PartonDensity.jl

A Novel Bayesian PDF Fitting Code

F. Capel, A. Caldwell, O. Schulz, A. Verbytskyi (Max Planck Institute for Physics)
M. Botje (Nikhef)
R. Aggarwal (USAR, Guru Gobind Singh Indraprastha University)

Presented by Allen Caldwell
Max Planck Institute for Physics

Main Features

- fully Bayesian approach, with all its advantages, including samples from the full posterior density distribution

- based on BAT.jl, a modern Bayesian Analysis Toolkit written in the Julia language

- general likelihood, not limited to χ^2 minimization, allowing for forward modeling approach to analysis

- standard output in useful format

PartonDensity.jl: a novel parton density determination code

Francesca Capel, Ritu Aggarwal, Michiel Botje, Allen Caldwell, Oliver Schulz et al. (Jan 31, 2024)
e-Print: 2401.17729 [hep-ph]

used in

Constraints on the Up-Quark Valence Distribution in the Proton
Ritu Aggarwal Pune U. Michiel Botje Nikhef, Amsterdam Allen Caldwell Munich, Max Planck Inst. Francesca Capel Munich, Max Planck Inst. Oliver Schulz Munich, Max Planck Inst.
Phys.Rev.Lett. 130 (2023) 14, 141901 • e-Print: 2209.06571 [hep-ph]

Reminder: Bayesian Formulation

$\mathscr{L}_D(\theta)$
likelihood

prior probability for the parameters θ

$$p(\theta|D,M) = \frac{p(D|\theta,M)\; p(\theta|M)}{p(D|M)}$$

posterior probability for the parameters θ

Normalization - can be ignored if not doing model comparison

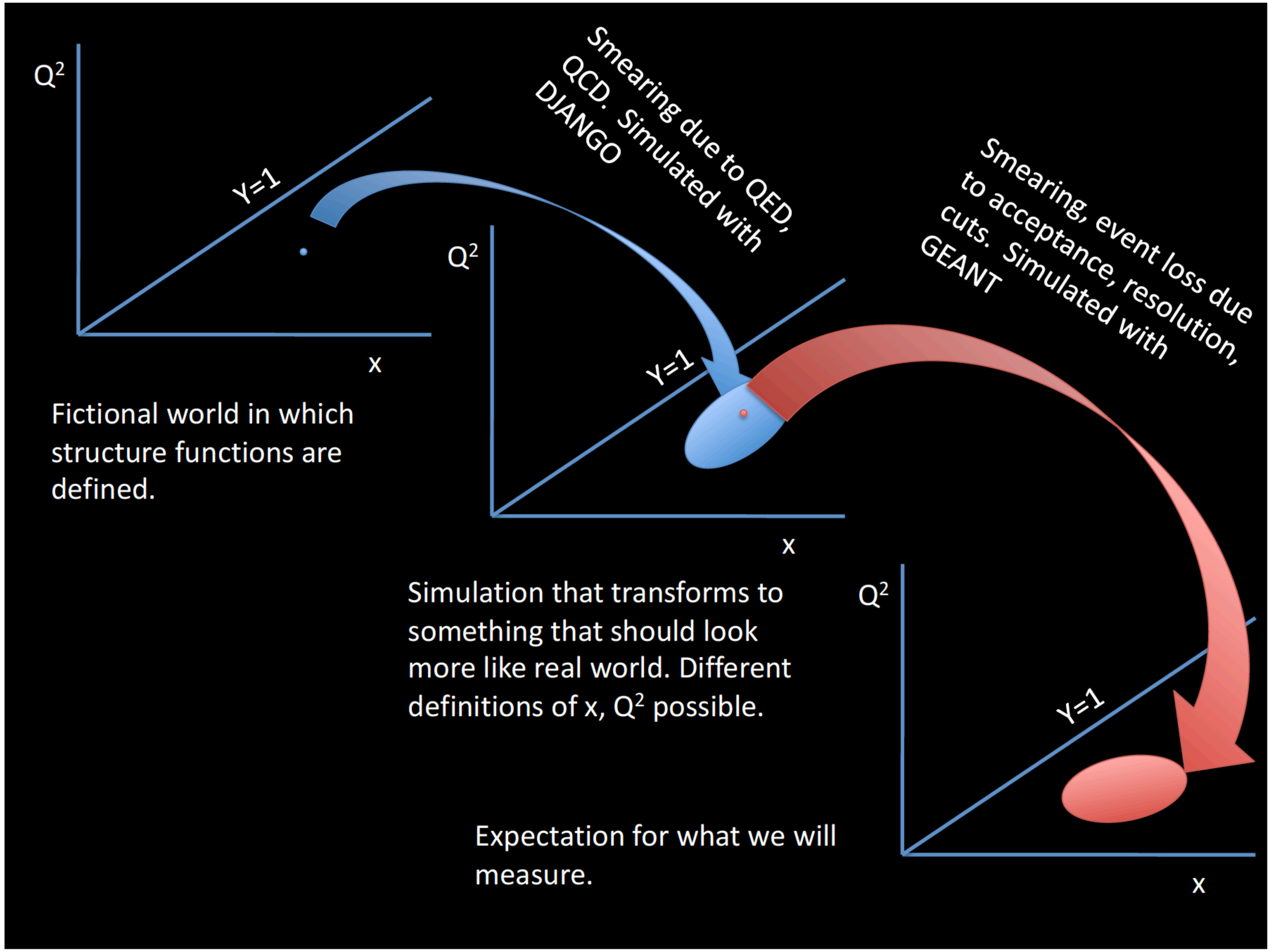
Simple expressions, but θ can be quite high dimensional making the solution difficult.

PartonDensity.jl solves Bayesian challenge using BAT.jl

For those interested in how numerical challenges are met - visit: https://github.com/bat/BAT.jl

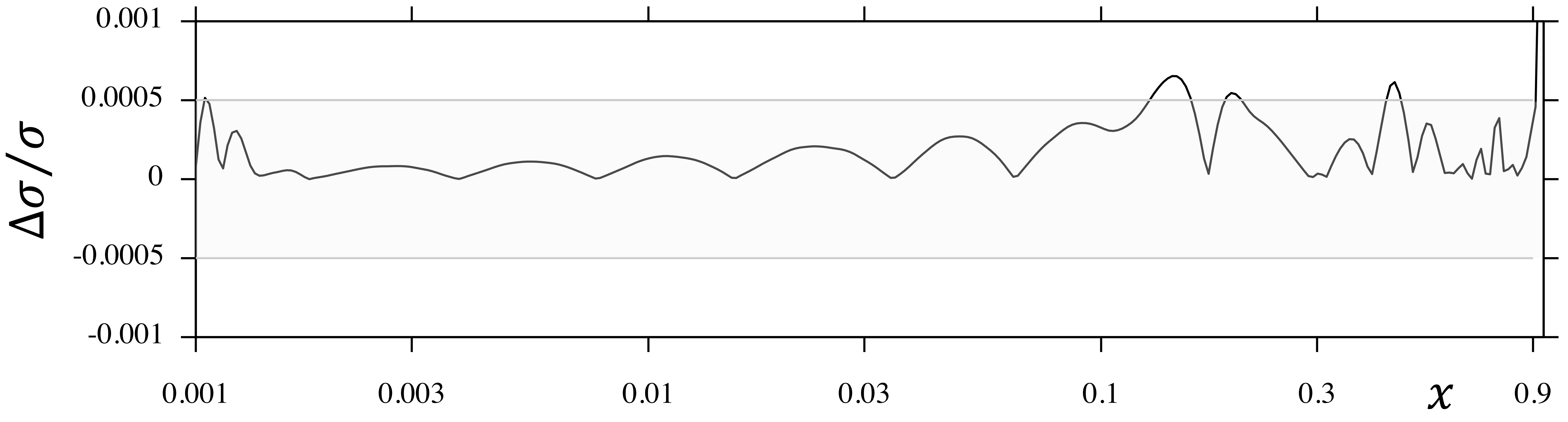
Forward Modeling

1. Define pdfs
2. Perform evolution: done with QCDNUM
3. Calculate differential cross section
   - for unfolded data, use this for performing analysis
4. Integrate over bins in which events counted for forward modeling. Use a fine (detector specified) binning
5. Apply detector specific response function to get event expectations in measurement bins
6. Calculate probability of observed event numbers given expectation



QCDNUM & SPLINT

SPLINT package constructs splines on selected set of QCDNUM (x, Q^2) evolution grid-points. Fast and accurate.



Compared to 2D Gauss integration routine for differently shaped bins. Relative difference $< 10^{-9}$ with SPLINT running about a factor of~300 faster than Gauss integration

2018 MacBook Pro with an Intel processor

Subtask	Grid	CPU time [ms]
Evolution	100×50	3.6
Structure function splines ($6\times$)	22×7	2.9
Cross section spline	100×25	2.2
Integration over 429 bins		0.8

Example

Use combinations of Beta functions, $xf(x) = A\, x^{\lambda}(1-x)^{K}$ to model PDFs

Parameters are λ, K and the integrated momenta, $\Delta = \int dx x f(x)$

For our example, we use parameters similar to those found for the ZEUS high-x data to generate simulated data.

Data simulated with luminosity as in ZEUS experiment as well as 100x ZEUS luminosity.

TABLE II. Parameter values used in the data simulation.

$\Delta \times 10^3$								
u^V	d^V	g^V	g^S	$2\bar{u}$	$2\bar{d}$	$2\bar{s}$	$2\bar{c}$	$2\bar{b}$
228	104	249	249	104	52	10	5	0.5

K_u	K_d	λ_g^{v}	λ_g^{s}	K_g	$\lambda_{\bar{q}}$	$K_{\bar{q}}$	$\boldsymbol{\beta}$
3.70	3.70	0.50	−0.50	5.0	−0.50	6.0	0

TABLE III. Priors used in the analysis of the pseudo-data sets. There are 9 parameters in the vector $\boldsymbol{\Delta}$ and 10 in $\boldsymbol{\beta}$. The normal distributions are truncated to the range indicated, and their mean and standard deviation are given in brackets.

	Prior	Range
$\boldsymbol{\Delta}$	Dir(20, 10, 20, 20, 5, 2.5, 1.5, 1.5, 0.5)	$[0, 1]$
K_u	Normal(3.5, 0.5)	$[1, 6.5]$
K_d	Normal(3.5, 0.5)	$[1, 6.5]$
λ_g^{v}	Uniform	$[0, 1]$
λ_g^{s}	Uniform	$[-1, -0.1]$
K_g	Normal(4, 1.5)	$[1, 8.5]$
$\lambda_{\bar{q}}$	Uniform	$[-1, -0.1]$
$K_{\bar{q}}$	Normal(4, 1.5)	$[1, 9.5]$
$\boldsymbol{\beta}$	Normal(0, 1)	$[-5, 5]$

Parameters used to model systematic uncertainties in modeling of detector response

Example Output

```
Sampling result
```

- Total number of samples: 180725
- Total weight of samples: 999996
- Effective sample size: between 1915 and 4848

```
Marginals
```

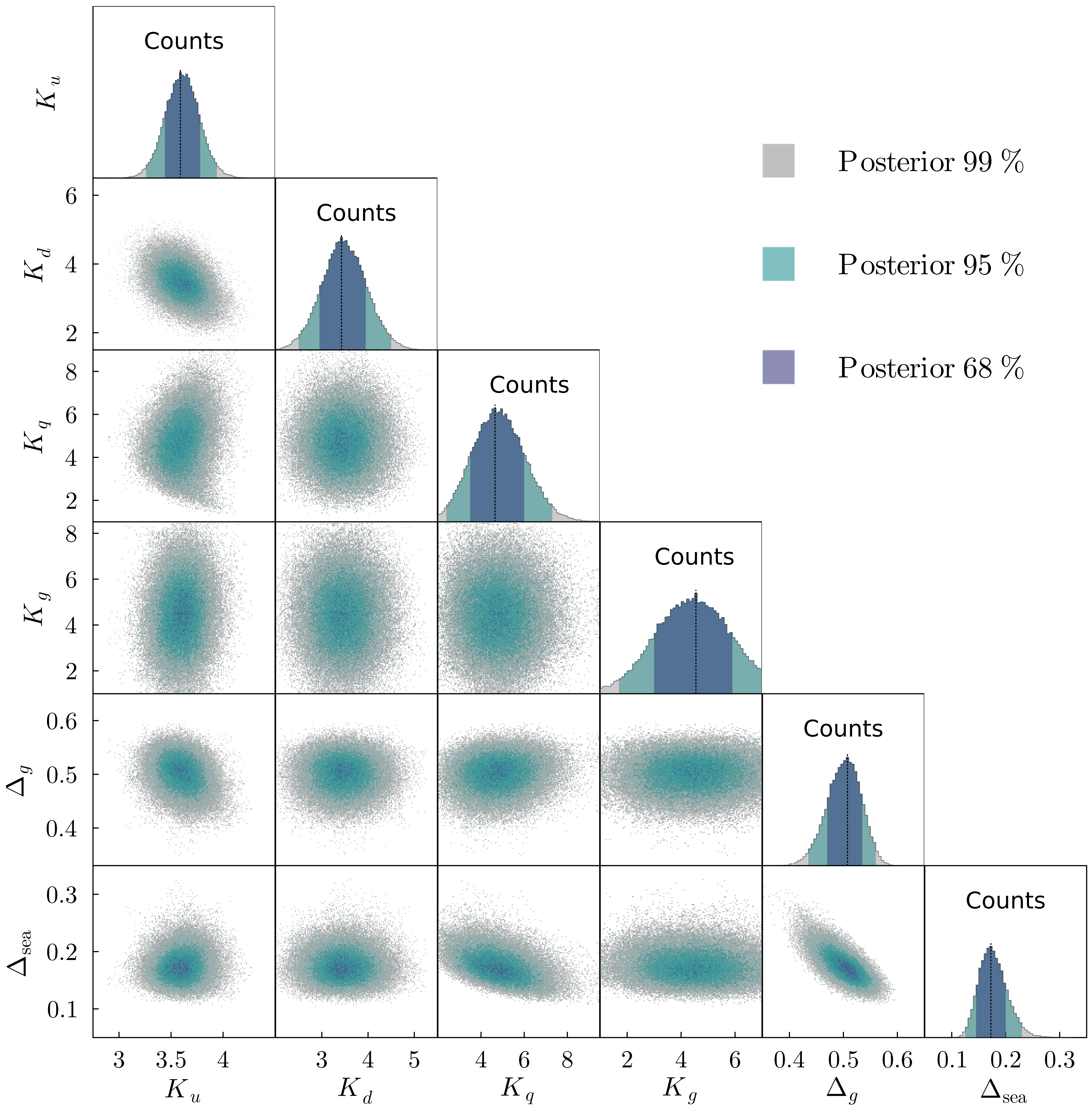
Parameter	Mean	Std. dev.	Gobal mode	Marg. mode	Cred. interval	Histogram
Δ_1	0.221272	0.0104517	0.214535	0.22175	0.212531..0.233155	0.172[… [0.254
Δ_2	0.135998	0.0296684	0.130628	0.1335	0.10425..0.164686	0.04[… [0.257
Δ_3	0.235809	0.0402725	0.212643	0.237	0.192969..0.274181	0.089[… [0.422
Δ_4	0.236329	0.0395947	0.280566	0.235	0.194414..0.274314	0.0995[… [0.394
Δ_5	0.0858189	0.0219218	0.110866	0.0895	0.0652377..0.108161	0.00805[… [0.203
Δ_6	0.0325197	0.0197449	0.014177	0.0225	0.00771978..0.0417355	0.000394[… [0.196
Δ_7	0.0196916	0.0159925	0.00706804	0.0065	0.00141507..0.0245333	2.05e-06[… [0.155
Δ_8	0.0259069	0.0180721	0.0295166	0.0115	0.0021522..0.0332662	3.37e-05[… [0.143
Δ_9	0.00665518	0.00952574	3.88412e-13	0.0005	3.88412e-13..0.00658574	3.88e-13[… [0.116
Ku	3.80003	0.199465	3.54776	3.775	3.59911..3.98004	3.02[… [4.78
Kd	3.50409	0.483632	3.77008	3.47	2.99302..3.96772	2[… [5
Kq	6.5525	1.27579	4.48005	6.375	5.1867..7.80312	3.01[… [10
λg_1	0.489073	0.28815	0.545423	0.0975	multiple	7.05e-06[… [1
λg_2	-0.506723	0.250449	-0.785071	-0.1975	-0.646814..-0.100004	-1[… [-0.1
λq	-0.491805	0.100082	-0.561121	-0.5075	-0.597922..-0.395554	-0.806[… [-0.122
Kg	4.84575	1.16458	4.83376	5.075	3.76593..6.27758	2[… [7
δ^1	0.501562	0.757327	-0.364463	0.525	-0.294443..1.2249	-2.59[… [3.89
δ^2	-0.236894	0.736022	-0.696325	-0.275	-1.01321..0.466908	-3.54[… [3.44
δ^0_1	0.0157033	1.00605	1.00191	-0.075	-0.939787..1.06978	-4.73[… [4.55
δ^0_2	0.00642715	0.995055	0.756425	0.125	-0.949105..1.05179	-4.36[… [4.5
δ^0_3	0.00928641	1.0063	0.989709	-0.075	-0.950165..1.07133	-4.09[… [4.21
δ^0_4	-3.89931e-5	1.00616	0.271427	0.025	-1.06209..0.959773	-4.33[… [4.05
δ^0_5	0.0206288	0.995915	2.67083	0.075	-1.02436..0.963545	-4.94[… [4.61
δ^0_6	0.0017444	0.997834	0.178345	-0.175	-0.943449..1.05423	-3.86[… [4.21
δ^0_7	-0.00418741	1.00237	0.35021	0.025	-1.05108..0.957236	-4.2[… [4.44
δ^0_8	0.00901734	0.993205	0.333895	-0.075	-1.03624..0.954911	-4.25[… [4.45

Example Output

Parameter Distributions

Full set of 1D and 2D distributions trivially available

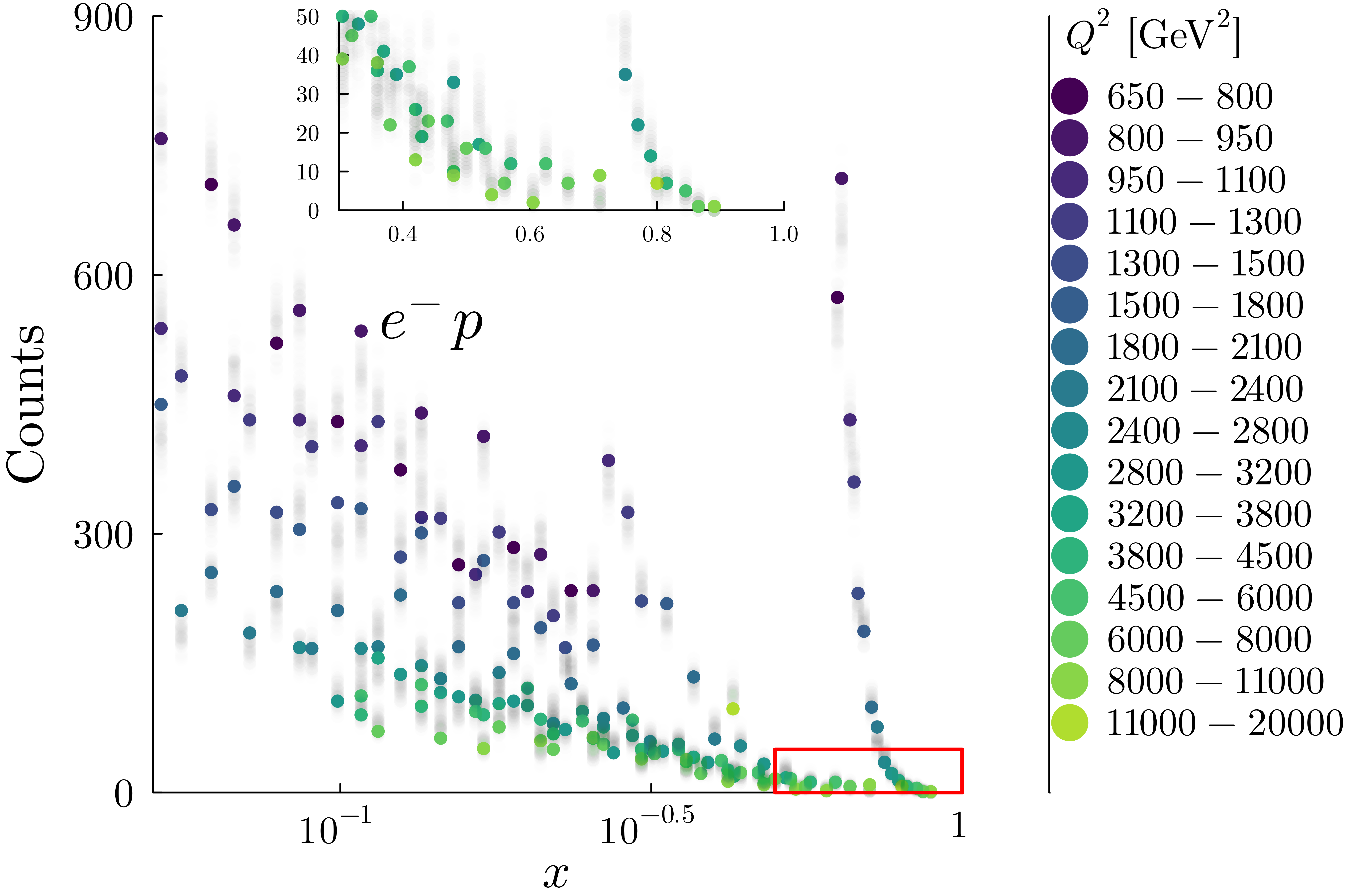
Functions to return correlation matrix, credible intervals, etc



Example Output

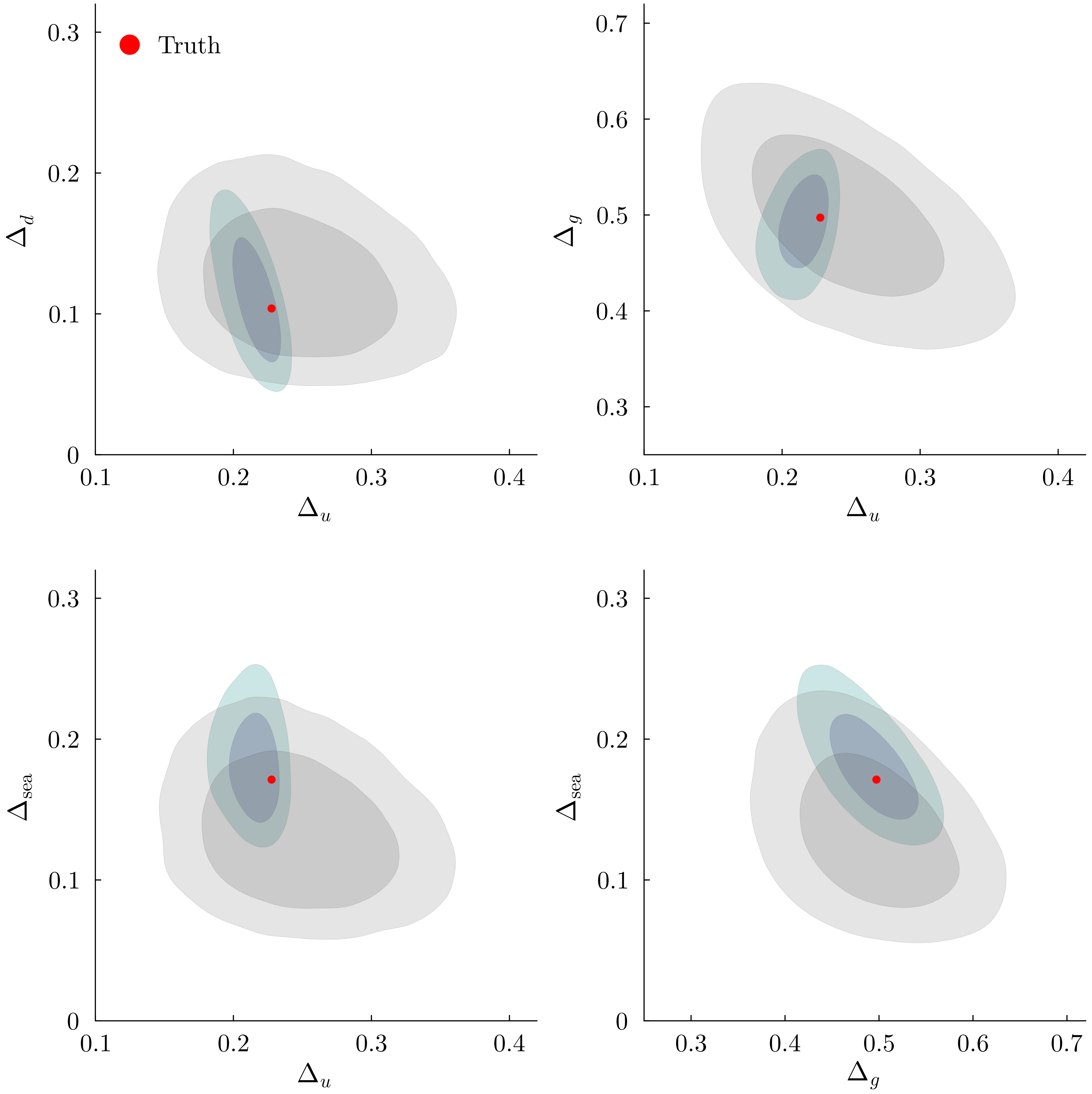
Posterior predictive check:
Use samples of the parameters from the posterior probability distribution to generate data and compare to fitted data.

This is a Bayesian goodness-of-fit test

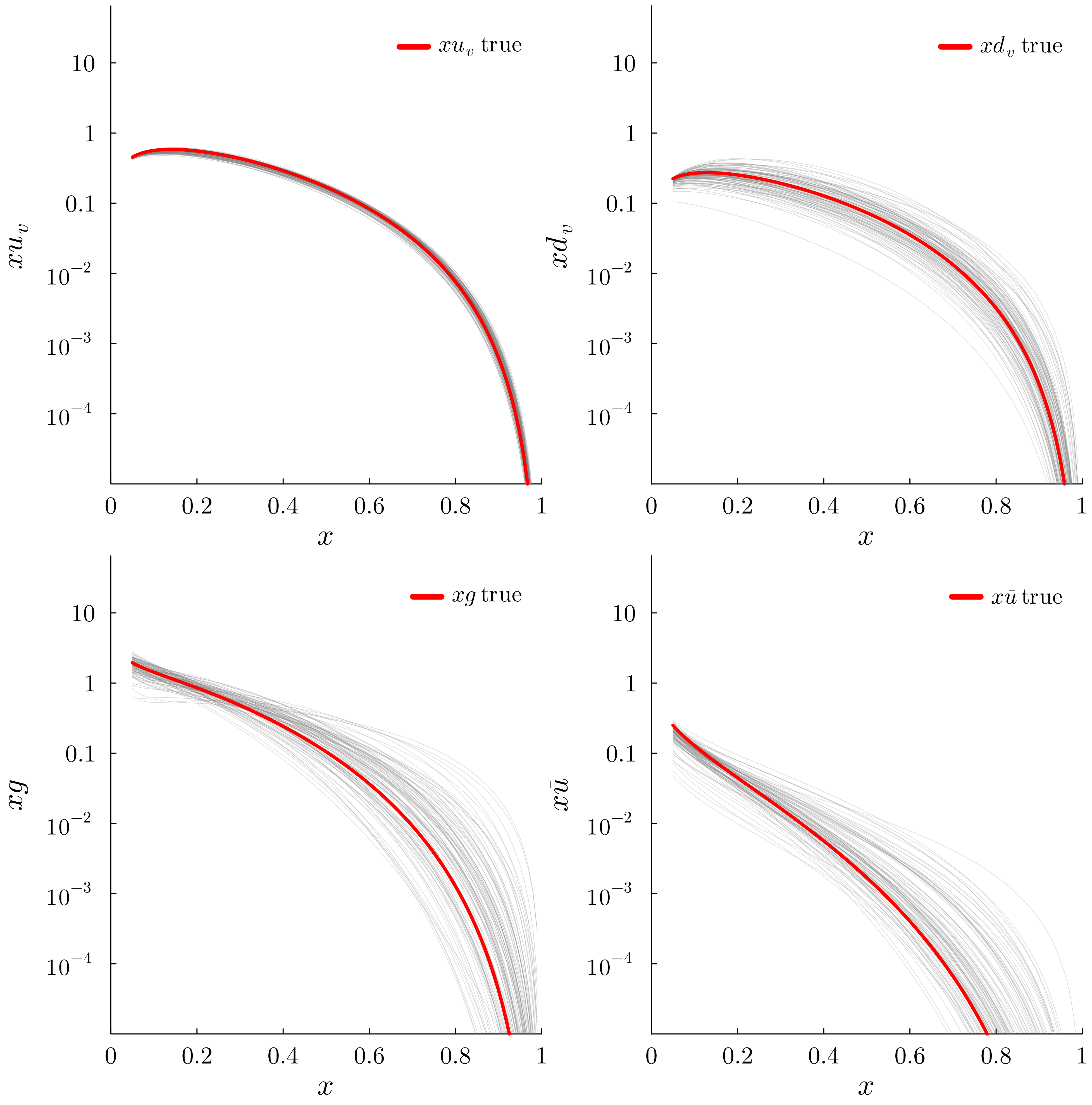


Analysis based on Samples

E.g., total Sea momentum Δ_{sea} and gluon momentum Δ_g calculated and correct probability distribution extracted

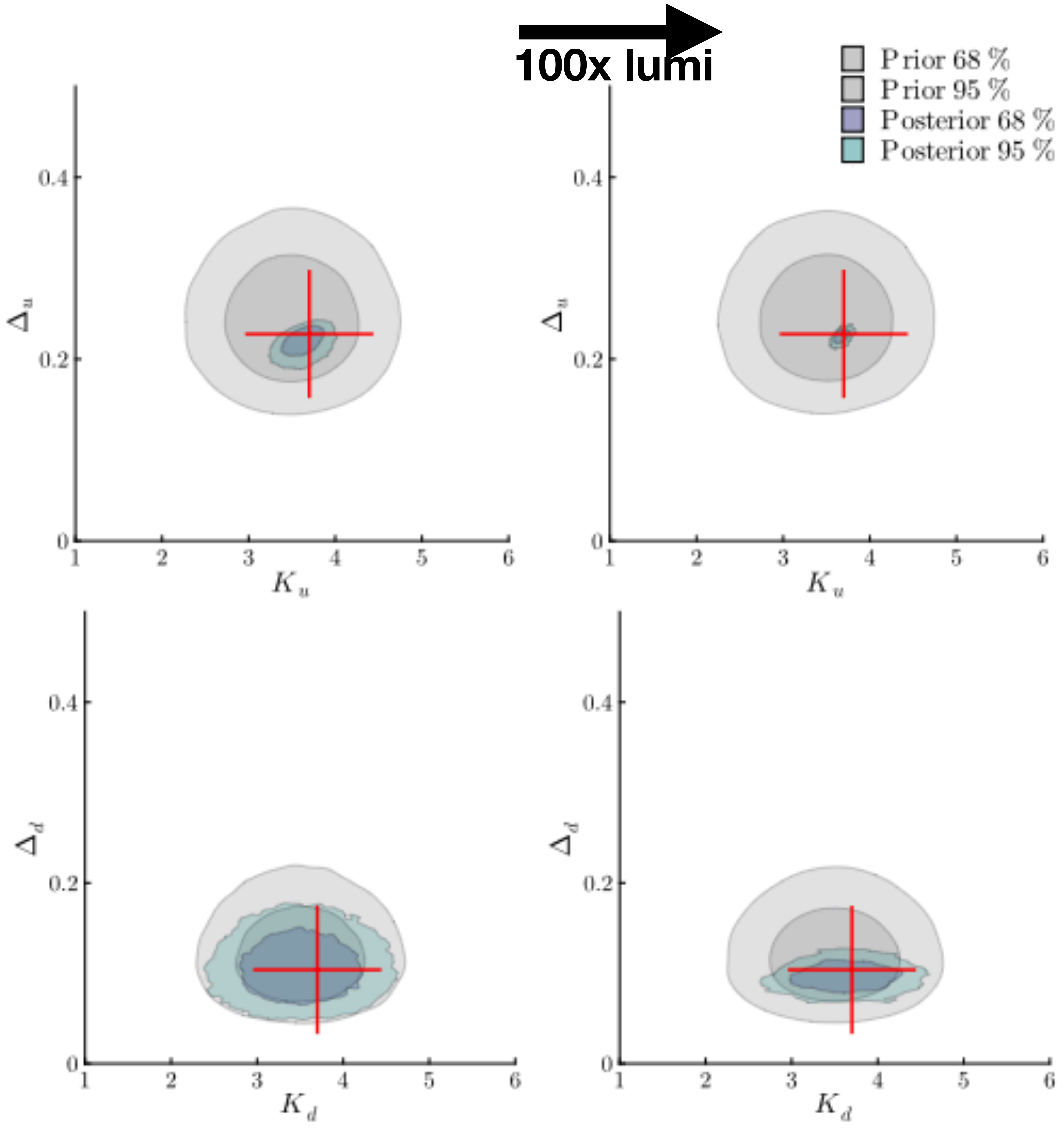


Any function of the parameters can be calculated and the probability distribution for the function plotted.

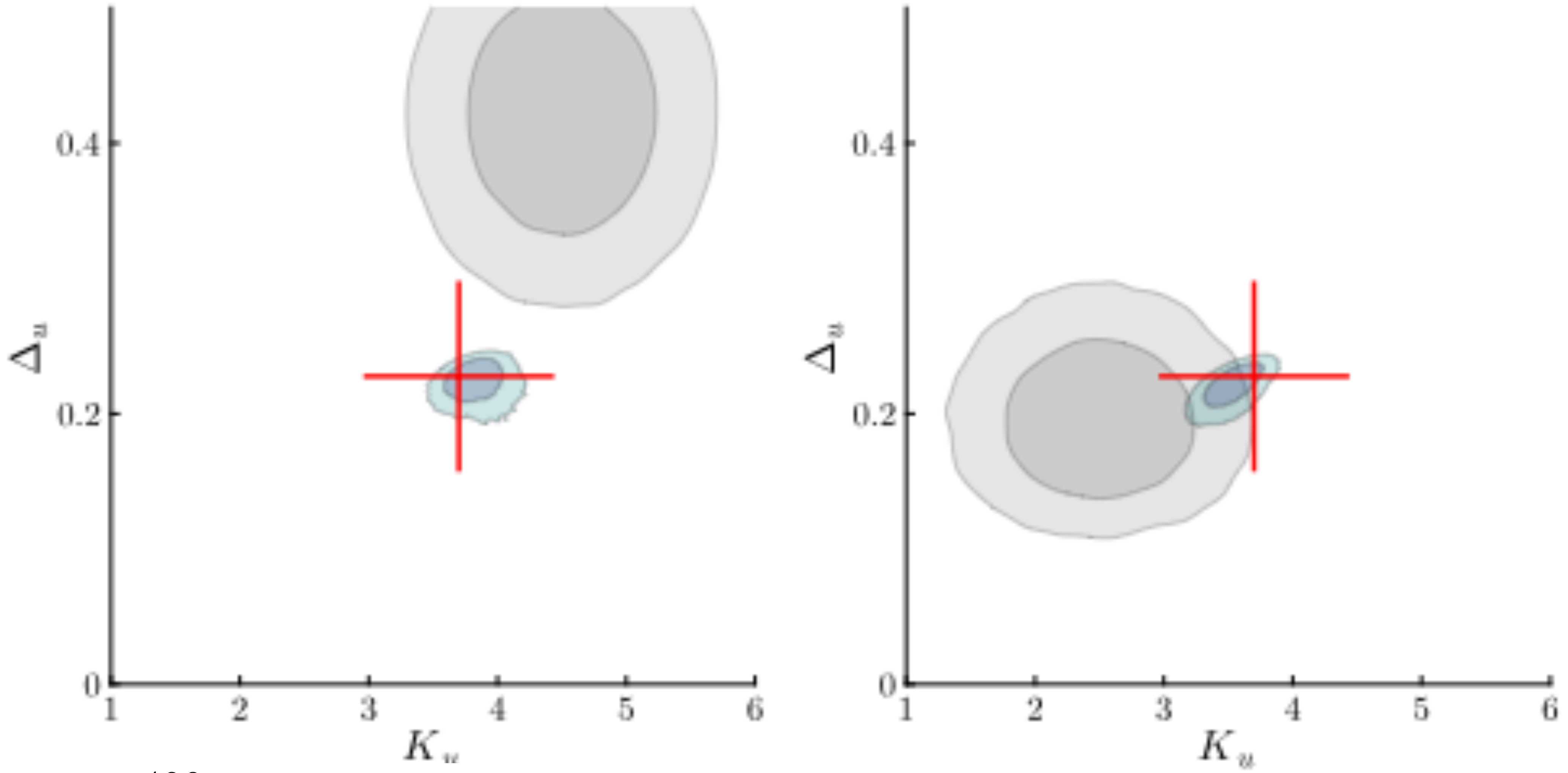


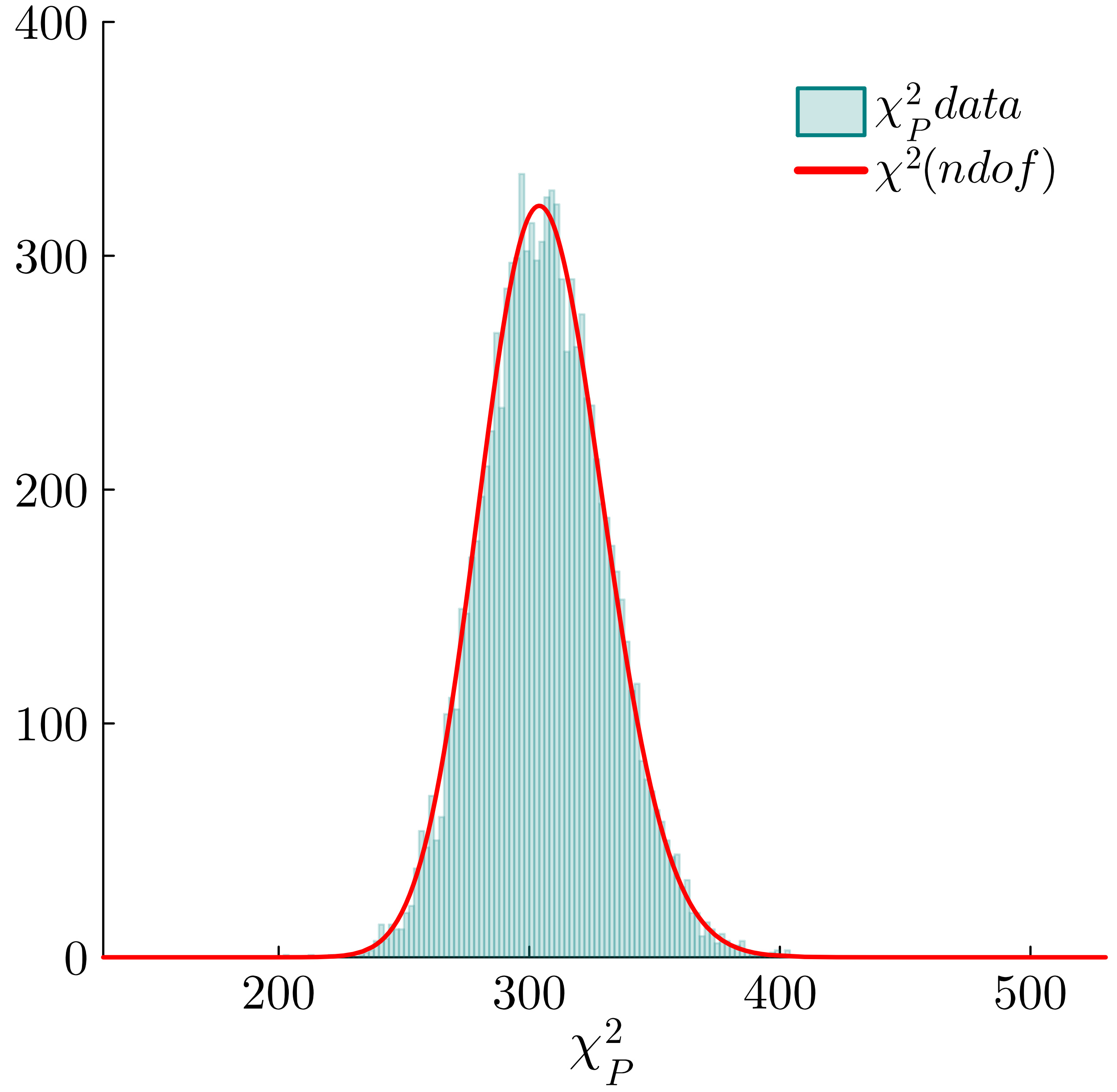
Testing

Scaling of uncertainties with data set size was verified



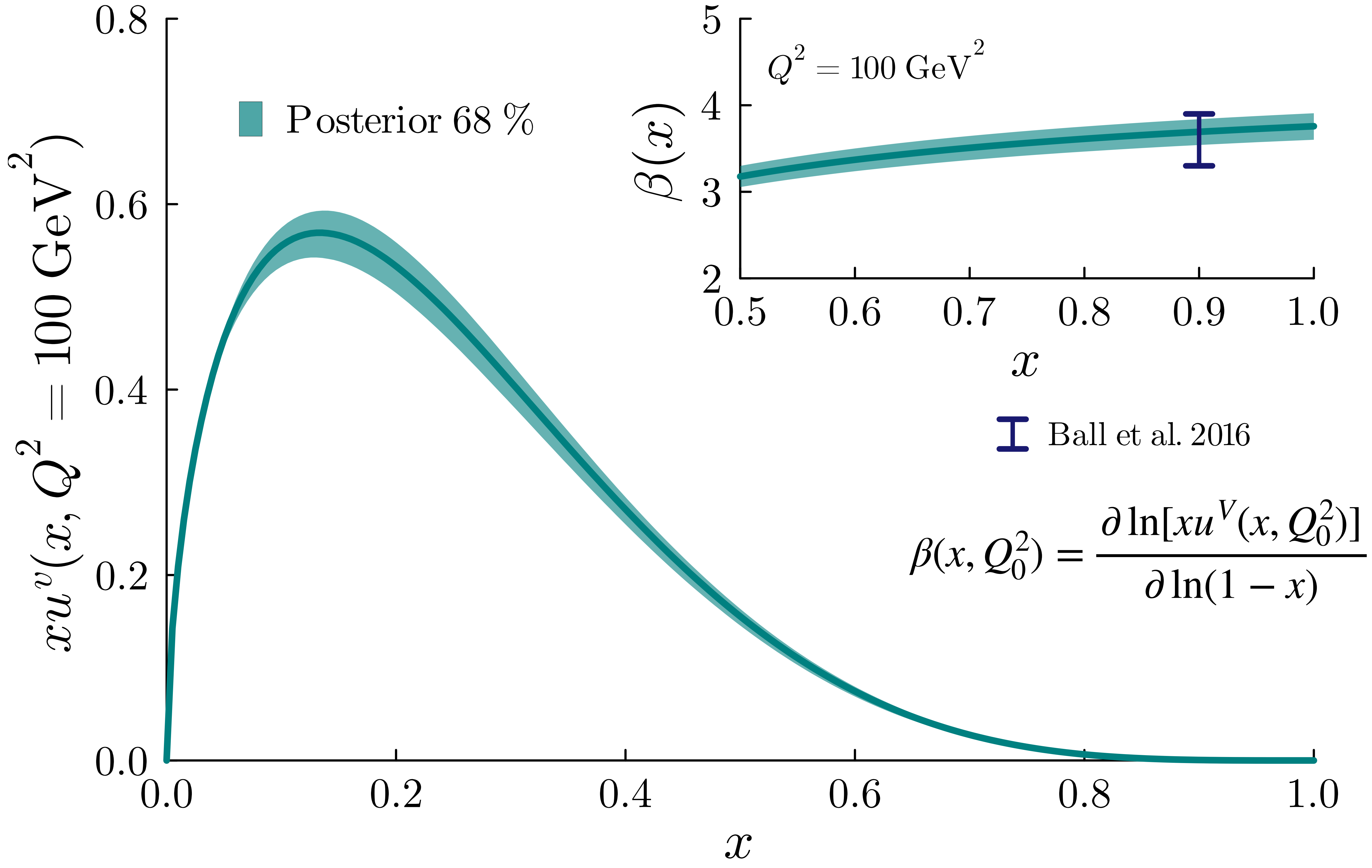
Effect of misspecified priors also studied in detail





Traditional goodness-of-fit tests also available. Here Pearson χ^2

ZEUS high-x Analysis



$$\beta(x, Q_0^2) = \frac{\partial \ln[xu^V(x, Q_0^2)]}{\partial \ln(1-x)}$$

TABLE II. Parameter values obtained from this analysis. For each parameter is given the value of the mode of the joint posterior and of its marginal distribution, with errors corresponding to the 68% smallest credible interval. The fit does not constrain the values of $\Delta_{\bar{s},\bar{c},\bar{b}}$, $\lambda_g^{\rm v}$ and $\lambda_g^{\rm s}$ (see text).

	Global mode	Marginal mode		Global mode	Marginal mode
Δ_u	0.225	$0.219^{+0.009}_{-0.009}$	K_u	3.89	$3.74^{+0.18}_{-0.13}$
Δ_d	0.084	$0.092^{+0.023}_{-0.026}$	K_d	3.18	$3.51^{+0.53}_{-0.42}$
$\lambda_{\bar{q}}$	−0.50	$-0.54^{+0.09}_{-0.09}$	$K_{\bar{q}}$	7.42	$6.38^{+1.17}_{-1.42}$
K_g	4.69	$5.02^{+1.21}_{-1.21}$			
$2\Delta_{\bar{u}}$	0.092	$0.100^{+0.026}_{-0.024}$	$2\Delta_{\bar{d}}$	0.032	$0.022^{+0.022}_{-0.014}$
$\Delta_g^{\rm v}$	0.250	$0.245^{+0.040}_{-0.044}$	$\Delta_g^{\rm s}$	0.275	$0.251^{+0.040}_{-0.045}$

Constraints on the Up-Quark Valence Distribution in the Proton
Ritu Aggarwal Pune U. Michiel Botje Nikhef, Amsterdam Allen Caldwell Munich, Max Planck Inst. Francesca Capel Munich, Max Planck Inst. Oliver Schulz Munich, Max Planck Inst.
Phys.Rev.Lett. 130 (2023) 14, 141901 • e-Print: 2209.06571 [hep-ph]

Summary

- We have developed a novel PDF analysis code allowing for a full Bayesian posterior probability determination
- Code supports a forward modeling approach.
- The open-source code has been thoroughly tested and is now available for distribution. https://github.com/cescalara/PartonDensity.jl
- Use of Julia allows I/O, plotting, analysis, …, in one high-level language
- So far, used exclusively for the analysis of high-x and high-Q^2 $e^{\pm}p$ NC scattering data. Look forward to extending the analysis to other data sets, including those reported as differential cross sections at the QED Born level.

Planned upgrades:

- Make the analysis code run much faster (e.g., by parallelizing computations through threading or forking, by improving the MCMC sampling efficiency and by speeding up the QCD evolution of the pdfs.
- Extend the framework to investigate more flexible pdf parameterizations using Bayesian model selection techniques.
- Help is welcome to make PartonDensity.jl more generally useable!