Bayesian inference and gaussian processes for PDF determination

Tommaso Giani

In collaboration with A. Candido, L.Del Debbio and G.Petrillo

Nikhef

vrije Universiteit

DIS2024, Grenoble, 10/04/2024

α_s from Z pT

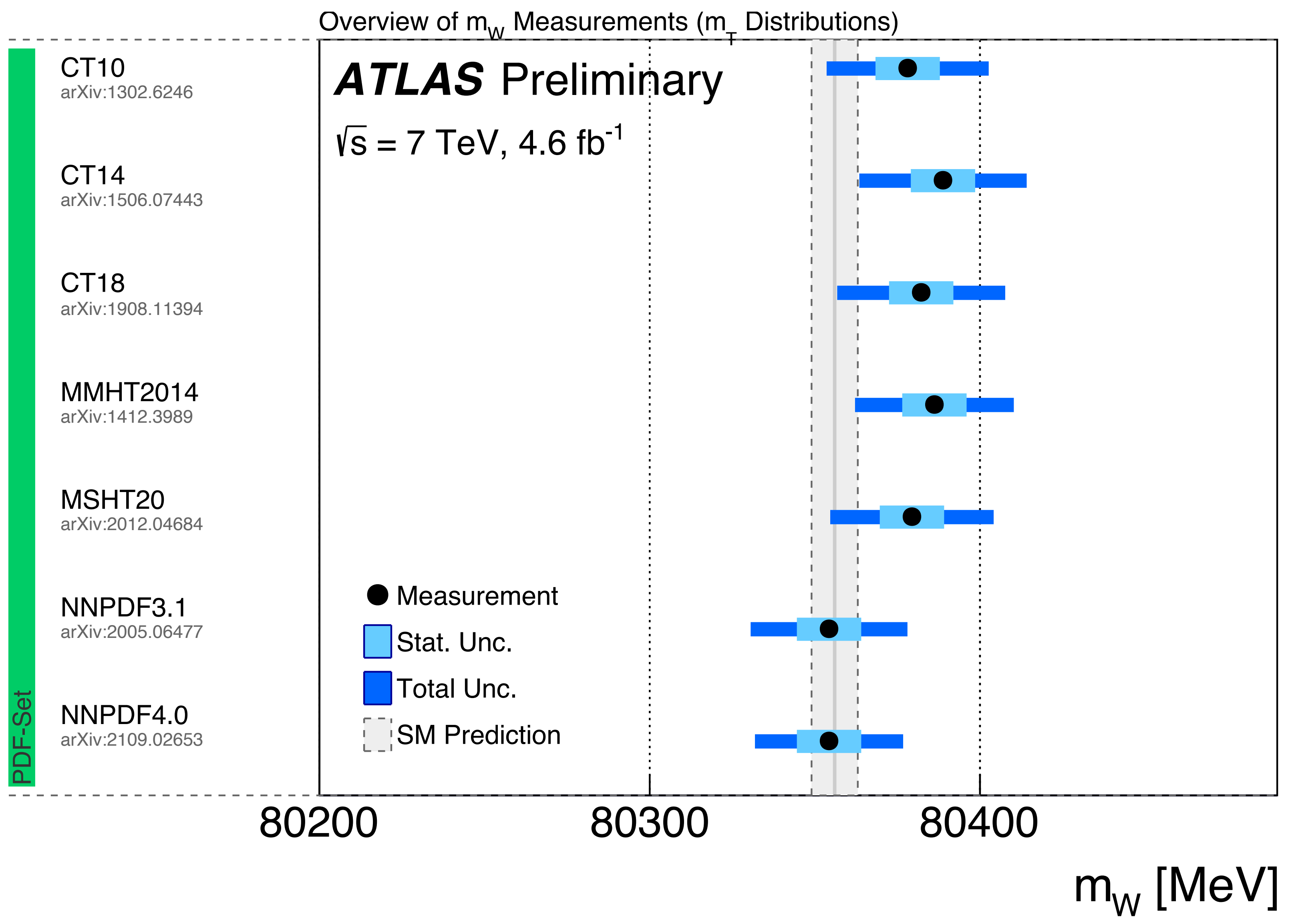
arXiv:2309.12986

$$\alpha_s(m_Z) = 0.11847 + 0.00091 - 0.00088$$

$\sim 0.76\,\%$

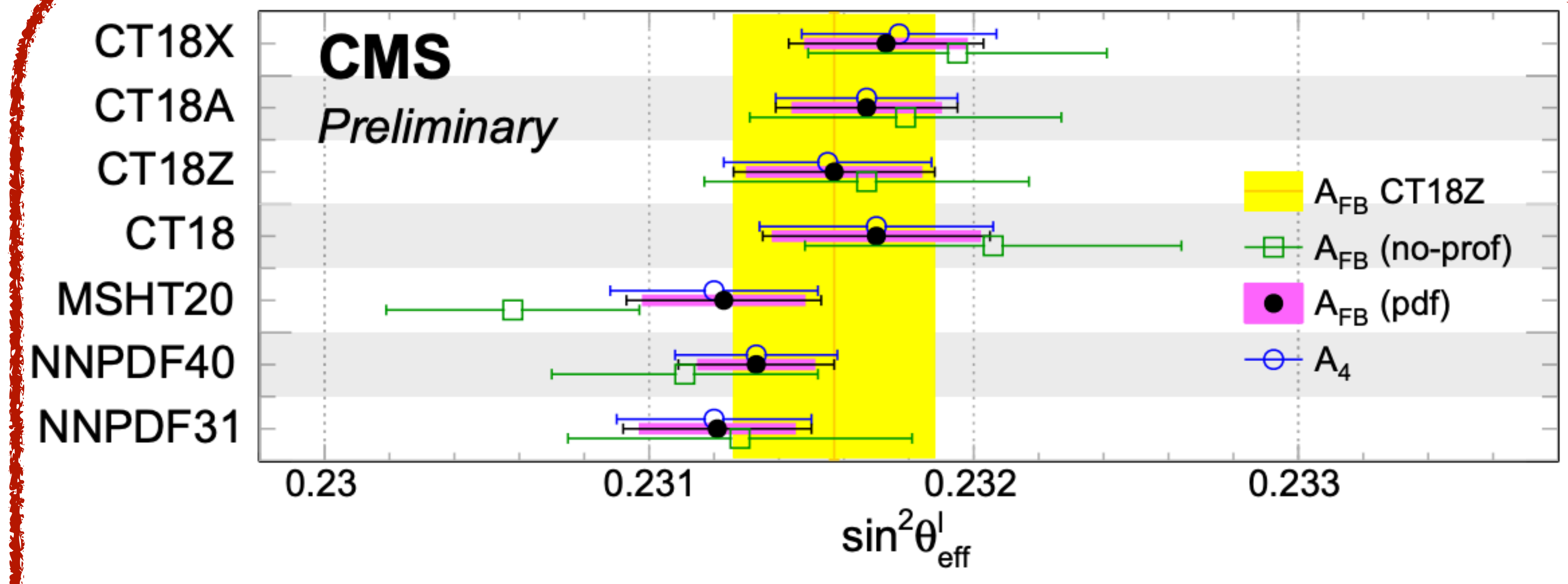
PDF set	$\alpha_s(m_Z)$	PDF uncertainty	$g\ [GeV^2]$	$q\ [GeV^4]$
MSHT20 [37]	0.11839	0.00040	0.44	−0.07
NNPDF4.0 [84]	0.11779	0.00024	0.50	−0.08
CT18A [29]	0.11982	0.00050	0.36	−0.03
HERAPDF2.0 [65]	0.11890	0.00027	0.40	−0.04

$\sim 1.7\,\%$



W mass determination

ATLAS-CONF-2023-004



weak mixing angle at 13 TeV

CMS-PAS-SMP-22-010

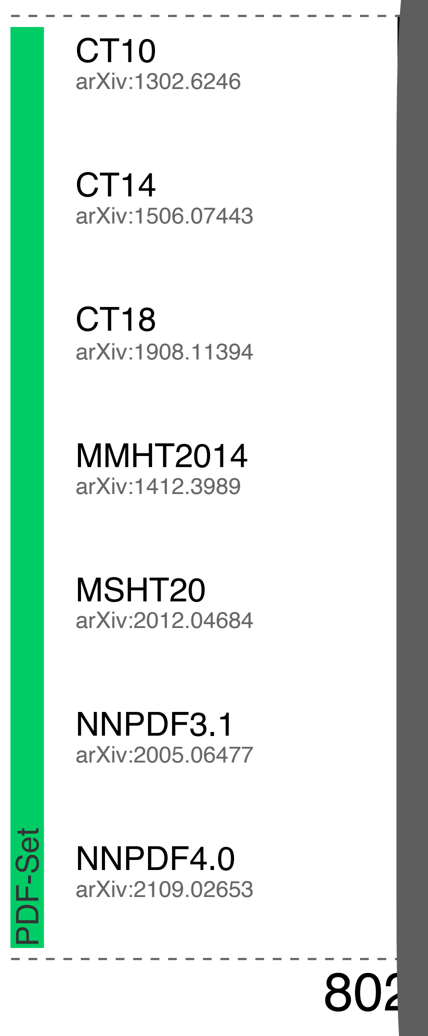
α_s from Z pT

arXiv:2309.12986

$$\alpha_s(m_Z) = 0.11847 + 0.00091 - 0.00088$$

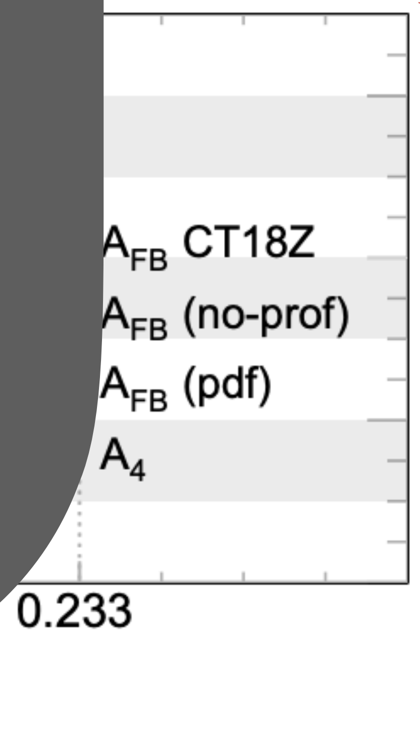
PDF set	$\alpha_s(m_Z)$	PDF uncertainty	$g\ [GeV^2]$	$q\ [GeV^4]$
MSHT20 [37]	0.11839	0.00040	0.44	−0.07
NNPDF4.0 [84]	0.11779	0.00024	0.50	−0.08
CT18A [29]	0.11982	0.00050	0.36	−0.03
HERAPDF2.0 [65]	0.11890	0.00027	0.40	−0.04

~ 1.7 %



Uncertainties in PDF determinations

- Experimental (from the data)
- Theoretical (missing higher order for theory predictions, …)
- Input SM parameters
- Methodological

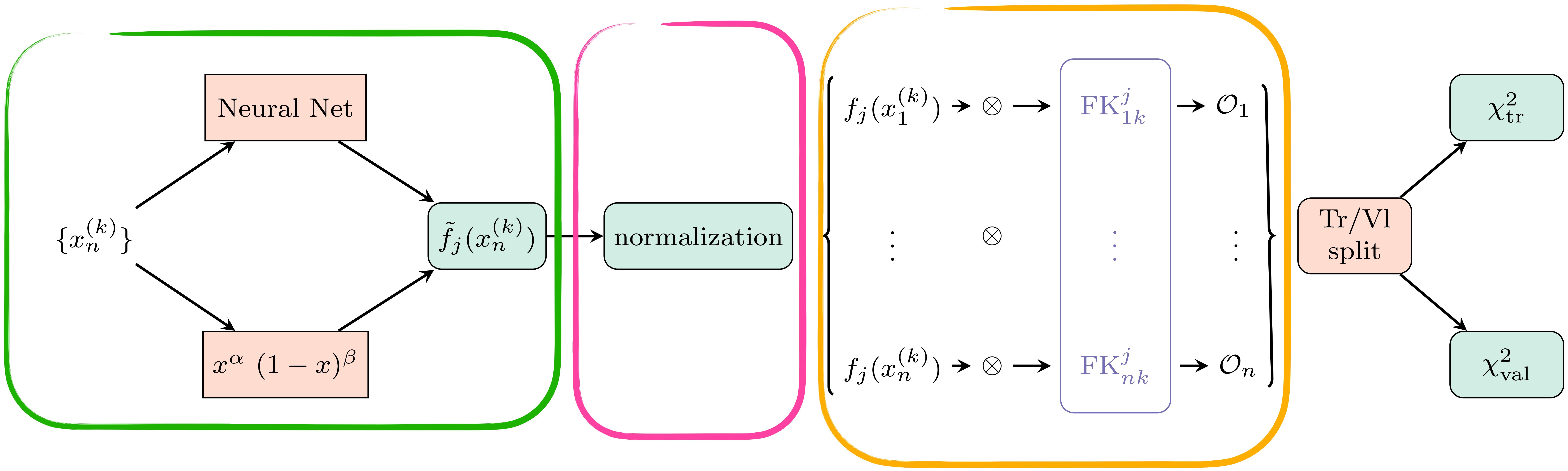


weak mixing angle at 13 TeV

CMS-PAS-SMP-22-010

Parametric regression

Build theory predictions for observables entering the fit



PDFs are parametrised at some initial scale $Q_0 = 1.65$ GeV. Sum rules are imposed with suitable normalisation

Use data to build χ^2 and minimise

Bayesian approach

- Start from a prior on the model $p\left(f\right)$
- Look at the data
- Get the posterior $p\left(f|D\right)$

Prior on the model

$$p\left(f|D\right) = \frac{p(D|f)p\left(f\right)}{p\left(D\right)}$$

Posterior of model given the data

→ Introduce probability distribution on a space of functions

→ Build a suitable prior

→ Use Bayes' theorem

Gaussian Processes

$$\mathbf{f} = \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_N) \end{pmatrix} \in \mathbb{R}^N$$

Parameters $\mathbf{f}$: stochastic variables representing values of the PDF on a grid of points

Kernel $\mathbf{K}$ and mean function $\mathbf{m}$: functions modelling the correlation between parameters

$$m\left(x_i;\theta\right) = \mathrm{E}\left(f\left(x_i\right)\right)$$

$$k\left(x_i, x_j;\theta\right) = \mathrm{cov}\left(f\left(x_i\right), f\left(x_j\right)\right)$$

Hyperparameters θ: set of parameters entering the definition of the kernel (they control some specific feature of the prior)

Joint probability distribution of $\mathbf{f}$ and θ: target of the analysis

$$p\left(\mathbf{f}, \theta \,|\, \mathrm{data}\right)$$

Some examples of application of GPs in physics

Gaussian process models—I. A framework for probabilistic continuous inverse theory FREE

Andrew P Valentine, Malcolm Sambridge

Geophysical Journal International, Volume 220, Issue 3, March 2020, Pages 1632–1647, https://doi.org/10.1093/gji/ggz520

Reconstructing QCD spectral functions with Gaussian processes

Jan Horak, Jan M. Pawlowski, José Rodríguez-Quintero, Jonas Turnwald, Julian M. Urban, Nicolas Wink, and Savvas Zafeiropoulos

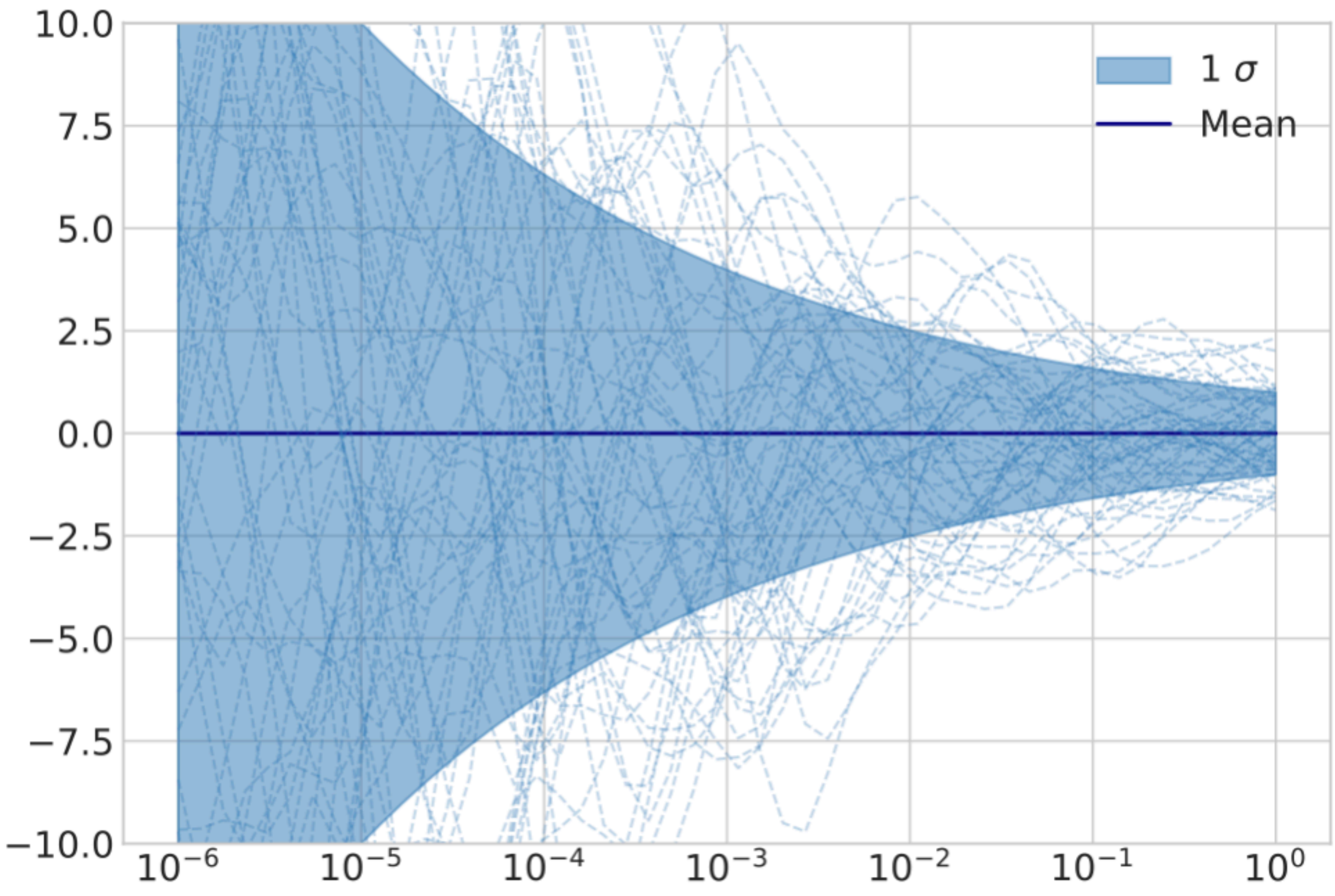
Phys. Rev. D **105**, 036014 – Published 23 February 2022

What about PDFs?

Prior for PDF

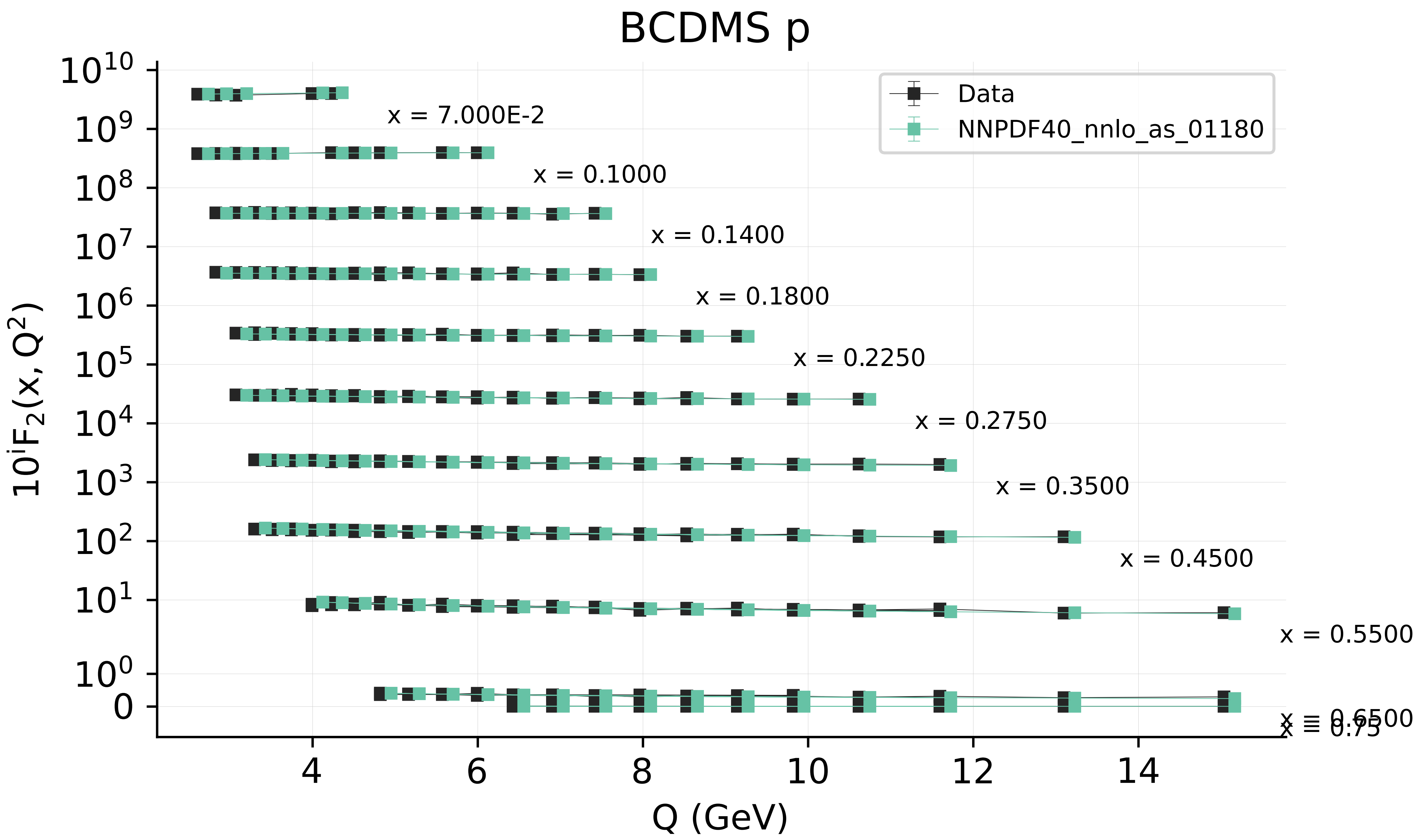
Gibbs Kernel

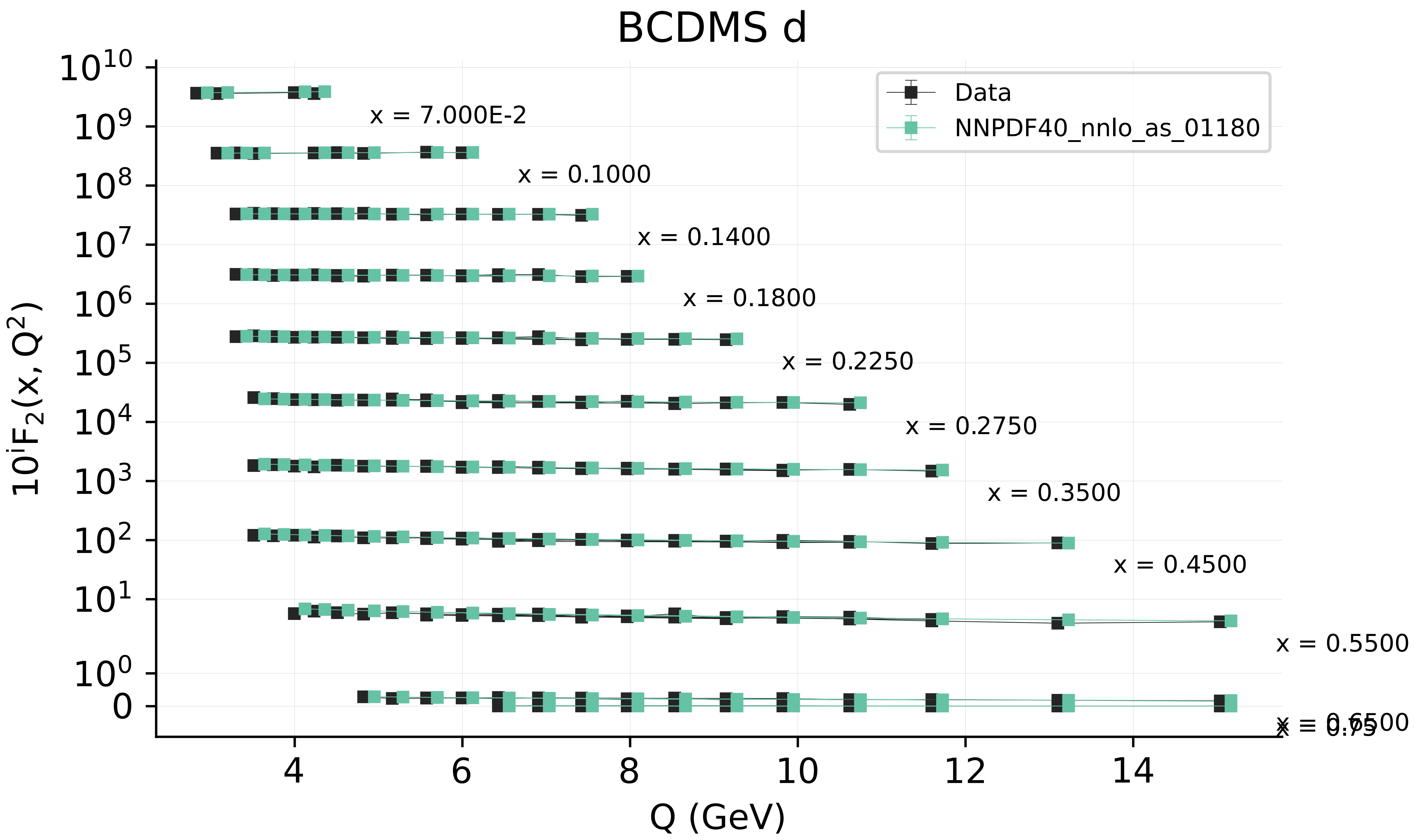
$$\tilde{k}(x,y) = x^{\alpha}y^{\alpha} \quad \sigma^2 \sqrt{\frac{2l(x)\,l(y)}{l^2(x)+l^2(y)}} \exp\left[-\frac{(x-y)^2}{l^2(x)+l^2(y)}\right] \quad \text{with} \quad l(x) = (x+\epsilon)\times l_0$$



3 hyperparameters controlling different features of the prior: α , l_0 , σ

Example: $u^+ - d^+$ from BCDMS





$$\mathcal{O} = F_2^p - F_2^d = C \otimes f \qquad f = u^+ - d^+$$

Introduce an interpolation basis for f

$$\mathcal{O}_i = \sum_\alpha (FK)_{i\alpha} f(x_\alpha) = FK\,\mathbf{f}$$

Gaussian inference

$\mathbf{f}$ Gaussian variable representing PDF on interpolation points $\mathbf{x}$ $\mathcal{O} = FK\,\mathbf{f}$	$\mathbf{f}^*$ Gaussian variable representing PDF on any set of points $\mathbf{x}^*$	$K\left(x, y; \theta\right)$ Function modelling correlation	$y, \quad \epsilon \sim N(0, C_y)$ Data and corresponding experimental error

$$\begin{pmatrix} \mathbf{f}^* \\ FK\,\mathbf{f} \end{pmatrix} \sim \mathcal{N}\left[0\,, \begin{pmatrix} K_{\mathbf{x}^*\mathbf{x}^*} & K_{\mathbf{x}^*\mathbf{x}}\,FK^T \\ FK\,K_{\mathbf{x}\mathbf{x}^*} & FK\,K_{\mathbf{x}\mathbf{x}}\,FK^T \end{pmatrix}\right]$$

$$p\left(\mathbf{f}^* \,|\, FK\,\mathbf{f} + \epsilon = y, \theta\right)$$

This is a gaussian distribution. Its mean and covariance can be computed analytically

Inference on the hyperparameters

Posterior on the hyperparamters given the data

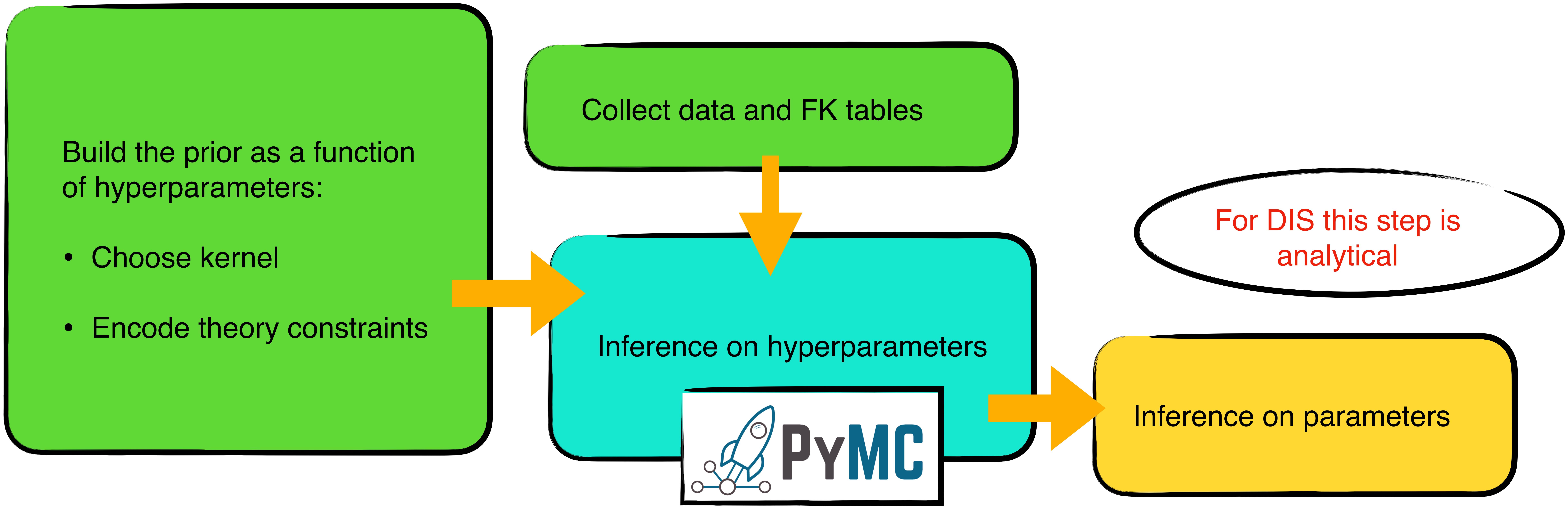
Joint probability distribution of $\mathbf{f}^*$ and θ

$$p\left(\mathbf{f}^*, \theta \,|\, \text{data}\right) = p\left(\mathbf{f}^* \,|\, \theta, \text{data}\right) p\left(\theta \,|\, \text{data}\right)$$

$$\propto p\left(\text{data} \,|\, \theta\right) p_\theta(\theta)$$

We can sample from $p\left(\theta \,|\, \text{data}\right)$ running a MCMC algorithm

Workflow


Build the prior as a function of hyperparameters:
• Choose kernel
• Encode theory constraints
Collect data and FK tables
Inference on hyperparameters
PyMC
For DIS this step is analytical
Inference on parameters

Posterior for hyperparameters

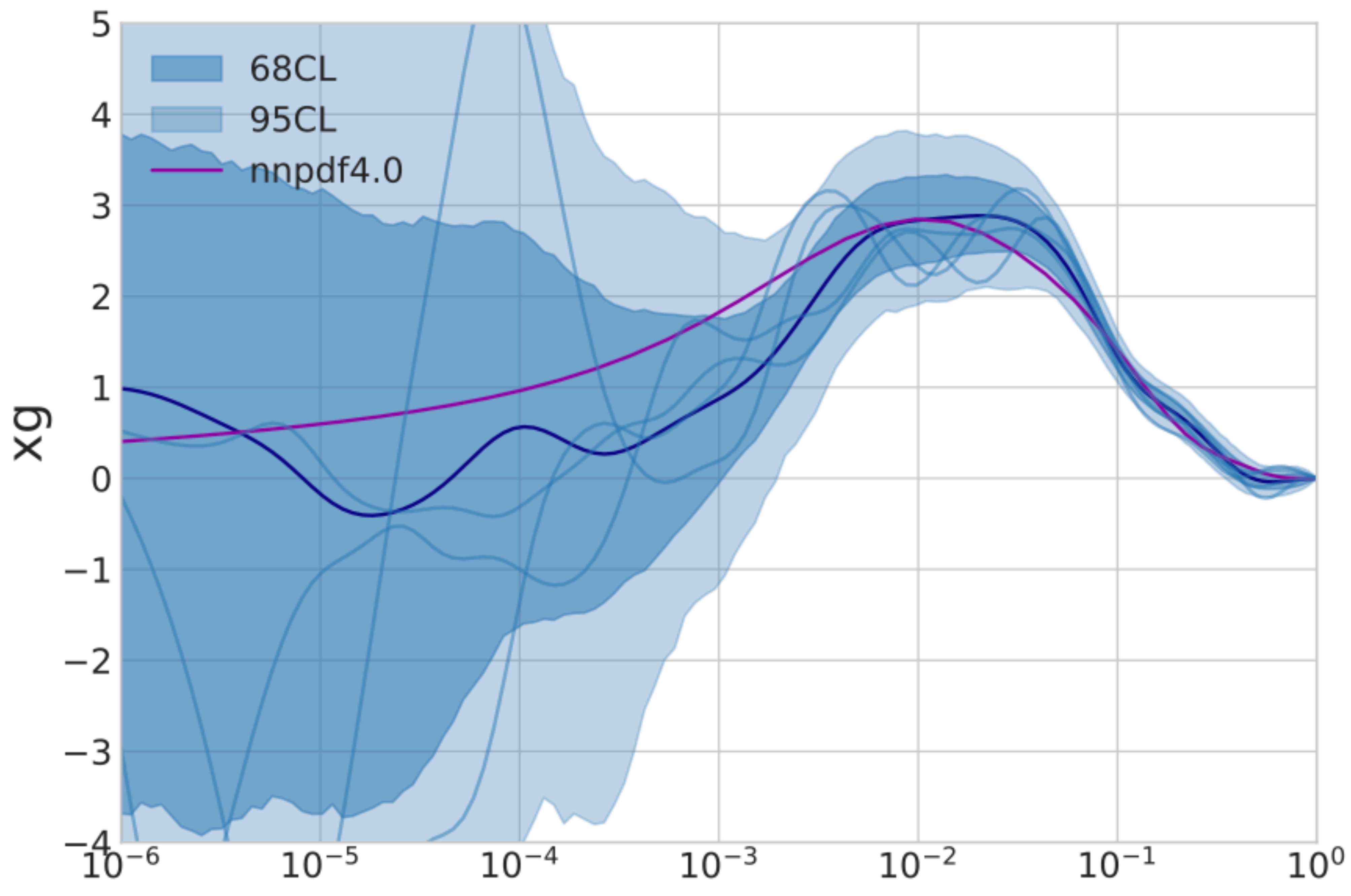
Samples from $p\left(\mathbf{f}^{*}, \theta \,|\, data\right)$

uncertainty due to hyperparameter selection is incorporated into the final PDF uncertainty

Towards a full DIS fit

- 9 flavours
- Extended set of DIS data
- Same formalism and simplifications apply
- Kinetic limit $f(1) = 0$
- Momentum and valence sumrules

Dataset	References	N_{dat}	x	Q [GeV]
NMC F_2^d/F_2^p	[33]	260 (121/121)	[0.012, 0.680]	[2.1, 10.]
NMC $\sigma^{\text{NC},p}$	[34]	292 (204/204)	[0.012, 0.500]	[1.8, 7.9]
SLAC F_2^p	[35]	211 (33/33)	[0.140, 0.550]	[1.9, 4.4]
SLAC F_2^d	[35]	211 (34/34)	[0.140, 0.550]	[1.9, 4.4]
BCDMS F_2^p	[36]	351 (333/333)	[0.070, 0.750]	[2.7, 15.]
BCDMS F_2^d	[36]	254 (248/248)	[0.070, 0.750]	[2.7, 15.]
CHORUS σ_{CC}^{ν}	[37]	607 (416/416)	[0.045, 0.650]	[1.9, 9.8]
CHORUS $\sigma_{CC}^{\bar{\nu}}$	[37]	607 (416/416)	[0.045, 0.650]	[1.9, 9.8]
NuTeV σ_{CC}^{ν} (dimuon)	[38,39]	45 (39/39)	[0.020, 0.330]	[2.0, 11.]
NuTeV $\sigma_{CC}^{\bar{\nu}}$ (dimuon)	[38,39]	45 (36/37)	[0.020, 0.210]	[1.9, 8.3]
[NOMAD $\mathcal{R}_{\mu\mu}(E_\nu)$] (*)	[111]	15 (–/15)	[0.030, 0.640]	[1.0, 28.]
[EMC F_2^c]	[44]	21 (–/16)	[0.014, 0.440]	[2.1, 8.8]
HERA I+II $\sigma_{\text{NC,CC}}^p$	[40]	1306 (1011/1145)	[$4\cdot10^{-5}$, 0.65]	[1.87, 223]
HERA I+II σ_{NC}^c (*)	[145]	52 (–/37)	[$7\cdot10^{-5}$, 0.05]	[2.2, 45]
HERA I+II σ_{NC}^b (*)	[145]	27 (26/26)	[$2\cdot10^{-4}$, 0.50]	[2.2, 45]



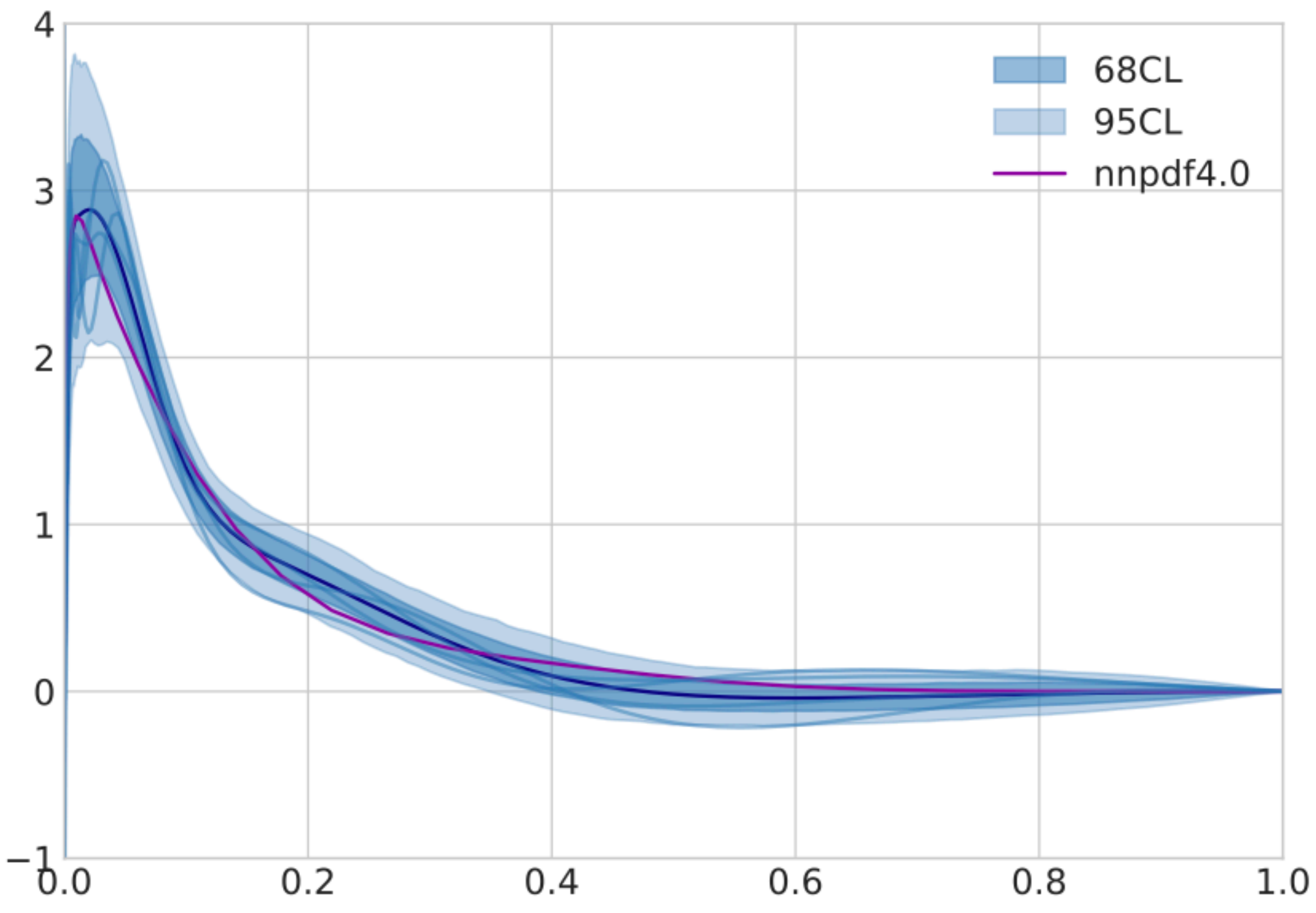


Table from Eur.Phys.J.C 82 (2022) 5

Towards a full DIS fit

- 9 flavours
- Extended set of DIS data
- Same formalism and simplifications apply
- Kinetic limit $f(1) = 0$
- Momentum and valence sumrules

Dataset	References	N_{dat}	x	Q [GeV]
NMC F_2^d/F_2^p	[33]	260 (121/121)	[0.012, 0.680]	[2.1, 10.]
NMC $\sigma^{\text{NC},p}$	[34]	292 (204/204)	[0.012, 0.500]	[1.8, 7.9]
SLAC F_2^p	[35]	211 (33/33)	[0.140, 0.550]	[1.9, 4.4]
SLAC F_2^d	[35]	211 (34/34)	[0.140, 0.550]	[1.9, 4.4]
BCDMS F_2^p	[36]	351 (333/333)	[0.070, 0.750]	[2.7, 15.]
BCDMS F_2^d	[36]	254 (248/248)	[0.070, 0.750]	[2.7, 15.]
CHORUS σ_{CC}^{ν}	[37]	607 (416/416)	[0.045, 0.650]	[1.9, 9.8]
CHORUS $\sigma_{CC}^{\bar{\nu}}$	[37]	607 (416/416)	[0.045, 0.650]	[1.9, 9.8]
NuTeV σ_{CC}^{ν} (dimuon)	[38,39]	45 (39/39)	[0.020, 0.330]	[2.0, 11.]
NuTeV $\sigma_{CC}^{\bar{\nu}}$ (dimuon)	[38,39]	45 (36/37)	[0.020, 0.210]	[1.9, 8.3]
[NOMAD $\mathcal{R}_{\mu\mu}(E_\nu)$] (*)	[111]	15 (–/15)	[0.030, 0.640]	[1.0, 28.]
[EMC F_2^c]	[44]	21 (–/16)	[0.014, 0.440]	[2.1, 8.8]
HERA I+II $\sigma_{\text{NC,CC}}^p$	[40]	1306 (1011/1145)	[$4\cdot10^{-5}$, 0.65]	[1.87, 223]
HERA I+II σ_{NC}^c (*)	[145]	52 (–/37)	[$7\cdot10^{-5}$, 0.05]	[2.2, 45]
HERA I+II σ_{NC}^b (*)	[145]	27 (26/26)	[$2\cdot10^{-4}$, 0.50]	[2.2, 45]

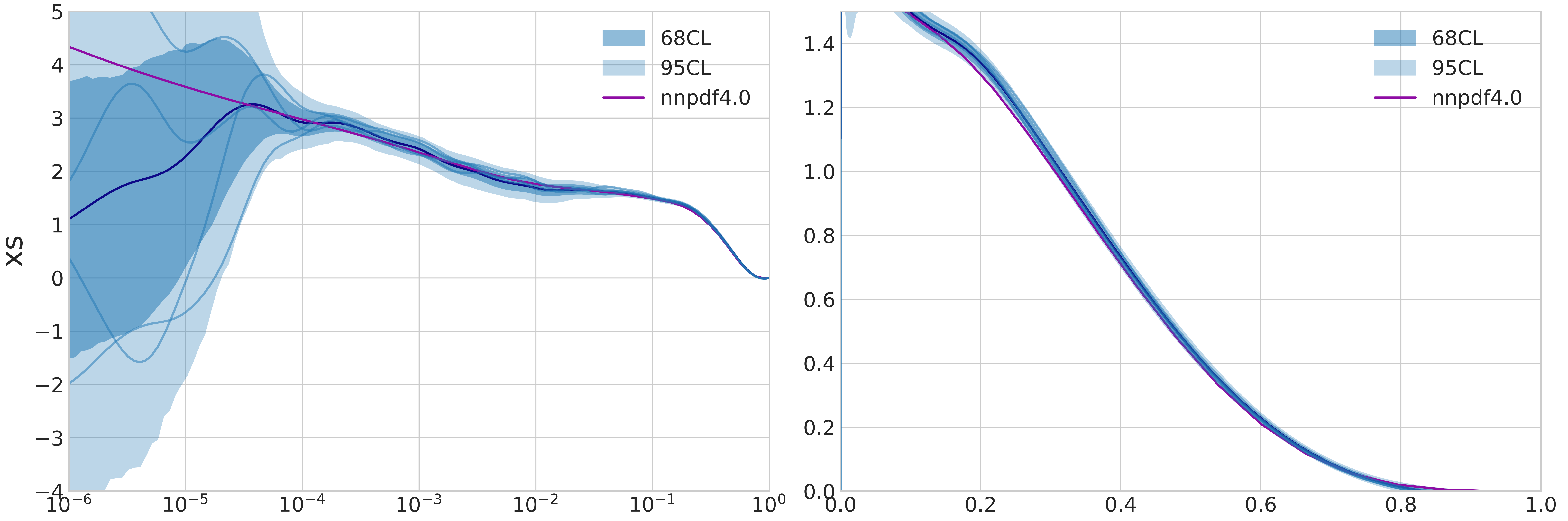
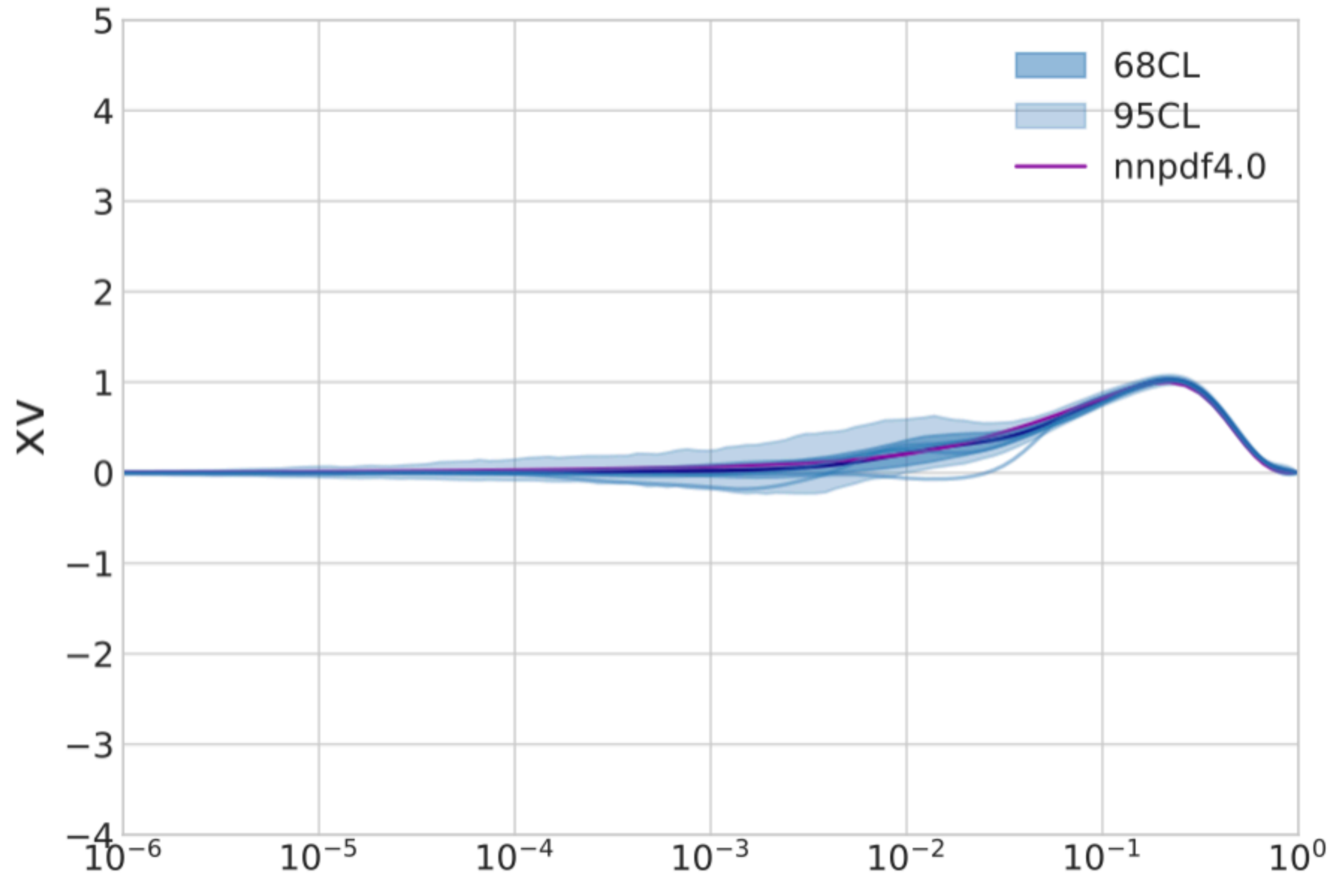


Table from Eur.Phys.J.C 82 (2022) 5

Towards a full DIS fit

- 9 flavours
- Extended set of DIS data
- Same formalism and simplifications apply
- Kinetic limit $f(1) = 0$
- Momentum and valence sumrules

Dataset	References	$N_{\rm dat}$	x	Q [GeV]
NMC F_2^d/F_2^p	[33]	260 (121/121)	[0.012, 0.680]	[2.1, 10.]
NMC $\sigma^{{\rm NC},p}$	[34]	292 (204/204)	[0.012, 0.500]	[1.8, 7.9]
SLAC F_2^p	[35]	211 (33/33)	[0.140, 0.550]	[1.9, 4.4]
SLAC F_2^d	[35]	211 (34/34)	[0.140, 0.550]	[1.9, 4.4]
BCDMS F_2^p	[36]	351 (333/333)	[0.070, 0.750]	[2.7, 15.]
BCDMS F_2^d	[36]	254 (248/248)	[0.070, 0.750]	[2.7, 15.]
CHORUS σ_{CC}^{ν}	[37]	607 (416/416)	[0.045, 0.650]	[1.9, 9.8]
CHORUS $\sigma_{CC}^{\bar{\nu}}$	[37]	607 (416/416)	[0.045, 0.650]	[1.9, 9.8]
NuTeV σ_{CC}^{ν} (dimuon)	[38,39]	45 (39/39)	[0.020, 0.330]	[2.0, 11.]
NuTeV $\sigma_{CC}^{\bar{\nu}}$ (dimuon)	[38,39]	45 (36/37)	[0.020, 0.210]	[1.9, 8.3]
[NOMAD $\mathcal{R}_{\mu\mu}(E_\nu)$] (*)	[111]	15 (–/15)	[0.030, 0.640]	[1.0, 28.]
[EMC F_2^c]	[44]	21 (–/16)	[0.014, 0.440]	[2.1, 8.8]
HERA I+II $\sigma_{\rm NC,CC}^p$	[40]	1306 (1011/1145)	[$4\cdot 10^{-5}$, 0.65]	[1.87, 223]
HERA I+II $\sigma_{\rm NC}^c$ (*)	[145]	52 (–/37)	[$7\cdot 10^{-5}$, 0.05]	[2.2, 45]
HERA I+II $\sigma_{\rm NC}^b$ (*)	[145]	27 (26/26)	[$2\cdot 10^{-4}$, 0.50]	[2.2, 45]



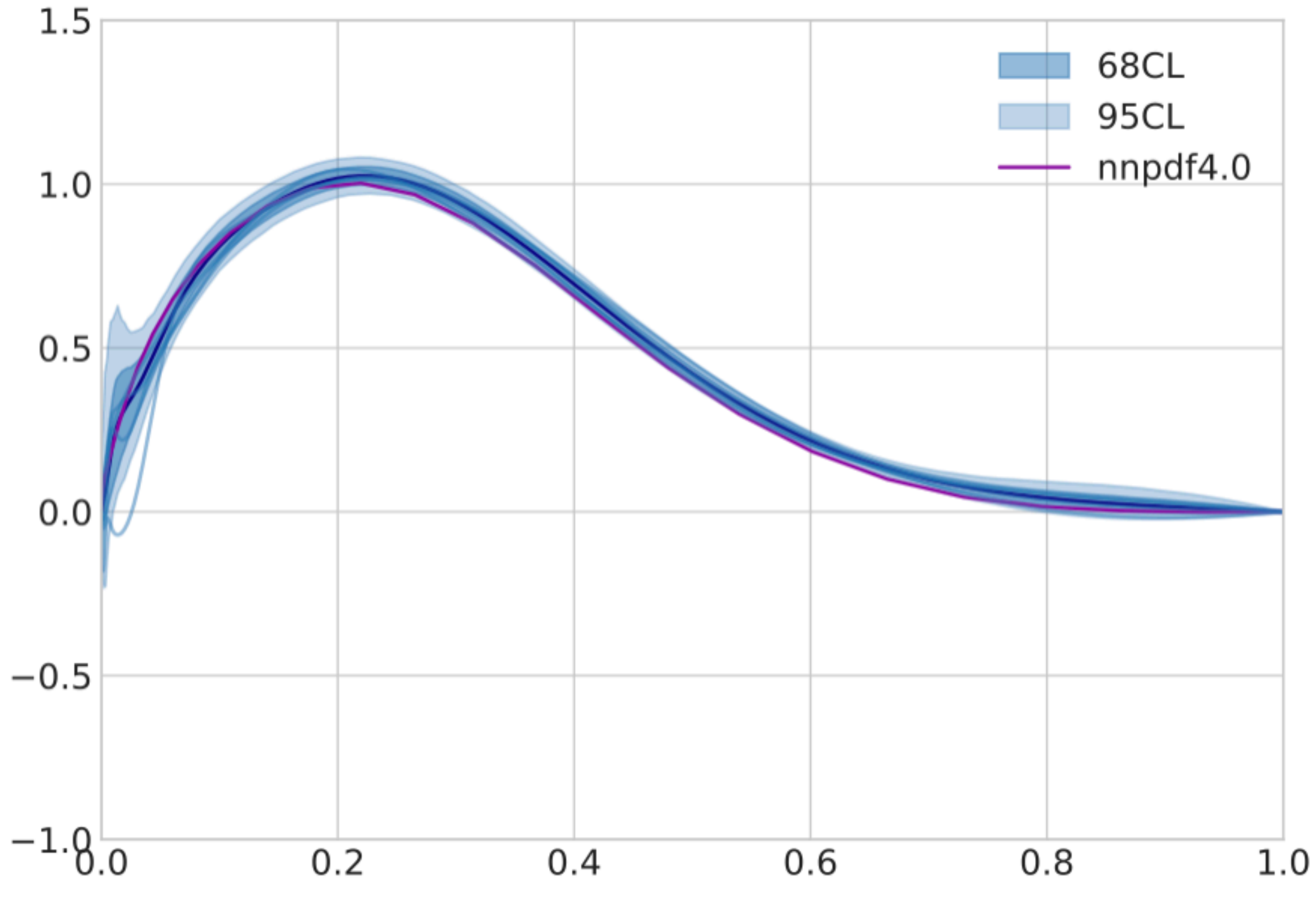


Table from Eur.Phys.J.C 82 (2022) 5

What changes in a global fit?

$$p\left(\mathbf{f},\theta\,|\,\mathsf{data}\right) = \underline{p\left(\mathbf{f}\,|\,\mathsf{data},\theta\right)}\, p\left(\theta\,|\,\mathsf{data}\right)$$

This bit is not a gaussian distribution any longer

To access the posterior we have to run a MCMC having dimension $\dim \mathbf{f} + \dim \theta$

TO DO :

- Provide a DIS only PDF fit
- Study dependence on the kernel
- Compare with existing methodology. Are there differences?

What changes in a global fit?

$$p\left(\mathbf{f},\theta\,|\,\text{data}\right) = p\left(\mathbf{f}\,|\,\text{data},\theta\right) p\left(\theta\,|\,\text{data}\right)$$

This bit is not a gaussian distribution any longer

To access the posterior we have to run a MCMC having dimension $\dim \mathbf{f} + \dim \theta$

TO DO :

- Provide a DIS only PDF fit
- Study dependence on the kernel
- Compare with existing methodology. Are there differences?

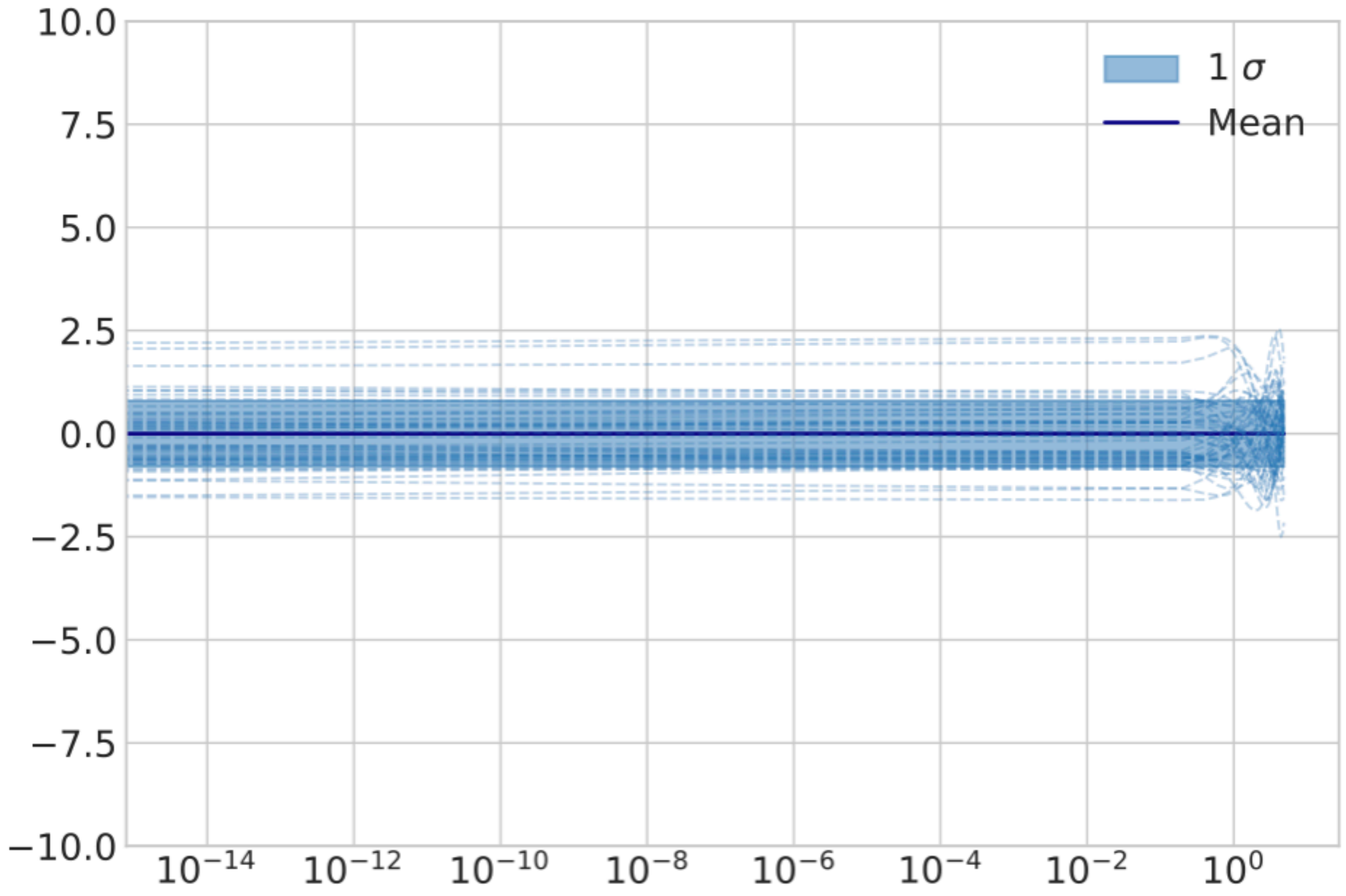
Thanks!

Backup slides

An example of a bad prior for PDFs

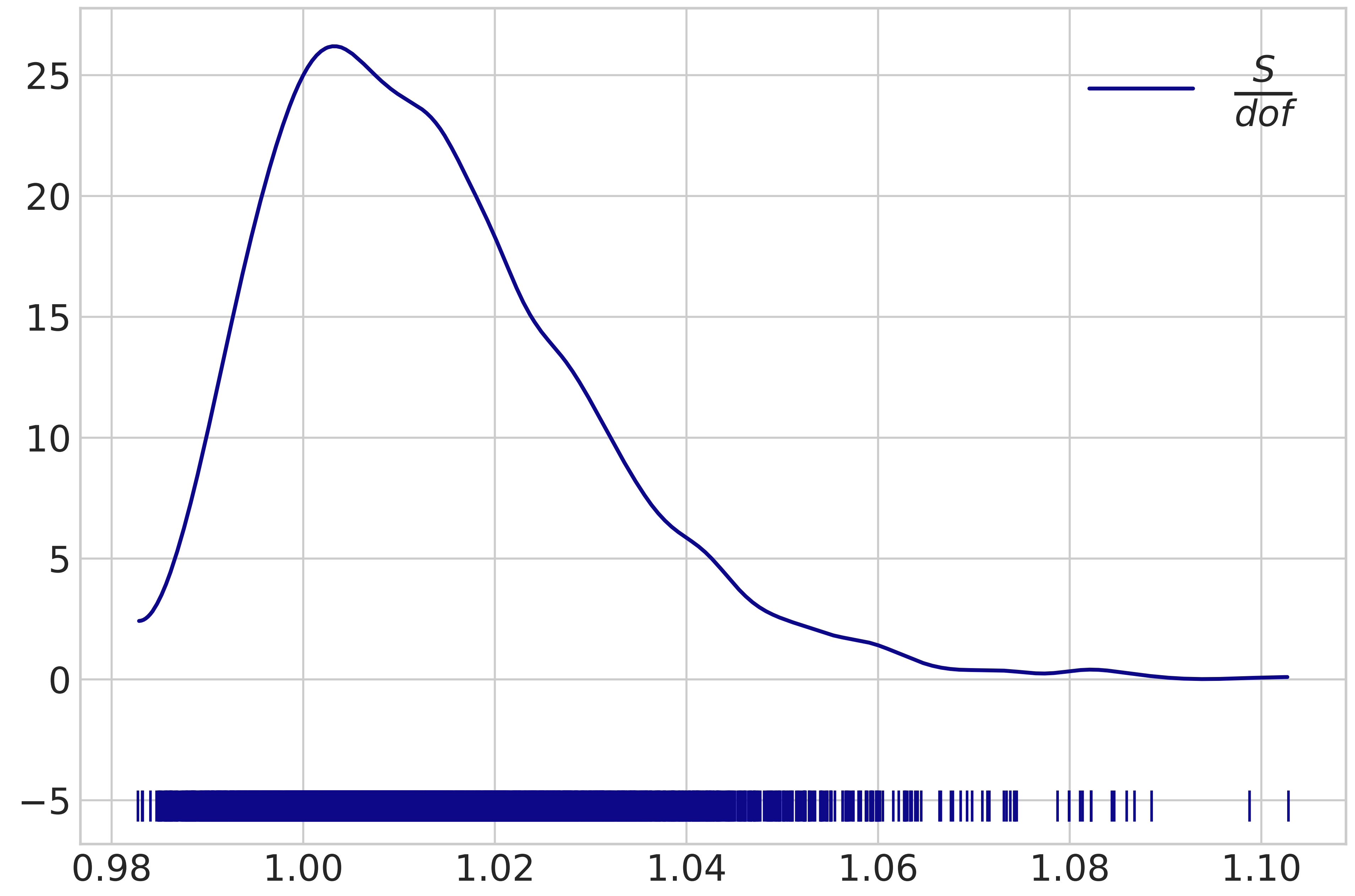
$$k(x, y) = \sigma^2 \exp\left[-\frac{(x-y)^2}{l^2}\right]$$

Exponential quadratic



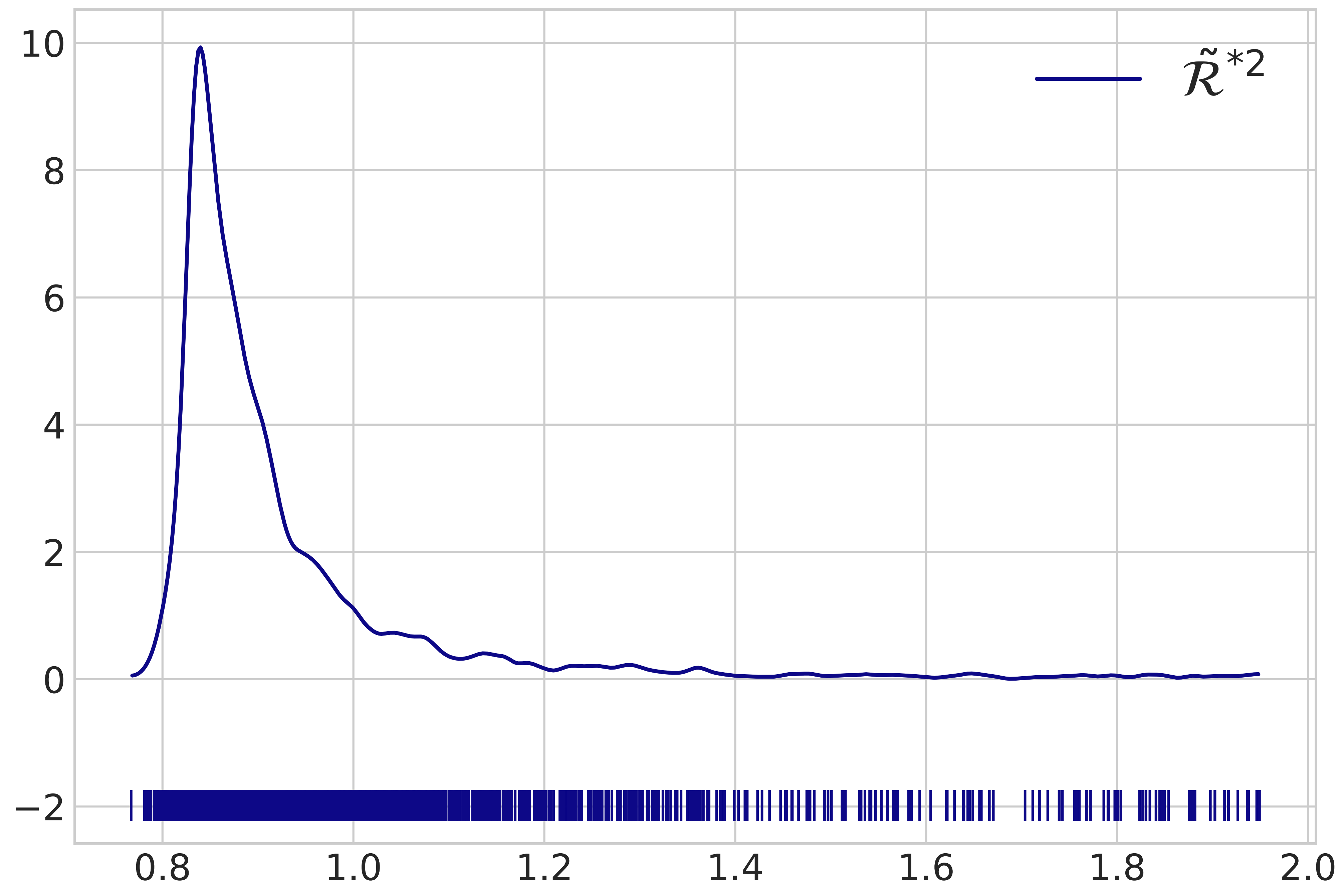
Fit quality

$$\frac{S}{dof} = \frac{1}{N_{\text{data}}}\left((\mathbf{m}-\tilde{\mathbf{m}})^T K_{xx}^{-1}(\mathbf{m}-\tilde{\mathbf{m}}) + (y - FK\,\tilde{\mathbf{m}})^T C_Y^{-1}(y - FK\,\tilde{\mathbf{m}})\right)$$



Generalisation on unseen data

$$\tilde{\mathscr{R}}^{*2} = \frac{1}{\dim(y^* \,|\, y)}(FK^*\,\tilde{\mathbf{m}} - y^*)^T\left(FK^*\,\tilde{K}_{xx}FK^{*T} + C_Y^*\right)^+(FK^*\,\tilde{\mathbf{m}} - y^*)$$



Gaussian inference

$\mathbf{f}$ Gaussian variable representing PDF on interpolation points $\mathbf{x}$ $\mathcal{O} = FK\,\mathbf{f}$	$\mathbf{f}*$ Gaussian variable representing PDF on any set of points $\mathbf{x}*$	$K\left(x, y; \theta\right)$ Function modelling correlation	$y, \quad \epsilon \sim N(0,C_y)$ Data and corresponding experimental error

$$\tilde{\mathbf{m}}* = \mathbf{m} + K_{\mathbf{x}*\mathbf{x}}FK^T\left(FK\,K_{\mathbf{xx}}FK^T + C_y\right)^+\left(\mathbf{y} - \mathbf{m}\right)$$

$$\tilde{K}* = K_{\mathbf{x}*\mathbf{x}*} - K_{\mathbf{x}*\mathbf{x}}FK^T\left(FK\,K_{\mathbf{xx}}FK^T + C_y\right)^+ FK\,K_{\mathbf{xx}*}$$

inference of the gaussian parameters $\mathbf{f}*$ can be done analytically

Further possible applications

- simultaneous fits of PDFs and Wilson coefficients

$$\sigma_{\rm eft}\left(c/\Lambda^2\right) = \sigma_{\rm SM} + \sum_i \tilde{\sigma}_i^{\rm LO/NLO}\frac{c_i}{\Lambda^2} + \sum_{i,j}\tilde{\sigma}_{ij}^{\rm LO/NLO}\frac{c_i c_j}{\Lambda^4}$$

The top quark legacy of the LHC Run II for PDF and SMEFT analyses

Zahari Kassabov,[a] Maeve Madigan,[a] Luca Mantani,[a] James Moore,[a] Manuel Morales Alvarado,[a] Juan Rojo[b,c] and Maria Ubiali[a]

JHEP 05 (2023) 205

- Inverse problems relevant for the lattice community

Reconstructing QCD Spectral Functions with Gaussian Processes

Jan Horak,[1] Jan M. Pawlowski,[1,2] José Rodríguez-Quintero,[3] Jonas Turnwald,[1] Julian M. Urban,[1,*] Nicolas Wink,[1] and Savvas Zafeiropoulos[4]

Phys.Rev.D 105 (2022) 3

Decomposition of PDF uncertainty

$$\tilde{K} = \left(I - R_{xx}\right) K_{\mathbf{xx}} \left(I - R_{xx}\right)^T + a_{xx}^T C_y a_{xx}$$

Methodology   Experimental error

$$a_{xx}^T = K_{\mathbf{xx}} FK^T \left(FK\, K_{\mathbf{xx}} FK^T + C_y \right)^+$$

$$R_{xx} = a_{xx}^T\, FK$$