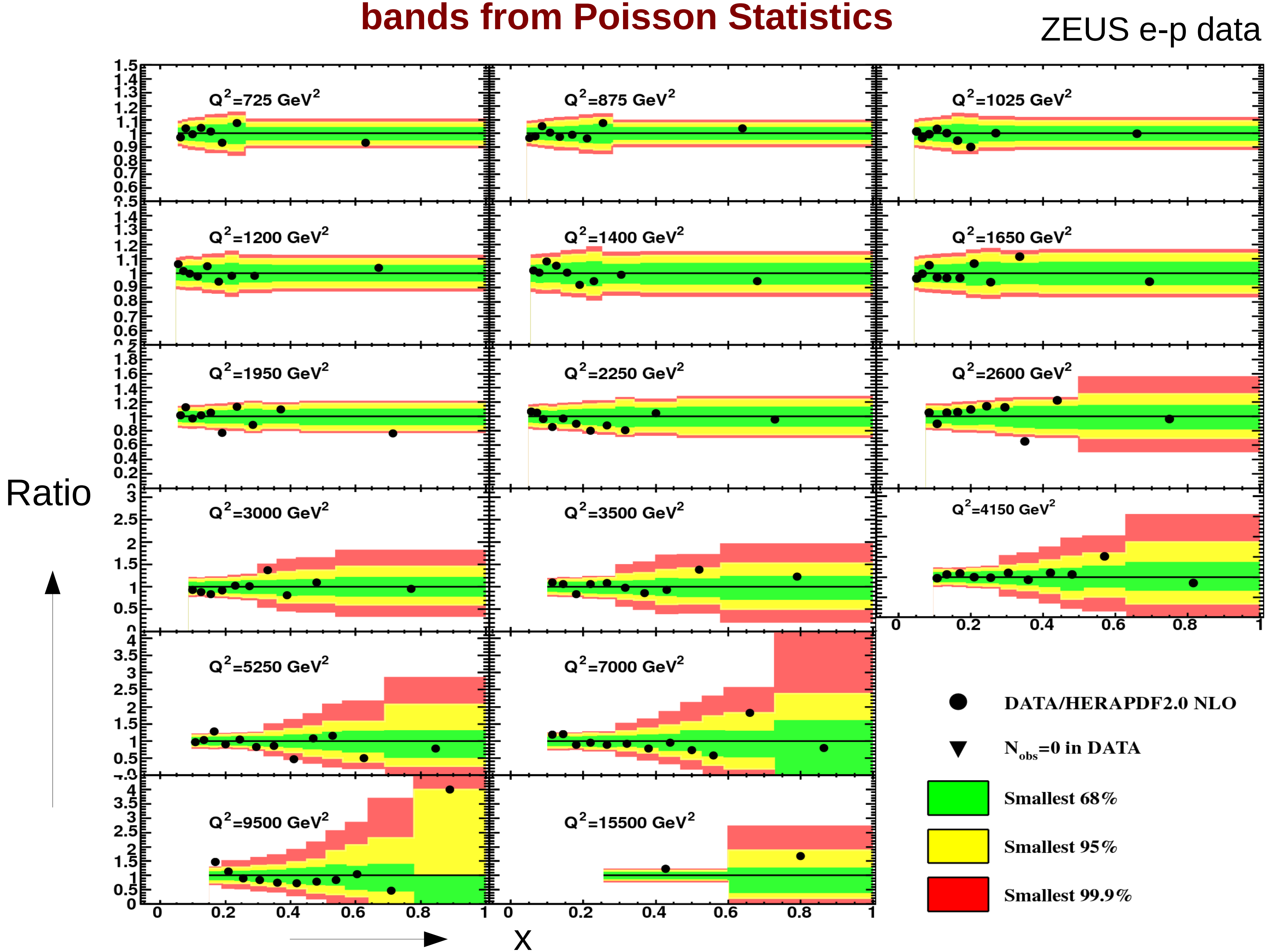
There is a normalization difference between HERAPDF & other PDFs!

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

ZEUS e+p data



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



Relative probability for data from different PDFs

$P(n|\nu) = (\nu^n \; e^{-\nu}) / n!$ Where $\nu_i = \Sigma_j A_{ij} m_j$ and reweighted to luminosity of data

Total probability for each PDF = $\prod_i P_i(n|\nu)$

BRIEF ARTICLE

	e^+p			e^-p		
PDF	$\ln P$	p-value	$\Delta\chi^2$	$\ln P$	p-value	$\Delta\chi^2$
CTEQ5D	−531.2	0.68	0	−577.0	0.079	0
HERAPDF1.5	−528.3	0.76	−5.8	−581.6	0.022	+9.2
HERAPDF2.0	−535.3	0.49	+8.2	−579.0	0.044	+4.0
MMHT2015	−527.3	0.70	−7.8	−599.8	0.00010	+45.6
CT10nlo	−528.4	0.66	−5.6	−597.2	0.00011	+40.4
NNPDF2.3	−529.0	0.64	−4.4	−600.4	0.00004	+46.8
HERAPDF2.0 NNLO	−535.6		+8.8	−576.98		−0.04
HERAPDF2.0 HiQ2 NNLO	−533.7		+5.0	−578.5		+3.0

TABLE 1. The results from comparisons of predictions using different PDF sets to the observed numbers of events. The log of the probability, the corresponding p-value, and the effective χ^2 difference relative to the CTEQ5D result are given. The results are shown separately for the e^+p and e^-p data sets. The results are for the full Bjorken-x range given in [?].

MMHT2014, CT10nlo, NNPDF2.3 much better than CTEQ5D for e^+P, much worse for e^-P. HERAPDF2.0 lower probability for both

Check 1 : Comparing N (calculated from Transfer Matrix) for different Pdfs
For high-x bins only (~20 bins)

Calculate Poisson probability P(n|ν) for each bin of [N] from different PDFs

$$\prod_i P(n|\nu) = \prod_i (\nu^n\ e^{-\nu}) / n!$$

Calculating the **relative Probablity wrt. CTEQ5D**

	e^+p		e^-p	
	$x < 0.6$	$x \geq 0.6$	$x < 0.6$	$x \geq 0.6$
PDF	$\ln P$	$\ln P$	$\ln P$	$\ln P$
CTEQ5D	−472.8	−58.3	−515.9	−61.0
HERAPDF1.5	−470.6	−57.7	−520.3	−61.0
HERAPDF2.0	−475.2	−60.0	−517.4	−61.4
MMHT2015	−472.4	−54.9	−537.2	−62.9
CT10nlo	−473.0	−55.3	−535.5	−62.0
NNPDF2.3	−473.8	−55.3	−538.7	−62.0

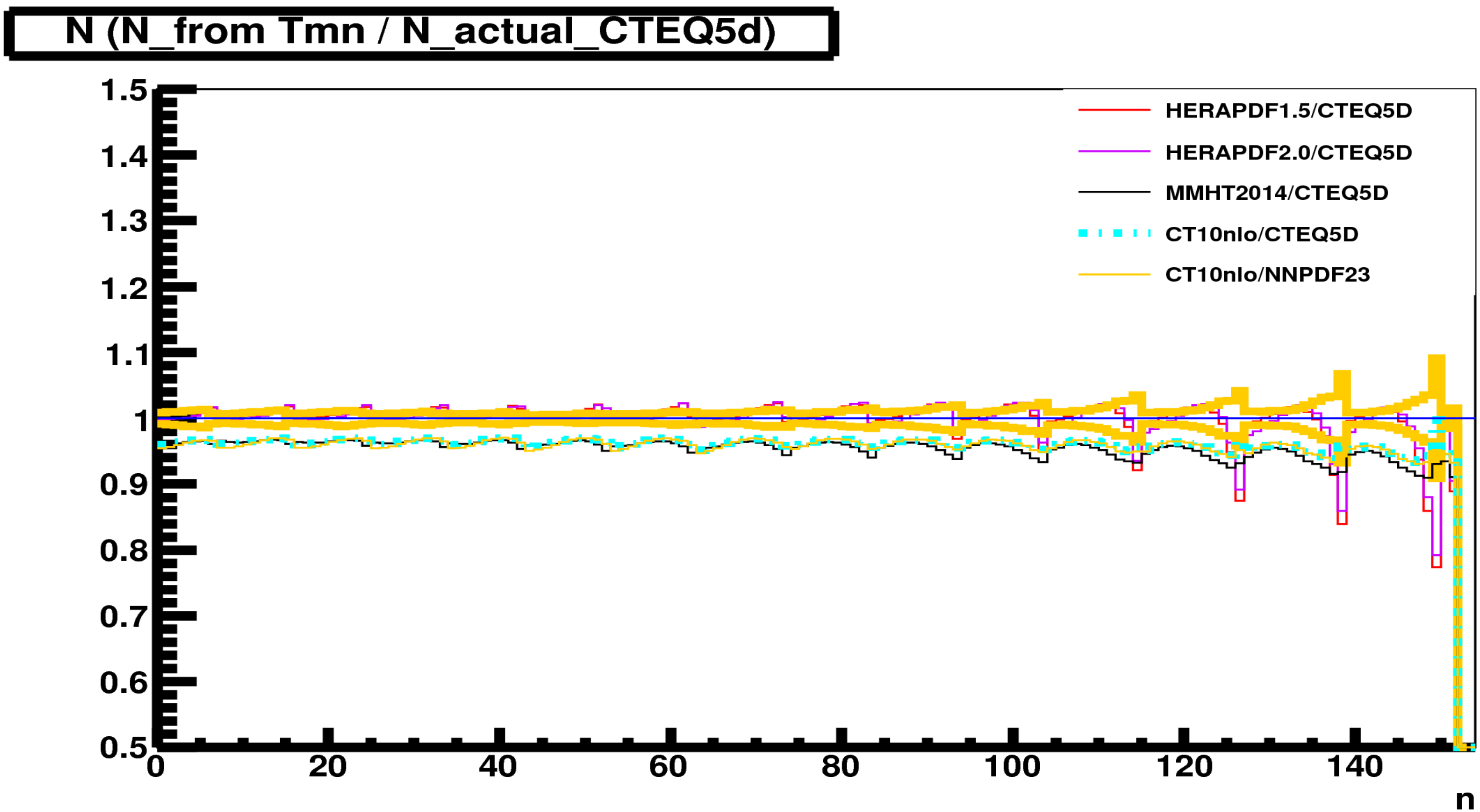
TABLE 2. The results from comparisons of predictions using different PDF sets to the observed numbers of events. The log of the probability is given for two different x ranges for the e^+p and e^-p data sets.

at high x, MMHT, CT, NNPDF better for e+P data. disagreement comes primarily from lower Q2 and e-p

Statistical Error : N (calculated from Transfer Matrix) for different Pdfs

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

Where Mi are the total number of events Generated in MC



For most of the bins with in 1%, increases to 6-10% in the highest x-bins at high Q2.

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

	+1.8 %			
	e^+p		e^-p	
PDF	$x < 0.6$ ln P	$x \geq 0.6$ ln P	$x < 0.6$ ln P	$x \geq 0.6$ ln P
CTEQ5D	−472.8	−58.3	−515.9	−61.0
HERAPDF1.5	−470.6	−57.7	−520.3	−61.0
HERAPDF2.0	−475.2	−60.0	−517.4	−61.4
MMHT2015	−472.4	−54.9	−537.2	−62.9
CT10nlo	−473.0	−55.3	−535.5	−62.0
NNPDF2.3	−473.8	−55.3	−538.7	−62.0
	−1.8 %			
	e^+p		e^-p	
PDF	$x < 0.6$ ln P	$x \geq 0.6$ ln P	$x < 0.6$ ln P	$x \geq 0.6$ ln P
CTEQ5D	−472.8	−58.3	−515.9	−61.0
HERAPDF1.5	−470.6	−57.7	−520.3	−61.0
HERAPDF2.0	−475.2	−60.0	−517.4	−61.4
MMHT2015	−472.4	−54.9	−537.2	−62.9
CT10nlo	−473.0	−55.3	−535.5	−62.0
NNPDF2.3	−473.8	−55.3	−538.7	−62.0

TABLE 4. 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 log of the probability is given for two different x ranges for the e^+p and e^-p data sets.

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

	e^+p		e^-p	
Systematic	$\ln P$ up : down	p-value up : down	$\ln P$ up : down	p-value up : down
Electron energy scale	$-533.4 : -531.6$	$0.62 : 0.60$	$-576.7 : -579.3$	$0.09 : 0.03$
Electron energy resolution	$-530.5 : -532.7$	$0.68 : 0.65$	$-578.6 : -576.3$	$0.05 : 0.09$
Electron isolation cut	$-532.7 : -528.9$	$0.61 : 0.74$	$-576.5 : -580.7$	$0.08 : 0.02$
Hadronic energy scale	$-531.4 : -531.2$	$0.66 : 0.67$	$-577.3 : -576.8$	$0.068 : 0.072$
FCAL alignment	$-530.8 : -531.5$	$0.69 : 0.66$	$-576.7 : -576.9$	$0.085 : 0.075$
F-BCal Crack cut	$-531.1 : -531.2$	$0.68 : 0.69$	$-575.2 : -578.5$	$0.052 : 0.108$
MEPS/Ariadne reweighting	$-530.8 : -532.0$	$0.69 : 0.64$	$-576.0 : -578.7$	$0.083 : 0.059$

TABLE 5. The log of the probability and the p-value for various systematic checks performed for the e^+p and e^-p data sets.

Nomial value for comparison	e^+p			e^-p		
PDF	$\ln P$	p-value	$\Delta\chi^2$	$\ln P$	p-value	$\Delta\chi^2$
CTEQ5D	-531.2	0.68	0	-577.0	0.079	0

Results

1)Technique of building Transfer Matrix Shown.

2) Transfer Matrix can be used to predict number of events in the given cross section bins in MC.

3) Transfer Matrix can be used to compare number of events reconstructed by different PDFs.

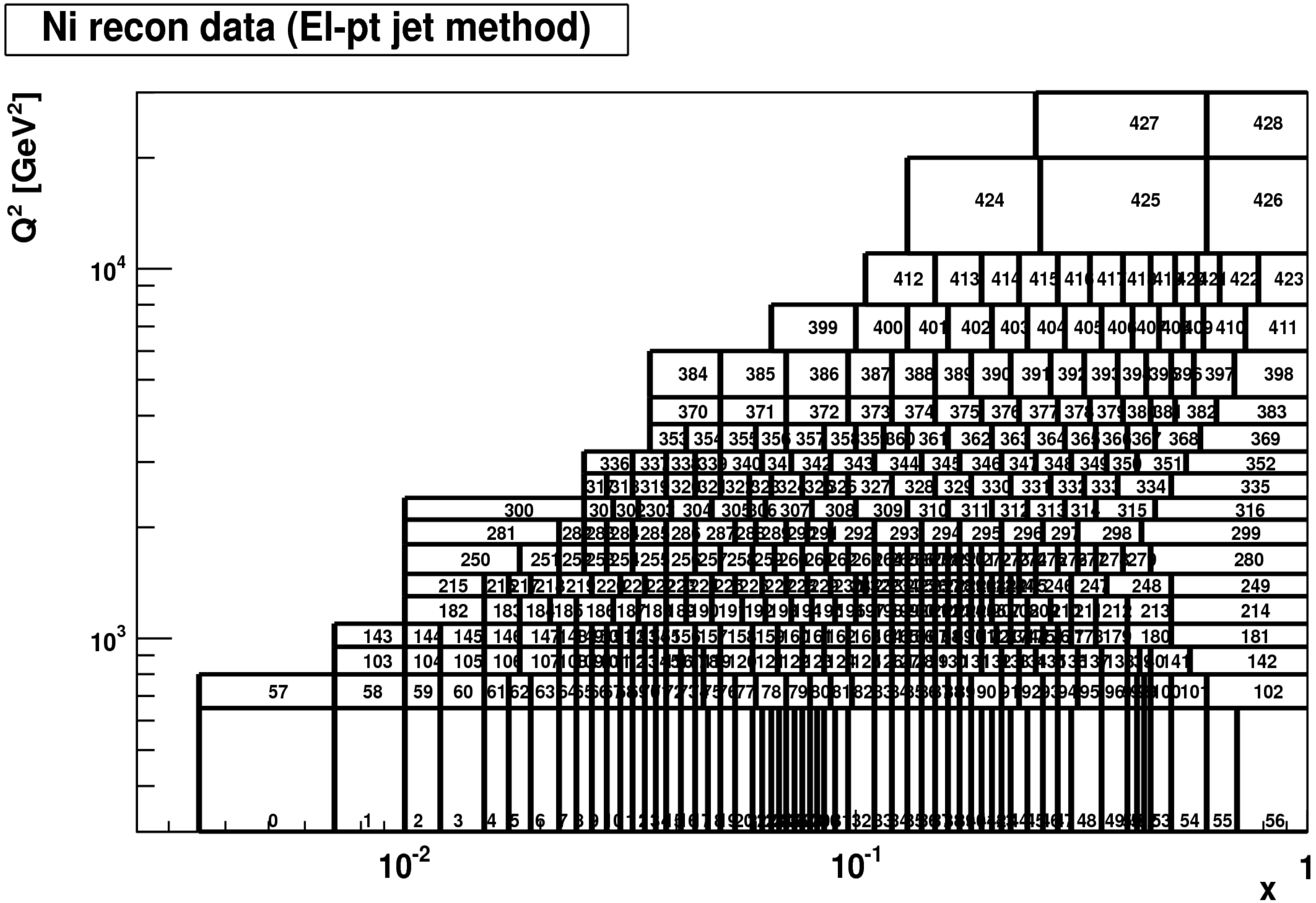
4) A comparison of different PDFs can be done on the basis of best explanation to the high-x data using Transfer Matrix.

5)Statistical, normalization and systemtic errors in a_ij checked.

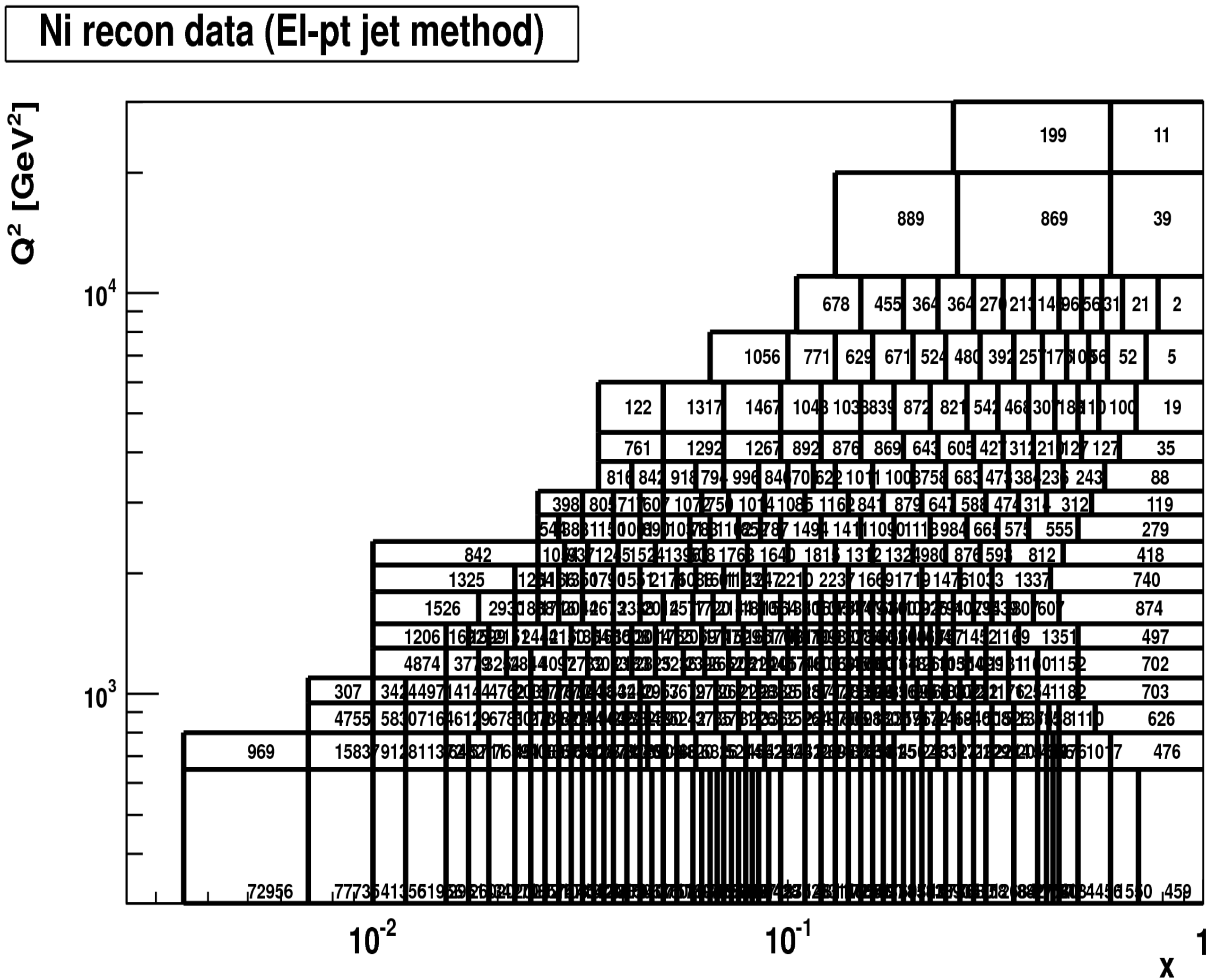
6)How to propagate them to the expectations : to be studied.

Back Up
(some Old slides)

extended binning Bin number (total 429 bins)



extended binning generated events (sum of weights)



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

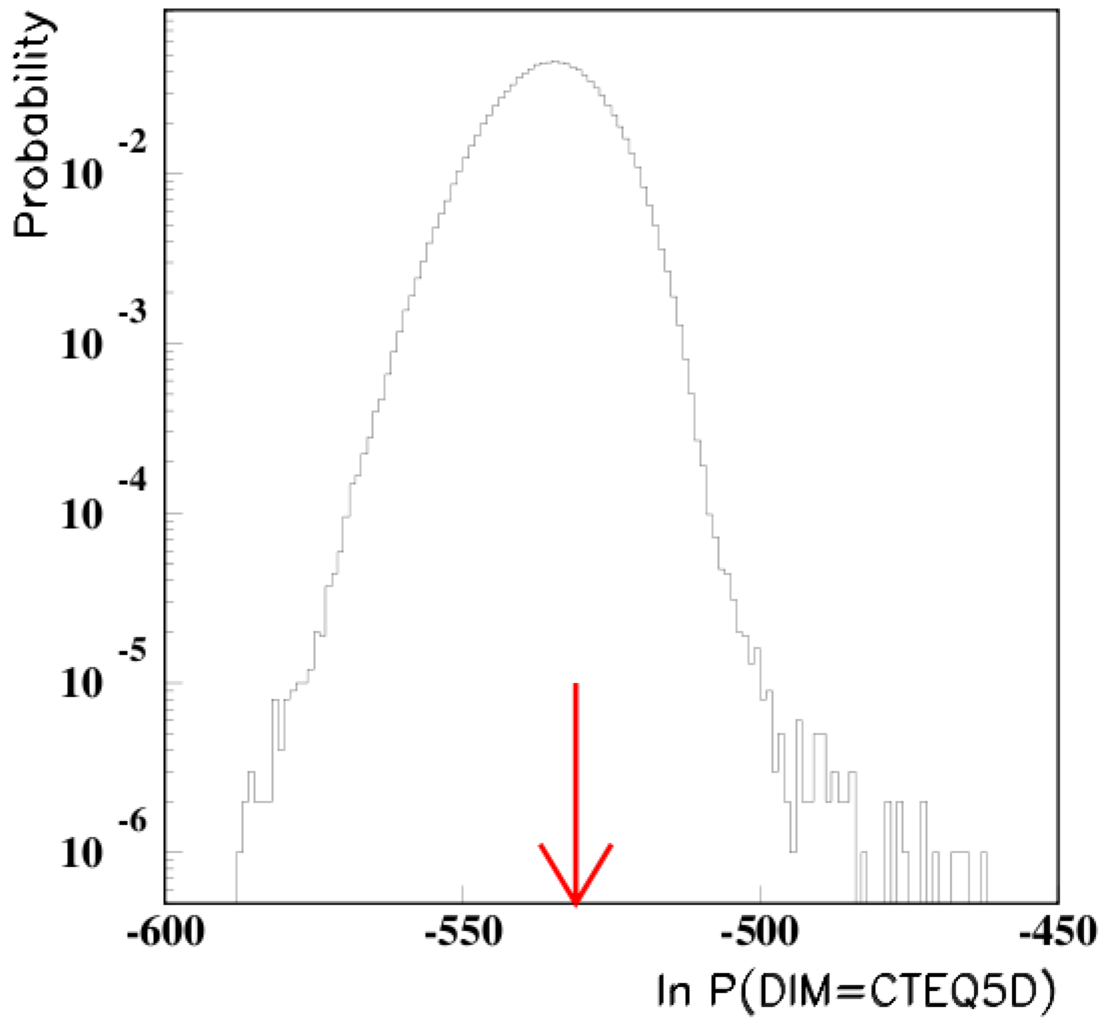
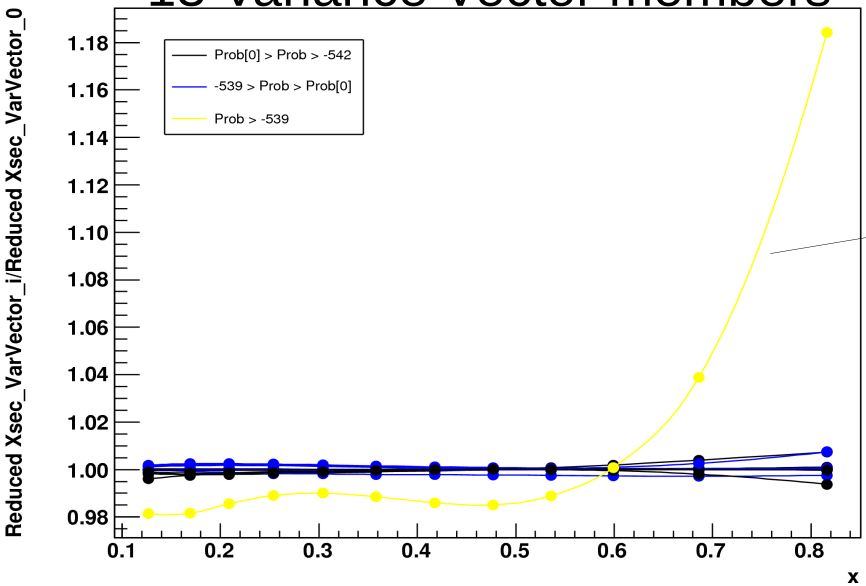


FIGURE 4. Distribution of expected values for $\ln P(D|M = CTEQ)$ for the e^+p data set. The arrow shows the value found in the data.

Study II - Check the Probability for each member in HERAPDF2.0 Error band For Q^2 ~ 9200 GeV2

13 Variance Vector members

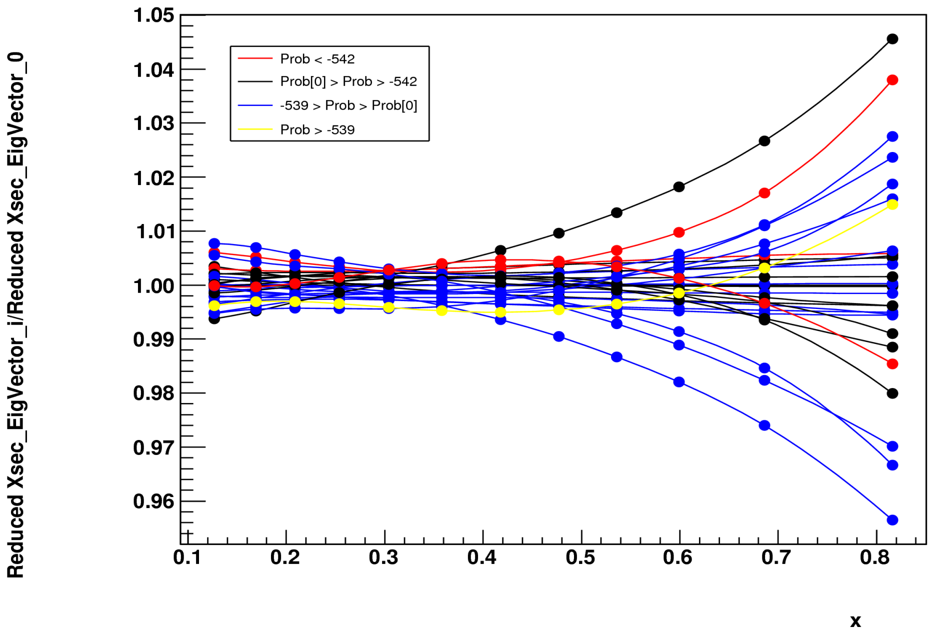


Prob[0] : Central value

13^{th} (last) member in the Vector
Corresponds to u-valence parameter

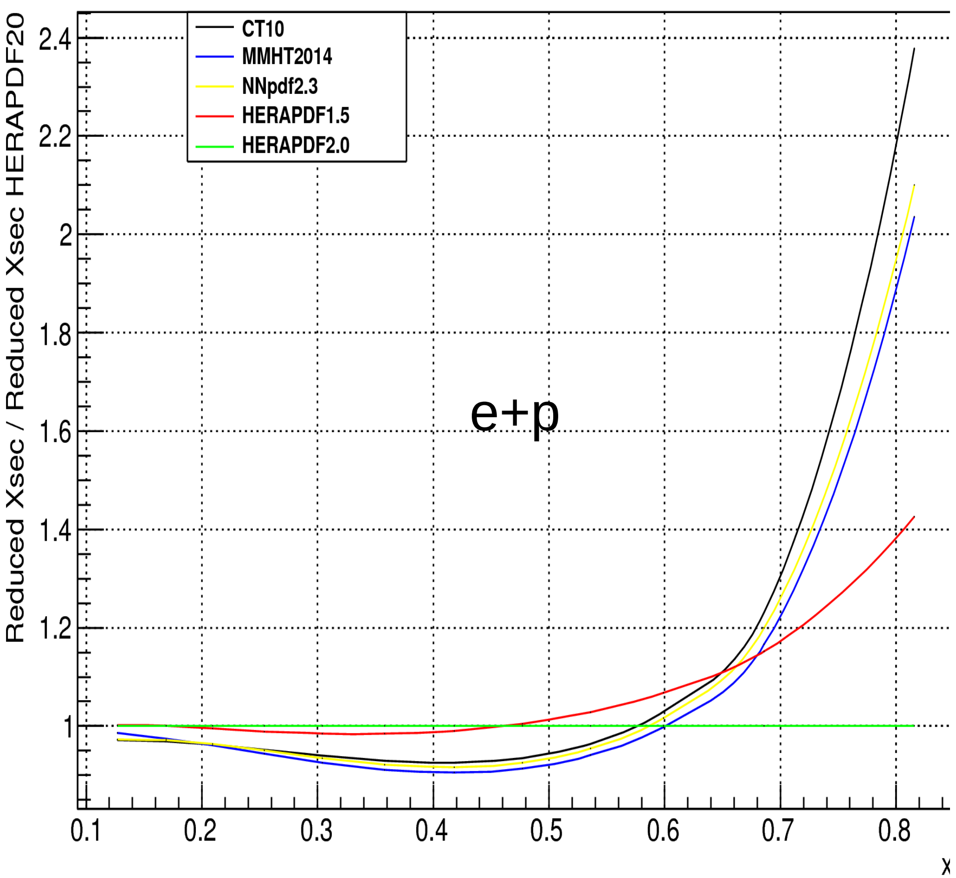
Variation in the PDFs within HERAPDF2.0 much smaller than the inter PDF variation. Big change comes from #13 which has different normalization!

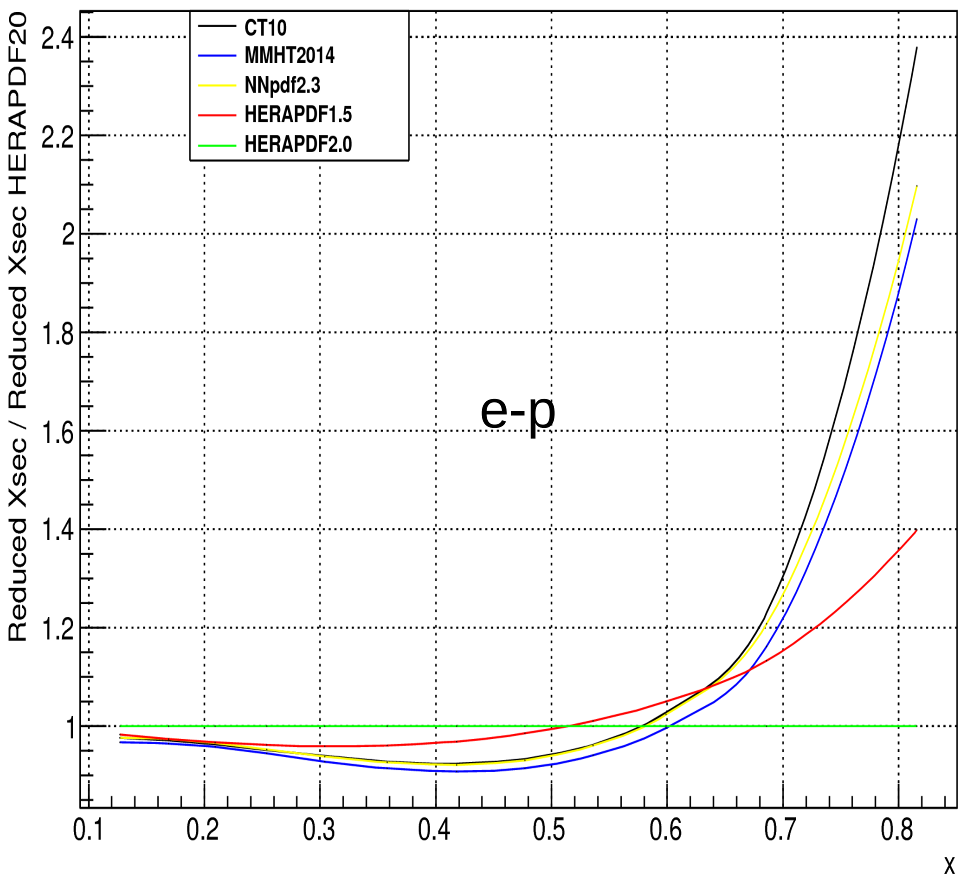
28 Eigen Vector members



How other PDFs behave in the same x-Q^2 region

For Q^2 ~ 9200 GeV2





Difference between PDFs much bigger than difference within HERAPDF set!

Check 1 : Comparing N (calculated from Transfer Matrix) for different Pdfs

For high-x bins only (~20 bins) e+p

Q2bin	x bin	N_data	CTEQ5D	CT10nlo	HERAPDF1.5	HERAPDF2.0	MMHT2014	NNPDF2.3
650 - 800 ,	0.26 - 1.00 ,	371,	405.72,	394.06,	410.30,	423.25,	389.24,	393.99
800 - 950,	0.28 - 1.00,	482,	489.51,	474.57,	495.66,	511.02,	468.23,	474.34
950 - 1100,	0.32 - 1.00,	281,	306.00,	295.75,	309.94,	319.02,	291.59,	295.59
1100 - 1300,	0.34 - 1.00,	275,	263.15,	253.57,	266.32,	273.48,	249.89,	253.43
1300 - 1500,	0.36 - 1.00,	146,	159.65,	153.43,	161.38,	165.35,	151.13,	153.32
1500 - 1800,	0.39 - 1.00,	115,	137.63,	131.77,	138.76,	141.57,	129.71,	131.65
1800 - 2100,	0.43 - 1.00,	62,	71.67,	68.33,	71.88,	72.89,	67.21,	68.25
2100 - 2400,	0.46 - 1.00,	31,	40.99,	38.95,	40.84,	41.15,	38.27,	38.89
2400 - 2800,	0.50 - 1.00,	27,	29.22,	27.68,	28.83,	28.79,	27.16,	27.62
2800 - 3200,	0.54 - 1.00,	13,	15.03,	14.20,	14.62,	14.43,	13.90,	14.15
3200 - 3800,	0.58 - 1.00,	11,	11.01,	10.41,	10.53,	10.26,	10.15,	10.35
3800 - 4500,	0.63 - 1.00,	6,	4.82,	4.57,	4.47,	4.26,	4.43,	4.53
4500 - 6000,	0.69 - 1.00,	3,	3.03,	2.92,	2.67,	2.47,	2.79,	2.86
6000 - 8000,	0.59 - 0.73,	1,	4.44,	4.16,	4.18,	3.98,	4.03,	4.11
6000 - 8000,	0.73 - 1.00,	0,	0.98,	0.96,	0.83,	0.75,	0.90,	0.93
8000 - 11000,	0.57 - 0.64,	2,	2.29,	2.13,	2.21,	2.13,	2.07,	2.11
8000 - 11000,	0.64 - 0.78,	0,	1.82,	1.72,	1.63,	1.49,	1.64,	1.68
8000 - 11000,	0.78 - 1.00,	0,	0.34,	0.35,	0.27,	0.23,	0.32,	0.33
11000 - 20000,	0.60 - 1.00,	3,	2.99,	2.82,	2.80,	2.60,	2.70,	2.75

Check 1 : Comparing N (calculated from Transfer Matrix) for different Pdfs

For high-x bins only (~20 bins) e-p

Q2bin	x bin	N_data	CTEQ5D	CT10nlo	HERAPDF1.5	HERAPDF2.0	MMHT2014	NNPDF2.3
650 - 800,	0.26 - 1.00,	504,	532.79,	517.39,	537.46,	555.91,	511.00,	517.30
800 - 950,	0.28 - 1.00,	671,	635.27,	615.70,	642.07,	663.36,	607.38,	615.41
950 - 1100,	0.32 - 1.00,	414,	407.28,	393.53,	412.05,	424.52,	388.02,	393.37
1100 - 1300,	0.34 - 1.00,	368,	348.28,	335.45,	352.39,	361.86,	330.60,	335.31
1300 - 1500,	0.36 - 1.00,	202,	210.08,	201.77,	212.44,	217.51,	198.75,	201.66
1500 - 1800,	0.39 - 1.00,	173,	181.26,	173.43,	182.95,	186.35,	170.75,	173.33
1800 - 2100,	0.43 - 1.00,	74,	95.75,	91.18,	96.29,	97.25,	89.70,	91.12
2100 - 2400,	0.46 - 1.00,	51,	53.00,	50.29,	53.01,	53.10,	49.43,	50.25
2400 - 2800,	0.50 - 1.00,	36,	37.61,	35.57,	37.30,	36.94,	34.90,	35.52
2800 - 3200,	0.54 - 1.00,	19,	20.34,	19.21,	19.95,	19.49,	18.80,	19.16
3200 - 3800,	0.58 - 1.00,	17,	14.32,	13.52,	13.81,	13.28,	13.18,	13.47
3800 - 4500,	0.63 - 1.00,	5,	6.32,	6.00,	5.93,	5.55,	5.80,	5.95
4500 - 6000,	0.69 - 1.00,	3,	4.34,	4.18,	3.88,	3.50,	3.98,	4.11
6000 - 8000,	0.59 - 0.73,	10,	5.88,	5.49,	5.53,	5.22,	5.32,	5.46
6000 - 8000,	0.73 - 1.00,	1,	1.47,	1.43,	1.26,	1.11,	1.34,	1.39
8000 - 11000,	0.57 - 0.64,	4,	4.05,	3.75,	3.86,	3.73,	3.64,	3.73
8000 - 11000,	0.64 - 0.78,	1,	2.46,	2.32,	2.21,	2.02,	2.21,	2.28
8000 - 11000,	0.78 - 1.00,	1,	0.32,	0.34,	0.24,	0.19,	0.30,	0.31
11000 - 20000,	0.60 - 1.00,	8,	5.28,	4.94,	4.82,	4.58,	4.75,	4.90