MAX-PLANCK-INSTITUT
FÜR PHYSIK

New methods for inclusive DIS

Allen Caldwell

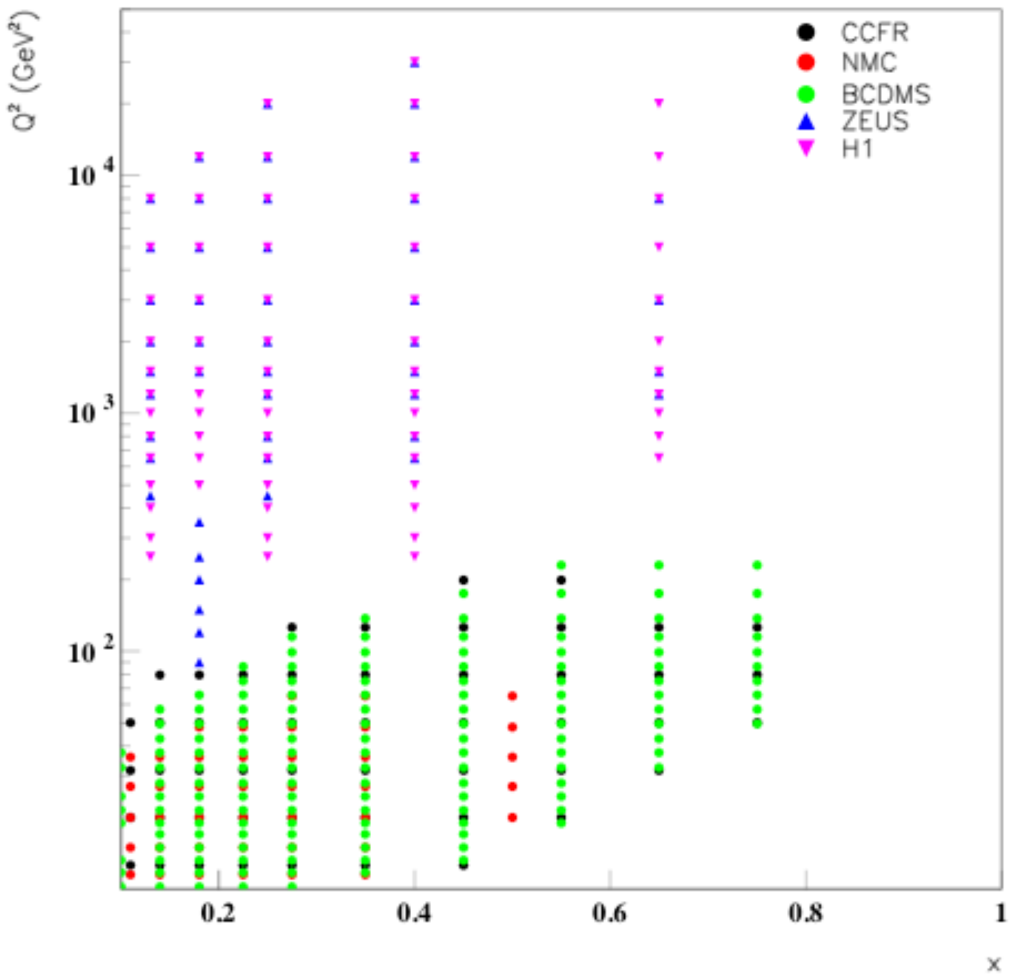
Workshop: Hera 4 EIC

June 8-10, 2022

1. Inclusive DIS measurements as $x \rightarrow 1$
2. Kinematic fitting for variable reconstruction
3. Reporting & analysis of data

Inclusive DIS measurements

Limited information used in global PDF fits for $x \to 1$



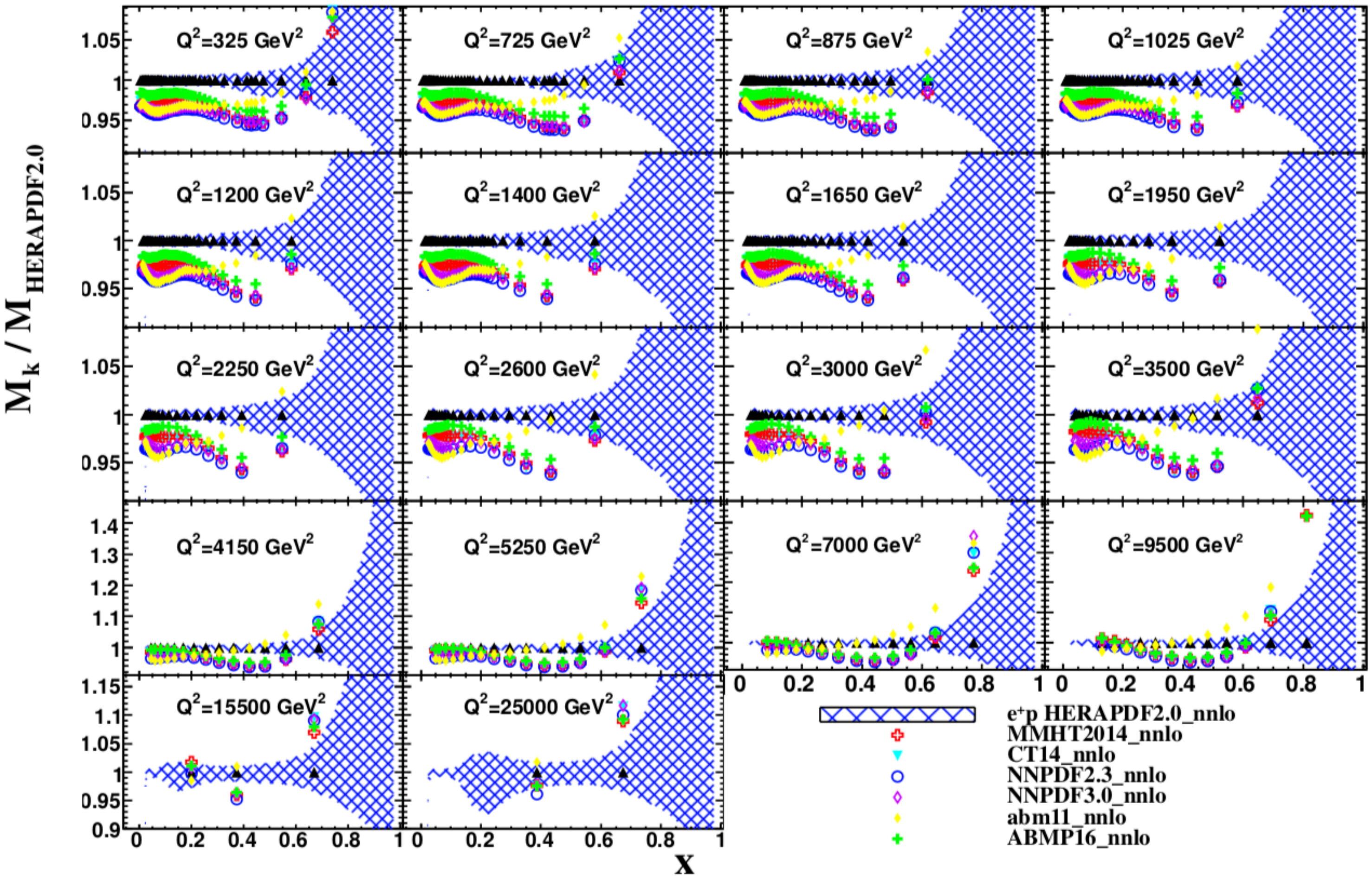
BCDMS has measured F_2 up to x=0.75

Combined H1, ZEUS measurements of F_2 up to x=0.65

ZEUS has measured up to x=1, but these data are not (yet) included in PDF fits.

Expectation: valence distribution behaves as $(1-x)^K$ according to quark counting rules, but would be good to test with more data.

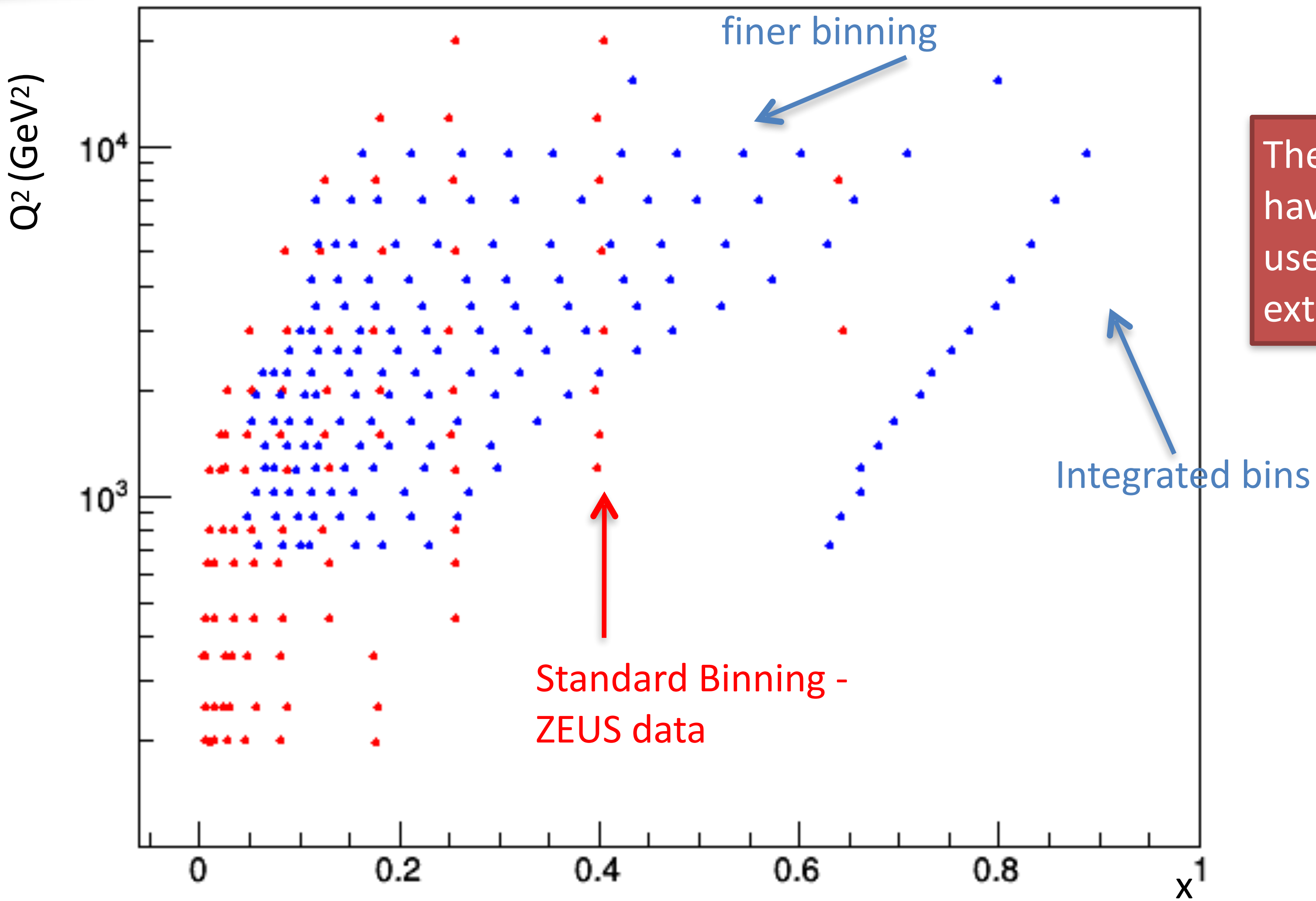
Comparisons of Parametrizations



Sizable differences in expectations (much bigger than quoted uncertainties) despite the fact that fits typically use similar parametrization $\propto (1-x)^K$. Is it possible to improve this situation ? (from I. Abt et al., ZEUS Collaboration, Phys. Rev. D **101** 112009).

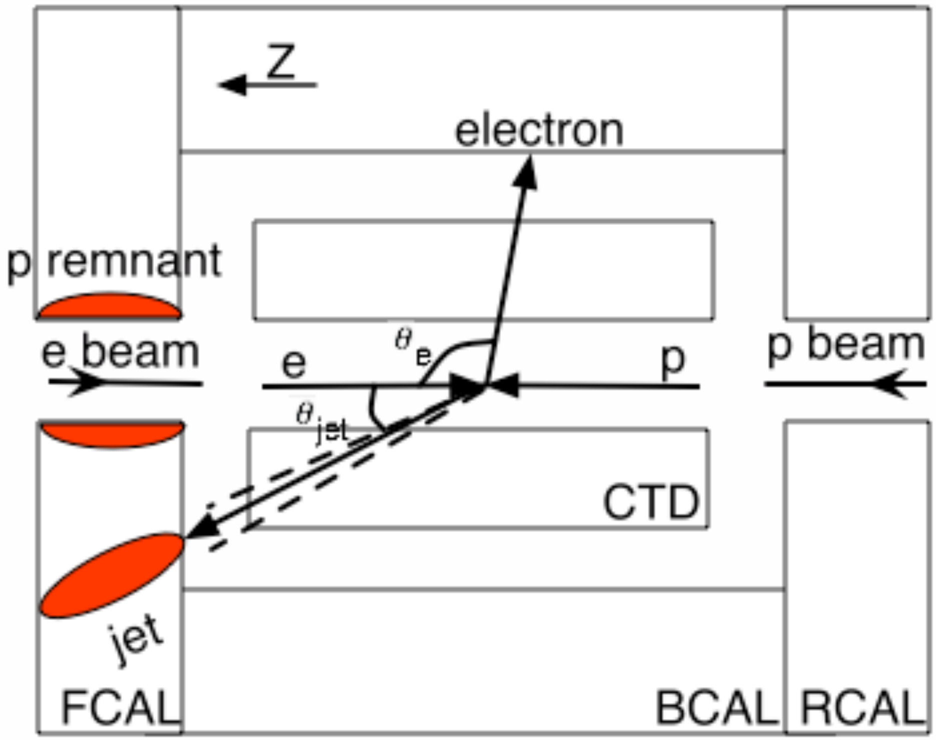
Measurement of neutral current $e^{\pm}p$ cross sections at high Bjorken x with the ZEUS detector

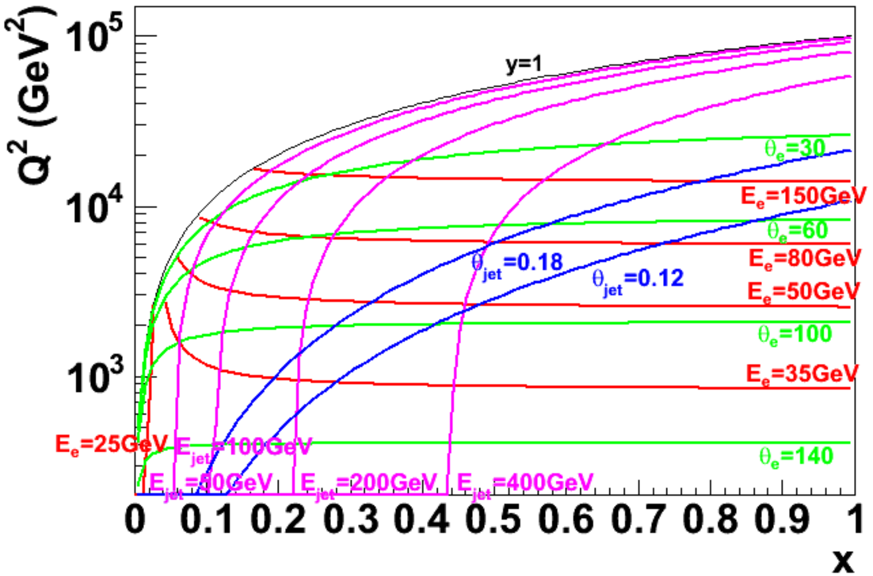
H. Abramowicz *et al.* (ZEUS Collaboration)
Phys. Rev. D **89**, 072007 – Published 8 April 2014



These data have not been used in PDF extractions

ZEUS high-x analysis

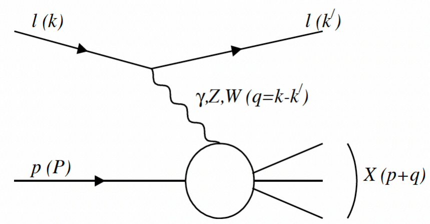




- At high Q^2, scattered electron seen with ≈100% acceptance
- For not too high x, measure x from hadronic system and count events in fine (x, Q^2) bins
- For x>x_{Edge}, count events and assigned to a bin ranging from $(x_{\mathrm{edge}}, 1)$ in well-defined Q^2 range

Kinematic Fitting for variable reconstruction

$$eP \rightarrow eX$$



$$\begin{aligned} s &= (k+P)^2 , \\ W^2 &= (q+P)^2 = p_{\prime}^2 , \\ x &= \frac{Q^2}{2P \cdot q} , \\ y &= \frac{q \cdot P}{k \cdot P} , \\ \nu &= \frac{q \cdot P}{m_N} . \end{aligned}$$

in principle the kinematics can be reconstructed from two variables (e.g., energy and angle of scattered electron)

Electron Method:

$$\begin{aligned} Q^2 &= 2E_e E_e'(1+\cos\theta_e) \\ y &= 1 - \frac{E_e'}{2E_e}(1-\cos\theta_e) \\ x &= \frac{Q^2}{sy} \end{aligned}$$

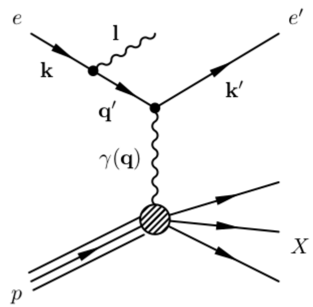
Hadron Method:

$$\begin{aligned} \delta_{had} &= \sum_{i=1}^{\#hadrons} E_i(1-\cos\theta_i) \\ &= E_{had} - p_{z\,had} \\ y &= \frac{\delta_{had}}{2E_e} \\ Q^2 &= \frac{p_{t\,had}^2}{1-y} \\ x &= \frac{Q^2}{sy} \end{aligned}$$

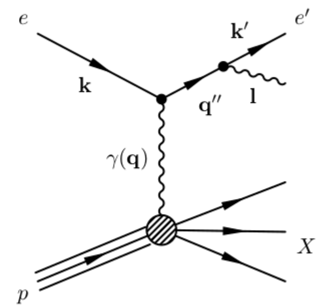
Double Angle Method:

$$\begin{aligned} \cos\gamma &= \frac{p_{t\,had}^2 - \delta_{had}^2}{p_{t\,had}^2 + \delta_{had}^2} \\ Q^2 &= 4E_e^2 \frac{\sin\gamma(1+\cos\theta_e)}{\sin\gamma + \sin\theta_e - \sin(\theta_e+\gamma)} \\ x &= \frac{E_e}{E_p} \frac{\sin\gamma + \sin\theta_e + \sin(\theta_e+\gamma)}{\sin\gamma + \sin\theta_e - \sin(\theta_e+\gamma)} \end{aligned}$$

Reconstructing the kinematics in the presence of radiation leads to errors when only two measured quantities taken into account



Initial State Radiation



Final State Radiation

Kinematic Fit: Use the information from the electron and hadronic system to reconstruct three pieces of information.
Bayesian approach - build in knowledge of distributions

$$P(x, y, E_\gamma | \mathbf{D}) \propto P(\mathbf{D} | x, y, E_\gamma) P_0(x, y, E_\gamma)$$

E_γ Energy of ISR photon

Q^2 calculated from x,y

$$A_r = A - E_\gamma$$

$$E = xyP + A_r(1-y)$$

$$F = x(1-y)P + yA_r$$

$$\cos\theta = \frac{xyP - A_r(1-y)}{xyP + A_r(1-y)}$$

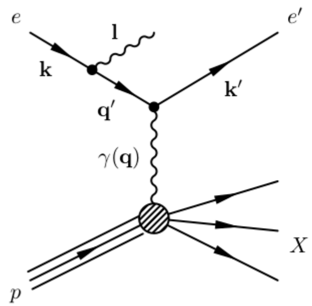
$$\cos\gamma = \frac{x(1-y)P - yA_r}{x(1-y)P + yA_r}$$

$$\mathbf{D} = \{E, \theta, P_T^{\text{had}}, \delta_{\text{had}}\}$$

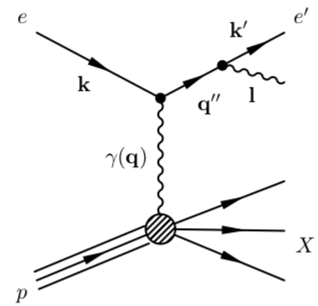
$$\left.\begin{aligned} P_T^{\text{had}} &= F\sin\gamma \\ \delta_{\text{had}} &= F(1-\cos\gamma) \end{aligned}\right\} \text{ from X}$$

$$x = \frac{Q^2}{2P\cdot q} \qquad y = \frac{Q^2}{s'x} \qquad s' = (k + P - E_\gamma)^2$$

Reconstructing the kinematics in the presence of radiation leads to errors when only two measured quantities taken into account



Initial State Radiation



Final State Radiation

Bayesian approach - build in knowledge of distributions

$$P(x, y, E_\gamma \,|\, \mathbf{D}) \propto P(\mathbf{D} \,|\, x, y, E_\gamma) P_0(x, y, E_\gamma)$$

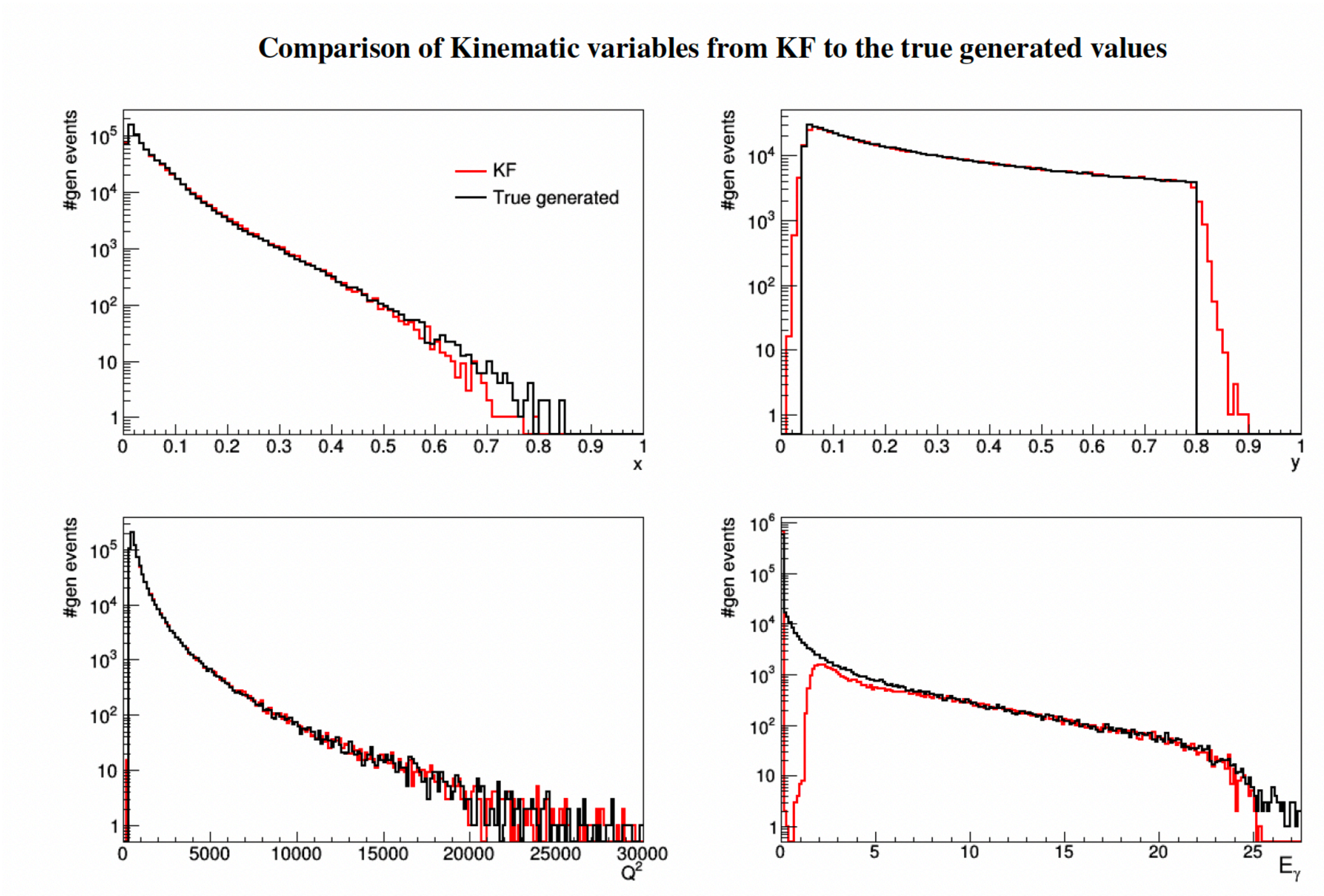
$$\begin{aligned} P_0(x, y, E_\gamma) &= P_0(x, y) P_0(E_\gamma) \\ P_0(x, y) &\propto \frac{(1-x)^5}{x^2 y^2} \\ P_0(E_\gamma) &\propto \frac{1 + (1 - E_\gamma/(E_e - E_\gamma))^2}{E_\gamma/(E_e - E_\gamma)} \end{aligned}$$

$$\begin{aligned} P(\mathbf{D} \,|\, x, y, E_\gamma) &= P(E, \theta, P_T^{\mathrm{had}}, \delta_{\mathrm{had}} \,|\, x, y, E_\gamma) \\ &= P(E, \theta \,|\, x, y, E_\gamma) P(P_T^{\mathrm{had}}, \delta_{\mathrm{had}} \,|\, x, y, E_\gamma) \end{aligned}$$

$$\approx P(E \,|\, x, y, E_\gamma) P(\theta \,|\, x, y, E_\gamma) P(P_T^{\mathrm{had}} \,|\, x, y, E_\gamma) P(\delta_{\mathrm{had}} \,|\, x, y, E_\gamma)$$

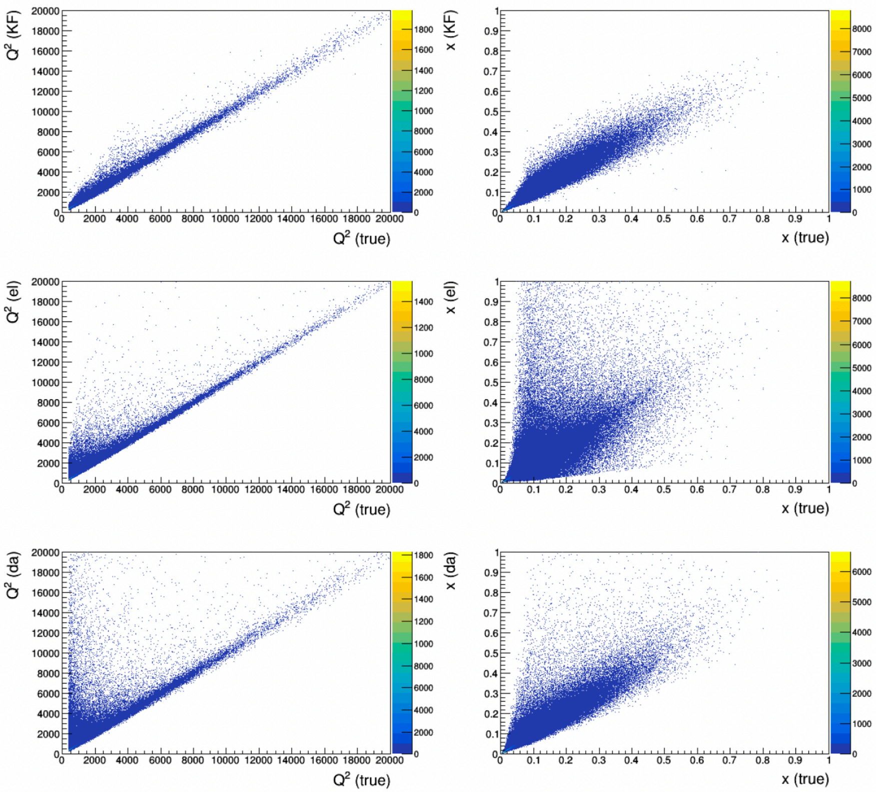
Each term taken initially as Normal distribution with measured value distributed around predicted value with a known resolution.

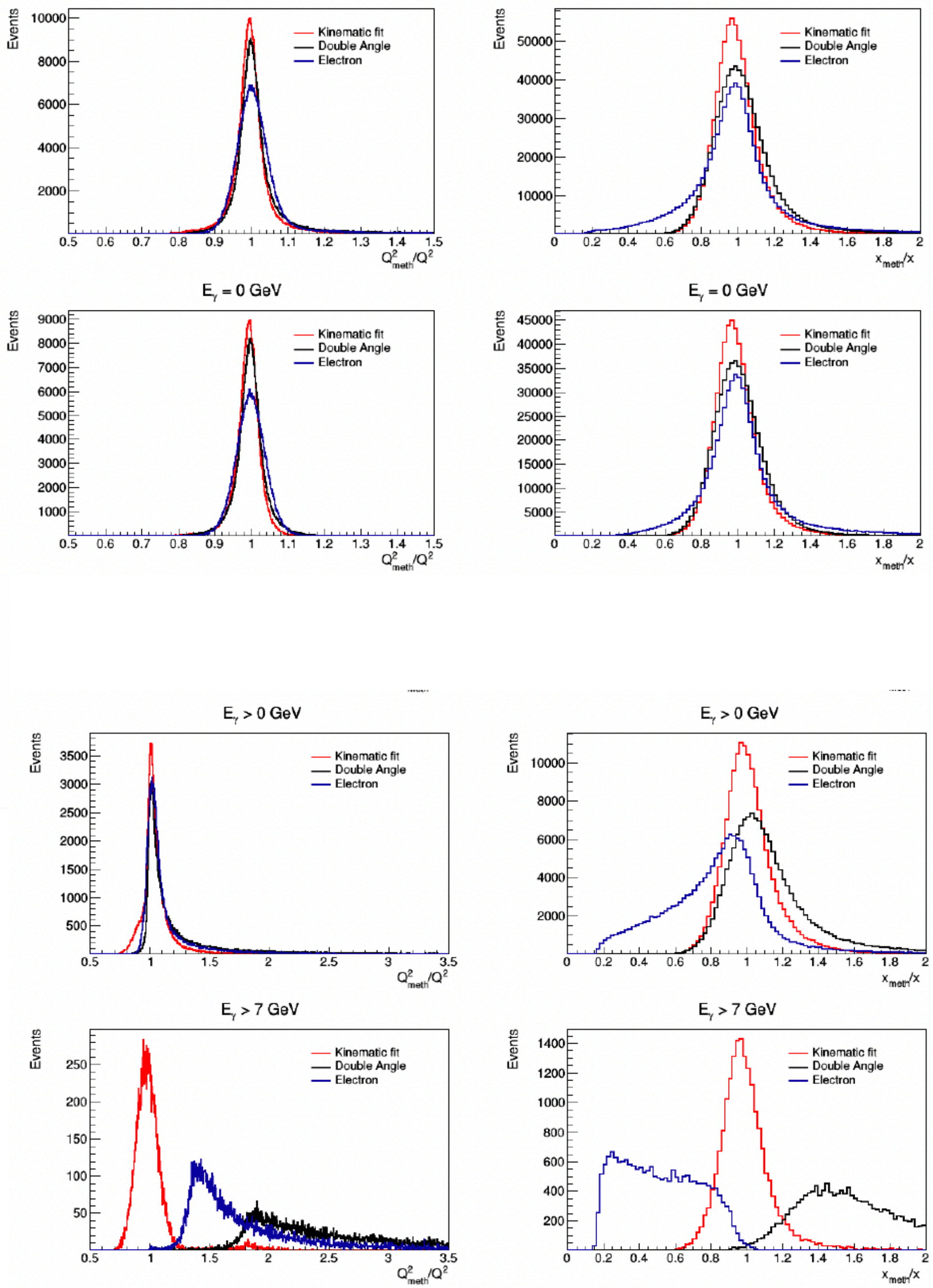
Correlations between electron, hadron variables should eventually be taken into account.



From simulation study of high Q^2 events (R. Aggarwal, Savitribai Phule Pune University, Pune)

Comparison of KF reconstruction to electron and double angle method





Used in the ZEUS high-x analysis so far.

The Kinematic Fitting approach should be further developed and, assuming it holds up, applied across the full kinematic plane. Could be tried on HERA data sets, and later at the EIC.

Inclusive DIS measurements

5250	0.62	5	1.76e − 04	$^{+55.2}_{-35.2}$	$^{+11.1}_{-10.5}$	$^{+9.2}_{-9.1}$	$^{-3.2}_{+4.6}$	$^{+0.1}_{+0.1}$	$^{+3.7}_{-3.7}$	$^{+0.6}_{-0.6}$
7000	0.12	93	1.61e − 02	$^{+10.4}_{-10.4}$	$^{+4.0}_{-5.1}$	$^{+3.3}_{-3.6}$	$^{-1.1}_{+0.8}$	$^{+0.8}_{-0.5}$	$^{-0.7}_{+0.7}$	$^{+0.0}_{-0.0}$
7000	0.14	89	1.25e − 02	$^{+10.6}_{-10.6}$	$^{+3.7}_{-5.2}$	$^{+3.4}_{-3.5}$	$^{-1.3}_{+1.2}$	$^{-0.5}_{+0.2}$	$^{-0.9}_{+0.9}$	$^{+0.0}_{-0.0}$
7000	0.18	68	7.02e − 03	$^{+12.1}_{-12.1}$	$^{+3.9}_{-3.6}$	$^{+3.4}_{-3.4}$	$^{-0.6}_{+0.6}$	$^{-0.6}_{+0.4}$	$^{-0.4}_{+0.4}$	$^{+0.0}_{-0.0}$
7000	0.22	56	5.60e − 03	$^{+13.4}_{-13.4}$	$^{+4.2}_{-4.2}$	$^{+3.9}_{-3.9}$	$^{-1.4}_{+1.1}$	$^{-0.4}_{+1.0}$	$^{-0.4}_{+0.4}$	$^{+0.0}_{-0.0}$
7000	0.26	49	3.79e − 03	$^{+14.3}_{-14.3}$	$^{+4.6}_{-4.8}$	$^{+3.9}_{-4.0}$	$^{-0.2}_{+2.1}$	$^{+0.2}_{-0.2}$	$^{-0.5}_{+0.5}$	$^{+0.0}_{-0.0}$
7000	0.32	41	2.70e − 03	$^{+15.6}_{-15.6}$	$^{+5.1}_{-4.7}$	$^{+5.3}_{-4.5}$	$^{-1.4}_{+0.8}$	$^{-0.4}_{+0.4}$	$^{-0.2}_{+0.2}$	$^{+0.0}_{-0.0}$
7000	0.38	23	1.52e − 03	$^{+20.9}_{-20.9}$	$^{+6.4}_{-6.2}$	$^{+5.5}_{-5.5}$	$^{-1.8}_{+1.7}$	$^{-0.7}_{+0.4}$	$^{+2.0}_{-2.0}$	$^{+0.0}_{-0.0}$
7000	0.44	17	1.15e − 03	$^{+27.2}_{-21.3}$	$^{+8.4}_{-7.9}$	$^{+7.1}_{-7.1}$	$^{-2.7}_{+2.7}$	$^{-0.0}_{+0.2}$	$^{-2.4}_{+2.4}$	$^{+0.0}_{-0.0}$
7000	0.50	8	5.38e − 04	$^{+41.8}_{-29.4}$	$^{+9.7}_{-10.3}$	$^{+9.5}_{-9.5}$	$^{-1.6}_{+1.4}$	$^{-0.3}_{+0.4}$	$^{+2.4}_{-2.4}$	$^{+0.1}_{-0.1}$
7000	0.56	4	2.37e − 04	$^{+63.2}_{-38.2}$	$^{+12.3}_{-11.8}$	$^{+11.3}_{-11.3}$	$^{-3.4}_{+3.4}$	$^{+0.1}_{-0.0}$	$^{+1.2}_{-1.2}$	$^{+0.2}_{-0.2}$
7000	0.66	10	2.30e − 04	$^{+36.7}_{-26.8}$	$^{+12.6}_{-13.6}$	$^{+12.1}_{-12.3}$	$^{-4.3}_{+2.5}$	$^{-0.3}_{-0.0}$	$^{+2.0}_{-2.0}$	$^{+0.9}_{-0.9}$
9500	0.17	76	6.77e − 03	$^{+11.5}_{-11.5}$	$^{+5.6}_{-7.7}$	$^{+4.9}_{-4.9}$	$^{-2.0}_{+2.3}$	$^{+0.2}_{-0.2}$	$^{-0.6}_{+0.6}$	$^{+0.0}_{-0.0}$
9500	0.21	53	3.87e − 03	$^{+13.7}_{-13.7}$	$^{+5.8}_{-5.1}$	$^{+4.3}_{-4.5}$	$^{-1.1}_{+1.8}$	$^{-0.7}_{+0.4}$	$^{-1.1}_{+1.1}$	$^{+0.0}_{-0.0}$
9500	0.25	40	2.27e − 03	$^{+15.8}_{-15.8}$	$^{+4.8}_{-4.9}$	$^{+4.5}_{-4.5}$	$^{-2.0}_{+1.5}$	$^{+0.2}_{+0.4}$	$^{+0.1}_{-0.1}$	$^{+0.0}_{-0.0}$
9500	0.31	27	1.50e − 03	$^{+19.2}_{-19.2}$	$^{+5.7}_{-8.1}$	$^{+5.2}_{-5.3}$	$^{-2.6}_{+1.5}$	$^{-0.5}_{+0.2}$	$^{+1.5}_{-1.5}$	$^{+0.0}_{-0.0}$
9500	0.36	19	8.89e − 04	$^{+25.5}_{-20.3}$	$^{+6.6}_{-6.1}$	$^{+5.9}_{-5.9}$	$^{-1.0}_{+1.9}$	$^{-0.4}_{+0.2}$	$^{-0.3}_{+0.3}$	$^{+0.0}_{-0.0}$
9500	0.42	12	5.64e − 04	$^{+33.1}_{-24.8}$	$^{+11.3}_{-7.5}$	$^{+13.4}_{-7.3}$	$^{-1.0}_{+2.4}$	$^{-0.8}_{+0.5}$	$^{-0.8}_{+0.8}$	$^{+0.0}_{-0.0}$
9500	0.48	8	3.63e − 04	$^{+41.7}_{-29.4}$	$^{+10.5}_{-10.4}$	$^{+9.2}_{-9.2}$	$^{-2.6}_{+2.4}$	$^{-0.4}_{+0.6}$	$^{-3.4}_{+3.4}$	$^{+0.0}_{-0.0}$
9500	0.54	5	2.31e − 04	$^{+55.2}_{-35.2}$	$^{+14.3}_{-13.7}$	$^{+12.5}_{-12.2}$	$^{-1.7}_{+3.6}$	$^{-0.2}_{+0.6}$	$^{+5.7}_{-5.7}$	$^{+0.1}_{-0.1}$
9500	0.61	4	1.39e − 04	$^{+63.3}_{-38.3}$	$^{+15.5}_{-15.4}$	$^{+14.6}_{-14.8}$	$^{-4.2}_{+4.2}$	$^{+0.0}_{-0.1}$	$^{+0.4}_{-0.4}$	$^{+0.4}_{-0.4}$
9500	0.71	1	1.50e − 05	$^{+158.0}_{-58.0}$	$^{+21.1}_{-19.8}$	$^{+18.9}_{-18.9}$	$^{-3.3}_{+4.5}$	$^{-0.4}_{+0.3}$	$^{+4.8}_{-4.8}$	$^{+1.3}_{-1.3}$

Not many events at high x

This uncertainty refers to how well we know the underlying cross section assuming that our only knowledge is the observed number of events. **Not the uncertainty that belongs in a fit.**

Inclusive DIS measurements

Need to use Poisson statistics in analyzing the data since event counts are small.

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

Probability of the data (likelihood)

$$\nu_{j,k} = \mathscr{L}\int_{(\Delta x, \Delta Q^2)_j} \left[\int T(x_{\text{rec}}, Q^2_{\text{rec}}|x, Q^2)\frac{d^2\sigma(x, Q^2|M_k)}{dxdQ^2}dxdQ^2\right] dx_{\text{rec}}dQ^2_{\text{rec}}$$

expectation

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

transfer matrix realization

$$t_{ij} = K_{il}a_{lj}$$

separate radiative & detectors/analysis effects

Radiative matrix from HERACLES - M generated number of events

$$K_{ii} = \frac{M_i}{\mathscr{L}^{\text{MC}}\sigma_{i,o}}$$

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

detector/analysis matrix from ZEUS simulation

Study of proton parton distribution functions at high x using ZEUS data

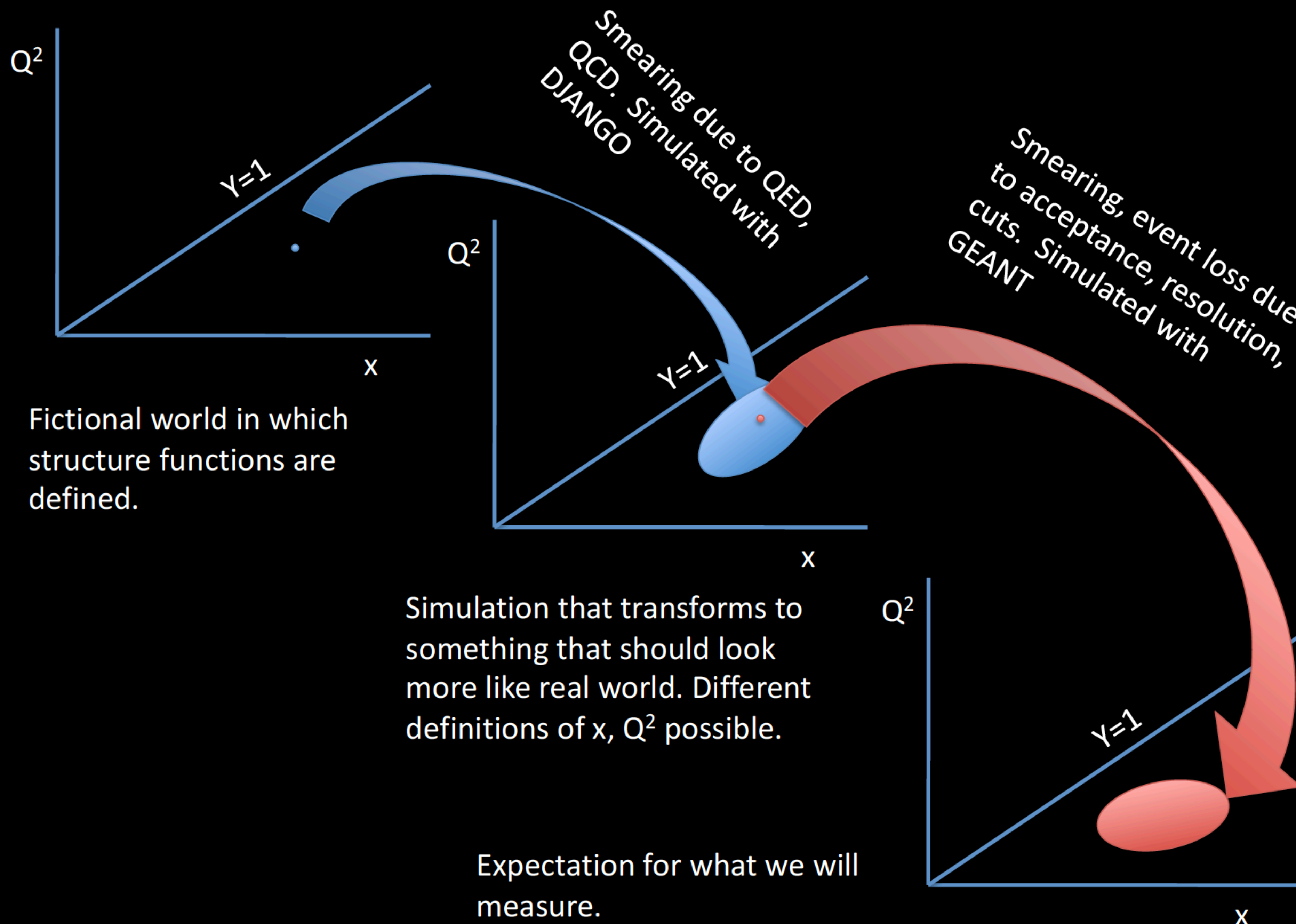
I. Abt *et al.* (ZEUS Collaboration)
Phys. Rev. D **101**, 112009 – Published 26 June 2020

Primary author: **(R. Aggarwal, Savitribai Phule Pune University, Pune)**

Described how to use a forward modeling for analysis of the data:

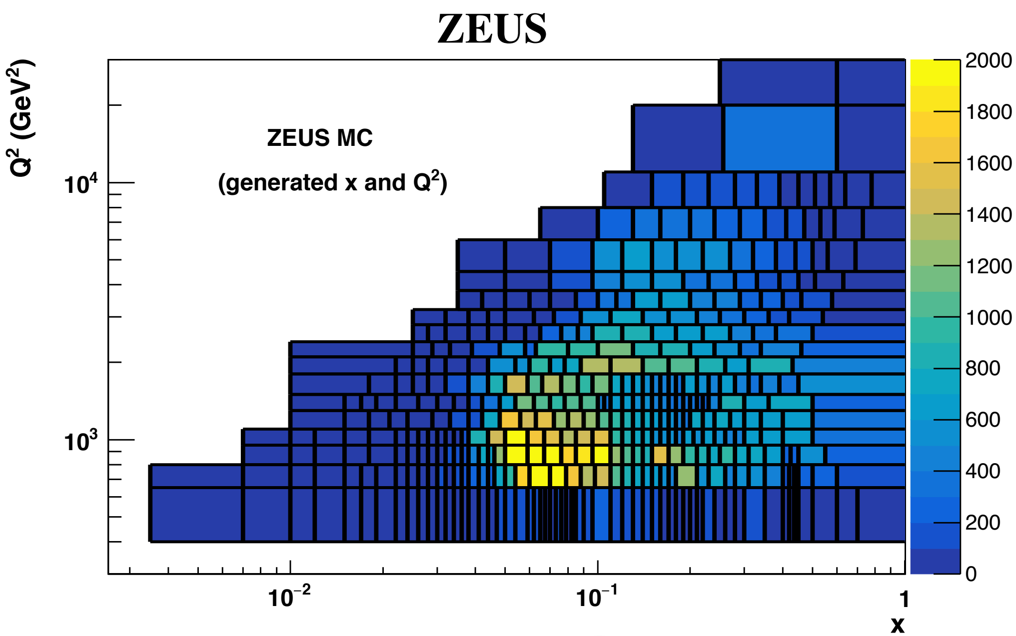
Define pdfs -> apply radiative effects
-> predict cross sections
-> apply detector/analysis effects
-> calculate expected number of events
-> calculate a Poisson probability

We are now developing a PDF fitting package to implement this scheme



Q²
Y=1
x
Fictional world in which structure functions are defined.
Smearing due to QED, QCD. Simulated with DJANGO
Q²
Y=1
x
Simulation that transforms to something that should look more like real world. Different definitions of x, Q² possible.
Smearing, event loss due to acceptance, resolution, cuts. Simulated with GEANT
Q²
Y=1
x
Expectation for what we will measure.

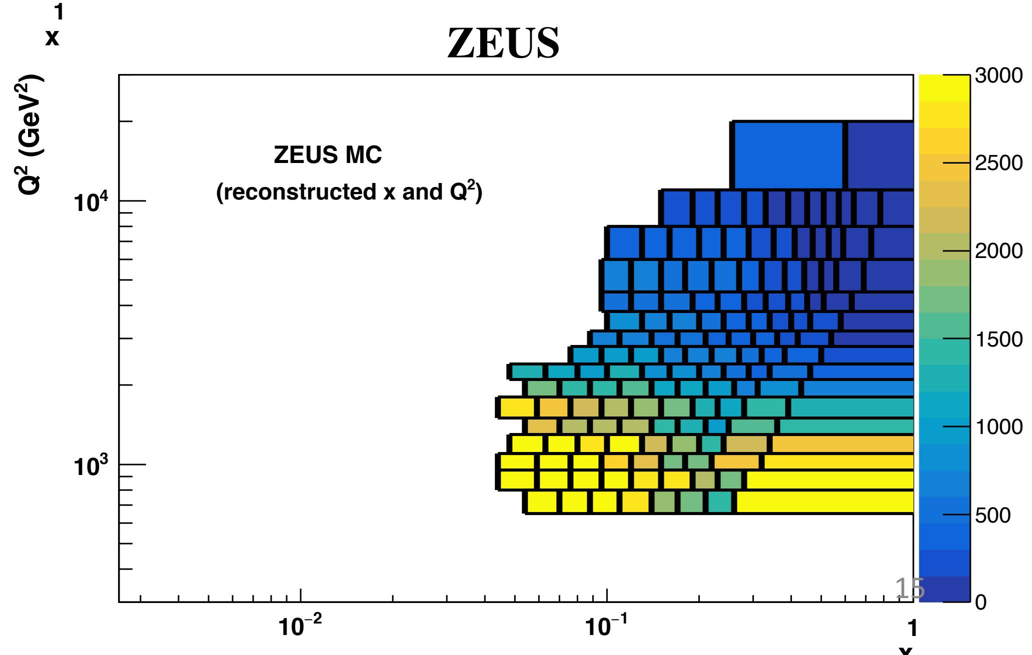
Transfer Matrices



generated variables/bins.

Separate transfer matrices exist for producing radiative cross sections and detector/ analysis effects.

transfer matrix



Reconstructed variables/bins

Procedure

- PDFs defined at a high scale: $Q_0^2 = 100\ \mathrm{GeV}^2$ in the Fixed Flavor number scheme (5 quarks)

- PDFs are evolved at NNLO using QCDNUM to cover the full range of the data

- Structure functions are computed with QCDNUM and represented by cubic splines. These are then used to form the differential cross section, which is also splined. This allows for a fast integration of the cross sections.

- The predictions at the observed level are then calculated using the transfer matrices

expected counts at generator level

$$\nu_j = (1 + 0.018 \cdot \beta_0^{+-}) \left[\sum_i \nu_i \cdot (a_{ij} + \sum_k \beta_k \delta_{ij}^k) \right]$$

normalization uncertainty

transfer matrix

systematic variations

$\beta' s$ are Unit Normal distributed nuisance parameters

The probability of observing the data is then calculated using the Poisson distribution

A first try

$Q_0^2 = 100\ \mathrm{GeV}^2$

$$\sum_i \int_0^1 x f_i(x) dx = \sum_i \Delta_i = 1$$

Densities & evolution in FFN (5) scheme & NNLO

$$\int_0^1 u(x) - \bar{u}(x) dx = 2$$

$$\int_0^1 d(x) - \bar{d}(x) dx = 1$$

$$\int_0^1 f(x) - \bar{f}(x) dx = 0 \qquad f \neq u, d, g$$

Parametrizations

$$xu_V(x) = xu(x) - x\bar{u}(x) = A_u x^{\lambda_u}(1-x)^{K_u}$$

$$xd_V(x) = xd(x) - x\bar{d}(x) = A_d x^{\lambda_d}(1-x)^{K_d}$$

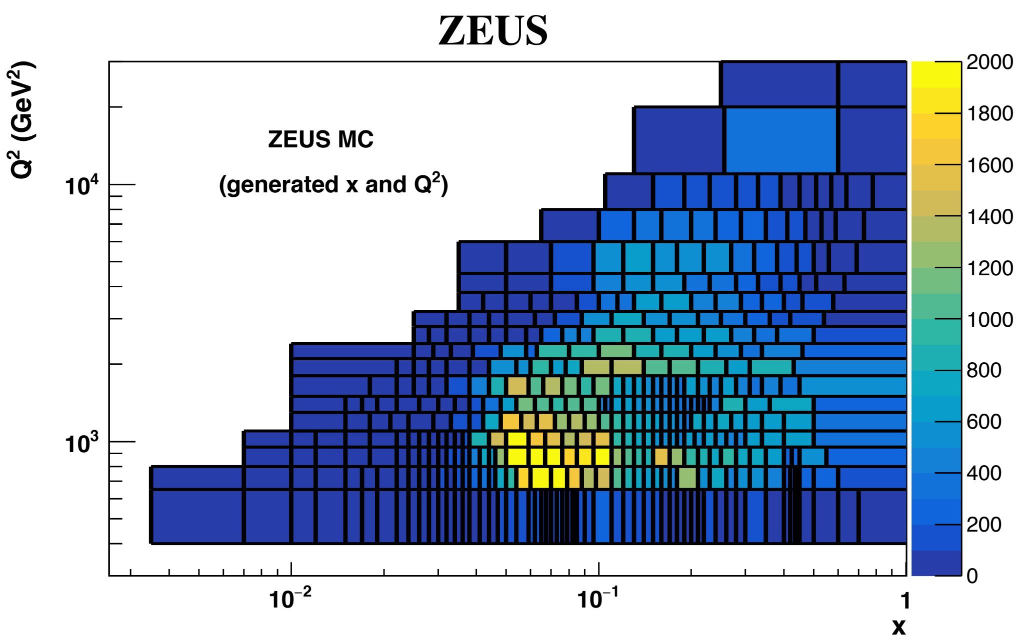
$$x\bar{u}(x) = A_{\bar{u}} x^{\lambda_q}(1-x)^{K_q}$$
$$x\bar{d}(x) = A_{\bar{d}} x^{\lambda_q}(1-x)^{K_q}$$
$$xg(x) = A_{g1} x^{\lambda_{g1}}(1-x)^{K_g} + A_{g2} x^{\lambda_{g2}}(1-x)^{K_q}$$
$$xs(x) = x\bar{s}(x) = A_s x^{\lambda_q}(1-x)^{K_q}$$
$$xc(x) = x\bar{c}(x) = A_c x^{\lambda_q}(1-x)^{K_q}$$
$$xb(x) = x\bar{b}(x) = A_b x^{\lambda_q}(1-x)^{K_q}$$

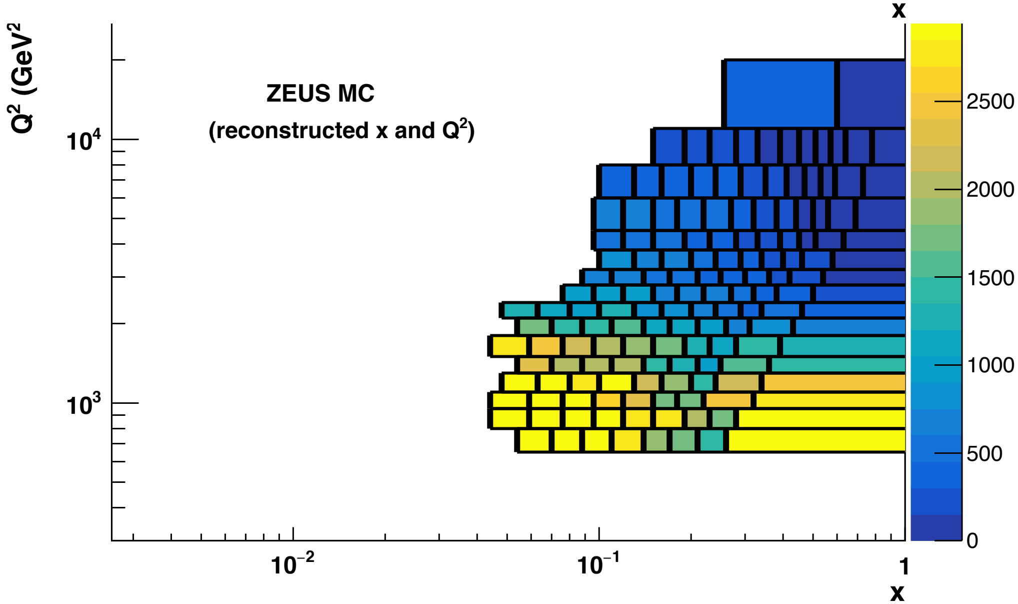
Fit parameters are
Δ_i's, $K_u, K_d, \lambda_{g1}, \lambda_{g2}, K_g, \lambda_q + \beta$

β are nuisance parameters (systematics)
$K_q = 5$ fixed (pdf zero as $x \to 1$)

2 free parameters for data normalization

A first try



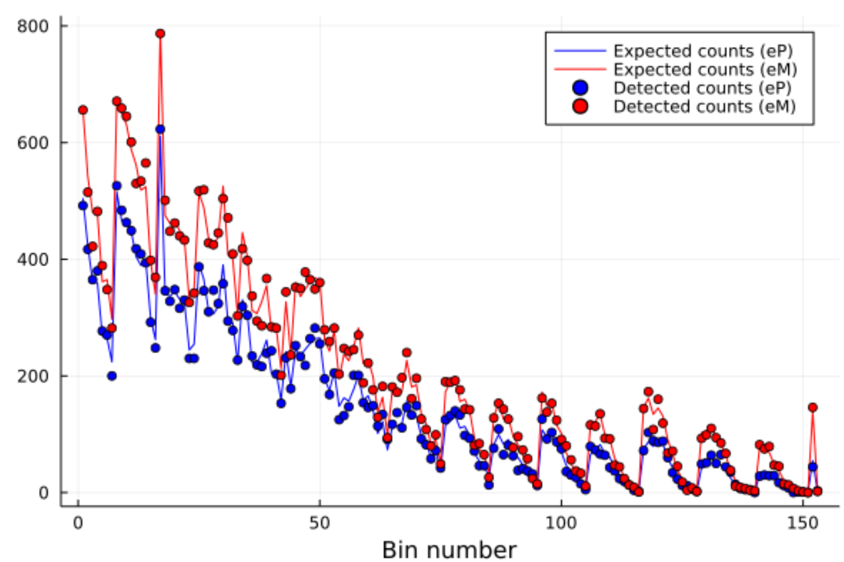


QCDNUM evolves PDFs to cover grid:

SPLINT package gives integrated cross sections in bins

transfer matrix used to get expected numbers of events in bins of observed quantities.

Poisson generated number of events.



A first try

Priors

Δ = Dirichlet([6., 3., 9., 4., 2.,1., 0.2, 0.2, 0.1]),

K_u= Uniform(3., 9.),

K_d= Uniform(3., 9.),

λ_{g1} = Uniform(1., 2.),

λ_{g2}= Uniform(-0.5, -0.1),

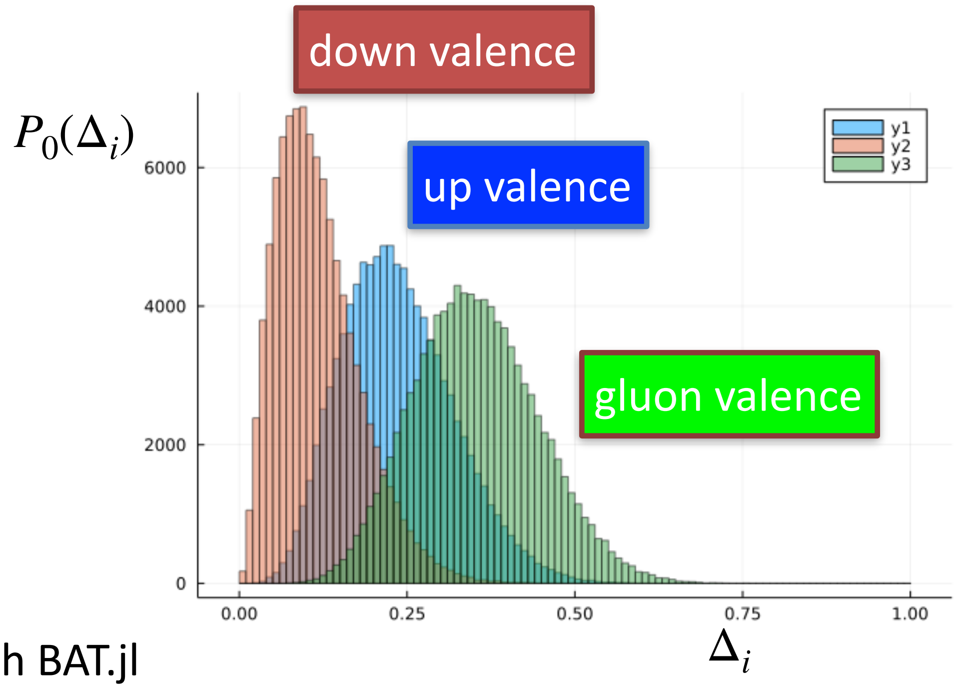
K_g= Uniform(3., 9.),

λ_q = Uniform(-0.5, -0.1),

β_0^+ = Truncated(Normal(0, 1), -5, 5),

β_0^-= Truncated(Normal(0, 1), -5, 5),

Some of the $\Delta's$



Markov Chain MC used to fit simulated data with BAT.jl

$$P(\boldsymbol{\Delta}, K_u, K_d, \lambda_{g1}, \lambda_{g2}, K_g, \lambda_q \,|\, D) \propto P(D \,|\, \boldsymbol{\Delta}, K_u, K_d, \lambda_{g1}, \lambda_{g2}, K_g, \lambda_q) P_0(\boldsymbol{\Delta}, K_u, K_d, \lambda_{g1}, \lambda_{g2}, K_g, \lambda_q)$$

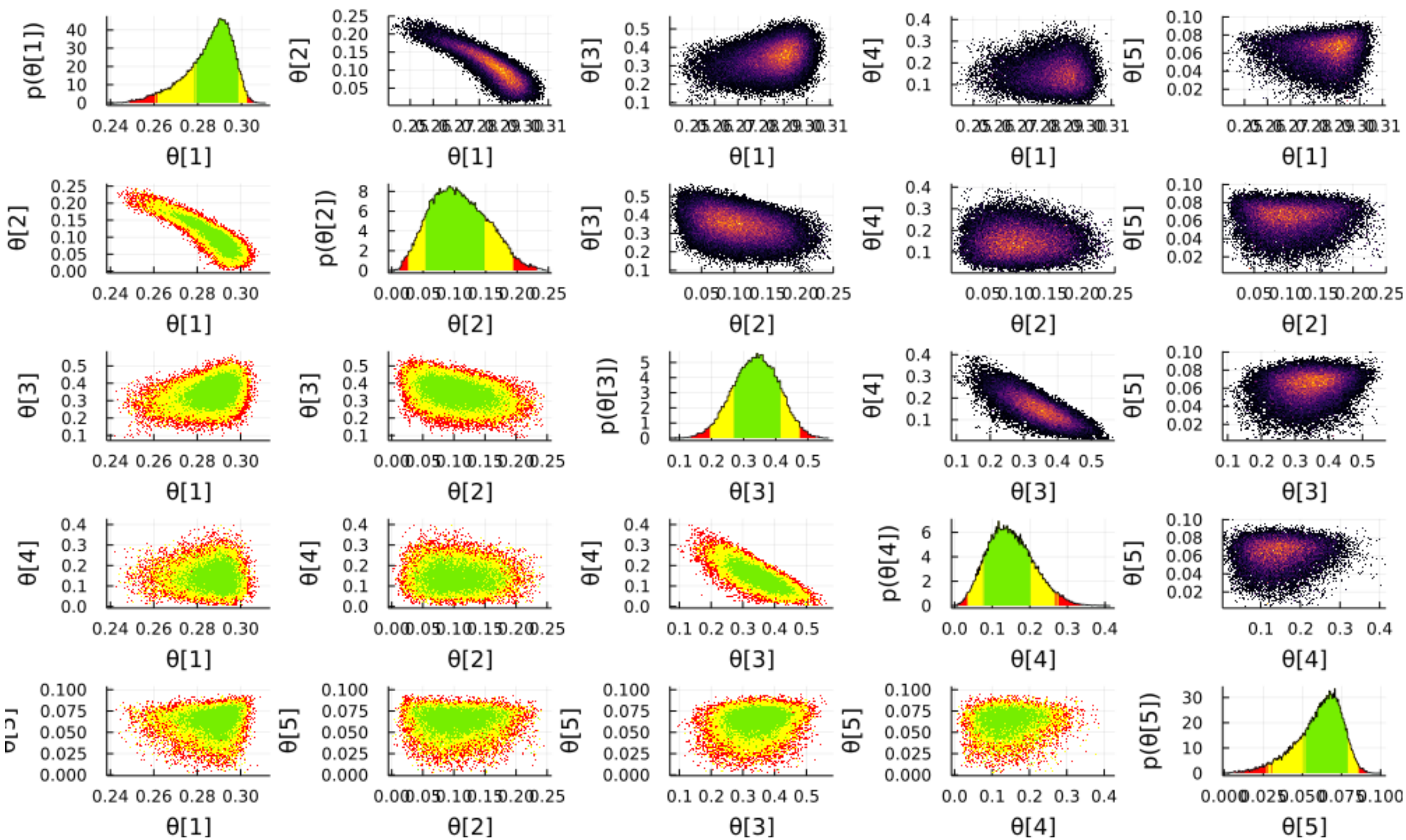
Some results …

Fitting code: **F. Capel** implemented fitting model, BAT.jl **O. Schulz** et al.)

MCMC Output

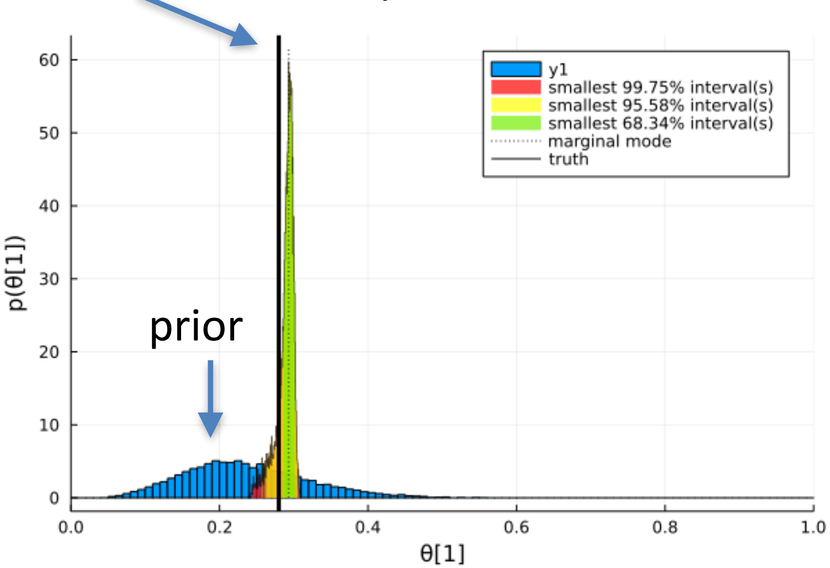
Output is $\{\boldsymbol{\Delta}, K_u, K_d, \lambda_{g1}, \lambda_{g2}, K_g, \lambda_q, \beta\}$ distributed $\propto P(\boldsymbol{\Delta}, K_u, K_d, \lambda_{g1}, \lambda_{g2}, K_g, \lambda_q, \beta \,|\, D)$.

BAT.jl outputs all 1,2D marginalized distributions. A small subset of possible plots.

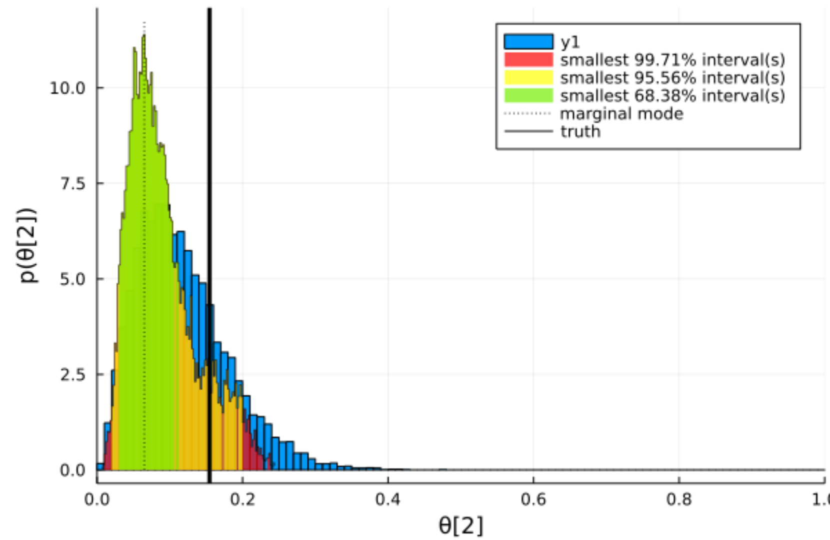


Momenta

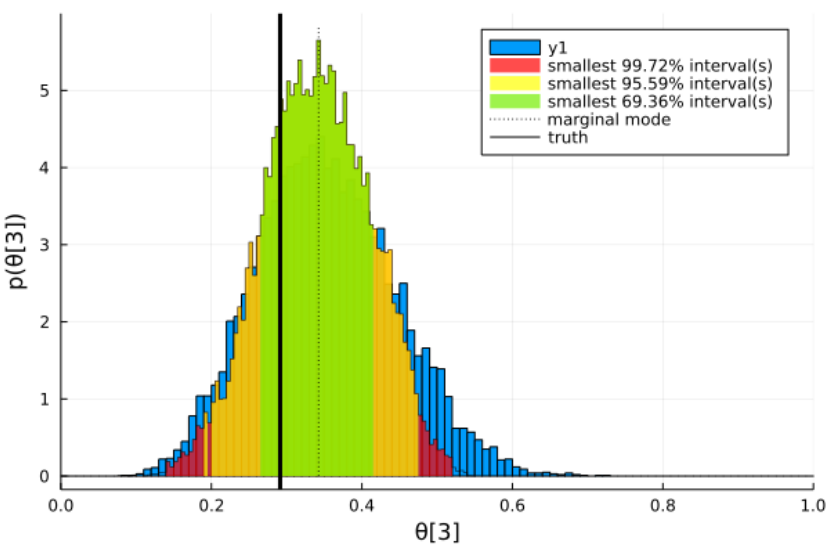
Up-valence



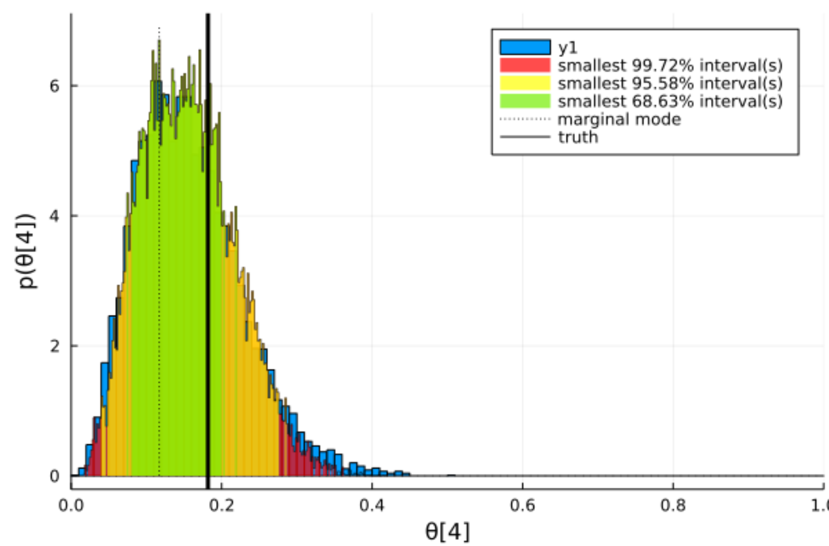
Down-valence



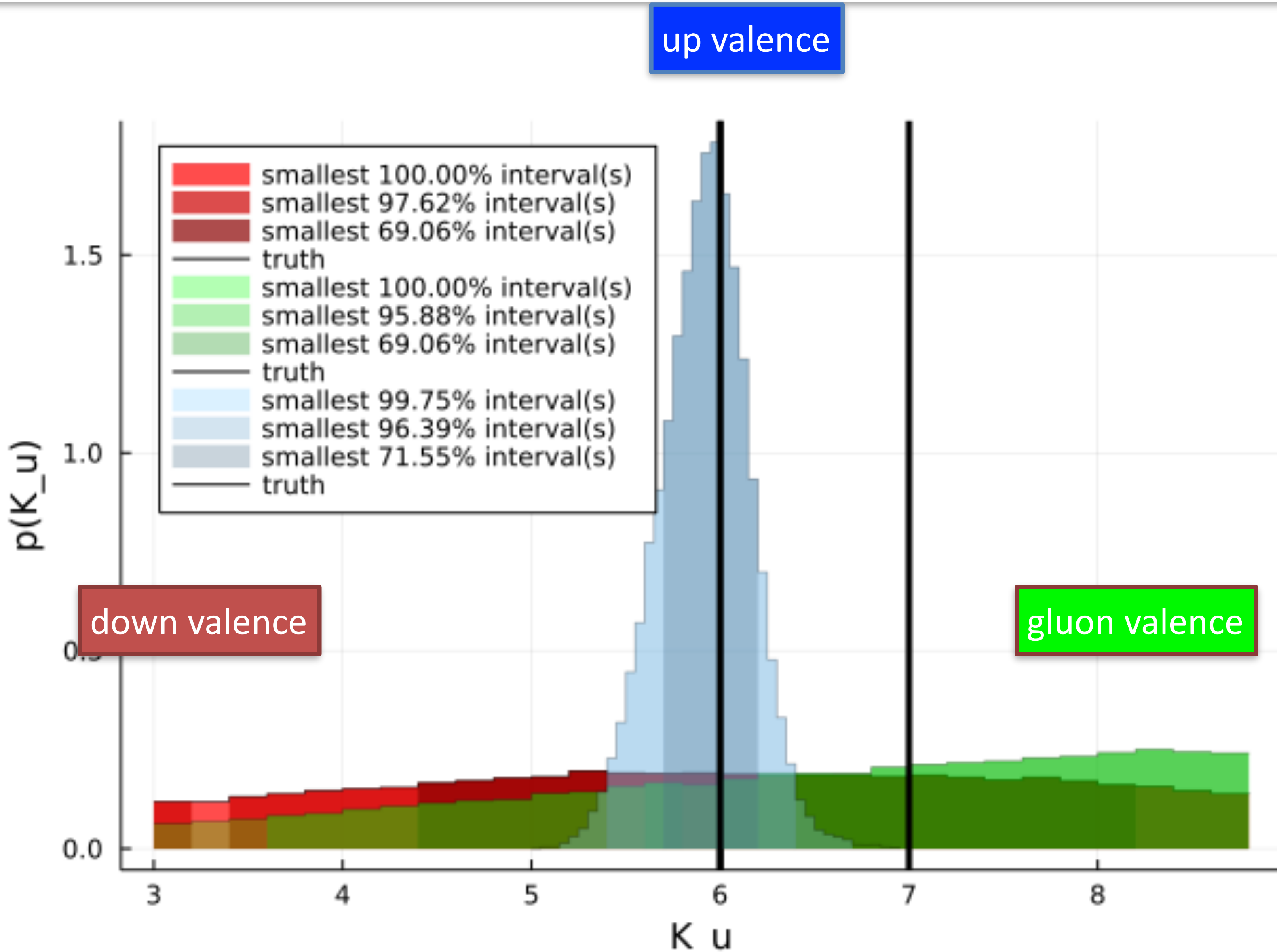
Glue-valence



Glue-sea



Shape Parameters



Shape (and momentum) of up-valence well constrained. Others weakly constrained.

Summary

- The kinematic range $x \to 1$ has not been fully exploited in the H1 data (as far as I know). Contains valuable information

- Kinematic reconstruction studies indicate that we could get better performance using new techniques. Could be applied to existing HERA data for updated cross section measurements.

- We could/should produce results that allow a forward analysis of the data. I.e., event numbers in bins and the information needed to get predictions from Born level cross sections. My dream a consistent analysis in this style over the full kinematic plane from the HERA data.

- Techniques for analyzing this data being developed.

Lots of room for important analyses!