Bayesian PDF fits to ZEUS high-x data

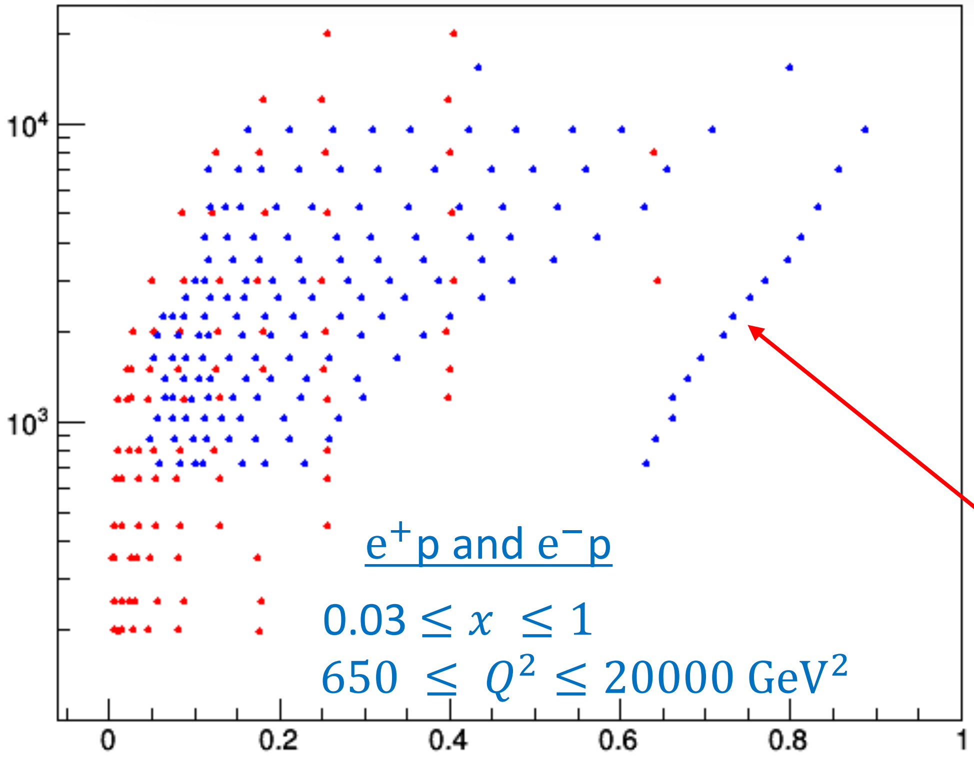
Michiel Botje for the Bayesian analysis team

Allen Caldwell	MPI Munich, Germany	Project leader
Oliver Schulz	MPI Munich, Germany	BAT.jl framework
Francesca Capel	MPI Munich, Germany	Analysis code
Ritu Aggarwal	Pune University, India	Forward model
Michiel Botje	Nikhef Amsterdam, The Netherlands	QCDNUM/SPLINT

arXiv:2209.06571 [hep-ph]

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

ZEUS collab., Phys.Rev. D **89**, 072007, hep-ex:1312.4438 (2014)



- Unique e^+p and e^-p data set extending to $x = 1$
- Data not used in any global pdf fit
- No higher twists which plague analysis of other high-x data
- Event numbers and differential cross-sections are given in 153 x-Q^2 bins for each data set
- At high x only integrated cross-sections are given

Use the full data set in a Bayesian pdf fit of bin-counts that can handle the Poisson statistics of low event numbers at very large x

Bayesian forward model approach

- Parameterise pdfs at some Q_0^2 and evolve at NNLO
- Compute at NNLO F_2, F_L and xF_3 and Born neutral current $\mathrm{e}^{\pm}\mathrm{p}$ cross-sections
- Integrate x-sections over bins in x-Q^2 and compute event numbers
- Apply radiative and detector effects to get predictions for the observed events n
- Compute Poisson likelihood $P(n|\theta)$ for the set of fit parameters θ
- Define prior probabilities $P(\theta)$ for the fit parameters
- Get posterior from Bayes theorem

$$P(\theta|n, \mathrm{model}) \propto P(n|\theta)P(\theta)$$

Posterior is also conditional on the choice of parameterisation

- **Bayesian approach is very attractive because …**
  - No Gaussian assumptions
  - Constraints are easily implemented (posterior cannot extend beyond prior range)
  - Badly constrained parameters do not spoil the fit
  - Uncertainties in badly constrained parameters (as encoded in the prior) are automatically propagated to the posterior of other parameters
  - Can easily judge information content of the data by comparing posterior to prior
  - Marginalisation of the posterior gives easy access to single-parameter distributions and correlations
- **But also …**
  - Priors should be chosen with care to not introduce bias in the posterior
  - Need lots of CPU to map-out the posterior in multi-dimensional parameter space

For this analysis we use the Bayesian Analysis Toolkit and QCDNUM

Bayesian Analysis Toolkit

O. Schulz *et al.*, SN computer Science **2**, 210 (2021)

- High-performance toolkit for Bayesian inference
  - Tools for definition of likelihoods, priors and posteriors
  - Provides MCMC sampling techniques to explore the posterior
  - Location and interval estimation, marginalisation, visualisation, *etc.*
  - And much more …
- Written in Julia (with Julia interface to QCDNUM)

BAT.jl	https://bat.github.io
Analysis	https://github.com/cescalara/PartonDensity.jl

QCDNUM

QCDNUM and SPLINT

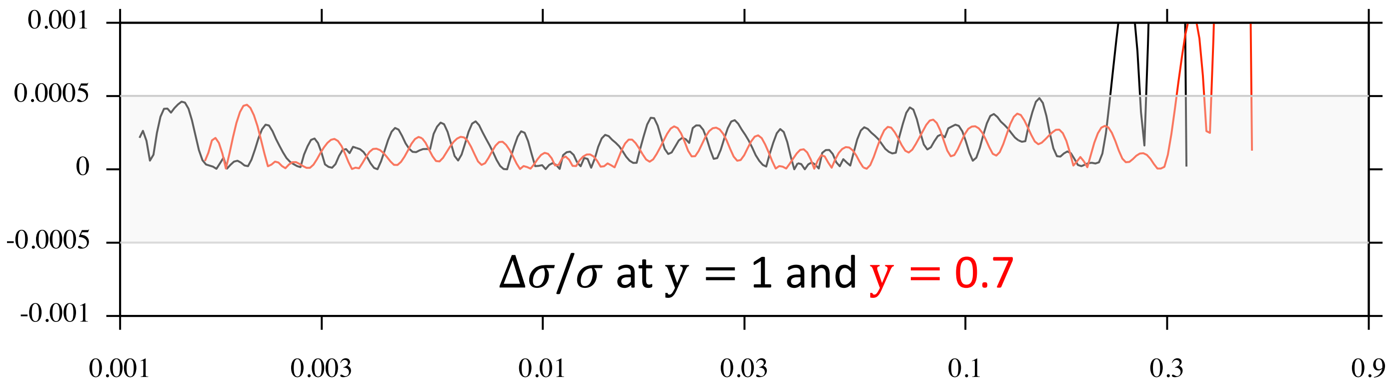
www.nikhef.nl/~h24/qcdnum

- Use QCDNUM for NNLO evolution and structure functions
- SPLINT add-on provides cubic spline interpolation and integration
  - Spline interpolation much faster than computing stfs and xsecs from scratch
  - Needs some tuning of spline-grid to balance speed vs accuracy
  - SPLINT provides fast integration over bins taking kinematic limit into account
  - SPLINT integration is factor 300 faster than 2-dim Gauss integration

Timing (MacBook Pro 2018)

	n_x	n_q	t [ms]
Evolution	100	50	3.6
6 Stf splines	22	7	2.9
Xsec spline	100	25	2.2
Integration	429 bins		0.8

$\Delta\sigma/\sigma \lesssim 5\times10^{-4}$ along the kinematic cut

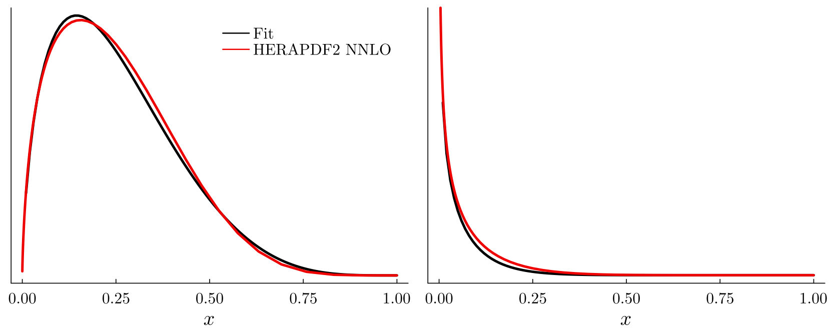


Parameterisation

- Parameterise pdf as beta distribution $xf(x) = A\,x^{\lambda}(1-x)^{K}$
- Integrable for $\lambda > -1$ $\quad \Delta \equiv \int_0^1 xf(x)\mathrm{d}x = A\,\frac{\Gamma(\lambda+1)\Gamma(K+1)}{\Gamma(\lambda+K+2)}$
- Replacement $\lambda \to \lambda - 1$ for valence pdf gives number density
- Integrable for $\lambda > 0$ giving $\quad A_i = N_i^{\mathrm{v}}\,\frac{\Gamma(\lambda_i+K_i+1)}{\Gamma(\lambda_i)\Gamma(K_i+1)} \quad \text{with} \begin{cases} N_d^{\mathrm{v}} = 1 \\ N_u^{\mathrm{v}} = 2 \end{cases}$
- Can fix valence λ for given N and Δ through $\quad \Delta_i = N_i^{\mathrm{v}}\,\frac{\lambda_i}{\lambda_i+K_i+1}$
- Easy to control low-x behaviour
  - Integrable and decreasing towards low x for $\lambda > 0$ (valence)
  - Integrable and increasing towards low x for $-1 < \lambda < 0$ (sea, gluon)

Parameterise 5 flavours at input scale $Q_0^2 = 100$ GeV^2

$$
\begin{aligned}
xd^{\mathrm{v}}(x) &= A_d\, x^{\lambda_d}(1-x)^{K_d} \\
xu^{\mathrm{v}}(x) &= A_u\, x^{\lambda_u}(1-x)^{K_u} \\
x\bar{q}(x) &= A_i\, x^{\lambda_{\bar{q}}}(1-x)^{K_{\bar{q}}} \quad i = \{\bar{d}, \bar{u}, \bar{s}, \bar{c}, \bar{b}\} \\
xg(x) &= A_g^{\mathrm{v}}\, x^{\lambda_g^{\mathrm{v}}}(1-x)^{K_g} + A_g^{\mathrm{s}}\, x^{\lambda_g^{\mathrm{s}}}(1-x)^{K_{\bar{q}}}
\end{aligned}
$$



Parameterise HERAPDF at $Q^2 = 100$ GeV^2 to check flexibility as shown here for u^{V} and $\bar{u}$

- Gluon with a valence and sea component
- Fix λ_d and λ_u via the quark counting rules
- All $x\bar{q}$ have the same shape but different normalisations
- Gluon sea component has same high-x power as the anti-quarks
- Do not fit normalisation constants but momentum fractions Δ which are more meaningful
- Momentum sum constraint $\Delta_u + \Delta_d + 2\sum_{\bar{q}} \Delta_{\bar{q}} + \Delta_g^{\mathrm{v}} + \Delta_g^{\mathrm{s}} = 1$

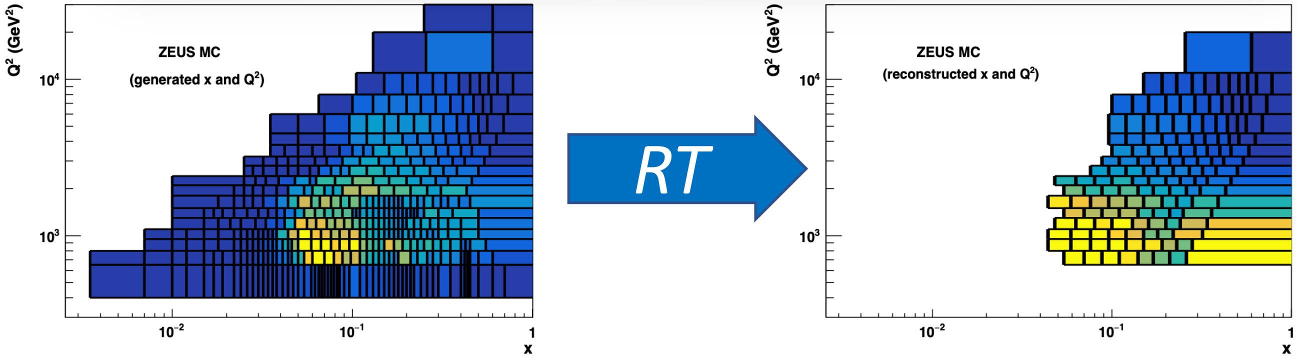
7 shape parameters + 9 momentum fractions with sum rule constraint contribute 15 degrees of freedom to the fit

Event predictions

- Evolve at NNLO with QCDNUM and compute neutral current $e^{\pm}p$ cross-sections
- Integrate over 429 bins and compute vector of event predictions $\vec{\nu}$
- Multiply by matrix RT to correct for radiative and detector effects: $\vec{u} = RT\vec{\nu}$
- This gives the observed event predictions $\vec{u}$ in 153 ZEUS bins for each data set
- For parameters θ the predictions $\vec{u}(\theta)$ give for the likelihood of observing the events $\vec{n}$:

$$\boxed{P(\boldsymbol{n}|\boldsymbol{\theta}) = \prod_{\text{bins}} \frac{u^n e^{-u}}{n!}}$$

ZEUS, Phys. Rev. D **101**, 11209, hep-ex:2003.08732 (2020)



- Added to RT is the weighted sum of 10 systematic matrices
- The 10 weights δ_i are left free parameters in the fit giving a total of 16+10=26 parameters

Priors

M. Betancourt, Proc. AIP Conf., **1443**, 157 (2012)

- Take 9-dimensional Dirichlet distribution for momentum prior
  - $\mathrm{dir}(\vec{\alpha})$ with 9 shape parameters α is multivariate extension of beta distribution
  - Lives on an 8-dimensional manifold in the space $\Delta_i \in [0,1]$ with $\sum \Delta_i = 1$

- Set Dirichlet shape parameters $\vec{\alpha}$ according to asymptotic expectations
  - $\Delta(\mathrm{gluon}) \approx \Delta(\mathrm{quarks}), \quad \Delta(u^{\mathrm{V}}) \approx 2\Delta(d^{\mathrm{V}}), \quad \Delta(s,c,b) \sim \mathrm{small}$

- Pdf shape priors set to truncated Normal or Uniform such that the pdfs are integrable and have the required low-x behaviour

- Priors of systematic δ parameters set to truncated Normal with zero mean and unit width

	Prior	Range
$\boldsymbol{\alpha}$	Dir(20, 10, 20, 20, 5, 2.5, 1.5, 1.5, 0.5)	$[0, 1]$
K_u	Normal(3.5, 0.5)	$[2, 5]$
K_d	Normal(3.5, 0.5)	$[2, 5]$
λ_g^{v}	Uniform	$[0, 1]$
λ_g^{s}	Uniform	$[-1, -0.1]$
K_g	Normal(4, 1.5)	$[2, 7]$
$\lambda_{\bar{q}}$	Uniform	$[-1, -0.1]$
$K_{\bar{q}}$	Normal(4, 1.5)	$[3, 10]$
$\boldsymbol{\delta}$	Normal(0, 1)	$[-5, 5]$

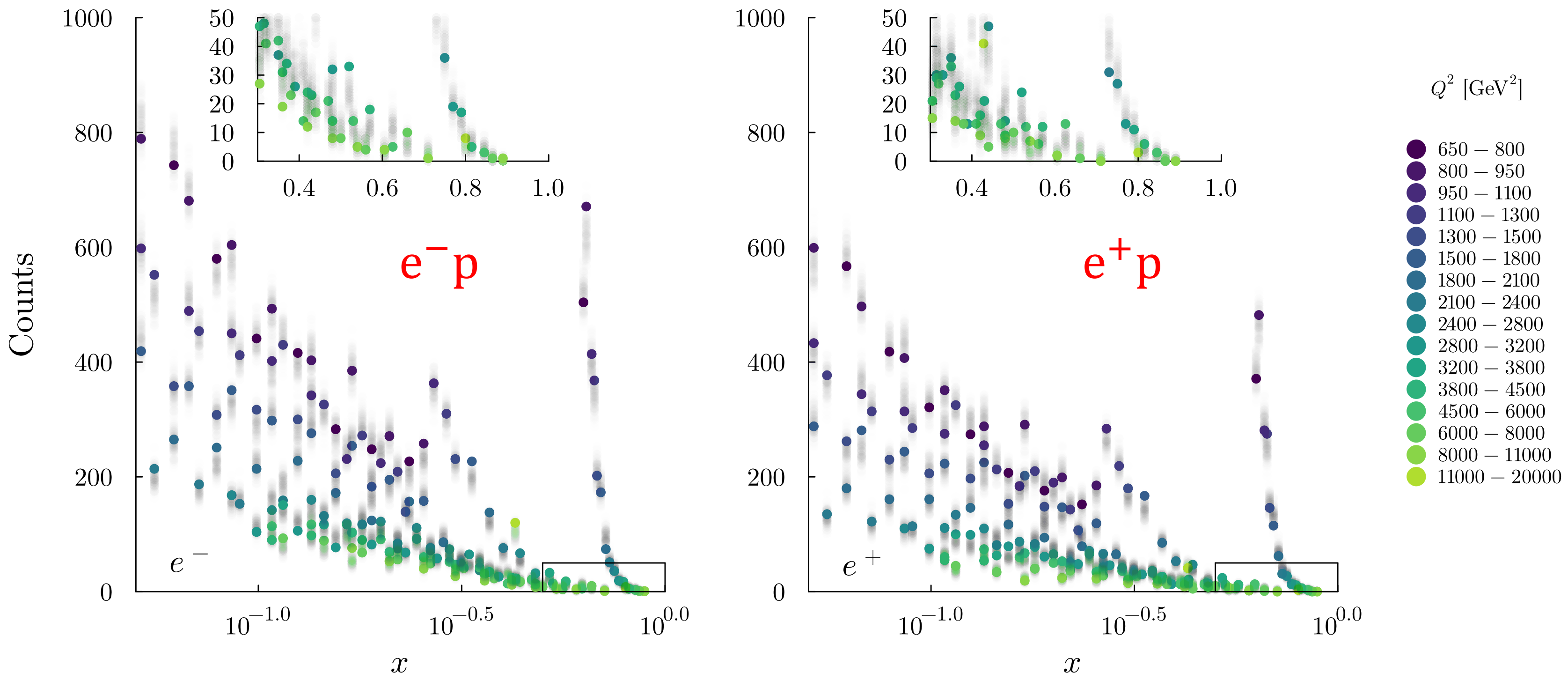
Results

- The real result of the analysis is the 26-dim posterior in parameter space
- Parameter values and errors are defined in two ways as
  - Position of the mode of the posterior in parameter space
  - Mode of the marginal parameter distribution with error corresponding to the smallest credible interval around the mode that contains 68% probability
- Here are the parameters that are reasonably well constrained by the data

	Global mode	Marginal mode		Global mode	Marginal mode
Δ_u	0.219	$0.219^{+0.008}_{-0.010}$	K_u	3.76	$3.78^{+0.14}_{-0.17}$
Δ_d	0.099	$0.088^{+0.028}_{-0.020}$	K_d	3.66	$3.69^{+0.33}_{-0.60}$
$\lambda_{\bar{q}}$	-0.55	$-0.52^{+0.06}_{-0.11}$	$K_{\bar{q}}$	6.01	$6.38^{+1.13}_{-1.40}$
K_g	4.92	$5.22^{+0.91}_{-1.57}$			
$2\Delta_{\bar{u}}$	0.126	$0.104^{+0.022}_{-0.027}$	$2\Delta_{\bar{d}}$	0.031	$0.024^{+0.020}_{-0.017}$
Δ_g^{v}	0.265	$0.239^{+0.043}_{-0.037}$	Δ_g^{s}	0.245	$0.241^{+0.047}_{-0.036}$

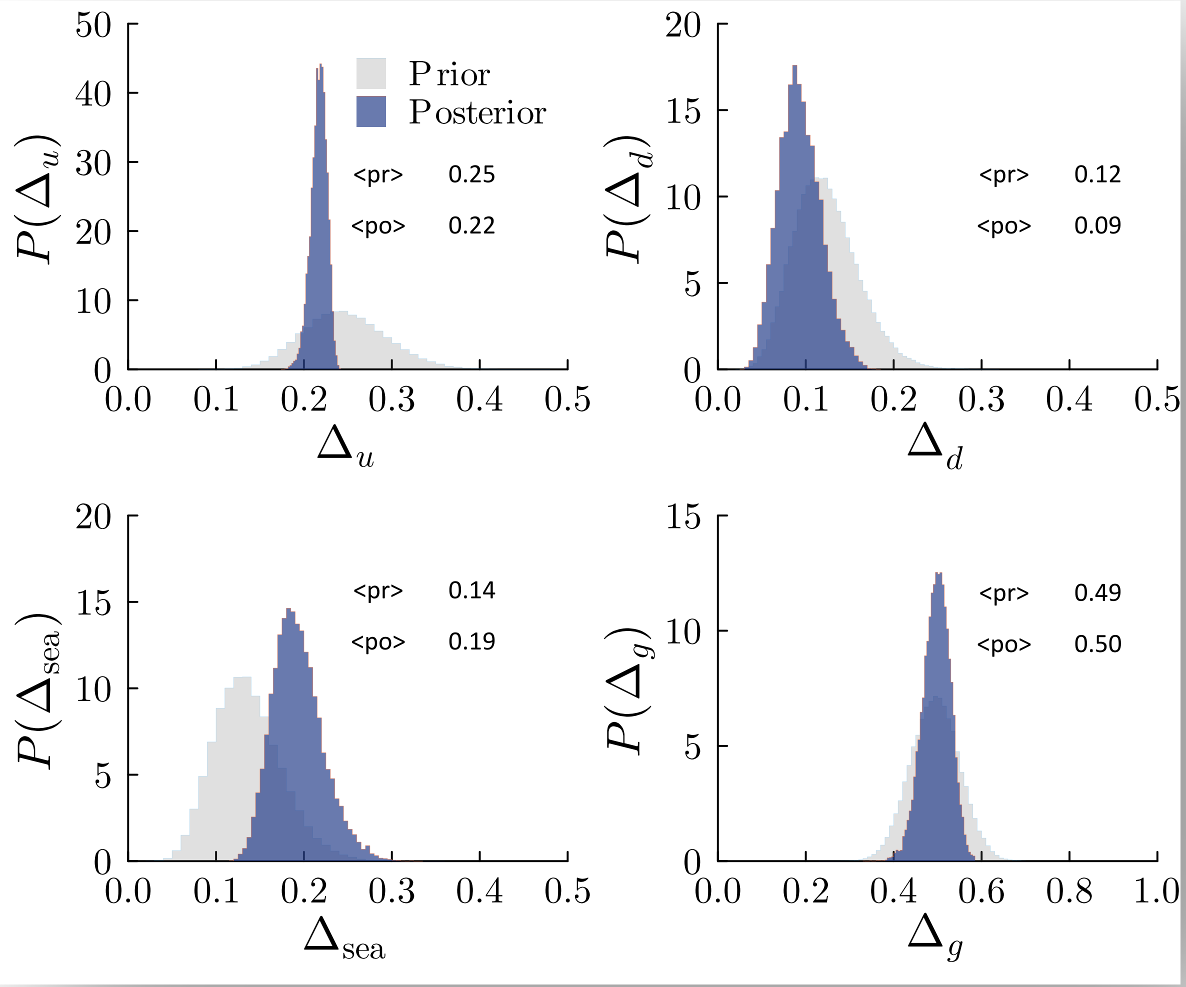
2×10^5 samples ~24h on MacBook Pro

Fitted event counts versus data



- Compute event predictions from posterior mode parameters (bands) and compare to observed event counts (dots), plotted at the bin centers
- Pierson chi-squared gives $\chi^2/\mathrm{pt} = 321/306$ with a p-value of 0.27

Our parameterisation yields an excellent description of the data



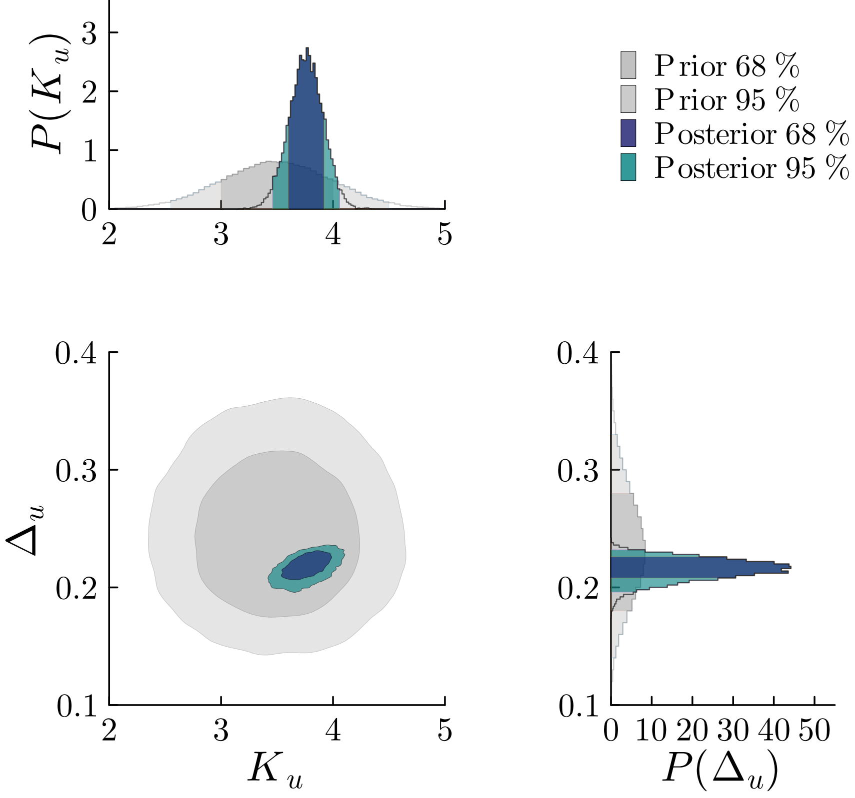
Momentum fraction priors and posteriors

- The data very much constrain the momentum carried by the up-valence
- Weaker constraints on the down-valence, sea and gluon
- From global mode we find

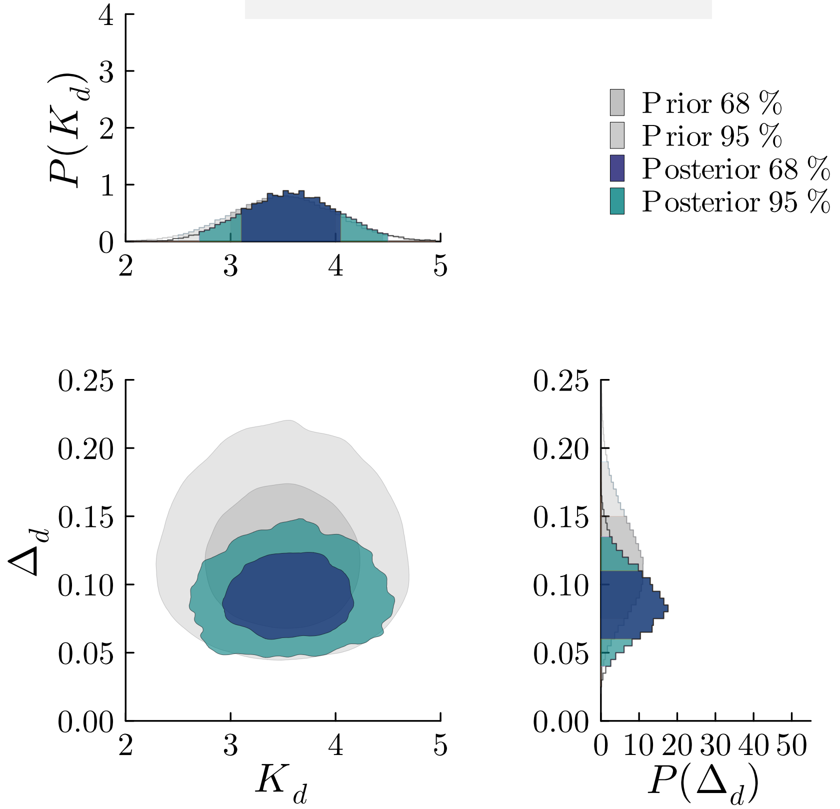
u_v	d_v	sea	gluon
0.22	0.10	0.17	0.51

Momentum Δ versus 1− x power K

up-valence

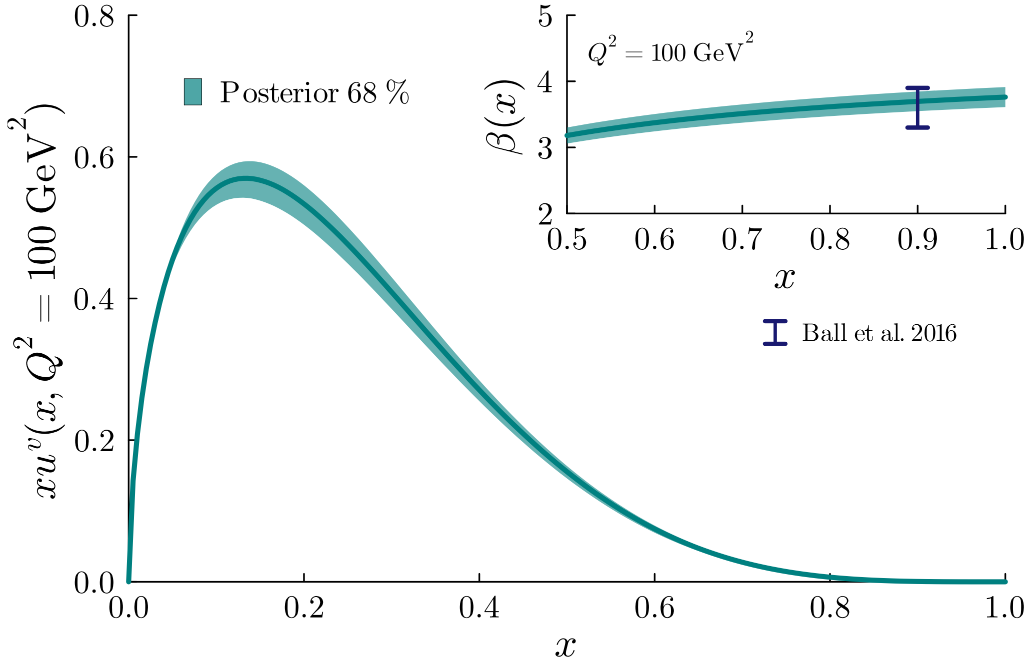


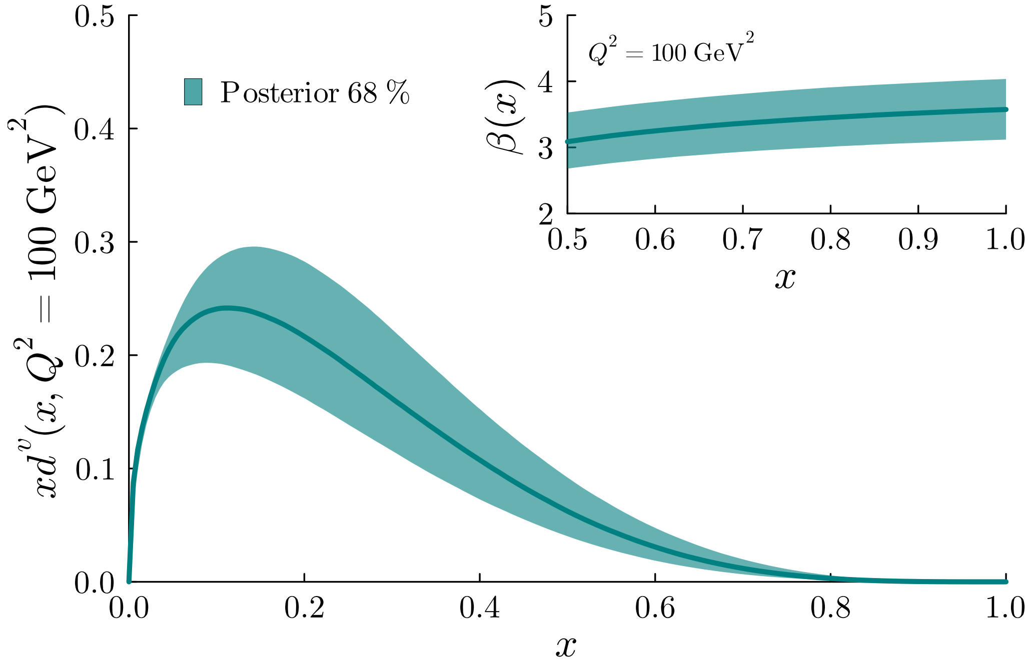
down-valence



Very strong constraint on the up-valence parameters

Up and down valence distributions



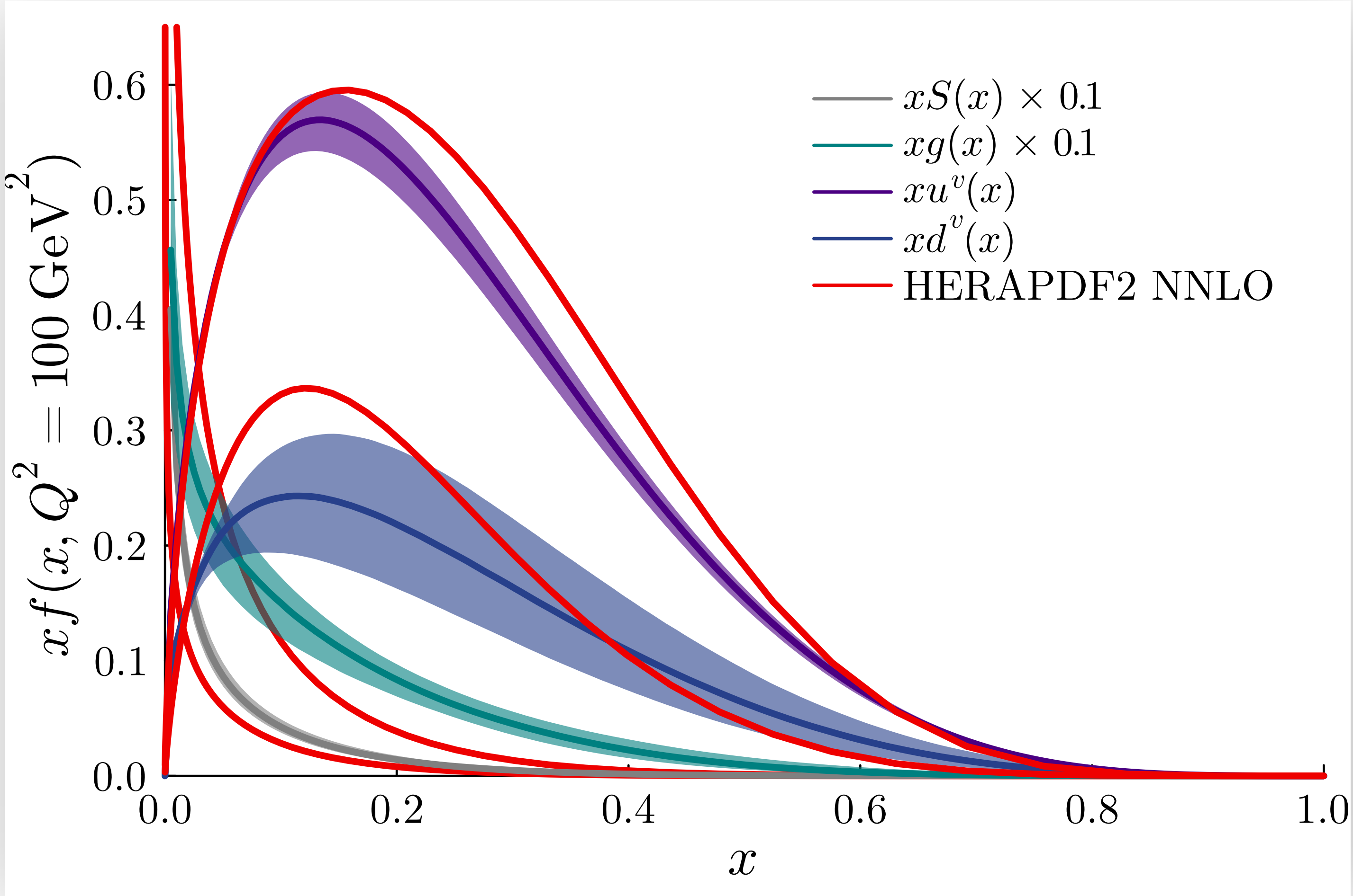


- The insets show the effective 1-x power $\beta(x) = \mathrm{d}\ln xf/\mathrm{d}\ln(1-x)$
- The β slope of u_v agrees well with a recent summary from Ball *et al.*

R. Ball, E. Nocera and J. Rojo, Eur. Phys. J. C **76** (2016) , 383 , hep-ph:1604.00024

Compare to HERAPDF

H. Abramowicz *et al.*, Eur. Phys. J. C **75**, (2015) 12, 580



- Analytic parameterisations strongly couple regions of small and large x
- HERAPDF parameterisation is similar to ours but fitted at much lower x (not using the ZEUS high-x data)
- This may explain at least part of the observed differences

Summary

- The Bayesian analysis of ZEUS high-x data shows that these data carry a lot of information on the up-valence distribution
- Given our parameterisation we obtain accurate results on the momentum Δ carried by the up-valence quark and its $(1-x)$ power K (marginal mode and 68% credible interval)

$$\Delta = 0.22^{+0.01}_{-0.01} \qquad K = 3.8^{+0.1}_{-0.2}$$

What is next ...

- Paper with detailed description of our analysis is in preparation
- Extend the analysis to investigate parameterisations with Bayesian model-selection techniques
- Exploit in the extended analysis the many opportunities for parallel computing

Bayesian pdf fitting is a viable, challenging, and highly interesting undertaking!