PAST PRESENT AND FUTURE CHALLENGES IN THE DETERMINATION OF THE STRUCTURE OF THE PROTON

LECTURE III

Maria Ubiali

University of Cambridge

Wrap up

- ➡ Monday's lecture about the foundation
  - ✓ Parametrisation of the proton in terms of structure functions
  - ✓ Parton model picture
  - ✓ QCD - Improved parton model
  - ✓ DGLAP evolution equations
  - ✓ Collinear Factorisation Theorem
- ➡ Tuesday's lecture mostly focussed on experimental data
  - ✓ Constraints for light quarks
  - ✓ Isospin and charge conjugation symmetries
  - ✓ The gluon PDFs

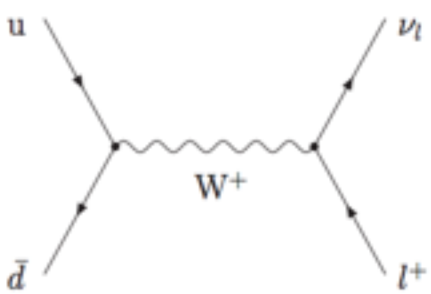
The name of the game

- Choose **experimental data** to fit and include all info on correlations ✅
- **Theory settings**: perturbative order, heavy quark mass scheme, EW corrections, intrinsic heavy quarks, α_s, quark masses value and scheme
- Choose a starting scale Q_0 where pQCD applies ✅
- **Parametrise** independent quarks and gluon distributions at the starting scale
- Solve **DGLAP equations** from initial scale to scales of experimental data and build up observables ✅
- **Fit** PDFs to data
- Provide PDF **error sets** to compute PDF uncertainties

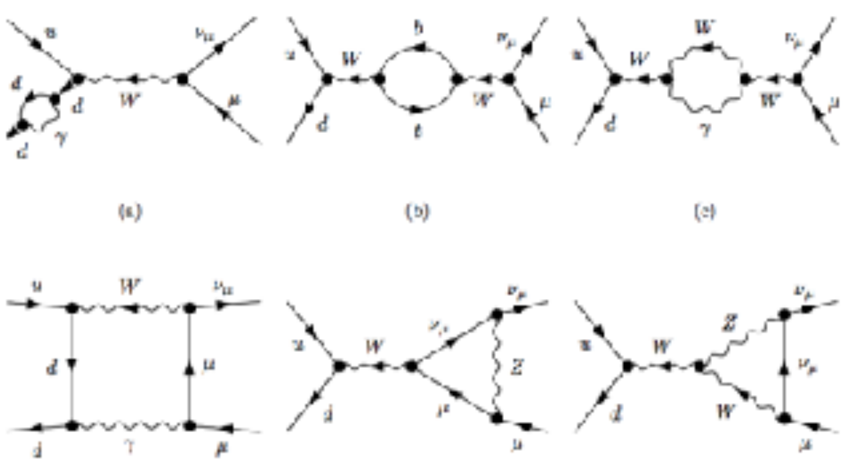
Photon and EW corrections

Electroweak corrections

- Because $\alpha(\mathrm{Mz}) \sim \alpha_S(\mathrm{Mz})/10 \Longrightarrow$ NLO EW corrections ~ NNLO QCD corrections



$$\mathcal{O}(\alpha^2)\,[1]$$



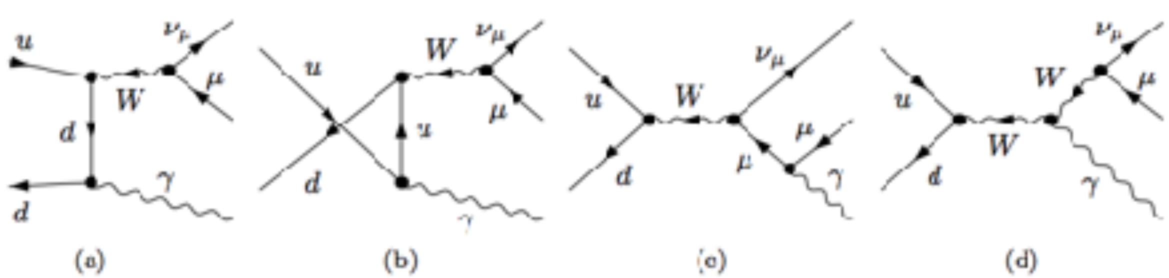
Virtual corrections

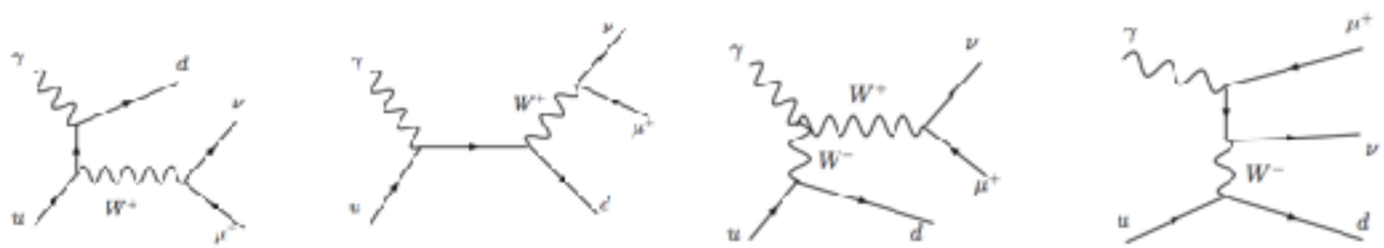
$$\mathcal{O}(\alpha^2)\,[1+\mathcal{O}(\alpha)]$$

$$+\frac{3\alpha}{\pi s_{\mathrm{W}}^2}\ln\left(\frac{s}{M_{\mathrm{W}}^2}\right)$$

At large s these logs can become large

Real corrections - quark initiated

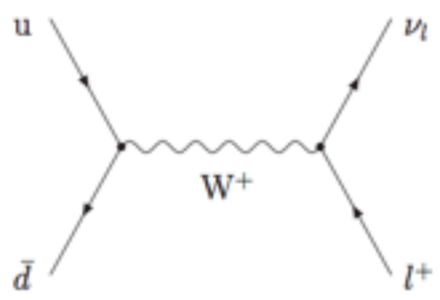




Real corrections - photon initiated

Electroweak corrections

- Because α(Mz) ~ α_S(Mz)/10 $\Longrightarrow$ NLO EW corrections ~ NNLO QCD corrections



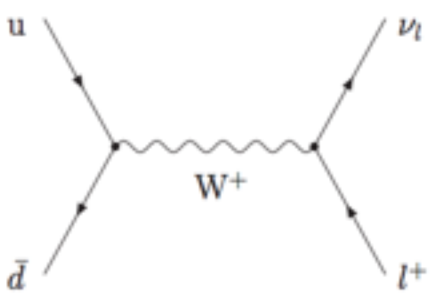
- NLO EW corrections become large in the large pT region of lepton but partially compensated by photon-initiated real corrections

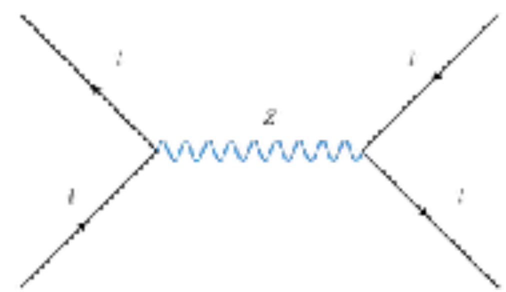
$p_{T,l}$/GeV	25–∞	50–∞	100–∞	200–∞	500–∞	1000–∞
$\delta_{e^+\nu_e}/\%$	−5.19(1)	−8.92(3)	−11.47(2)	−16.01(2)	−26.35(1)	−37.92(1)
$\delta_{\mu^+\nu_\mu}/\%$	−2.75(1)	−4.78(3)	−8.19(2)	−12.71(2)	−22.64(1)	−33.54(2)
$\delta_{rec}/\%$	−1.73(1)	−2.45(3)	−5.91(2)	−9.99(2)	−18.95(1)	−28.60(1)
$\delta_{\gamma q}/\%$	+0.071(1)	+5.24(1)	+13.10(1)	+16.44(2)	+14.30(1)	+11.89(1)

Dittmaier, Krämer

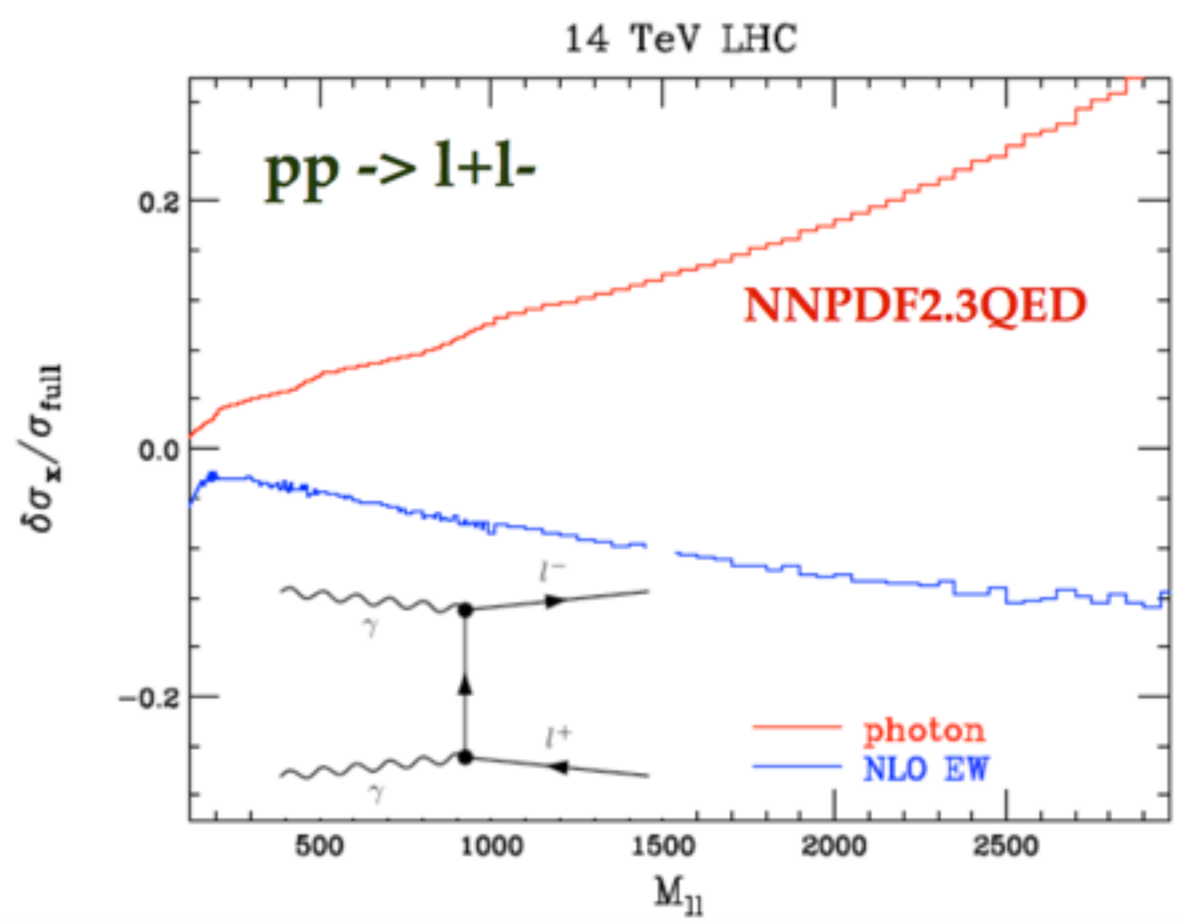
Electroweak corrections

- Because $\alpha(\mathrm{Mz}) \sim \alpha_S(\mathrm{Mz})/10 \Longrightarrow$ NLO EW corrections ~ NNLO QCD corrections





- NLO EW corrections become large in the large pT region of lepton but partially compensated by photon-initiated real corrections



Boughezal et al Phys.Rev. D89 (2014)3, 034030

Modified DGLAP

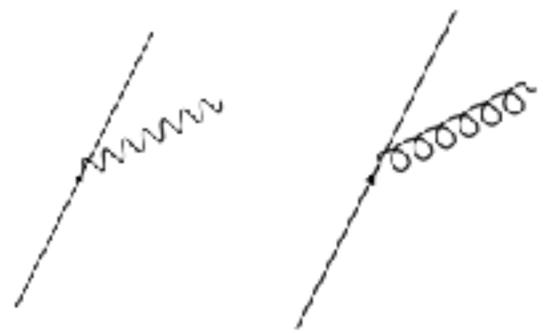
- How are PDFs modified by inclusion of initial photon PDF?

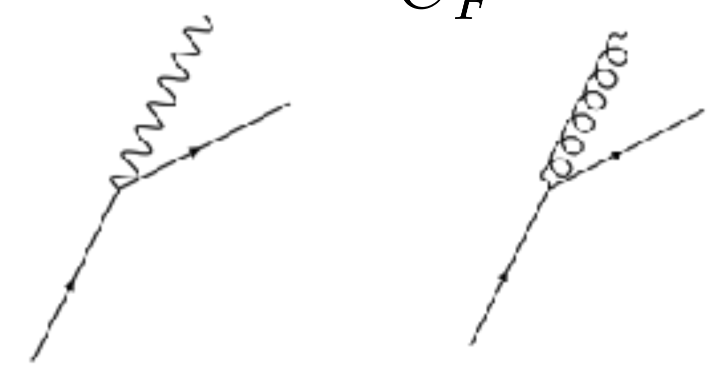
$$\boxed{\begin{aligned}
Q^2\frac{\partial}{\partial Q^2}g(x,Q^2) &= \sum_{q,\bar{q},g} P_{ga}(x,\alpha_s(Q^2))\otimes f_a(x,Q^2) + P_{g\gamma}(x,\alpha_s(Q^2))\otimes\gamma(x,Q^2),\\
Q^2\frac{\partial}{\partial Q^2}q(x,Q^2) &= \sum_{q,\bar{q},g} P_{qa}(x,\alpha_s(Q^2))\otimes f_a(x,Q^2) + P_{q\gamma}(x,\alpha_s(Q^2))\otimes\gamma(x,Q^2),\\
Q^2\frac{\partial}{\partial Q^2}\gamma(x,Q^2) &= P_{\gamma\gamma}\otimes\gamma(x,Q^2) + \sum_{q,\bar{q},g} P_{\gamma a}(x,\alpha_s(Q^2))\otimes f_a(x,Q^2).
\end{aligned}}$$

- DGLAP splitting functions expanded in powers of α_s and $\boldsymbol{\alpha}$

$$P_{ij} = \sum_{m,n}\left(\frac{\alpha_S}{2\pi}\right)^m\left(\frac{\alpha}{2\pi}\right)^n P_{ij}^{(m,n)}$$

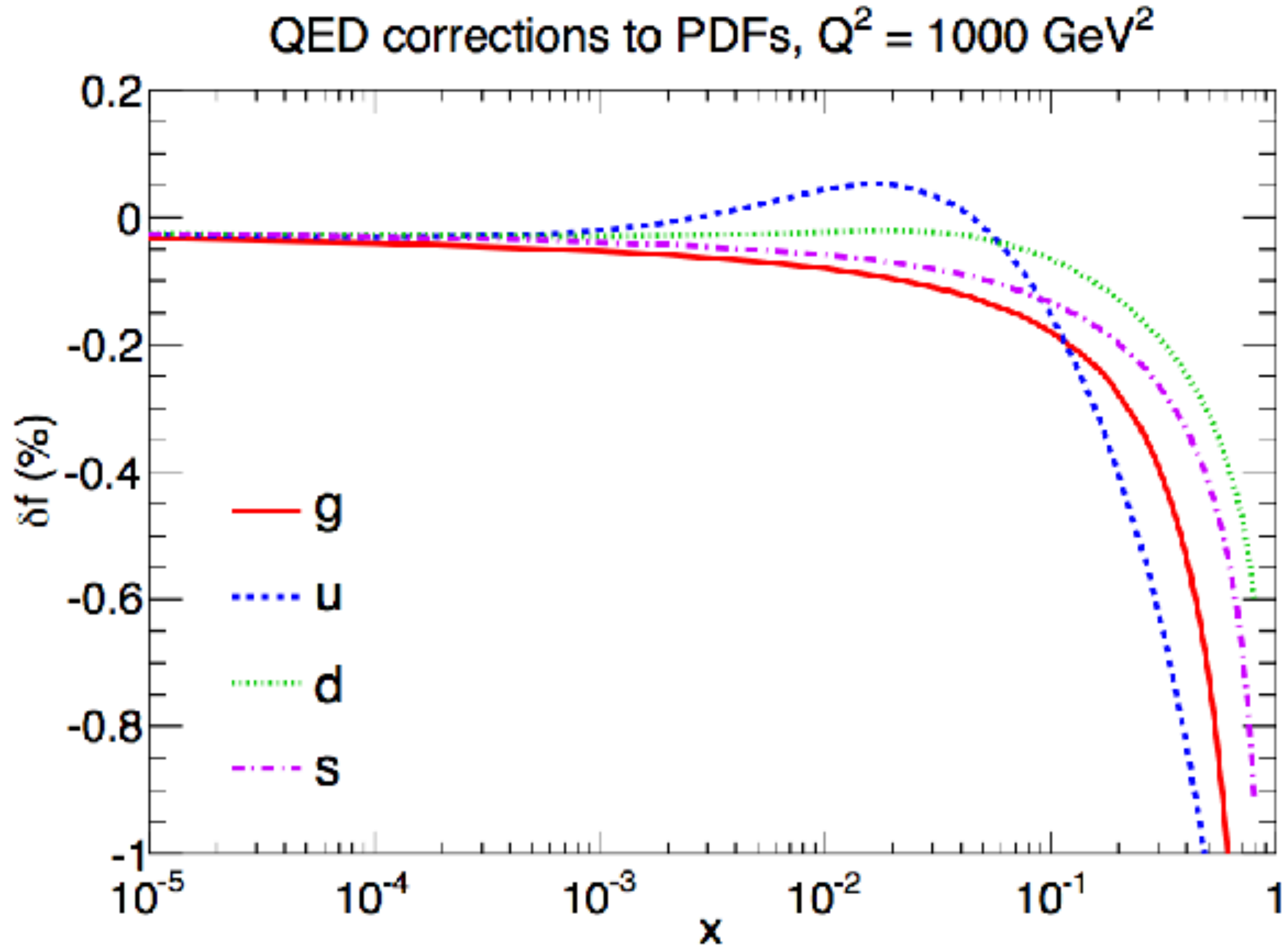
$$P_{qq}^{(0,1)} = \frac{e_q^2}{C_F}P_{qq}^{(1,0)} \qquad P_{q\gamma}^{(0,1)} = \frac{e_q^2}{T_R}P_{qg}^{(1,0)} \qquad P_{\gamma q}^{(0,1)} = \frac{e_q^2}{C_F}P_{gq}^{(1,0)}$$





Modified DGLAP

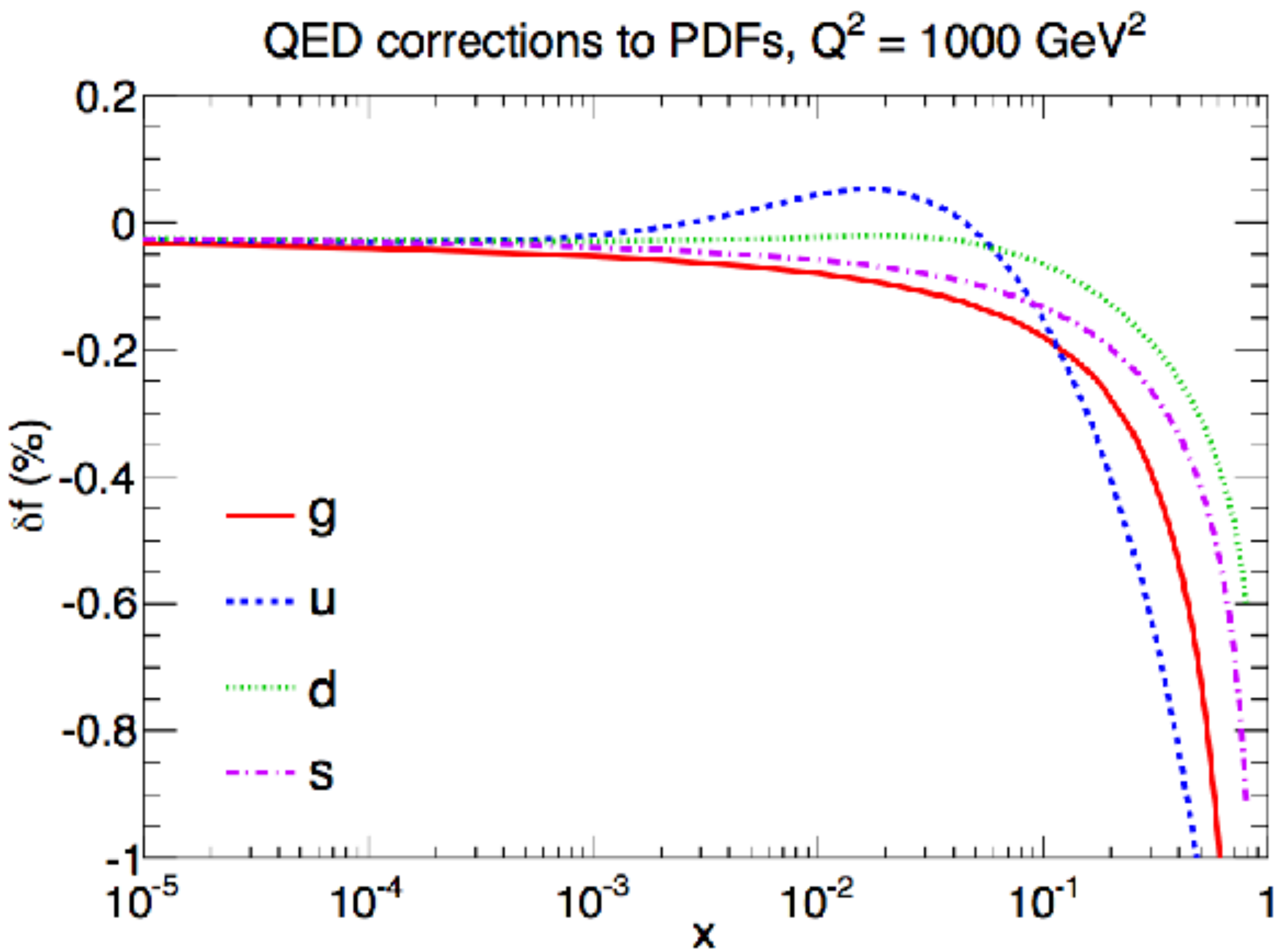
- Quark and gluon PDFs change up to 1% at large x



Ball et al, Nucl.Phys. B877 (2013) 290-320

Modified DGLAP

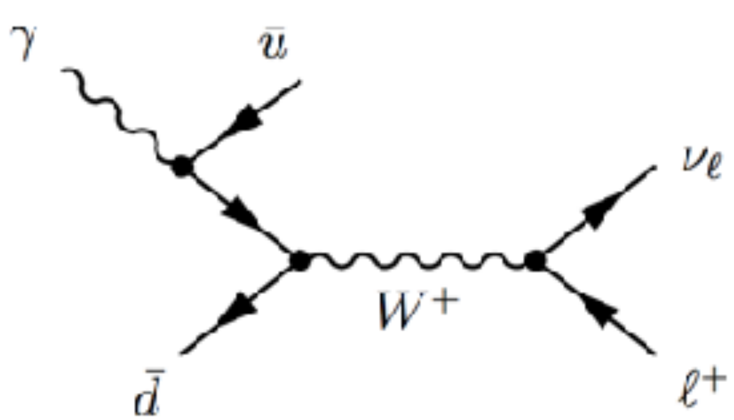
- Quark and gluon PDFs change up to 1% at large x
- How do we determine the photon PDF?
- Two ways in the next slides: from data or from theory
- In the best possible world: theory input and data input together

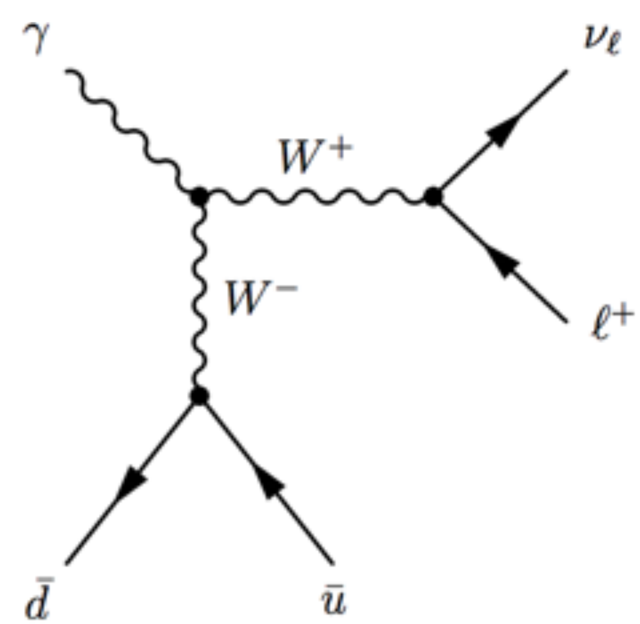


Ball et al, Nucl.Phys. B877 (2013) 290-320

Photon PDFs

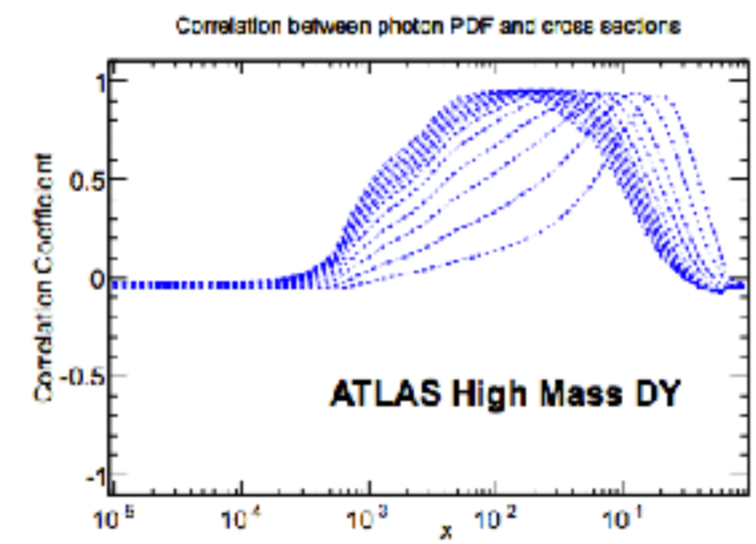
- Largest correlations between photon PDFs and pp cross sections are for Drell-Yan processes, but also for top pair production and VV production

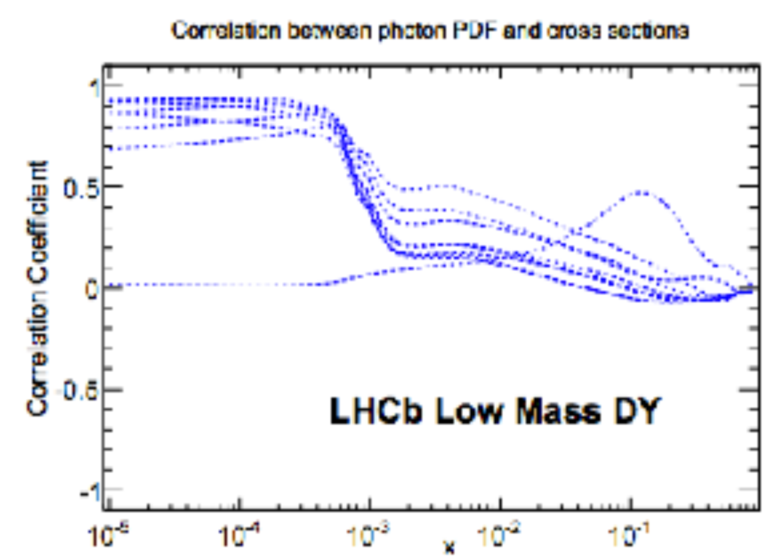




Photon-induced Drell-Yan

$Q = 100\,\mathrm{GeV}$

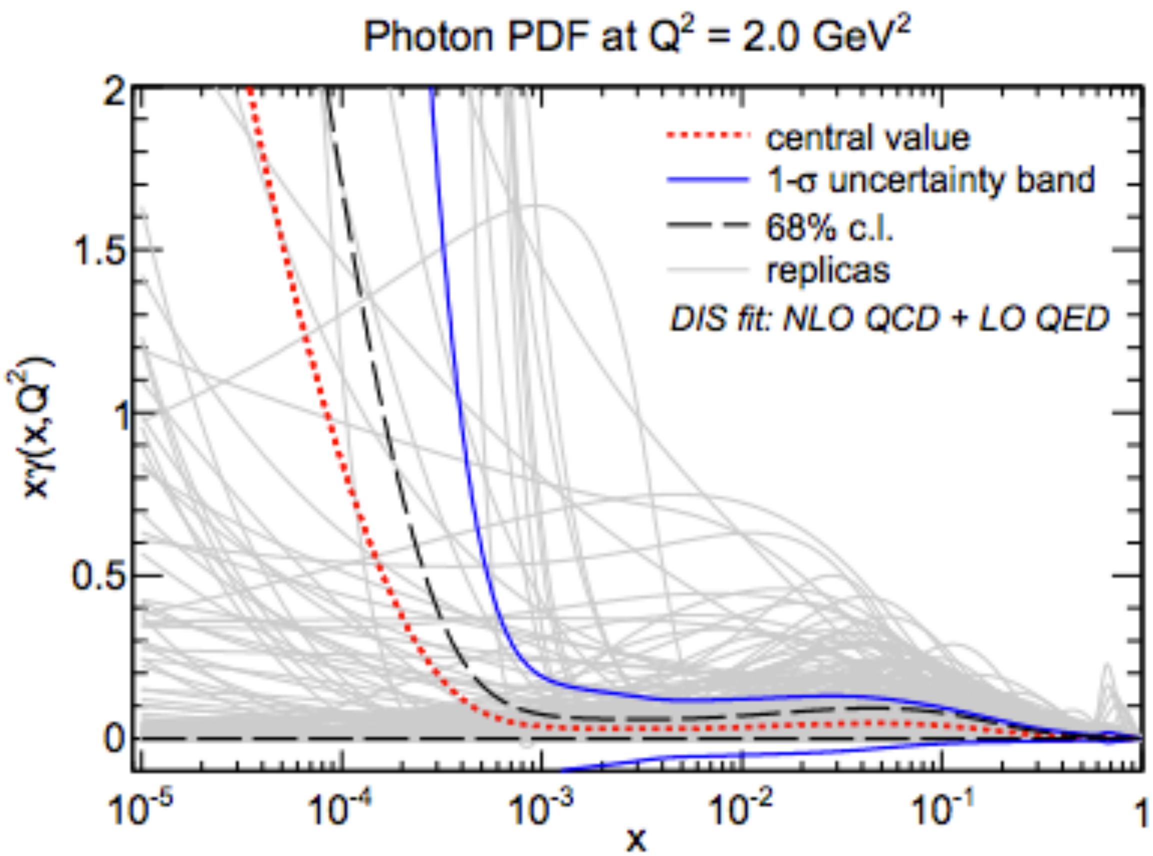


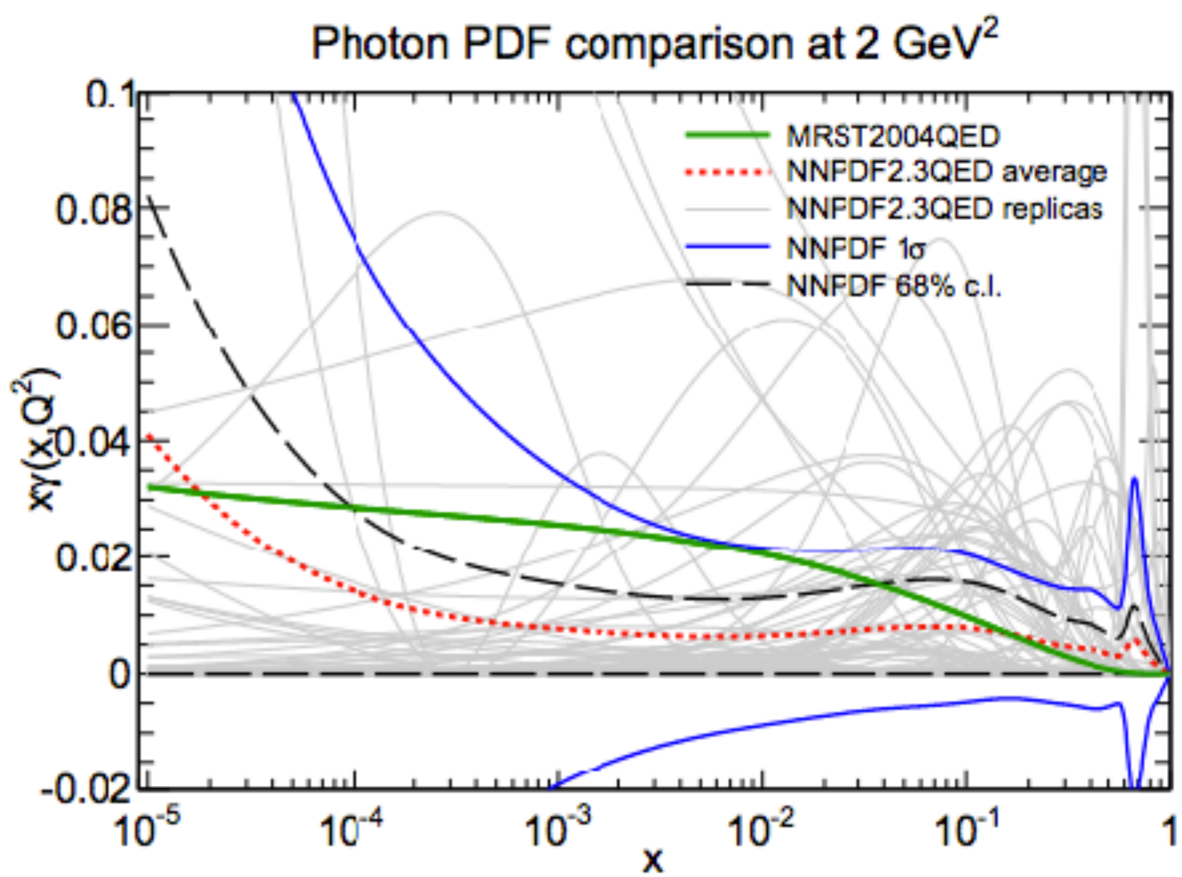


Ball et al, Nucl.Phys. B877 (2013) 290-320

Photon PDFs

Data-driven knowledge

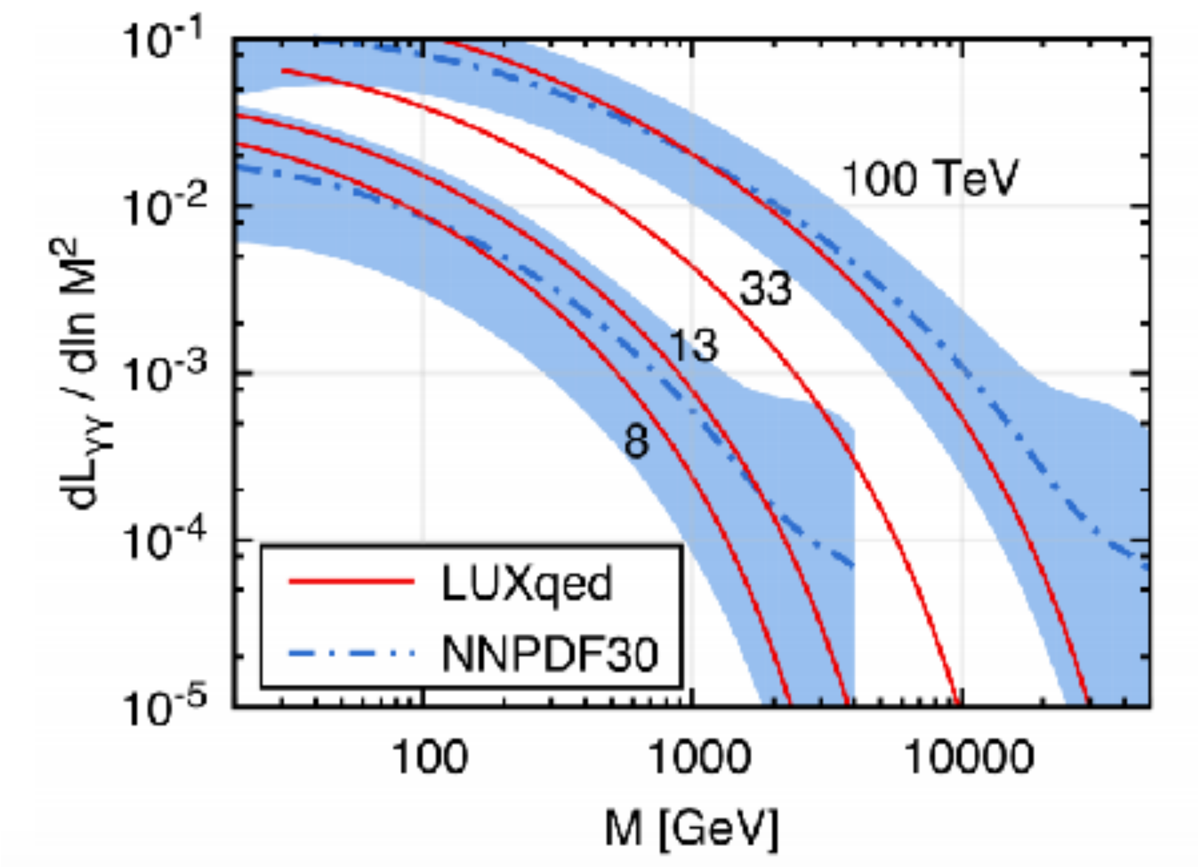


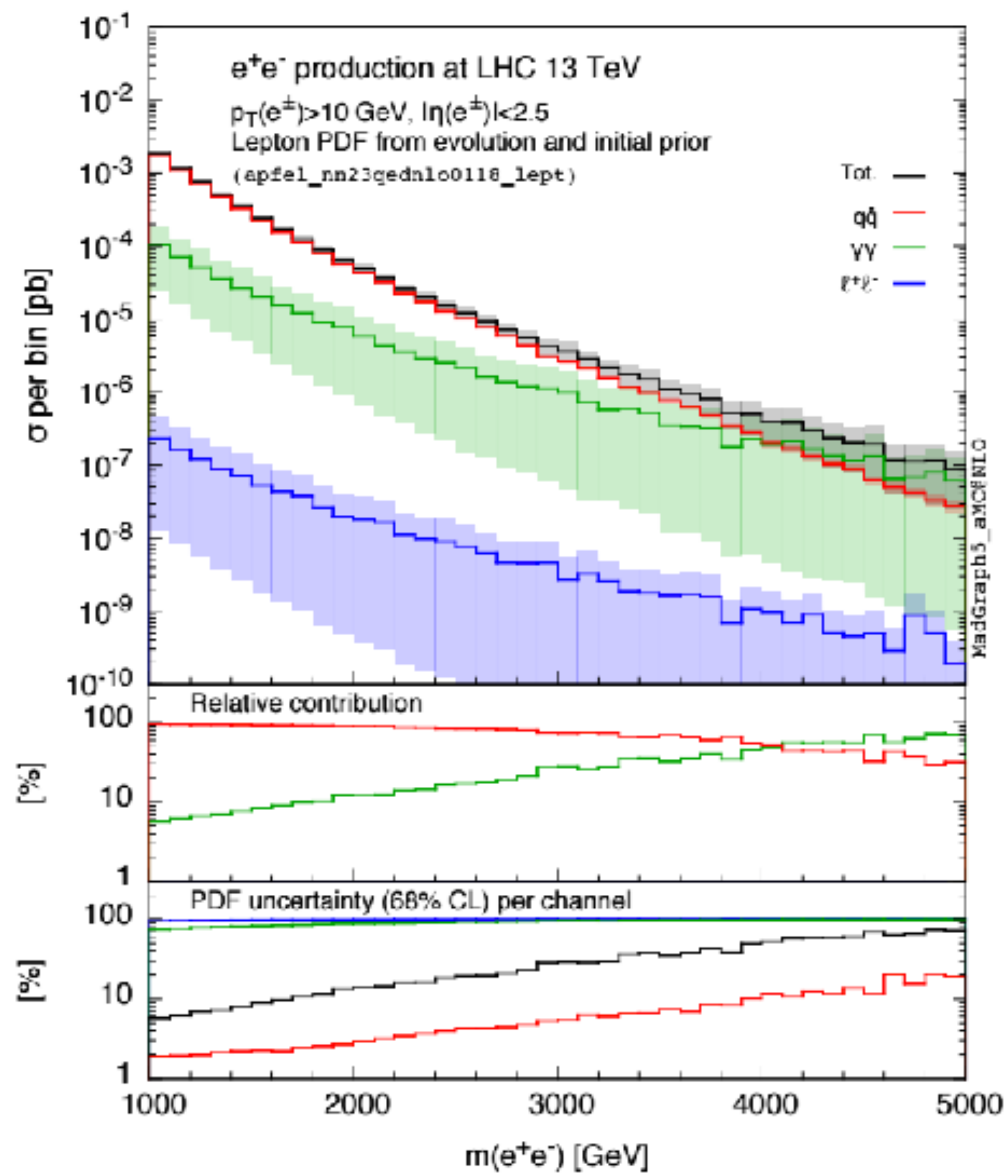


Ball et al, Nucl.Phys. B877 (2013) 290-320

Photon PDFs

- Data-driven approach associated with a large uncertainty on photon PDF
- Theory breakthrough: LUX PDF [Manohar, Nason, Salam, Zanderighi, 1607.04266]





Bertone et al, 1508.07002

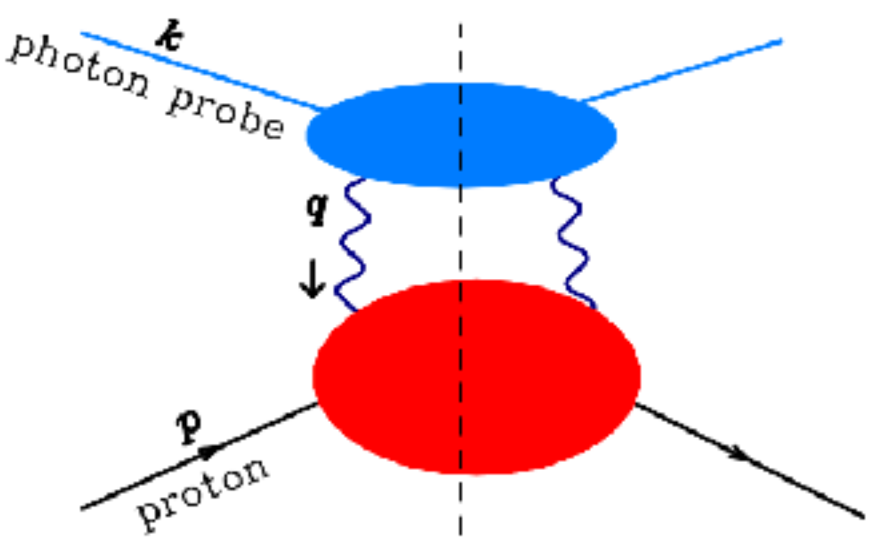
Photon PDFs

- QED is perturbative down to low scales $\Longrightarrow$ The photon must be computable is the input mark substructure is known
- Manohar et al: write the cross section for a chosen BSM process, e.g. production of heavy supersymmetric lepton L in ep collision (Drees, Zeppenfeld 1989)

$$\sigma = \frac{1}{4p\cdot k}\int \frac{d^4q}{(2\pi)^4 q^4} e^2_{\rm ph}(q^2)\left[4\pi W_{\mu\nu}(p,q)\, L^{\mu\nu}(k,q)\right]\; 2\pi\delta((k-q)^2 - M^2)$$

$$\sigma = c_0 \sum_a \int_x^1 \frac{dz}{z}\,\hat{\sigma}_a(z,\mu^2)\,\frac{M^2}{zs} f_{a/p}\left(\frac{M^2}{zs},\mu^2\right)$$

$$l(k) + p(p) \to L(k') + X$$



$$\sigma = \frac{c_0}{2\pi}\int_x^{1-\frac{2xm_p}{M}} \frac{dz}{z}\int_{Q^2_{\rm min}}^{Q^2_{\rm max}} \frac{dQ^2}{Q^2}\alpha^2_{\rm ph}(-Q^2)\Bigg[\Bigg(2-2z+z^2 + \frac{2x^2m_p^2}{Q^2} + \frac{z^2Q^2}{M^2} - \frac{2zQ^2}{M^2} - \frac{2x^2Q^2m_p^2}{M^4}\Bigg)F_2(x/z,Q^2) + \left(-z^2 - \frac{z^2Q^2}{2M^2} + \frac{z^2Q^4}{2M^4}\right)F_L(x/z,Q^2)\Bigg], \quad (3)$$

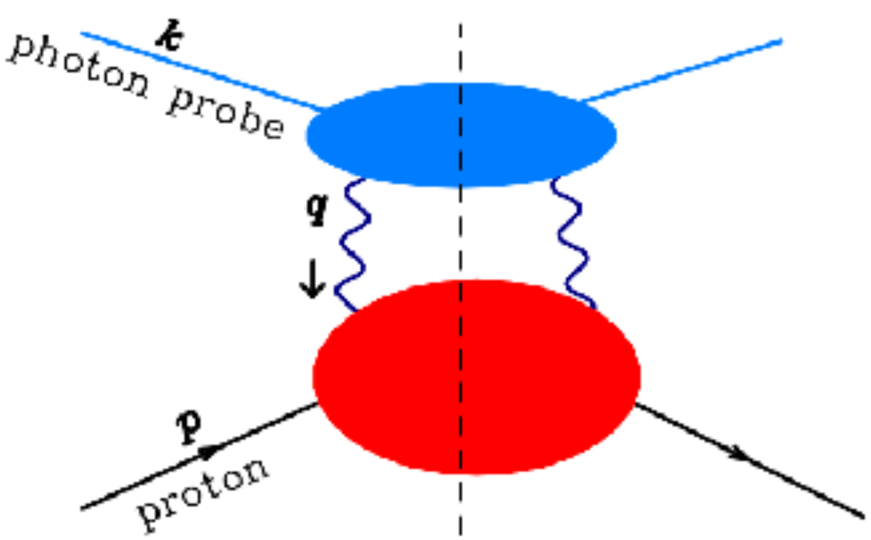
Manohar et al 1607.04266

<u>Theory-driven knowledge</u>

Photon PDFs

- QED is perturbative down to low scales $\Longrightarrow$ The photon must be computable is the input mark substructure is known
- Manohar et al: write the cross section for a chosen BSM process, e.g. production of heavy supersymmetric lepton L in ep collision (Drees, Zeppenfeld 1989)
- Equate the two expressions and find analytically the PDF of the photon

⇒ PDFs expressed in terms of the structure functions integrated over all scales, including elastic form factors (in the x →1 region)

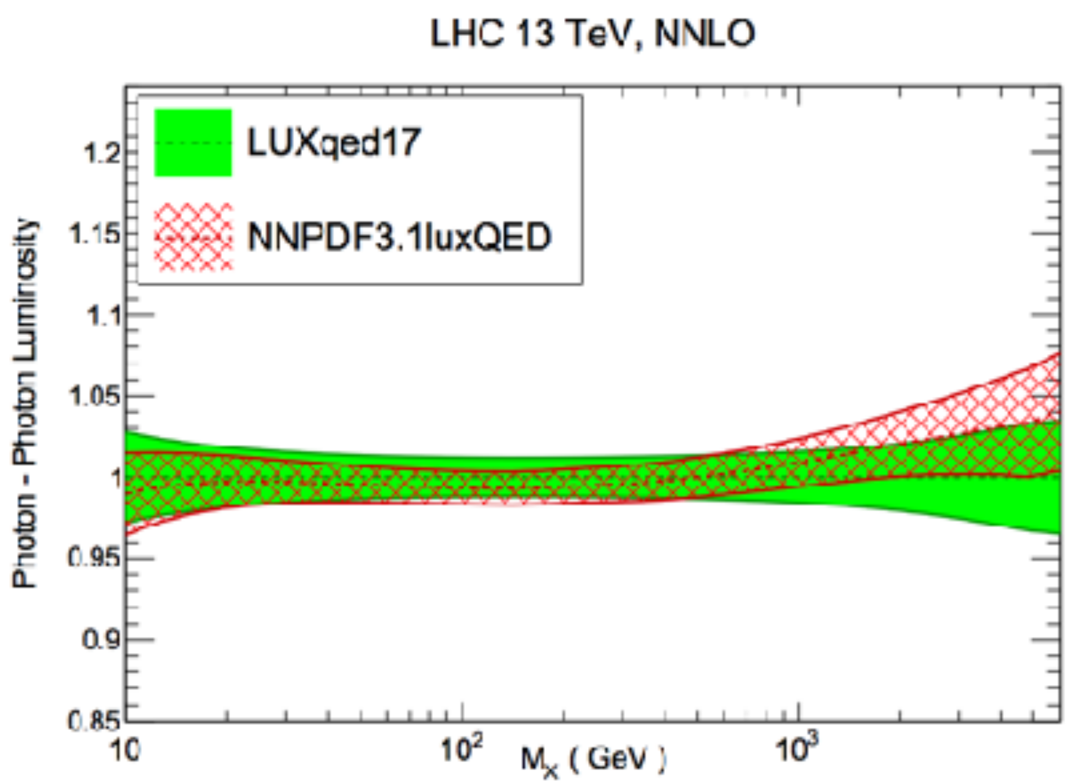


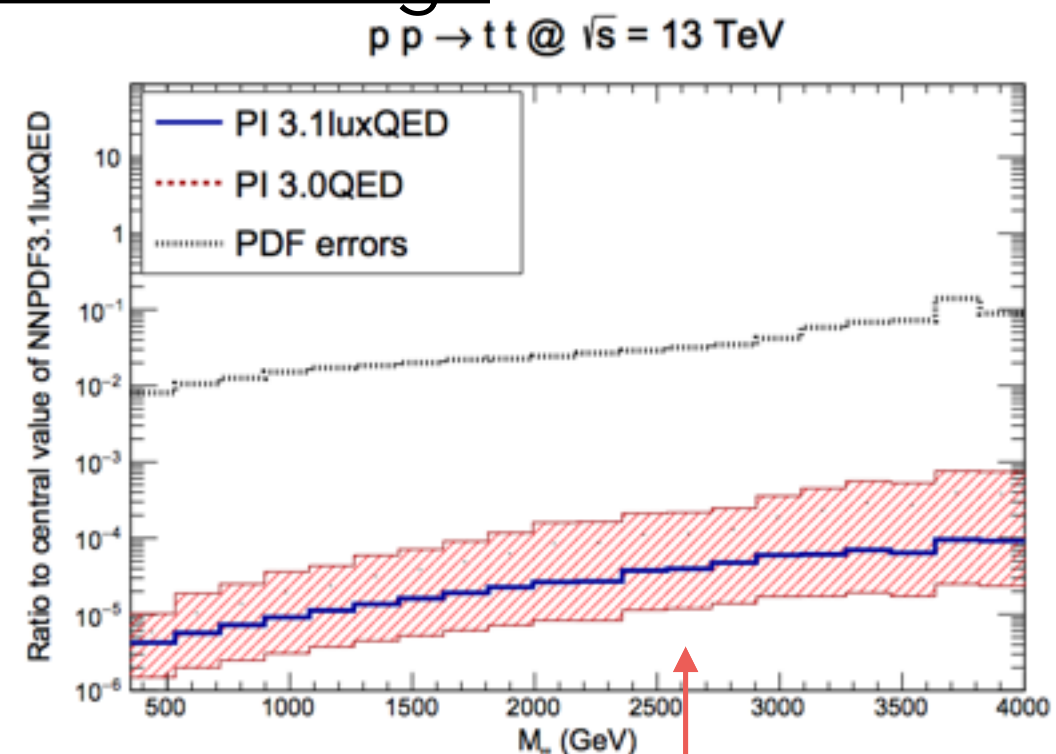
$$
x f_{\gamma/p}(x,\mu^2) = \frac{1}{2\pi\alpha(\mu^2)} \int_x^1 \frac{dz}{z} \left\{ \int_{\frac{x^2 m_p^2}{1-z}}^{\frac{\mu^2}{1-z}} \frac{dQ^2}{Q^2} \alpha^2(Q^2) \left[\left(z p_{\gamma q}(z) + \frac{2x^2 m_p^2}{Q^2} \right) F_2(x/z, Q^2) - z^2 F_L\left(\tfrac{x}{z}, Q^2\right) \right] - \alpha^2(\mu^2) z^2 F_2\left(\tfrac{x}{z}, \mu^2\right) \right\},
$$

<u>Theory-driven knowledge</u>

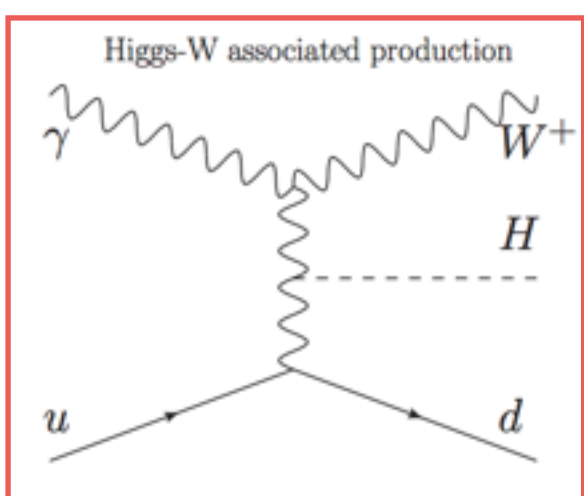
Photon PDFs

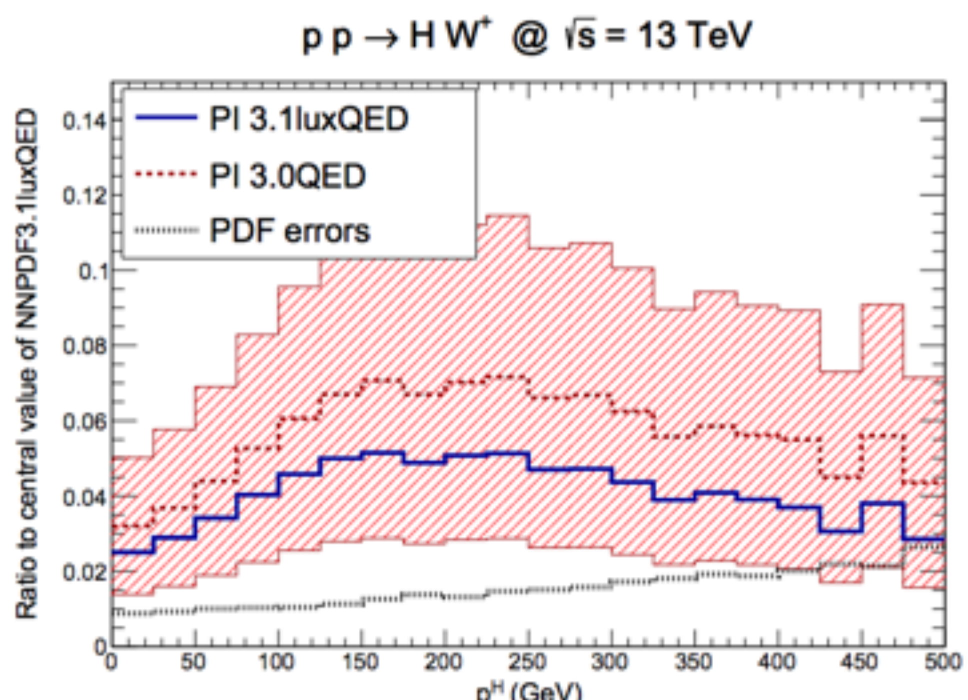
(Data+Theory)-driven knowledge

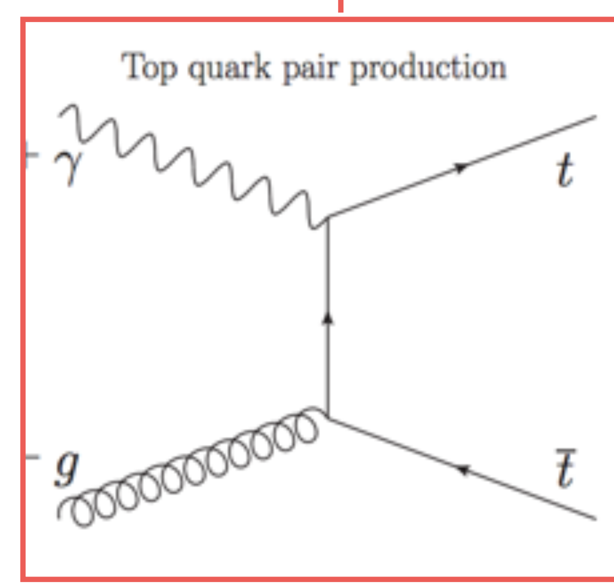




Bertone et al, 1712.07053







What happens at very high energy?

- What happens at scales much above the EWK scale (100 GeV)? SU(3) x SU(2) x U(1) unbroken!

$$f_i(x,\mu) = x\int \frac{dy}{2\pi}\, e^{-i\,2x\bar{n}\cdot p y}\langle p|\, \bar{\psi}^{(i)}(y)\, \bar{n}\!\!\!/\, \psi^{(i)}(-y)|p\rangle\,,$$

$$f_{\bar{i}}(x,\mu) = x\int \frac{dy}{2\pi}\, e^{-i\,2x\bar{n}\cdot p y}\langle p|\, \psi^{(i)}(y)\, \bar{n}\!\!\!/\, \bar{\psi}^{(i)}(-y)|p\rangle\,,$$

- 42: 8 quark PDFs and 4 lepton PDFs for each generation
- 1 gluon PDF

$$\begin{pmatrix} f_\gamma \\ f_Z \\ f_{\gamma Z}\end{pmatrix} = \begin{pmatrix} c_W^2 & s_W^2 & c_W s_W \\ s_W^2 & c_W^2 & -c_W s_W \\ -2c_W s_W & 2c_W s_W & c_W^2 - s_W^2 \end{pmatrix}\begin{pmatrix} f_B \\ f_{W_3} \\ f_{BW}\end{pmatrix}$$

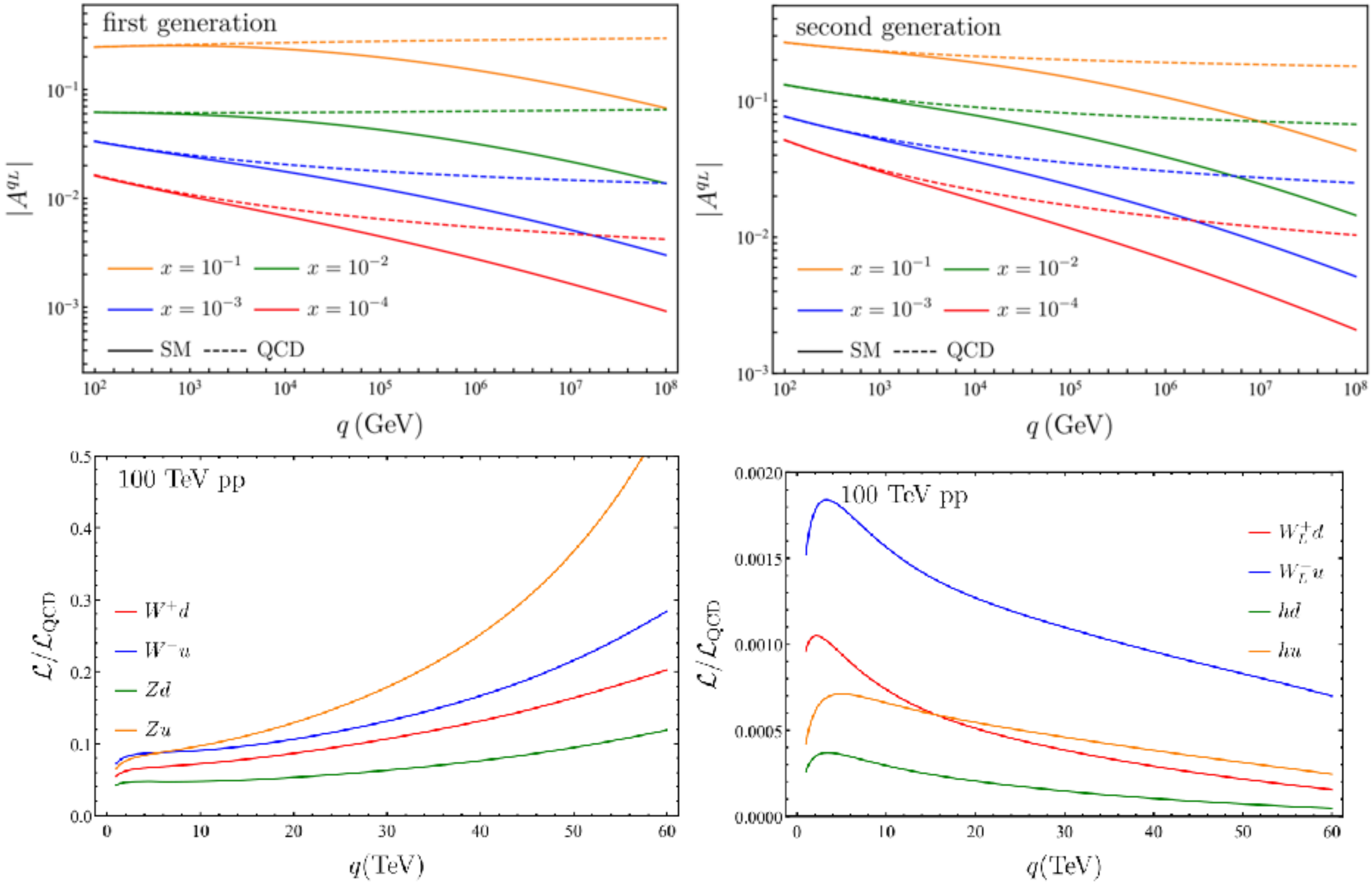
3 V PDFs (mixed BW)

$$f_H(x) = x\int \frac{dy}{2\pi}\, e^{-i\,2x\bar{n}\cdot p y}\,\langle p|\,\Phi(y)\Phi(-y)|p\rangle$$

4 PDFs for the Higgs

Bauer et al,1703.08562

What happens at very high energy?

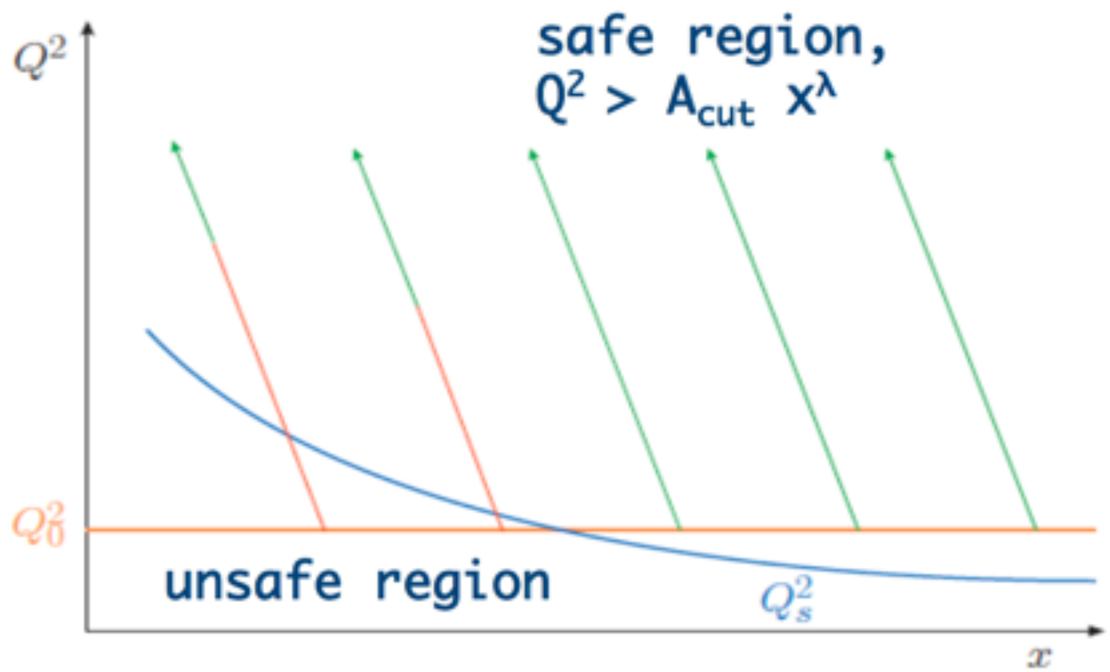


Bauer et al, 1703.08562

Beyond Collinear Factorisation

Beyond DGLAP

- In DGLAP formalism there is an implicit approximation: the transverse momentum of the emitted partons in the initial state is much smaller than hard scale
- It works well for inclusive processes with one hard scale and for not-too-small x



Possible effects beyond DGLAP

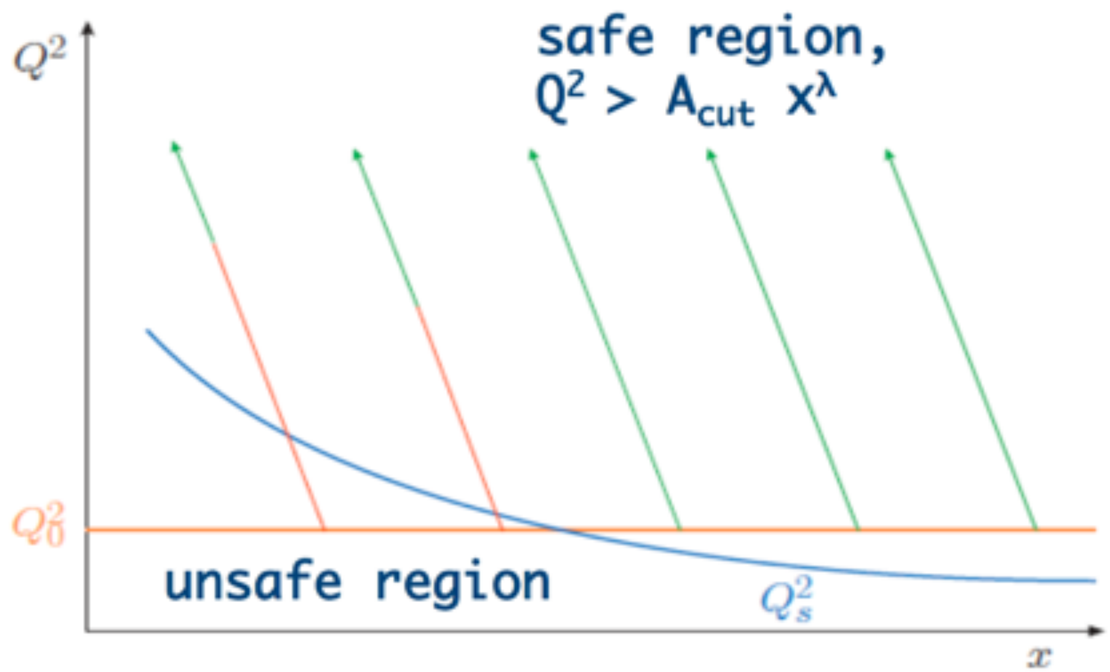
(i) Leading-twist small-x perturbative resummation

(ii) Non-linear evolution and saturation

(iii) Higher twist effects

Beyond DGLAP

- In DGLAP formalism there is an implicit approximation: the transverse momentum of the emitted partons in the initial state is much smaller than hard scale
- It works well for inclusive processes with one hard scale and for not-too-small x



Possible effects beyond DGLAP

➜ (i) Leading-twist small-x perturbative resummation

➜ (ii) Non-linear evolution and saturation

(iii) Higher twist effects

(i) Small-x resummation

- In DGLAP formalism there is an implicit approximation: the transverse momentum of the emitted partons in the initial state is much smaller than hard scale
- It works well for inclusive processes with one hard scale and for not-too-small x
- If s >> M^2 (the high-energy limit or small-x limit) then there are enhanced small-x logarithms in the DGLAP Pqg and Pgg splitting functions that spoil the perturbative expansion in α_s
- These large logs are resumed by BFKL evolution equations

DGLAP Evolution in Q^2

$$\frac{\partial}{\partial \ln Q^2} f_i(x, Q^2) = \int_x^1 \frac{dz}{z} P_{ij}\left(\frac{x}{z}, \alpha_s(Q^2)\right) f_j(z, Q^2)$$

BFKL Evolution in x

$$\frac{\partial}{\partial \ln 1/x} f_+(x, Q^2) = \int_0^\infty \frac{d\nu^2}{\nu^2} K\left(\frac{Q^2}{\nu^2}, \alpha_s(Q^2)\right) f_+(x, \nu^2)$$

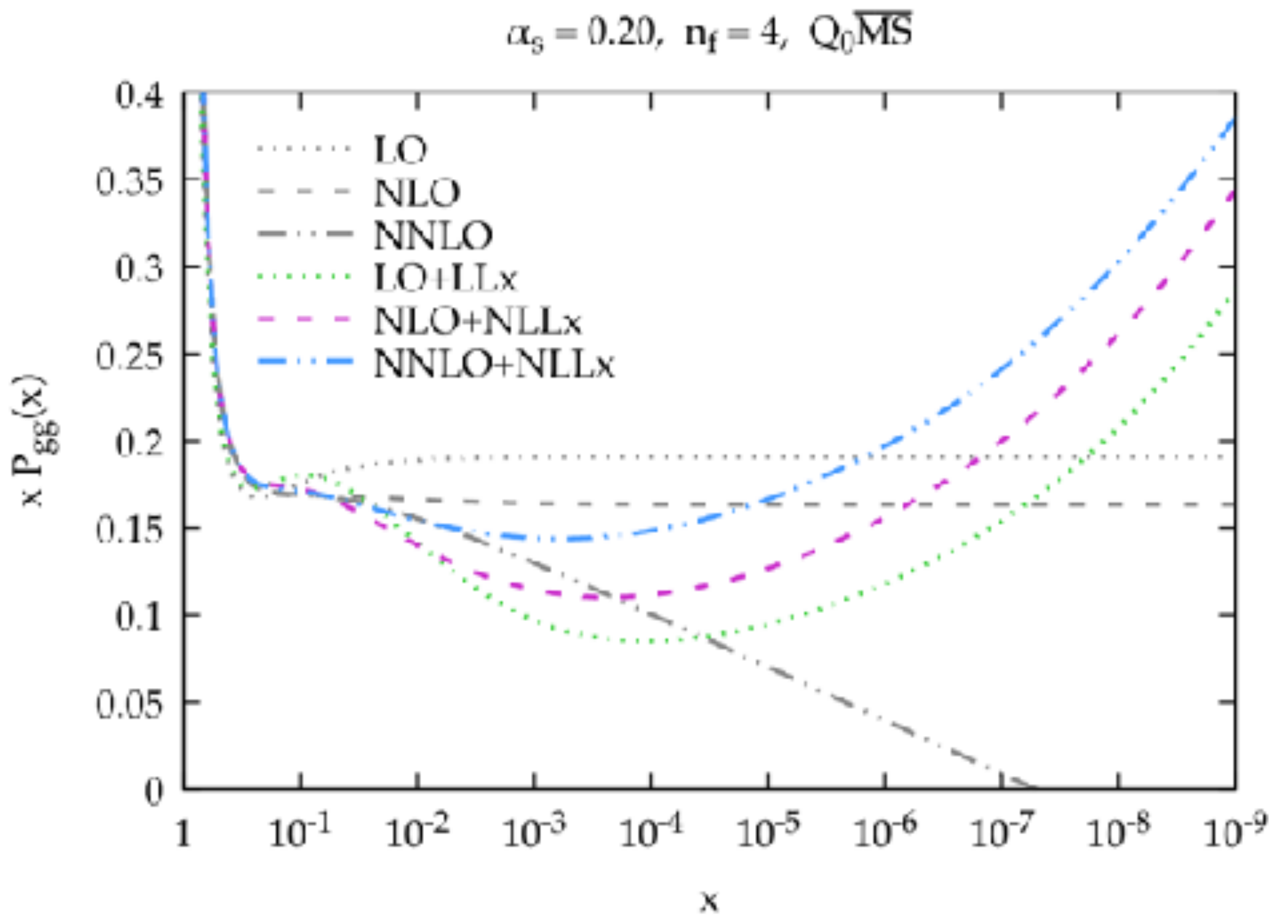
Valid only at small-x

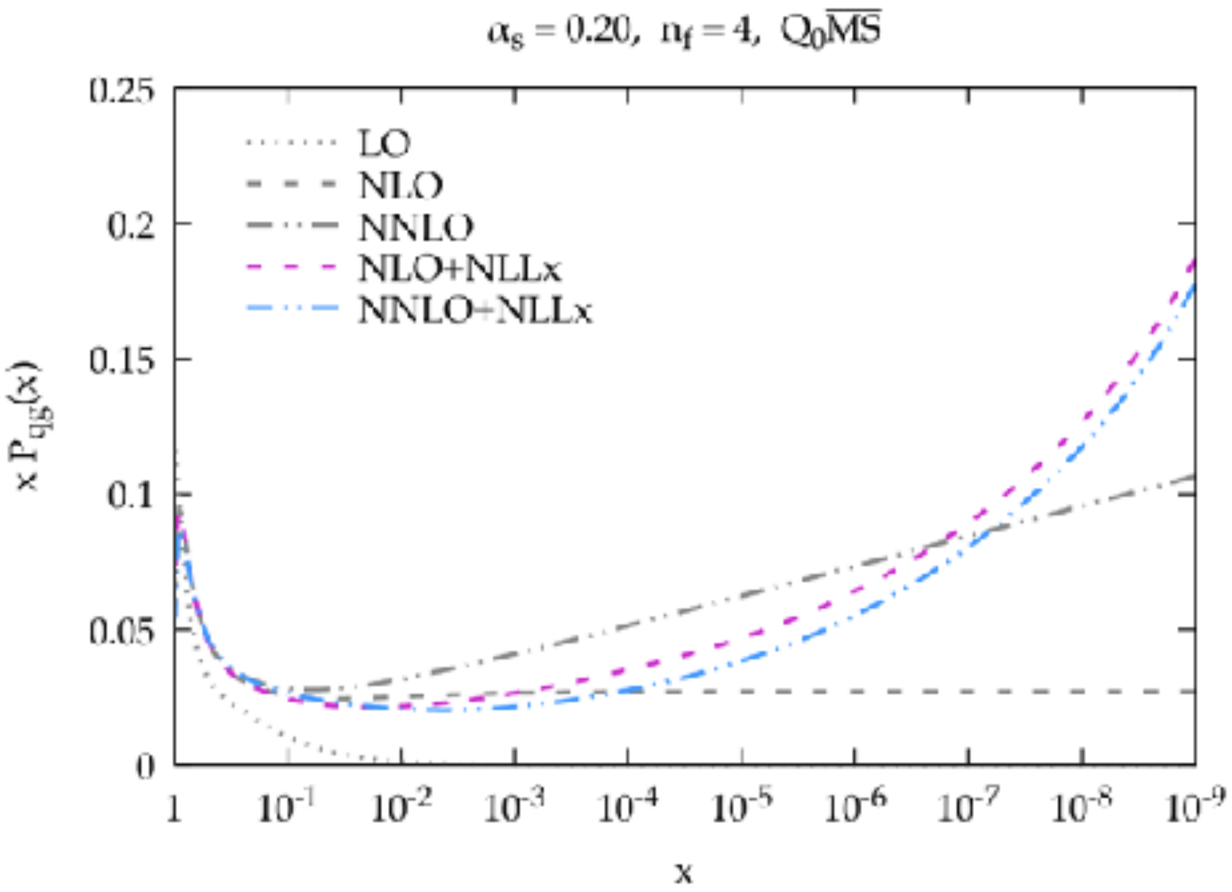
- There are way of combining DGLAP & BFKL - what are the effects?

Balitsky Fadin Kuraev Lipatov 1978, Lipatov Fadin 1998

(i) Small-x resummation

- Small-x resummation stabilises splitting function behaviour at small x

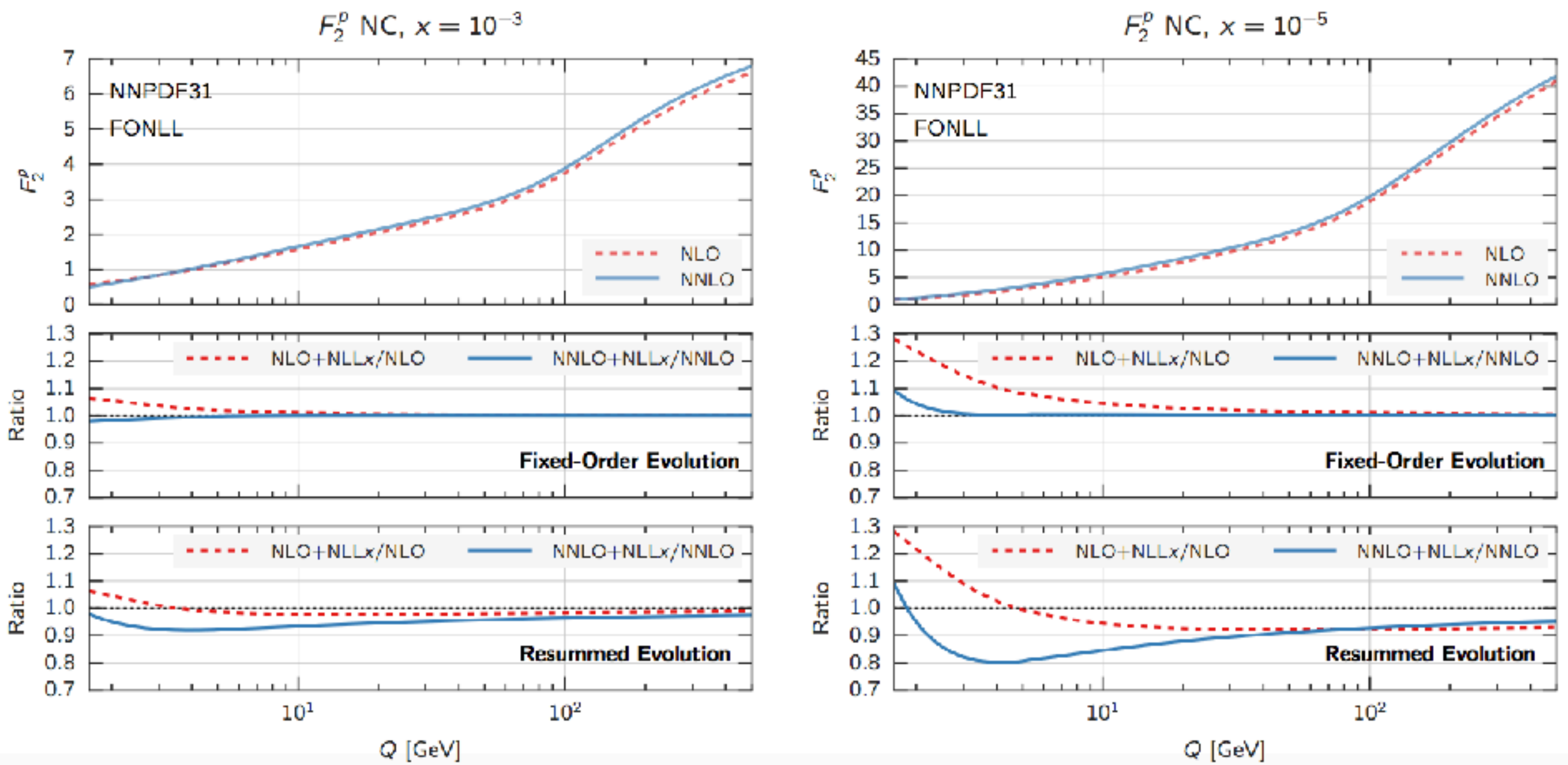




Ball et al, 1710.05935

(i) Small-x resummation

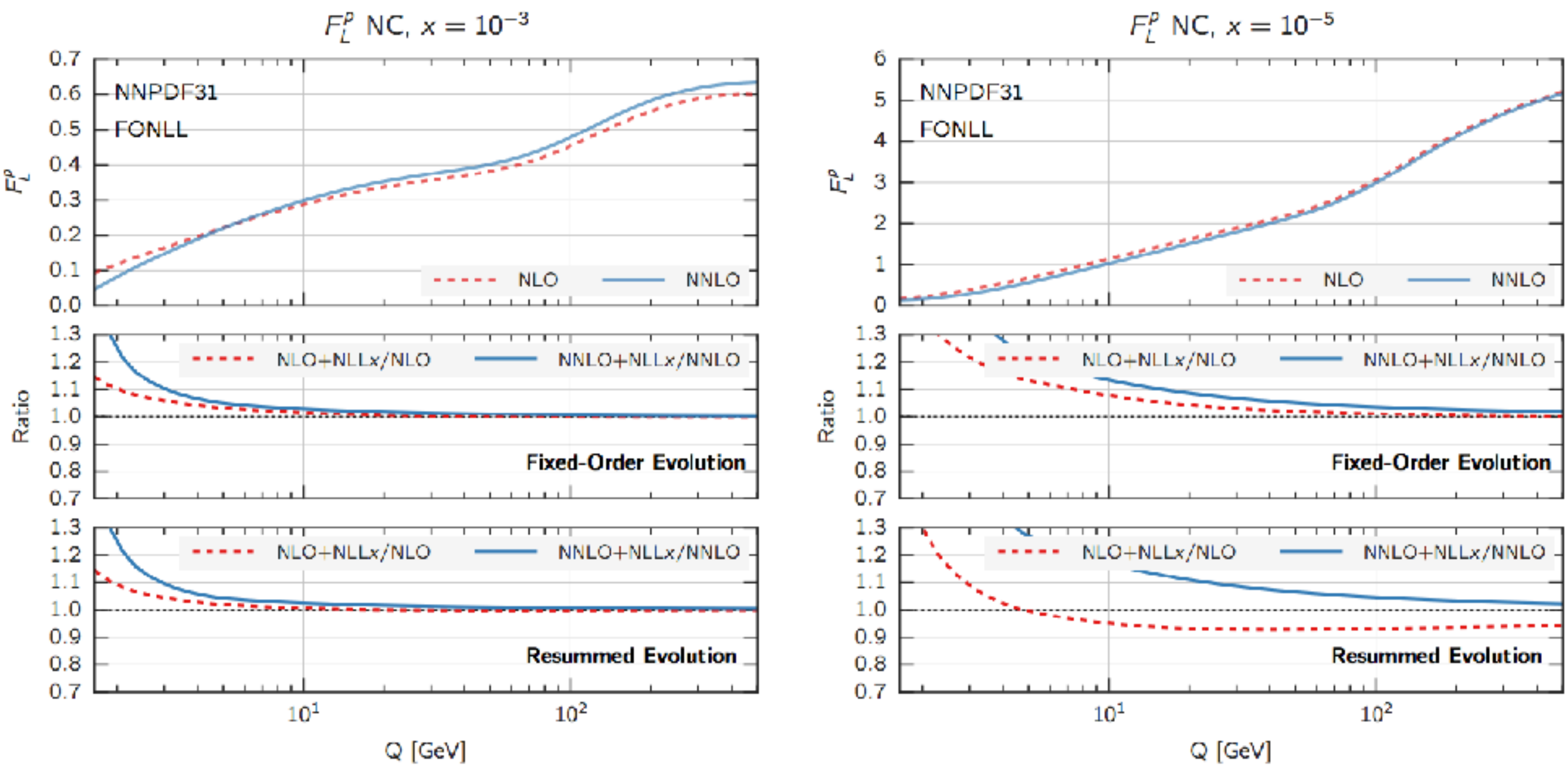
- Large corrections at small Q2 and small-x



Ball et al, 1710.05935

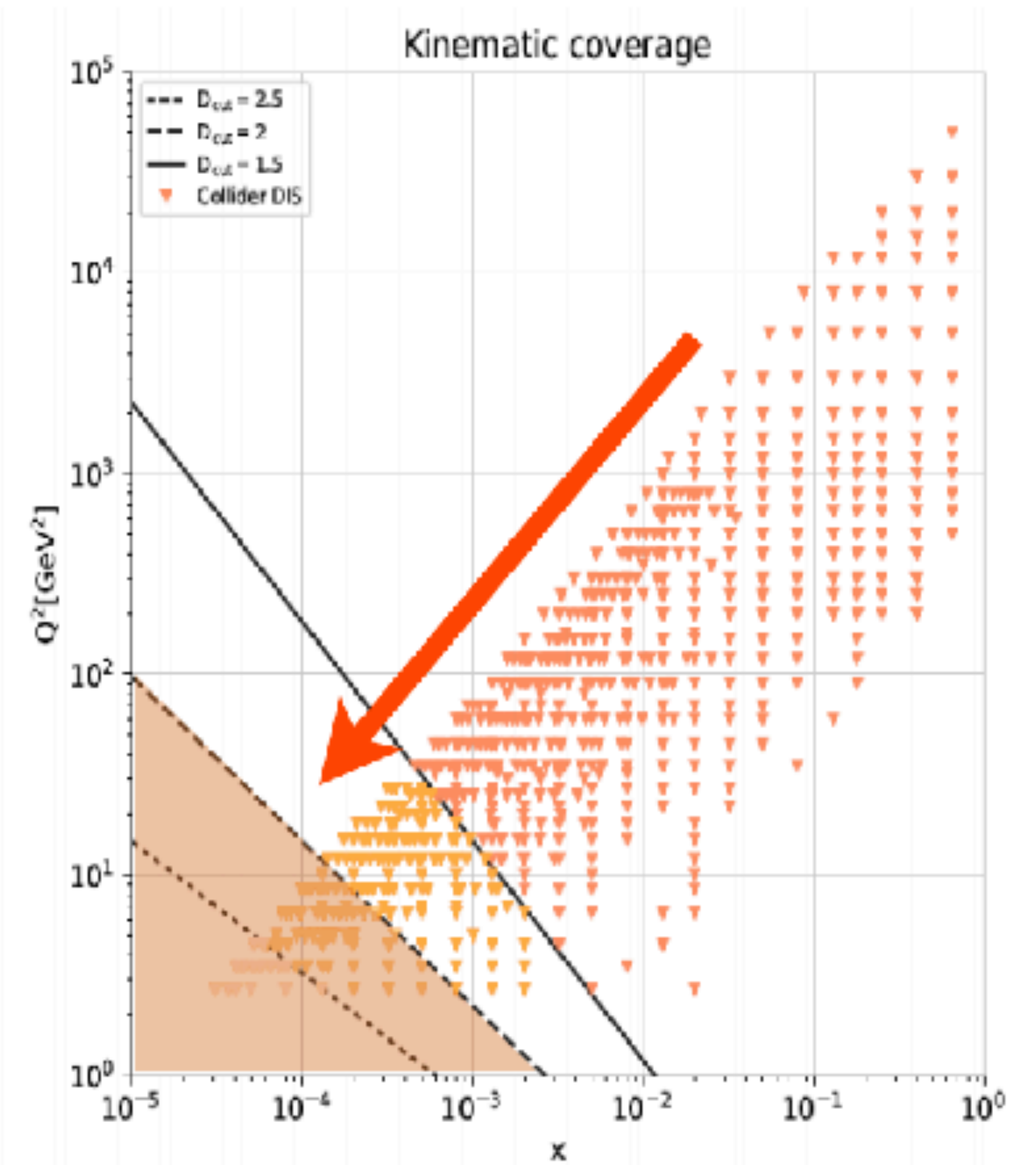
(i) Small-x resummation

- Large corrections at small Q2 and small-x, especially for FL

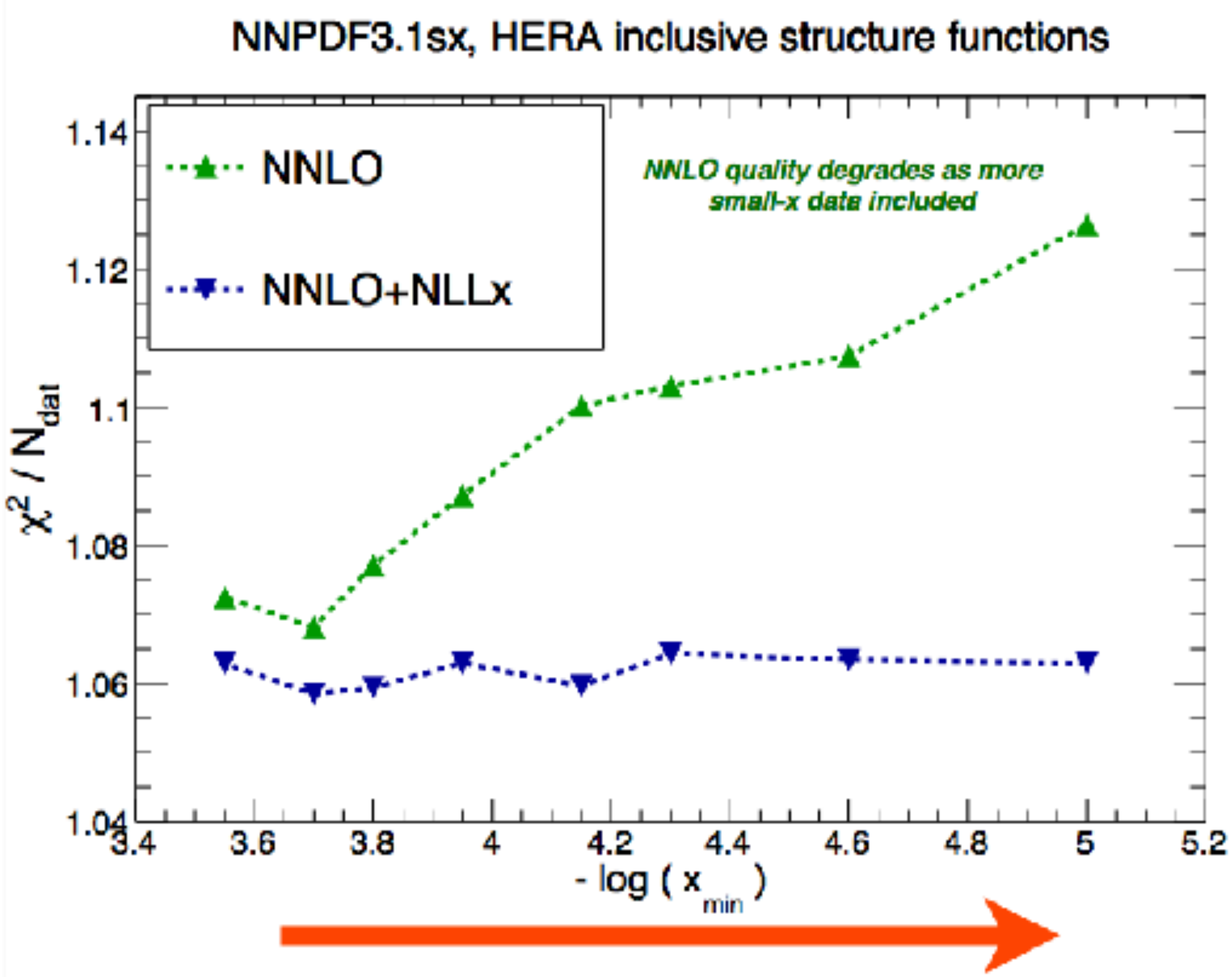


Ball et al, 1710.05935

(i) Small-x resummation



Monitor the **fit quality** as one includes more data from the **small-x region**



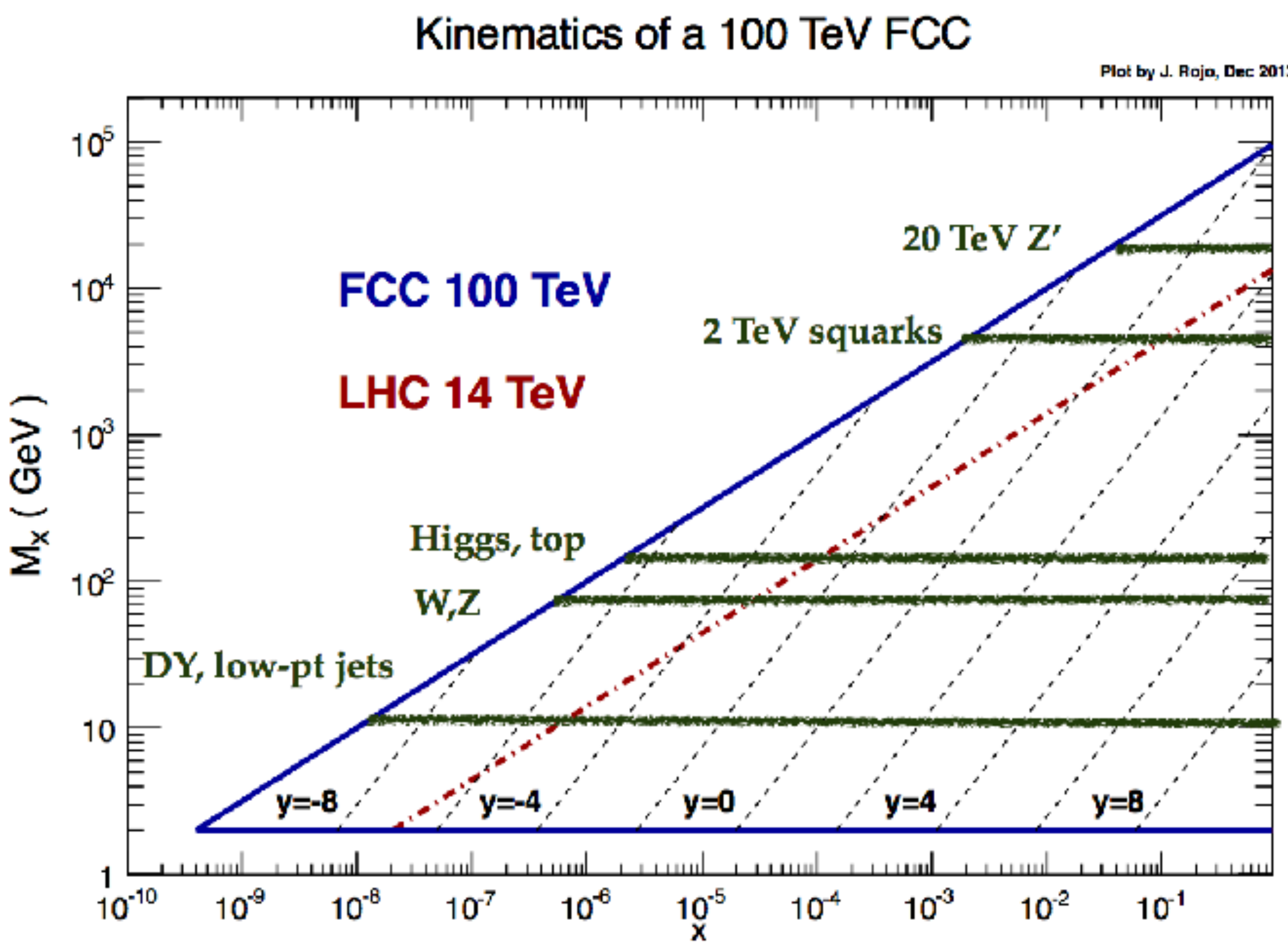
Best description of **small-*x* HERA data** only possible with **BFKL effects!**

J Rojo, BNL talk

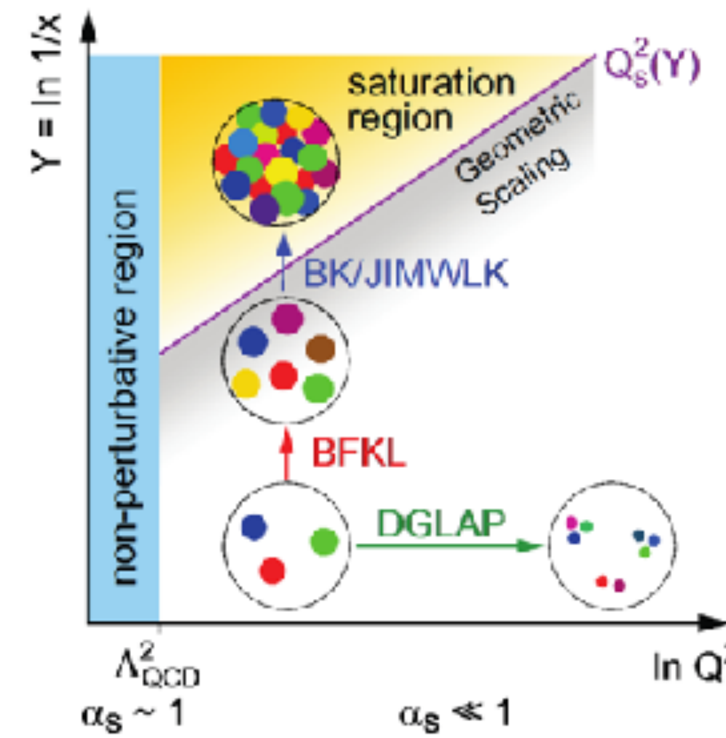
Ball et al, 1710.05935

(i) Small-x resummation

- Will this be enough when we will reach even smaller values of x?

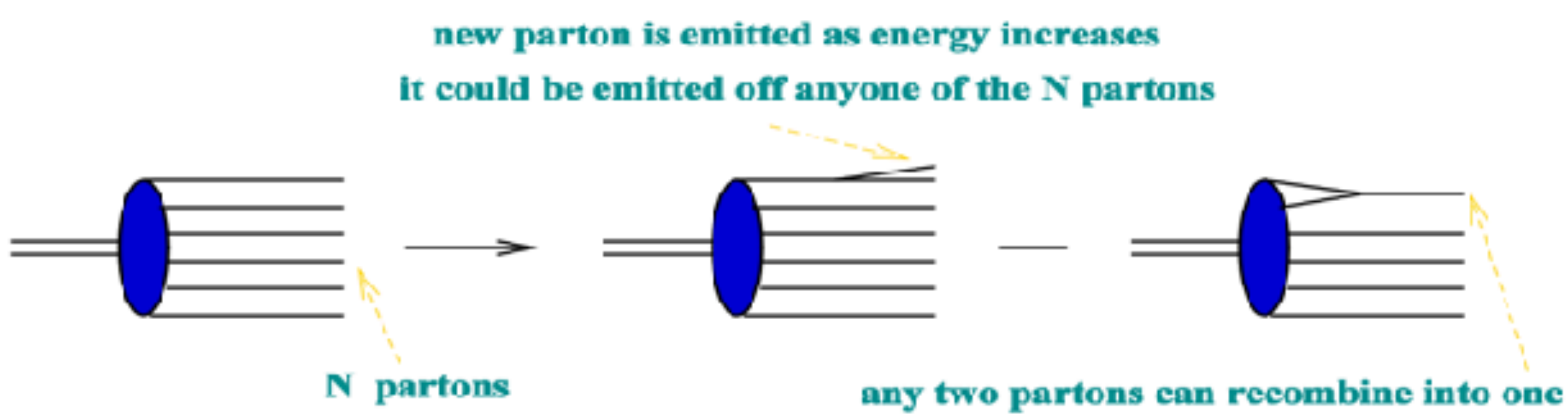


(ii) Non-linear evolution and saturation



$$Q_s^2 \sim A^{1/3} s^{\lambda}$$

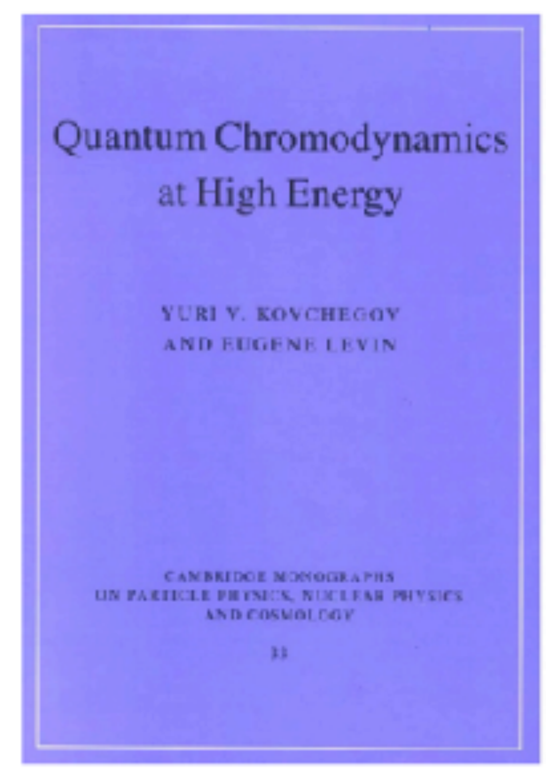
Important for ion-ion collisions



$$\frac{\partial}{\partial Y} N(x, k_T^2) = \alpha_s \, K_{BFKL} \otimes N(x, k_T^2) - \alpha_s \, [N(x, k_T^2)]^2$$

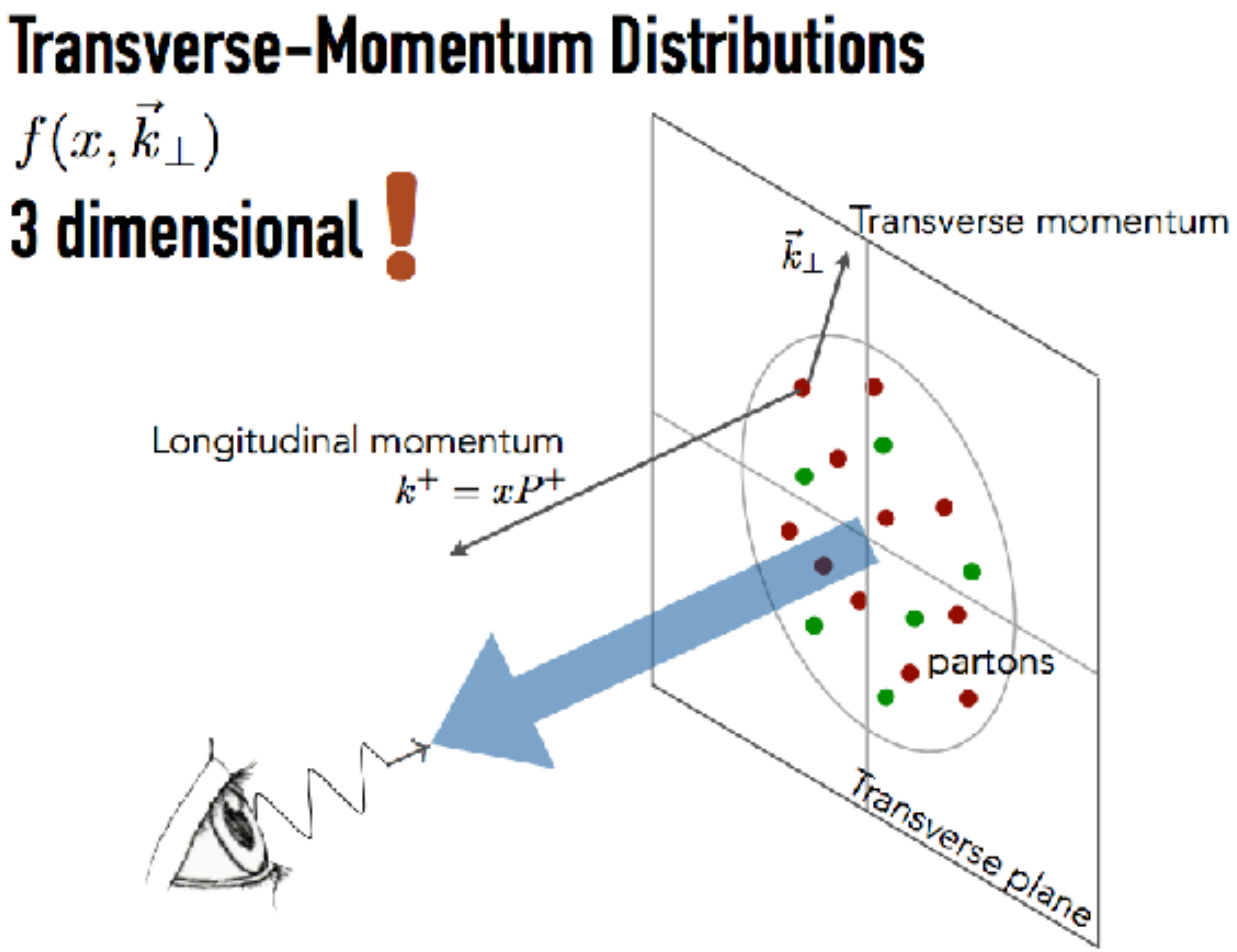
Number of parton pairs ~ N^2

I. Balitsky '96 (effective Lagrangian)
Yu. K. '99 (large N_c QCD)
JIMWLK '98-'01 (beyond large-N_c)



Beyond DGLAP: TMDs

- Much less mature field (universality and factorisation not well established but lots of interesting developments)



A. Bacchetta, talk at DIS2017

Outline

- First lecture (Monday)
    - Motivation: the big picture
    - Parton Model and QCD
    - Collinear Factorisation and definition of PDFs
- Second lecture (Tuesday)
    - Experimental Data
    - Disentangling proton's components
    - Heavy quarks and photons
- **Third lecture (today)**
    - Statistics and methodology
    - Parametrisation issues
    - Error propagation
    - State-of-the-art PDFs
- Fourth lecture (Thursday)
    - New frontiers and challenges

Methodological issues

A quite complicated game

- A single quantity: 1σ error
- Multiple quantities: 1σ contours
- Functions: 1σ "error band" in the space of functions
  = find the probability density in the space of functions f(x)
  Expectation values are functional integrals

Not as simple as it may look...

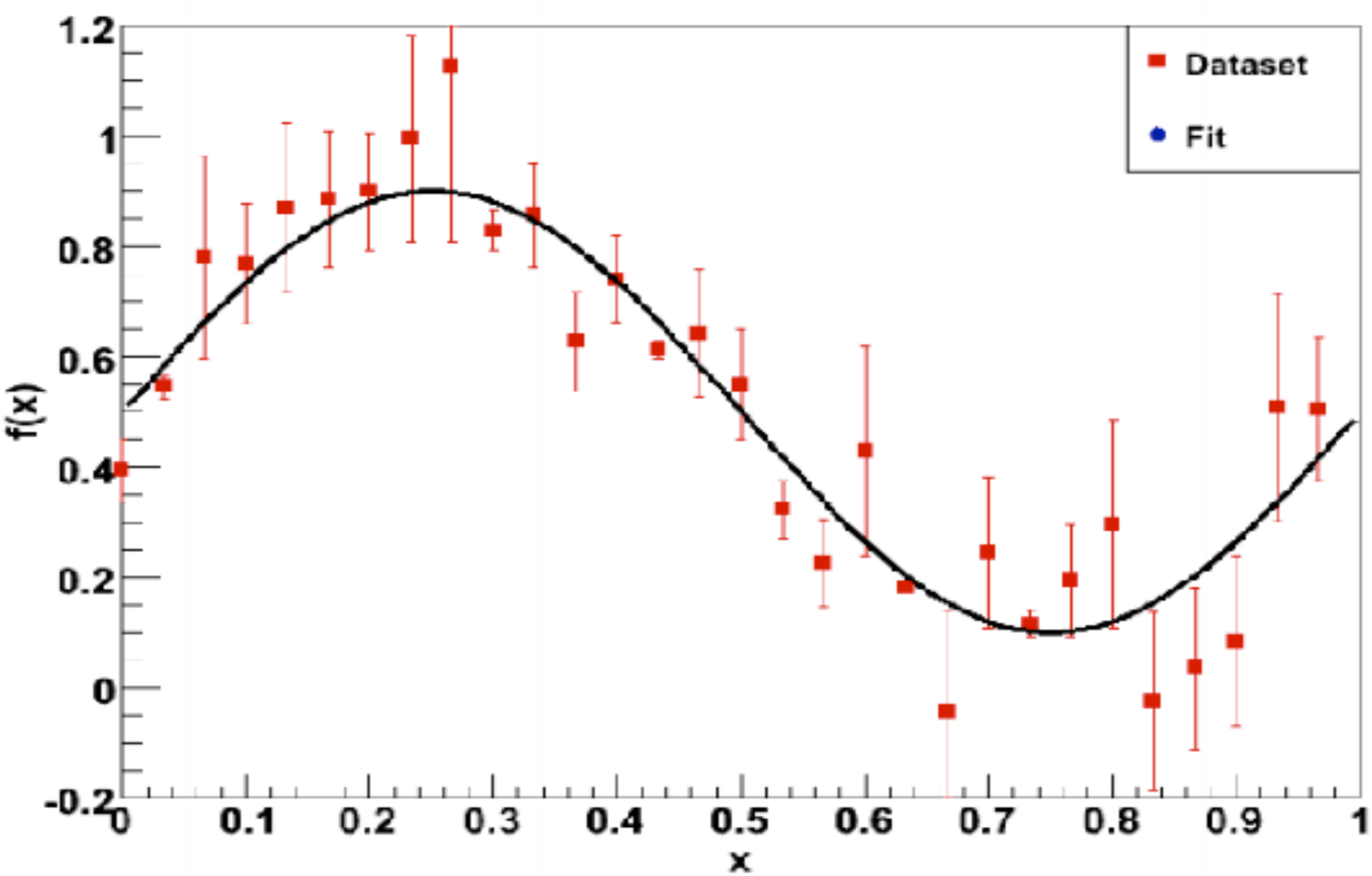
$$\langle \mathcal{O}[\{f\}] \rangle = \int [\mathcal{D}f]\, \mathcal{O}[\{f\}] \mathcal{P}[\{f\}]$$

- Given a finite number of experimental data points want a set of functions
- Want to find a infinite-dimensional object from a finite number of information

A toy model

A toy-model:

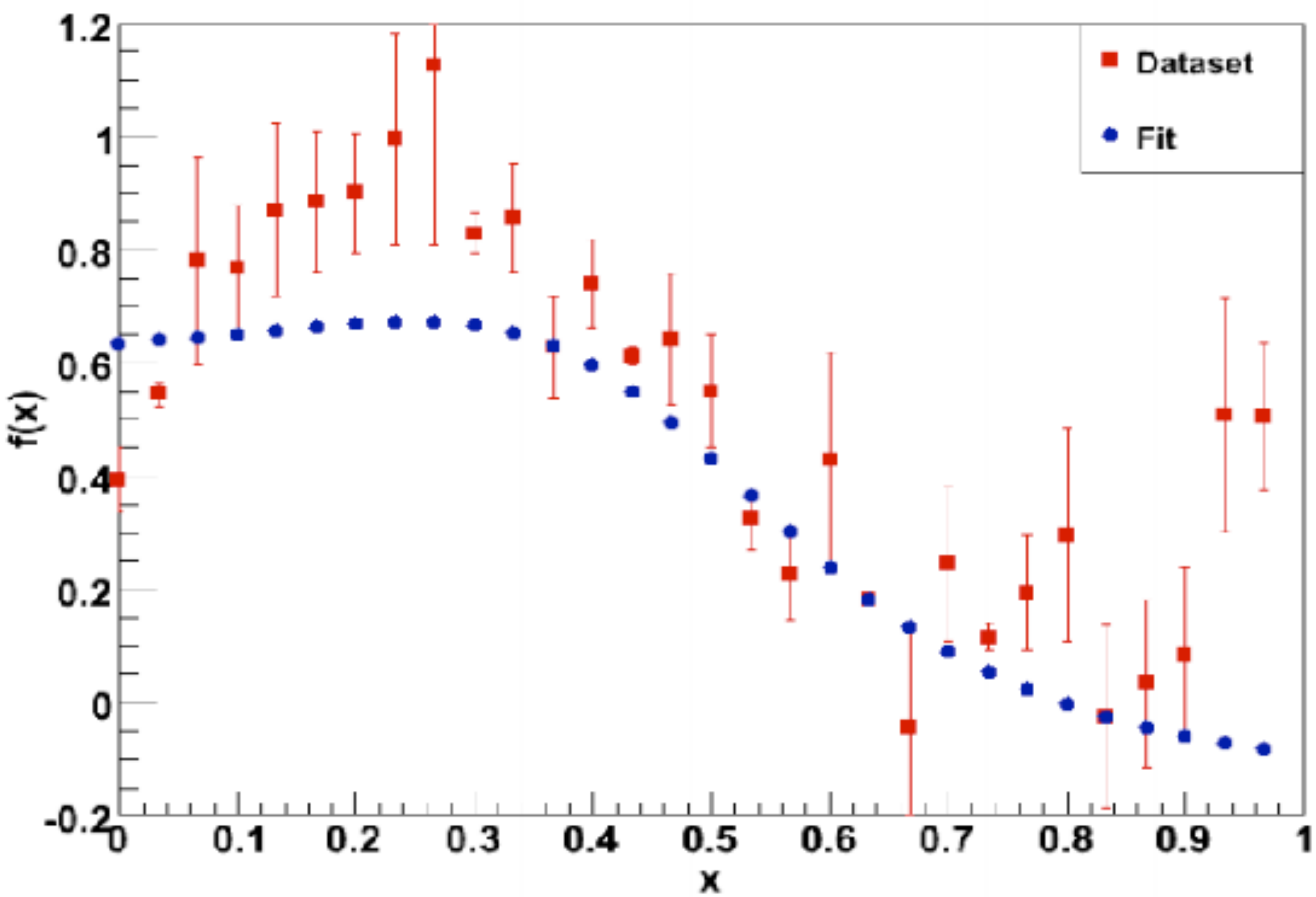
1) Imagine that we have a set of uncorrelated measurements of a quantity f(x) at different x. The underlying law that Nature established for this quantity is a sinusoidal, but we don't know anything about that and try to guess it with a fit.



A toy model

A toy-model:

2) Choose a parametrisation for f(x) and perform a fit by minimising a function, a figure of merit, like the χ^2



$$\chi^2 = \sum_{i=1}^{N_{\rm data}} \frac{(D_i - T_i)^2}{\sigma_i^2}$$

χ^2/d.o.f. » 1

We are not quite there...

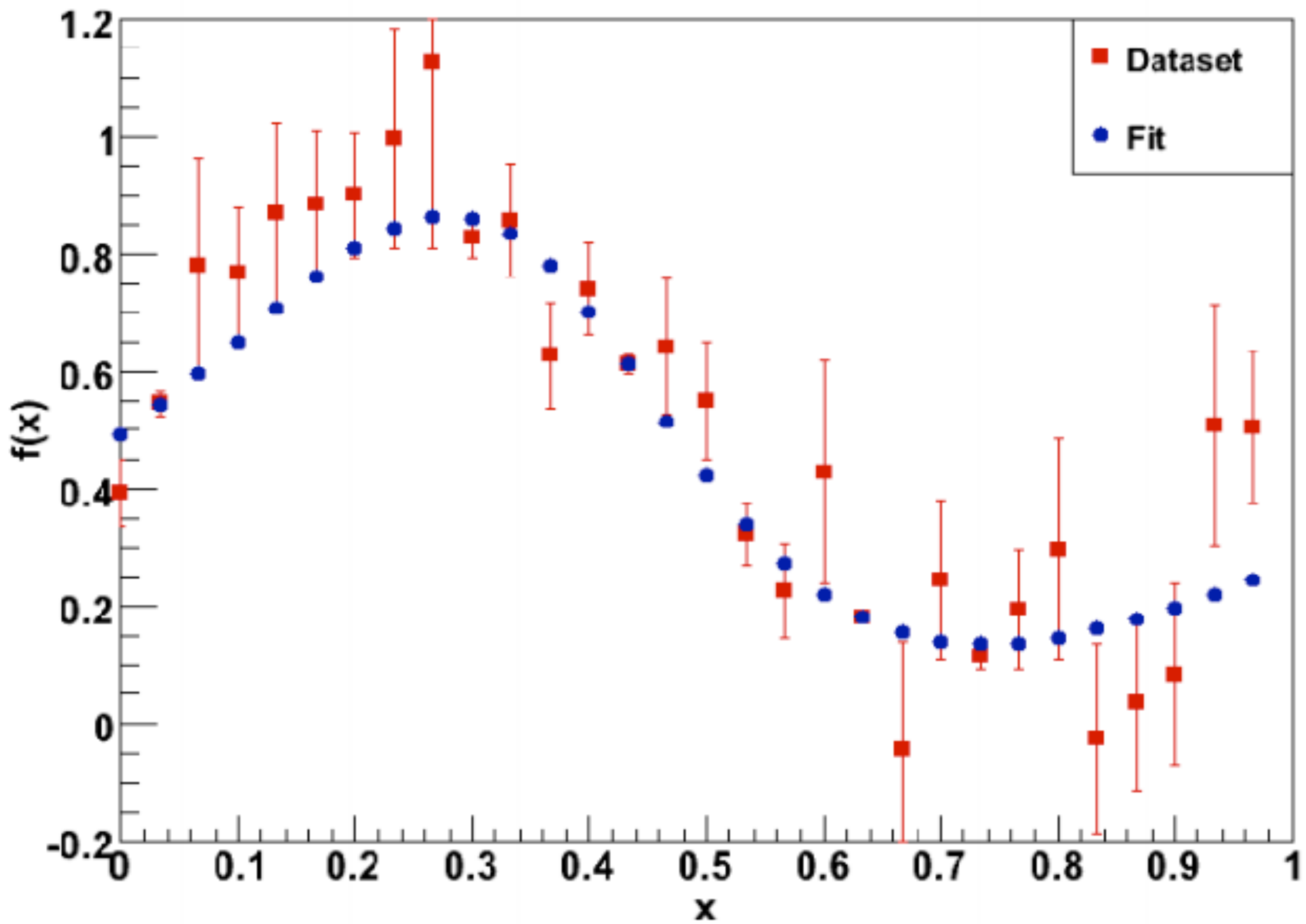
under-learning

A toy model

A toy-model:

2) Choose a parametrisation for f(x) and perform a fit by minimising a function, a figure of merit, like the χ^2

$$\boxed{\chi^2 = \sum_{i=1}^{N_{\rm data}} \frac{(D_i - T_i)^2}{\sigma_i^2}}$$



χ^2/d.o.f. ~ 1

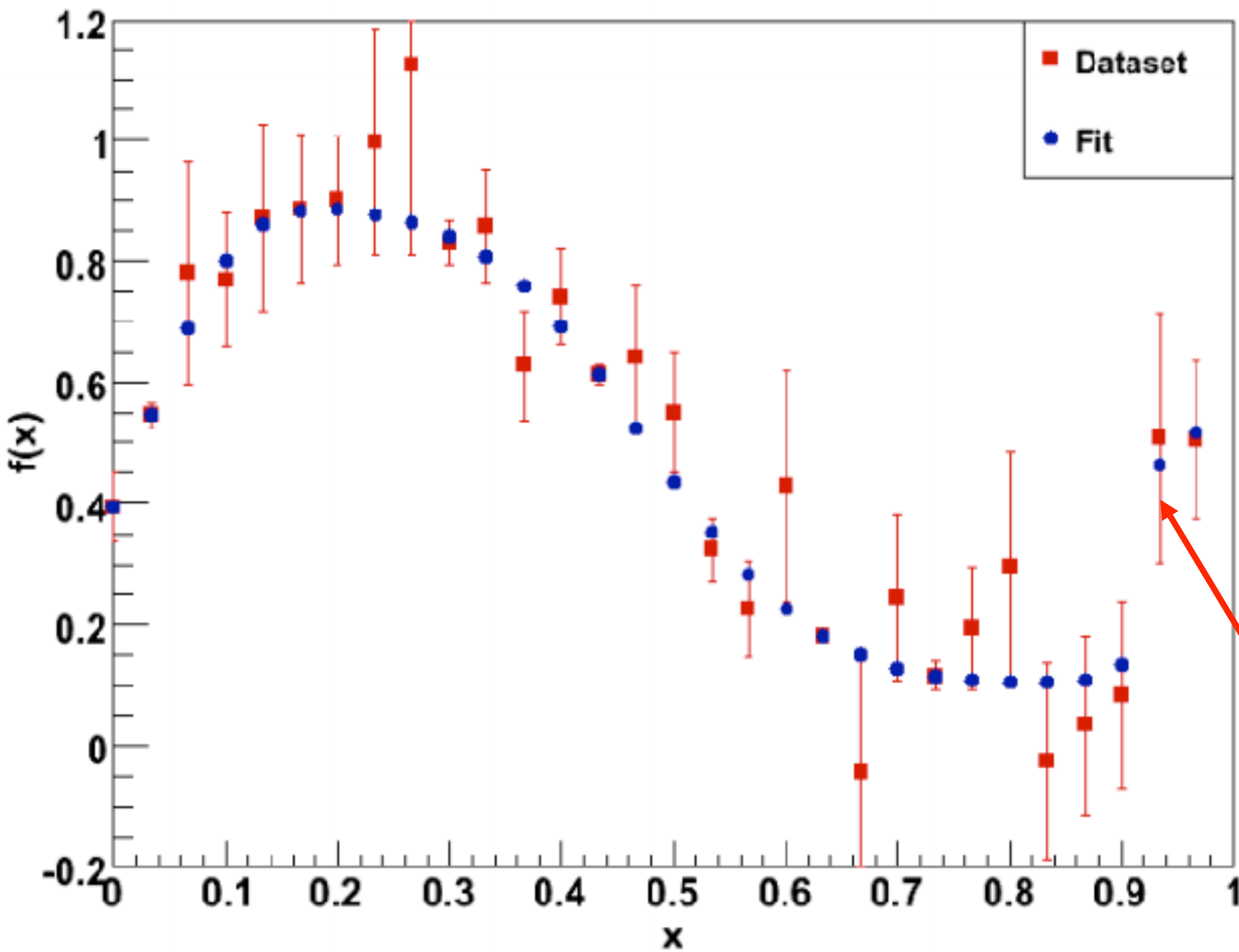
We are there...

[proper learning]()

A toy model

A toy-model:

2) Choose a parametrisation for f(x) and perform a fit by minimising a function, a figure of merit, like the χ^2



$$\chi^2 = \sum_{i=1}^{N_{\rm data}} \frac{(D_i - T_i)^2}{\sigma_i^2}$$

χ^2/d.o.f. → 0

We went too far...

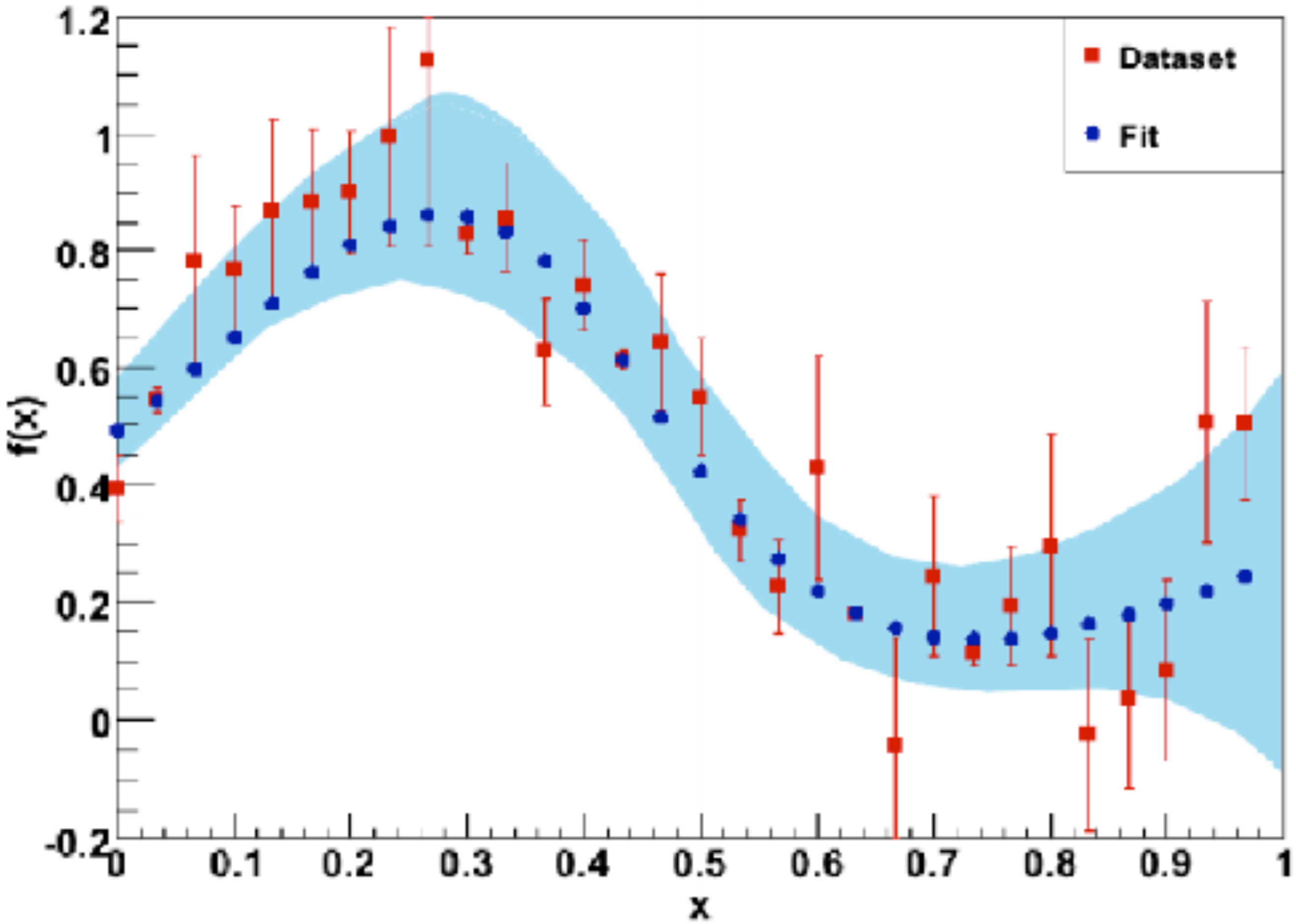
over-learning

start fitting the statistical noise

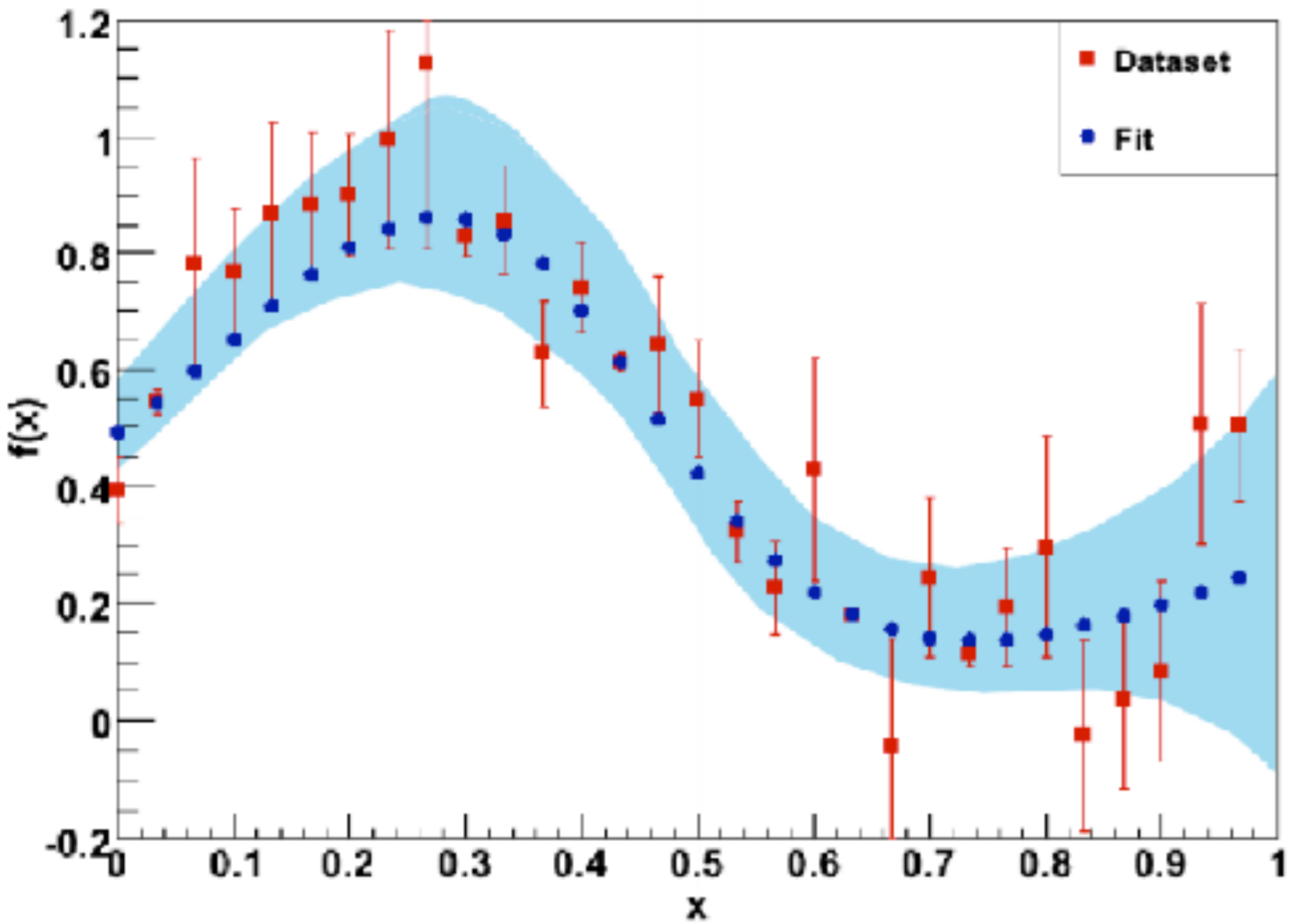
A toy model

A toy-model:

3) Determine the error of our fit, which corresponds to the lack of information that the data provide. In the limit of infinite and infinitely-precise and compatible data, the error band tends to 0



The actual game



The actual games is more complicated since we have 6+6+1 functions (actually 3+3+1) and errors to determine which are not directly measured. They enter in the measured observables according to different combinations. But still…

- ✓ Need to choose a clever and flexible parametrisation
- ✓ Need a way to stop the fit before over-learning sets in to avoid fitting statistical noise
- ✓ Need a reliable error estimate

Parametrisation

Choice of parametrisation

Usually one parametrises independently the gluon, light quarks and anti-quarks, strange and anti-strange (not everybody), while heavy quarks are generated perturbatively from light quarks and gluons*

The ideal parametrisation

- Too rigid

Global fit might not have flexibility to describe data or inadequate small uncertainties where there are no data

- Too flexible

Difficult minimisation and it might develop artefacts driven by statistical fluctuations of the data

Sum rules

- From baryon number conservation → Valence Sum Rules

$$\int_0^1 dx\,(u(x,Q^2) - \bar{u}(x,Q^2)) = 2$$
$$\int_0^1 dx\,(d(x,Q^2) - \bar{d}(x,Q^2)) = 1$$
$$\int_0^1 dx\,(s(x,Q^2) - \bar{s}(x,Q^2)) = 0$$

- From momentum conservation → Momentum Sum Rule

$$\int_0^1 dx\,(x\Sigma(x,Q^2) + xg(x,Q^2)) = 1$$

$$\text{with } \Sigma = \sum_{i=1}^{n_F} q_i + \bar{q}_i$$

Traditional (parametrical) approach

- Introduce a simple functional form with enough free parameters

$$f_i(x, Q_0^2) = a_0\, x^{a_1}\, (1-x)^{a_2} P(x, a_3, a_4, \ldots)$$

- Typically about 20-25 free parameters for 7 independent functions

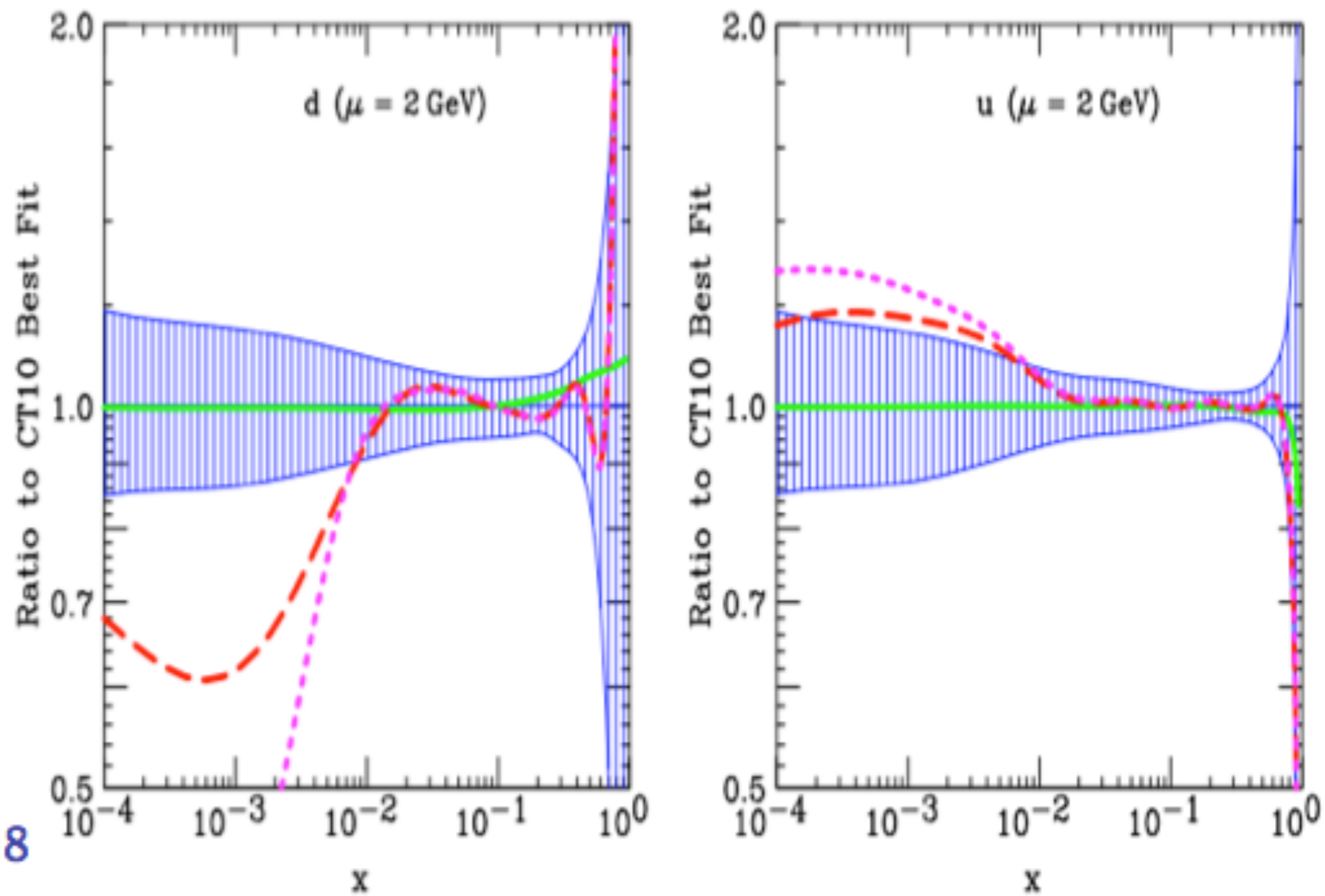
MSTW2008

20 free parameters

$$\begin{aligned}
xu_v(x, Q_0^2) &= A_u\, x^{\eta_1}(1-x)^{\eta_2}(1+\epsilon_u\sqrt{x}+\gamma_u\, x),\\
xd_v(x, Q_0^2) &= A_d\, x^{\eta_3}(1-x)^{\eta_4}(1+\epsilon_d\sqrt{x}+\gamma_d\, x),\\
xS(x, Q_0^2) &= A_S\, x^{\delta_S}(1-x)^{\eta_S}(1+\epsilon_S\sqrt{x}+\gamma_S\, x),\\
x\Delta(x, Q_0^2) &= A_\Delta\, x^{\eta_\Delta}(1-x)^{\eta_S+2}(1+\gamma_\Delta\, x+\delta_\Delta\, x^2),\\
xg(x, Q_0^2) &= A_g\, x^{\delta_g}(1-x)^{\eta_g}(1+\epsilon_g\sqrt{x}+\gamma_g\, x)+A_{g'}\, x^{\delta_{g'}}(1-x)^{\eta_{g'}},\\
x(s+\bar{s})(x, Q_0^2) &= A_+\, x^{\delta_S}\,(1-x)^{\eta_+}(1+\epsilon_S\sqrt{x}+\gamma_S\, x),\\
x(s-\bar{s})(x, Q_0^2) &= A_-\, x^{\delta_-}(1-x)^{\eta_-}(1-x/x_0),
\end{aligned}$$

Traditional (parametrical) approach

- Possible issues:
  What is the error associated to a given functional form?

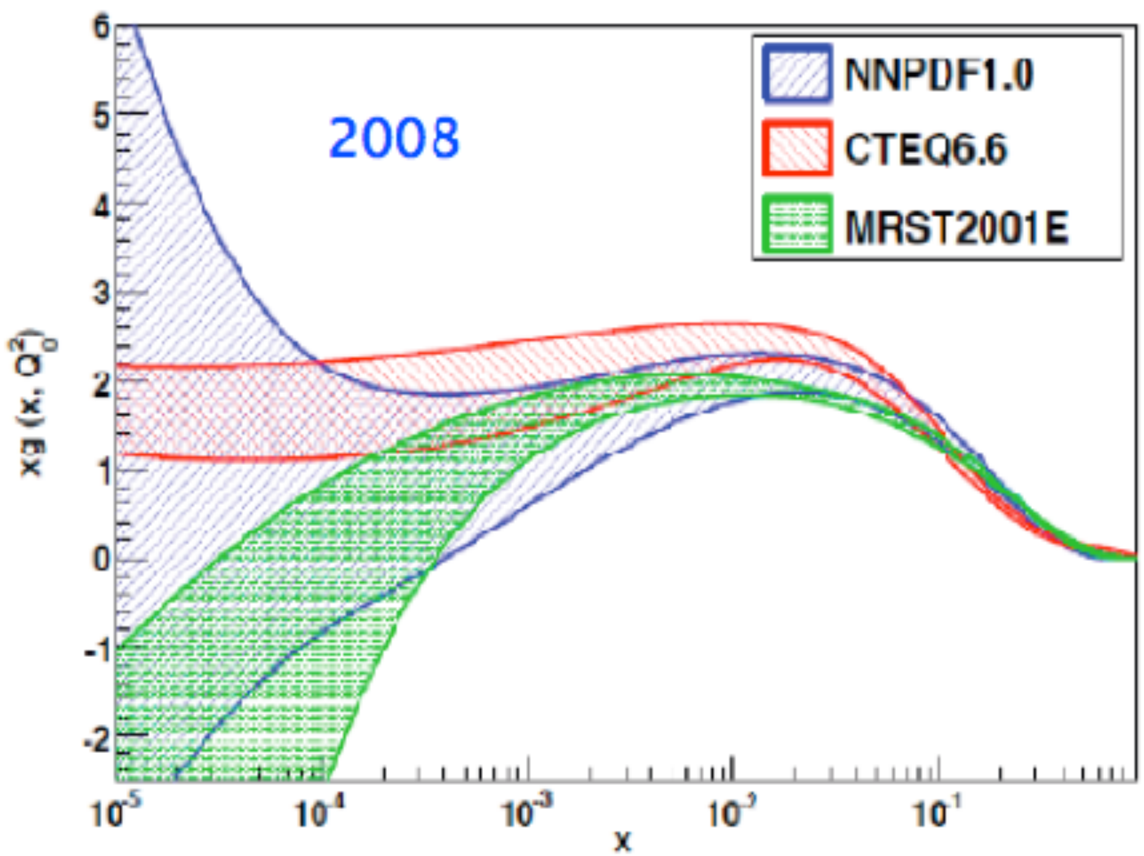


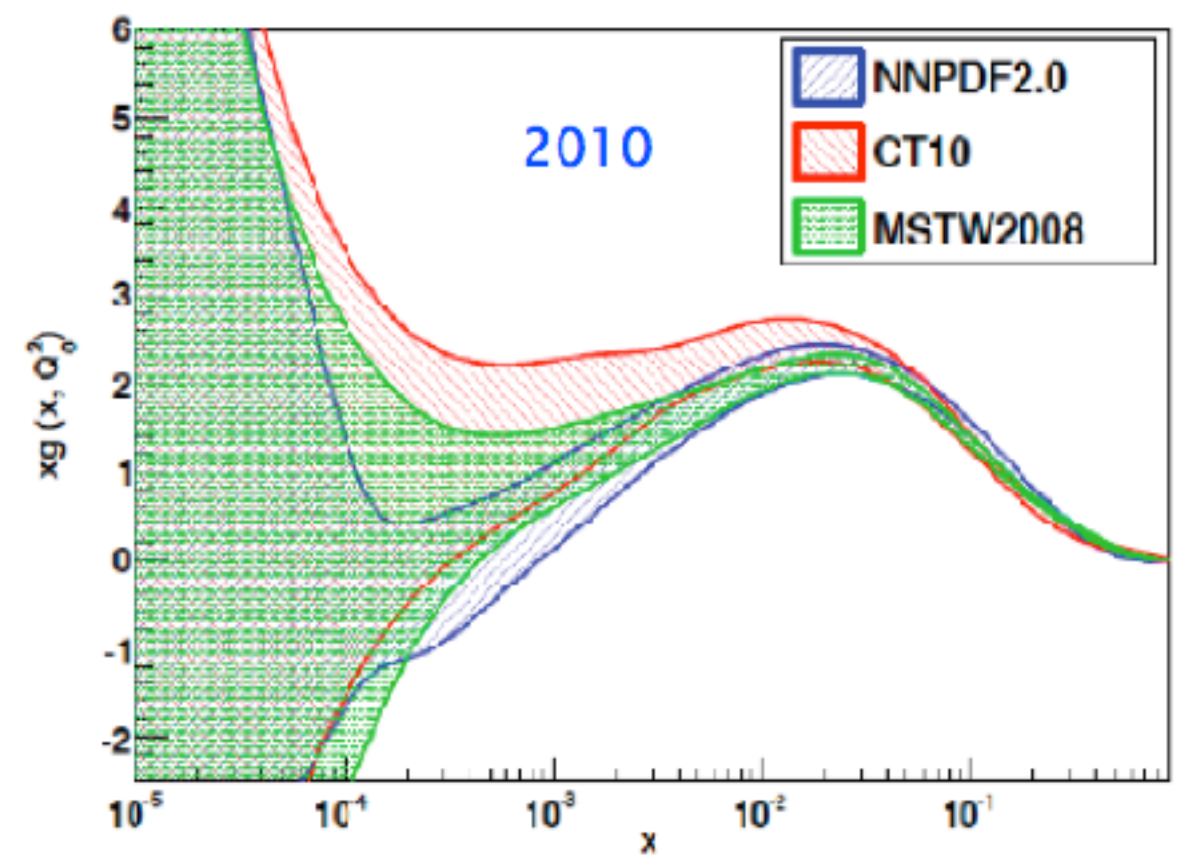
Pink and red curves give same good description of data but outside error bar

J. Pumplin ArXiv:0909.0268

Traditional (parametrical) approach

- Possible issues:
 If functional form not flexible enough PDFs may present unrealistically small errors where data do not constrain PDF uncertainties

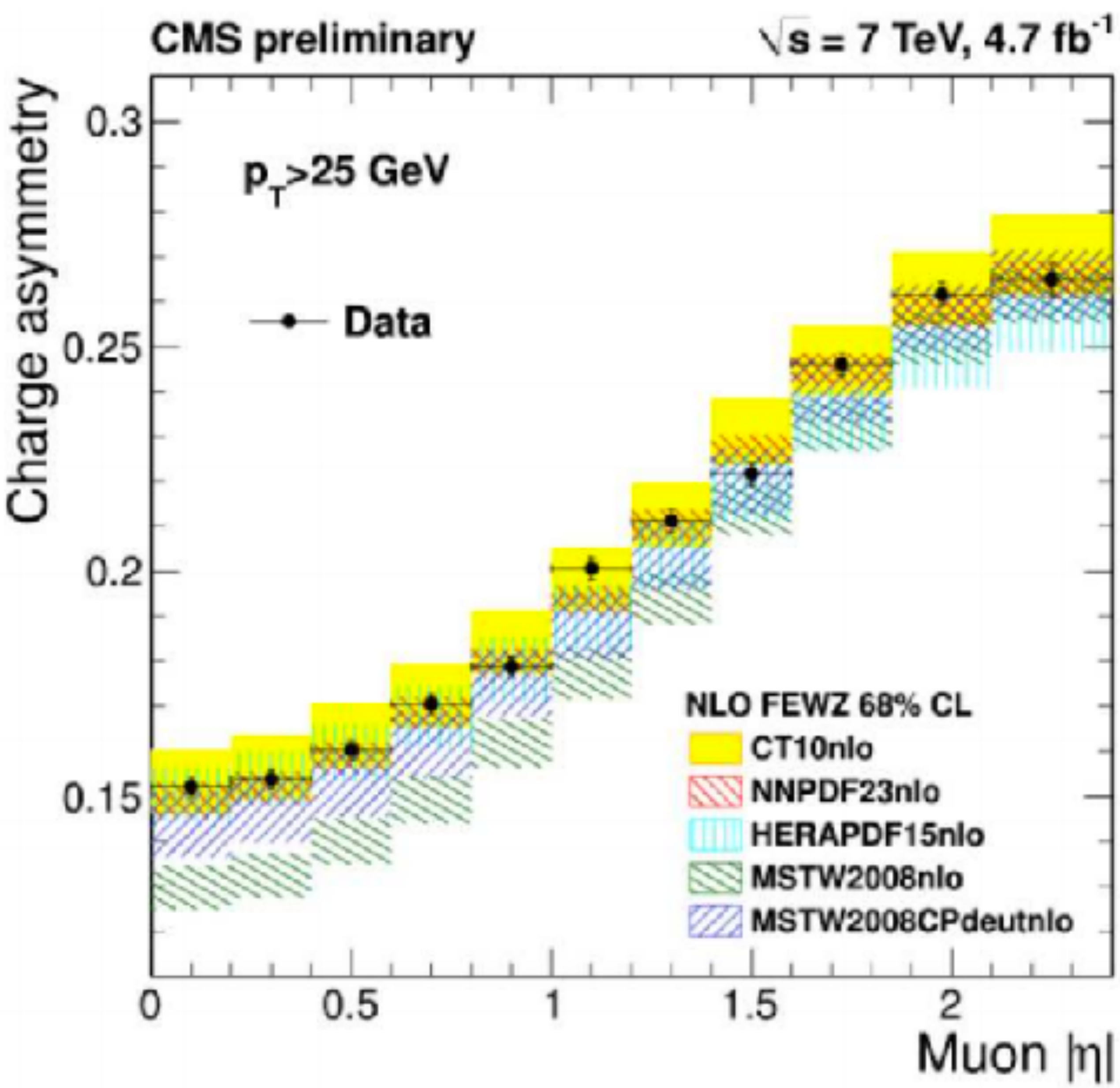


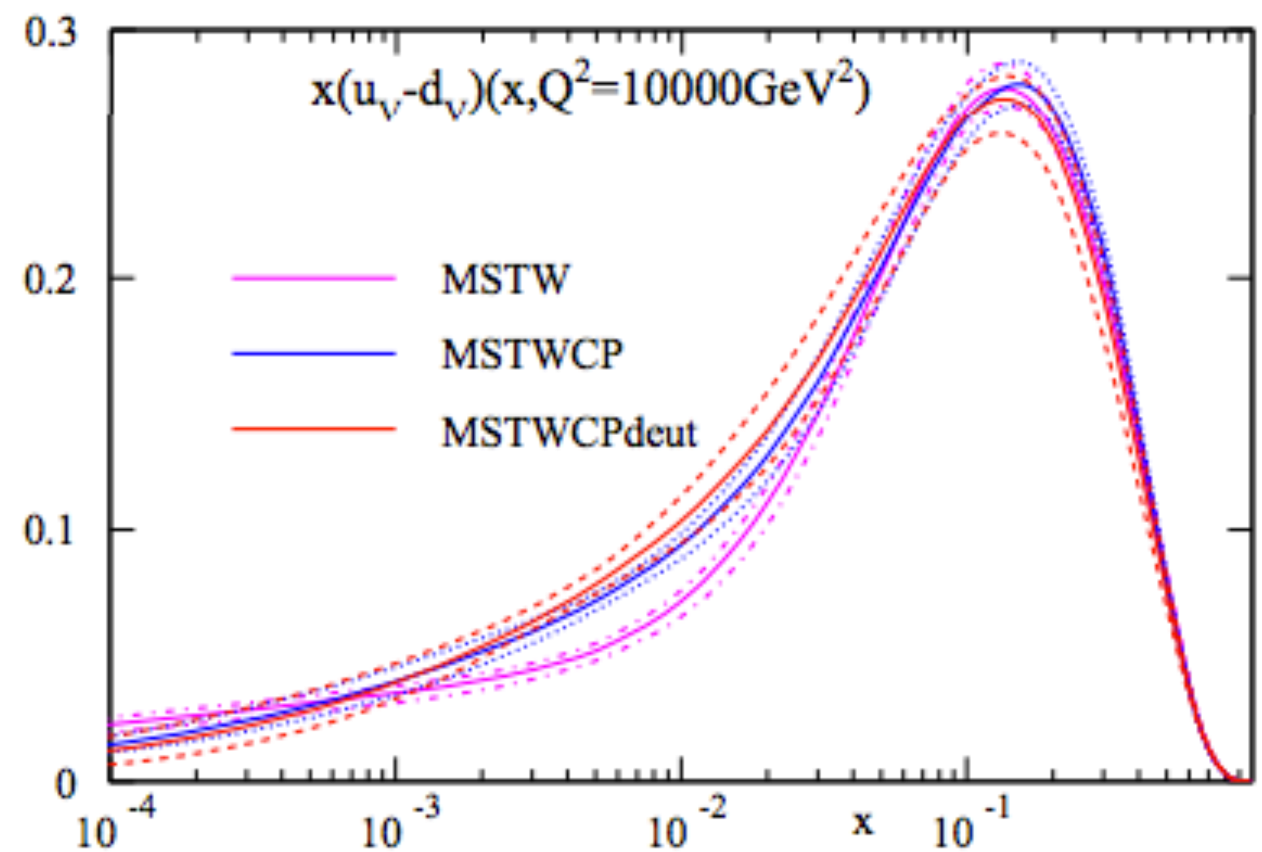


$$xg = A_g\, x^{\delta_g} (1-x)^{\eta_g} (1 + \epsilon_g \sqrt{x} + \gamma_g\, x) + A_{g'}\, x^{\delta_{g'}} (1-x)^{\eta_{g'}}$$

Traditional (parametrical) approach

- Possible issues:
  If functional form not flexible enough PDFs may be not able to adapt to new data





Martin et al
EPJC73 (2013) 2, 2318

Data-driven progress

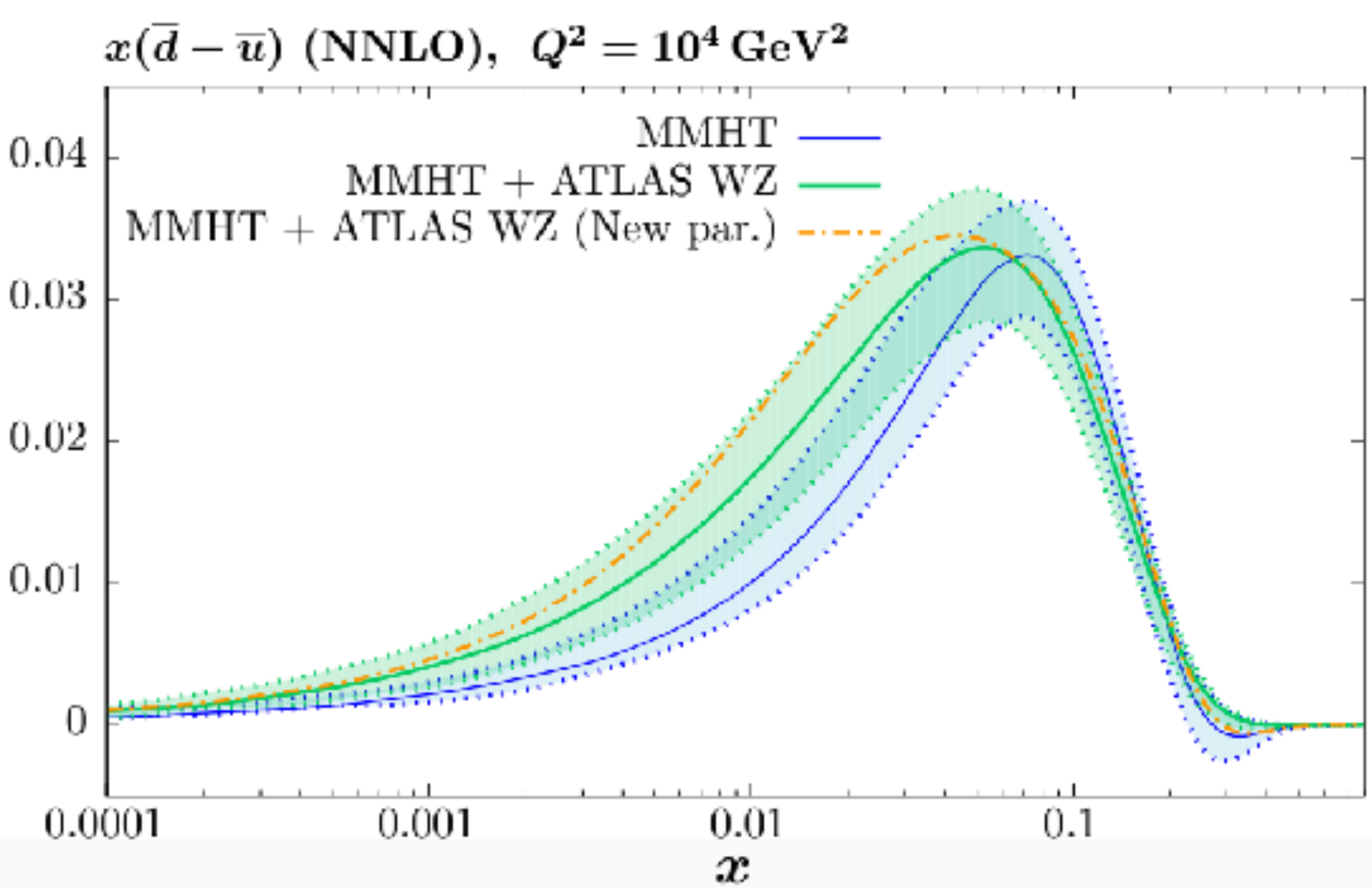
R. Thorne, talk last week at EW-LHC precision working group

For $(\bar{d}-\bar{u})(x,Q_0^2)$ by default use 4 parameters,

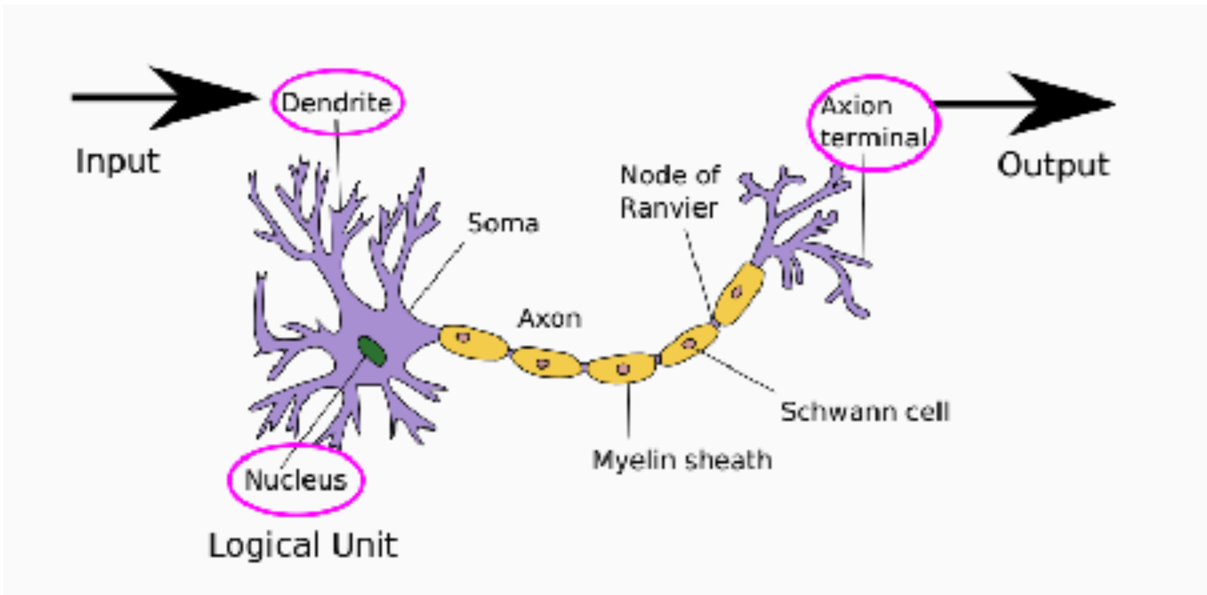
$$(\bar{d}-\bar{u})(x,Q_0^2) = A(1-x)^{\eta_{sea}+2}x^{\delta}(1+\gamma x+\Delta x^2),$$

Extend to $(\bar{d}-\bar{u})(x,Q_0^2) = A(1-x)^{\eta_{sea}+2}x^{\delta}(1+\sum_{i=1}^{4} a_i T_i(1-2x^{\frac{1}{2}})),$

So 6 free parameters. Easily allows multiple turning points. Improves fit by >10 points - eases ATLAS W, Z and DY ratio tension.



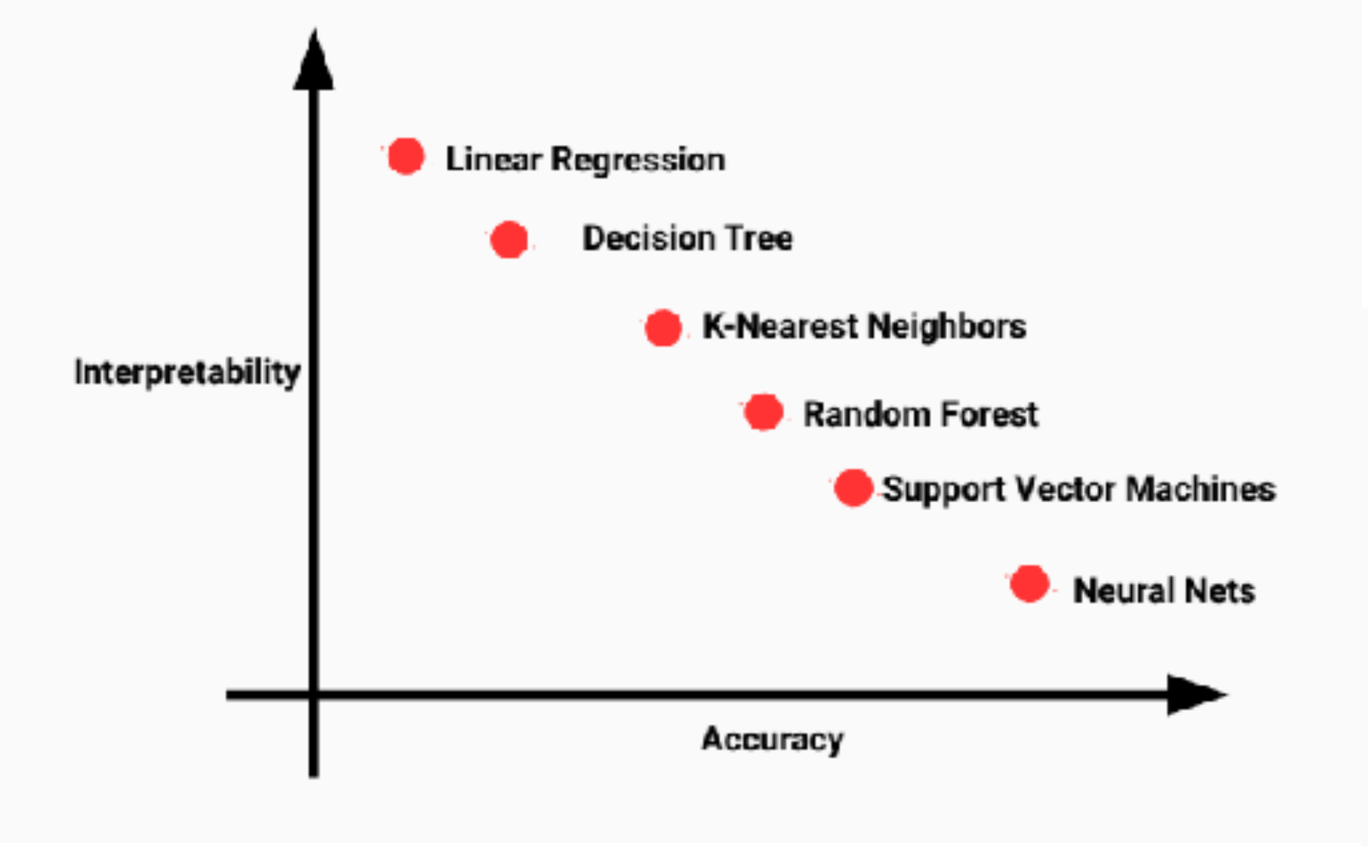
Neural networks and ML



S. Carrazza, Colloquium, S. Paolo, N3PDF

- Artificial neural networks are computer systems inspired by the biological neural networks in the brain
- Data communication pattern
- Currently state-of-the-art for several Machine Learning Applications

Neural networks and ML



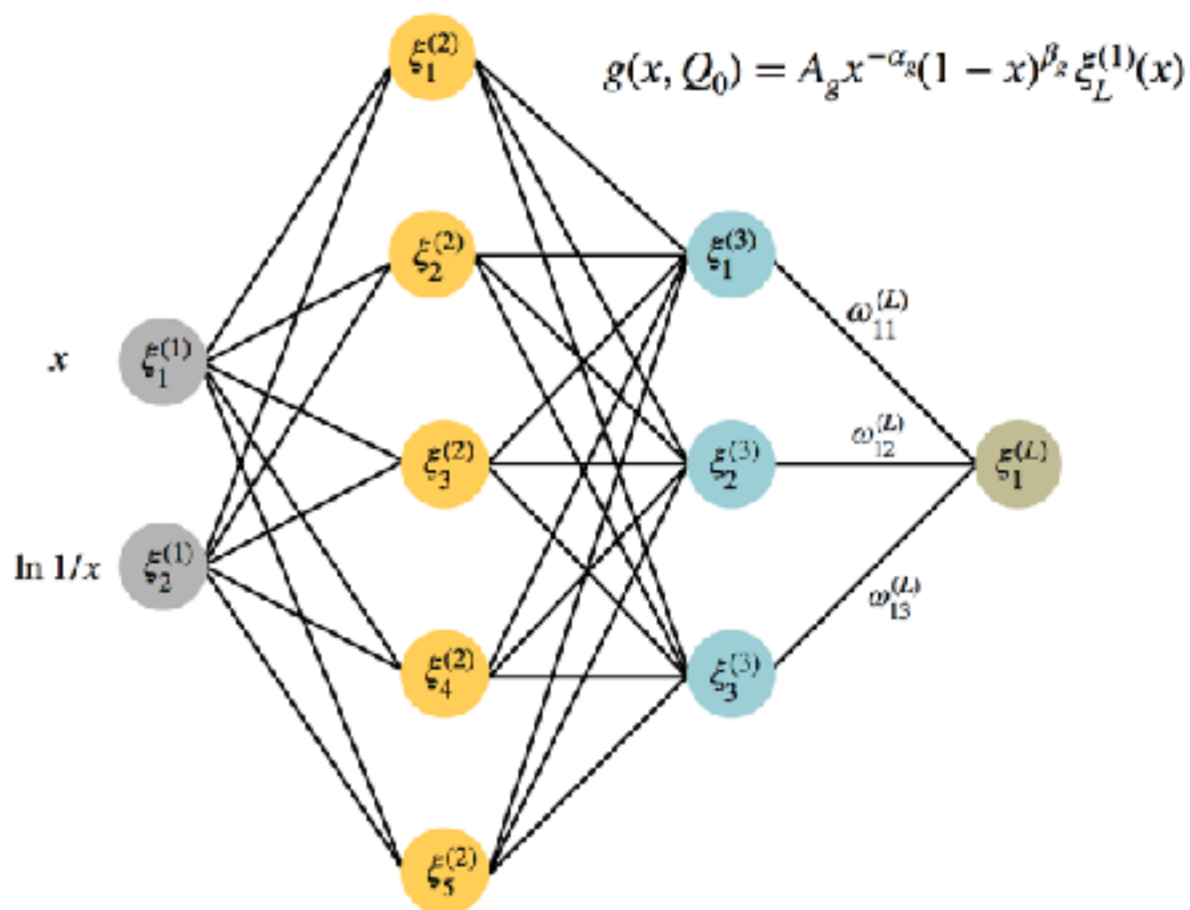
N3PDF
Machine Learning • PDFs • QCD

Example of Machine Learning in theoretical High Energy Physics

- **Supervised learning:**
  - The structure of the proton at the LHC
    - parton distribution functions
  - Theoretical prediction and combination
  - Monte Carlo reweighting techniques
    - neural network Sudakov
  - BSM searches and exclusion limits
- **Unsupervised learning:**
  - Clustering and compression
    - PDF4LHC15 recommendation
  - Density estimation and anomaly detection
    - Monte Carlo sampling

Neural network parametrisation

Fully connected multi-layer perceptron



For a 1-2-1 feedforward neural network can write explicitly functional form

$$\xi_1^{(3)}(\xi_1^{(1)}) = \frac{1}{1 + e^{\theta_1^{(3)} - \frac{\omega_{11}^{(2)}}{1+e^{\theta_1^{(2)} - \xi_1^{(1)}\omega_{11}^{(1)}}} - \frac{\omega_{12}^{(2)}}{1+e^{\theta_2^{(2)} - \xi_1^{(1)}\omega_{21}^{(1)}}}}}$$

- Neural Networks: all independent PDFs are associated to an unbiased and flexible parametrisation: O(300) parameters versus O(30) in polynomial parametrisation

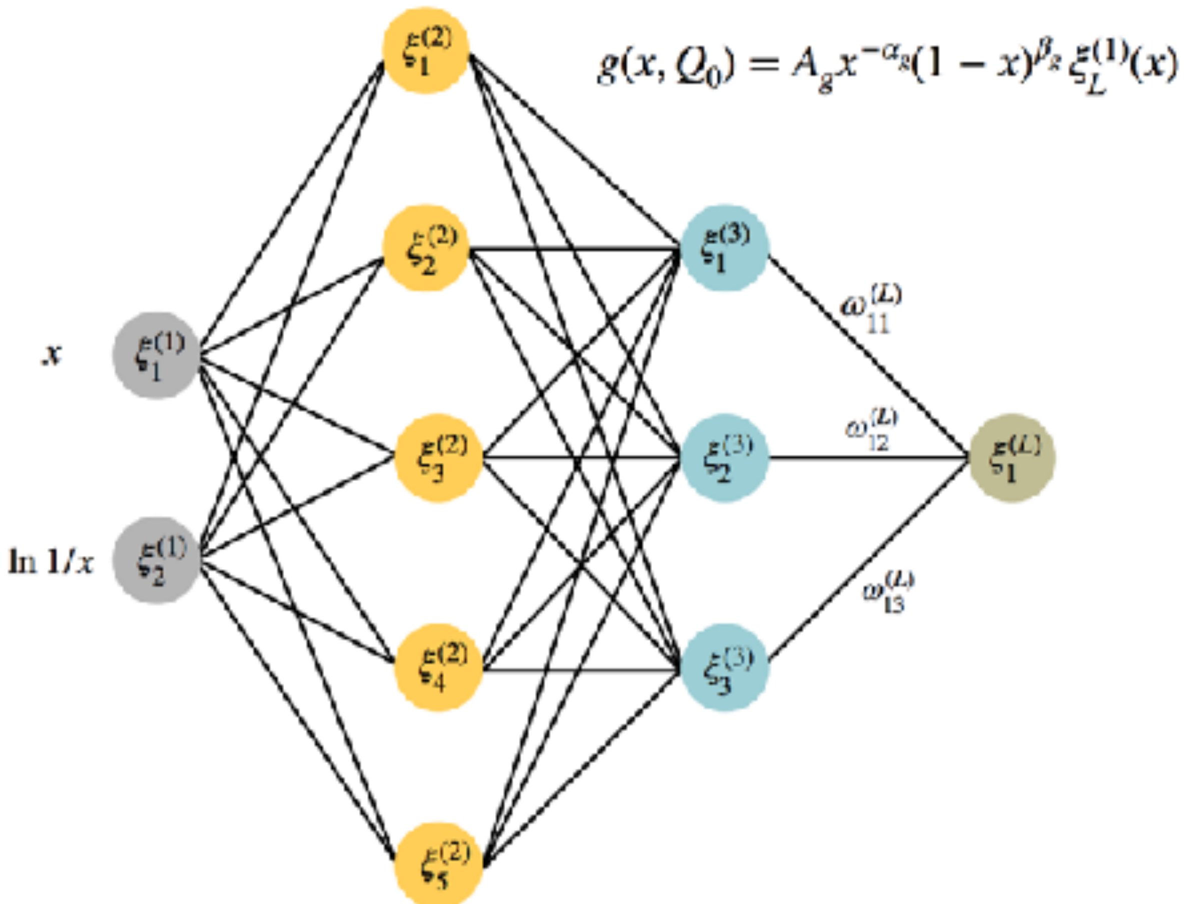
- 2-5-3-1 Neural network associated to each independent PDF (gluon, up, anti-up, down, anti-down, strange, anti-strange and charm)

$$\xi_i = g\left(\sum_j \omega_{ij}\xi_j - \theta_i\right)$$

$$g(x) = \frac{1}{1+e^{-x}}$$

Neural network training

Fully connected multi-layer perceptron



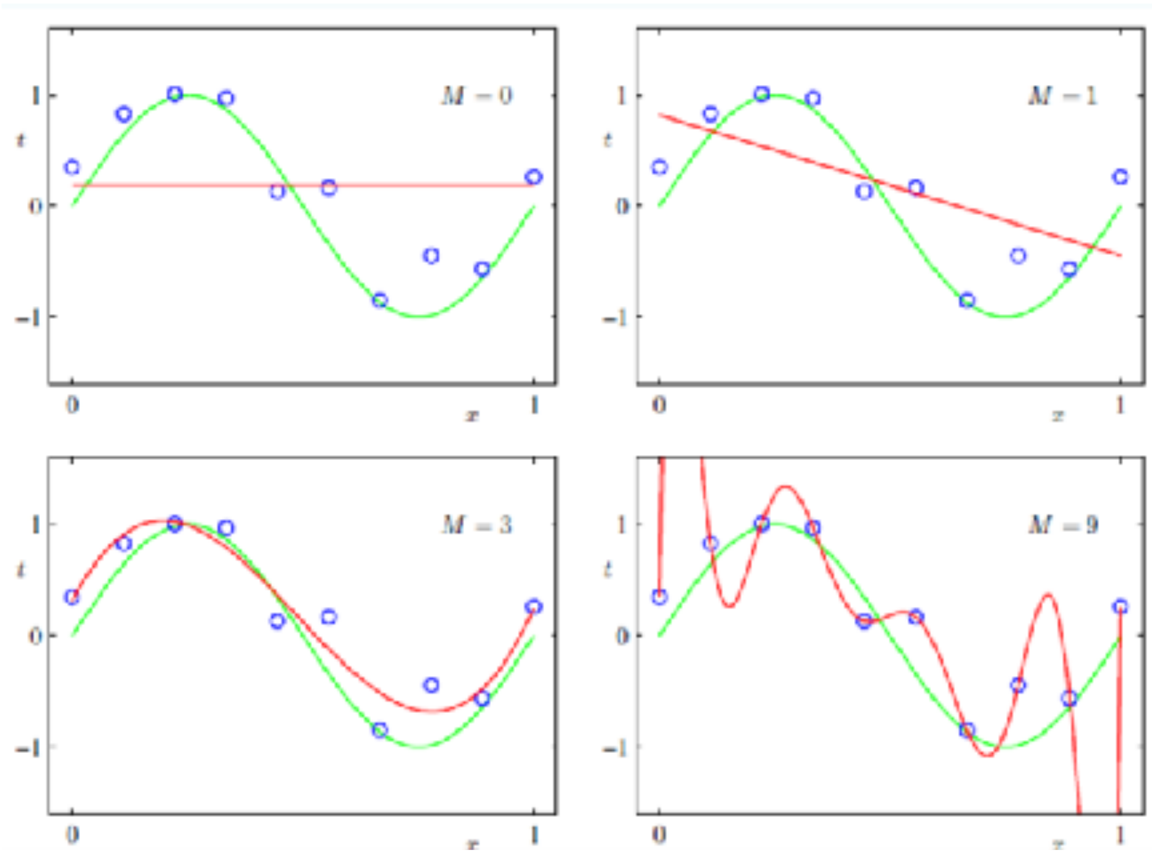
How do we train the 7(8) independent NN?

Minimise the cost function:

$$\chi^2 = \sum_{i,j=1}^{N_{\rm dat}} (D_i - T_i)(\mathrm{cov})^{-1}_{ij}(D_j - T_j)$$

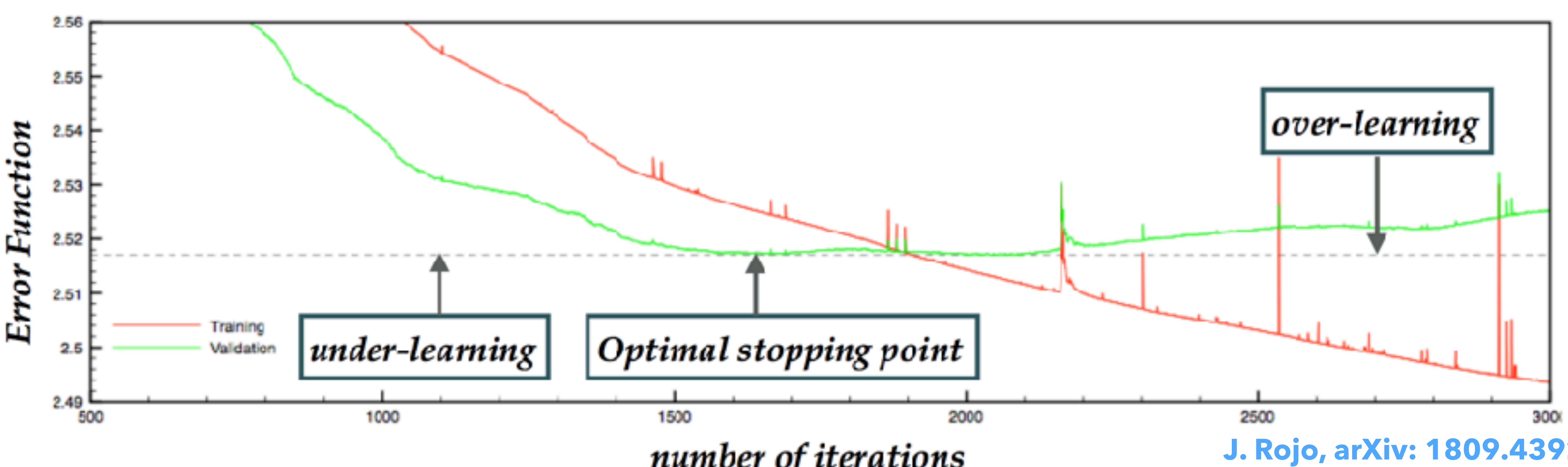
- D_i experimental measurement for the point i
- T_i theoretical prediction for the point i(depending on PDF parameters $\sigma_h = \sigma_{12} \otimes f_1 \otimes (f_2)$)
- (cov)ij is the covariance matrix between point i and j with corrections for normalisation uncertainties
- Supplemented by additional penalty for positive observables

Neural network training



- Large parameter space: need an algorithm that is able to explore it without getting trapped in local minima such as genetic algorithm

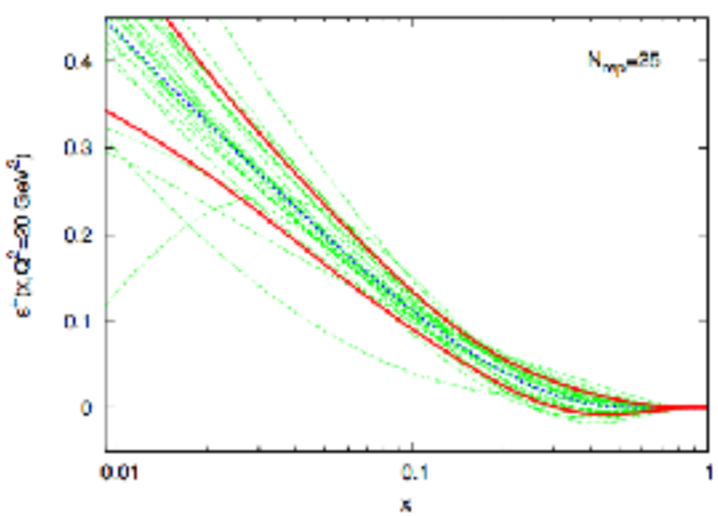
- Redundant parametrization: risk of over-fitting. Cross-validation necessary.



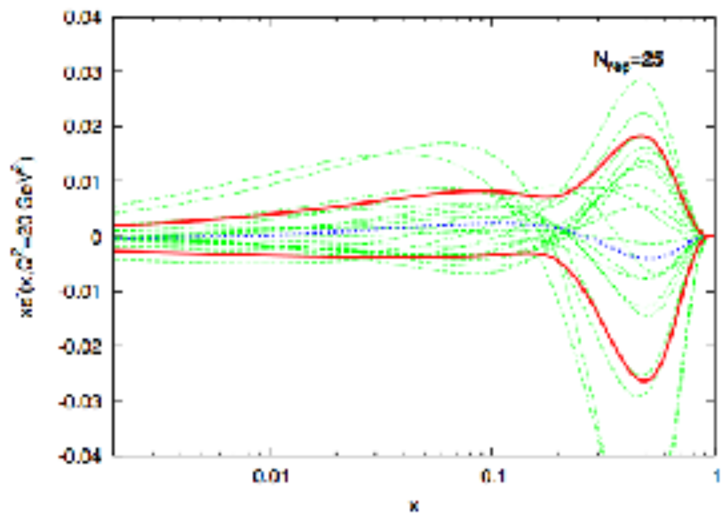
J. Rojo, arXiv: 1809.4392

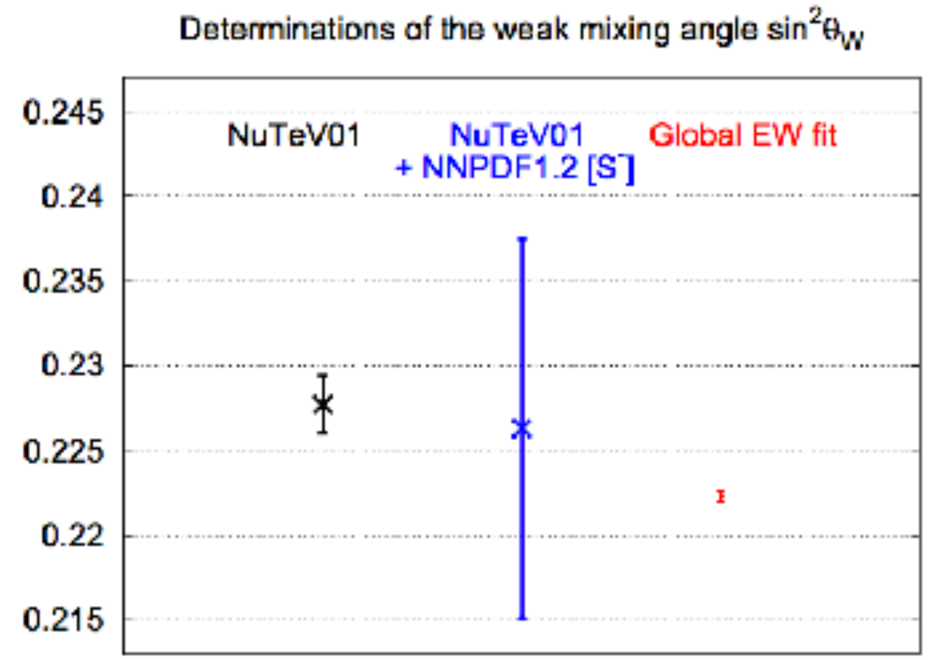
E.g. the NuTeV anomaly

Total strangeness ↓



Strange valence ↓





EW fit

$\sin^2\theta_W = 0.2223 \pm 0.0002$

NuTeV

$\sin^2\theta_W = 0.2276 \pm 0.0014$

$$\sin^2\theta_W\Big|_{\rm NuTeV} - \sin^2\theta_W\Big|_{\rm EW} = 0.0053$$

$$[F] = \int_0^1 dx\, x\, f(x, Q^2).$$

- $>3\sigma$ discrepancy between EW fits and NuTeV measurements
- Unbiased parametrisation of strangeness (2010) solved NuTeV anomaly

$$\delta_s \sin^2\theta_W \sim -0.240\frac{[S^-]}{[Q^-]}$$

$$\delta_s \sin^2\theta_W = -0.0005 \pm 0.0096^{\rm PDFs} \pm sys$$

Ball et al, 0906.1958

Error propagation

Error propagation

$$\langle \mathcal{O}[\{f\}]\rangle = \int [\mathcal{D}f]\, \mathcal{O}[\{f\}]\mathcal{P}[\{f\}].$$

- Given a finite number of experimental data points want a set of functions
- Want to find a infinite-dimensional object from a finite number of information

Option a) Project into a n-dimensional space of parameters which parametrise PDFs and use linear approximation around minimum χ^2

$$\langle \mathcal{O}[\{f\}]\rangle \simeq \int da_1 da_2 ... da_{N_{par}}\, \mathcal{O}[\vec{a}]\, \mathcal{P}[\vec{a}]$$

Hessian Method

Option b) Choose a parametrisation and perform a Monte Carlo sampling of probability density in functional space

$$\langle \mathcal{O}[\{f\}]\rangle \simeq \frac{1}{N_{\rm rep}} \sum_{i=1}^{N_{\rm rep}} \mathcal{O}[f_i],$$

Monte Carlo Method

Hessian method

➡ Used by most PDF fitters (CTEQ/TEA, MSTW/MMHT, HERAPDF, ABM)

➡ Pick a functional form and project problem in the N_{par}-dimensional space of parameters (typically 15 - 25)

➡ Determine best fit values of parameters $\{\vec{a}_0\}$

➡ Shift $\vec{a} \to \vec{a} - \vec{a}_0$

➡ Determine error on PDFs and any observable depending on PDFs (all denoted by X) by propagation of the error in the parameter space

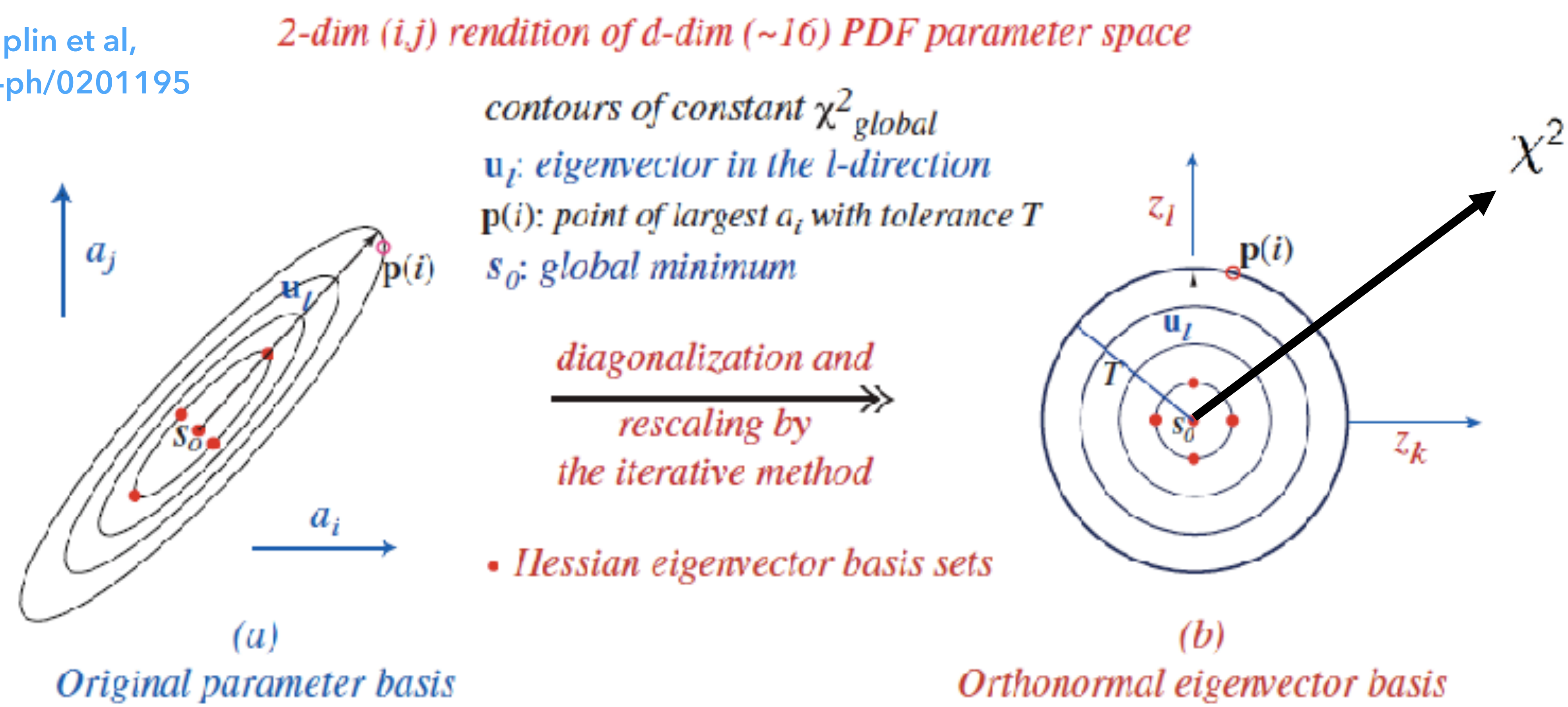
Assuming linear prop: $X(\vec{a}) \simeq X(\vec{0}) + a_i \partial_i X(\vec{a})\big|_{\vec{a}=\vec{0}}$

Variance: $\sigma_X^2 = (\mathrm{cov})_{ij}\, \partial_i X \partial_j X$ $(\mathrm{cov})_{IJ}$ covariance matrix in param, space

Maximum likelihood: $(\mathrm{cov})_{ij} = (H)_{ij} = \left.\frac{\partial^2 \chi^2(\vec{a})}{\partial_i a \partial_j a}\right|_{\vec{a}=\vec{0}}$ cov $\Longleftrightarrow$ Hessian at the minimum of χ^2

Hessian method

Pumplin et al, hep-ph/0201195



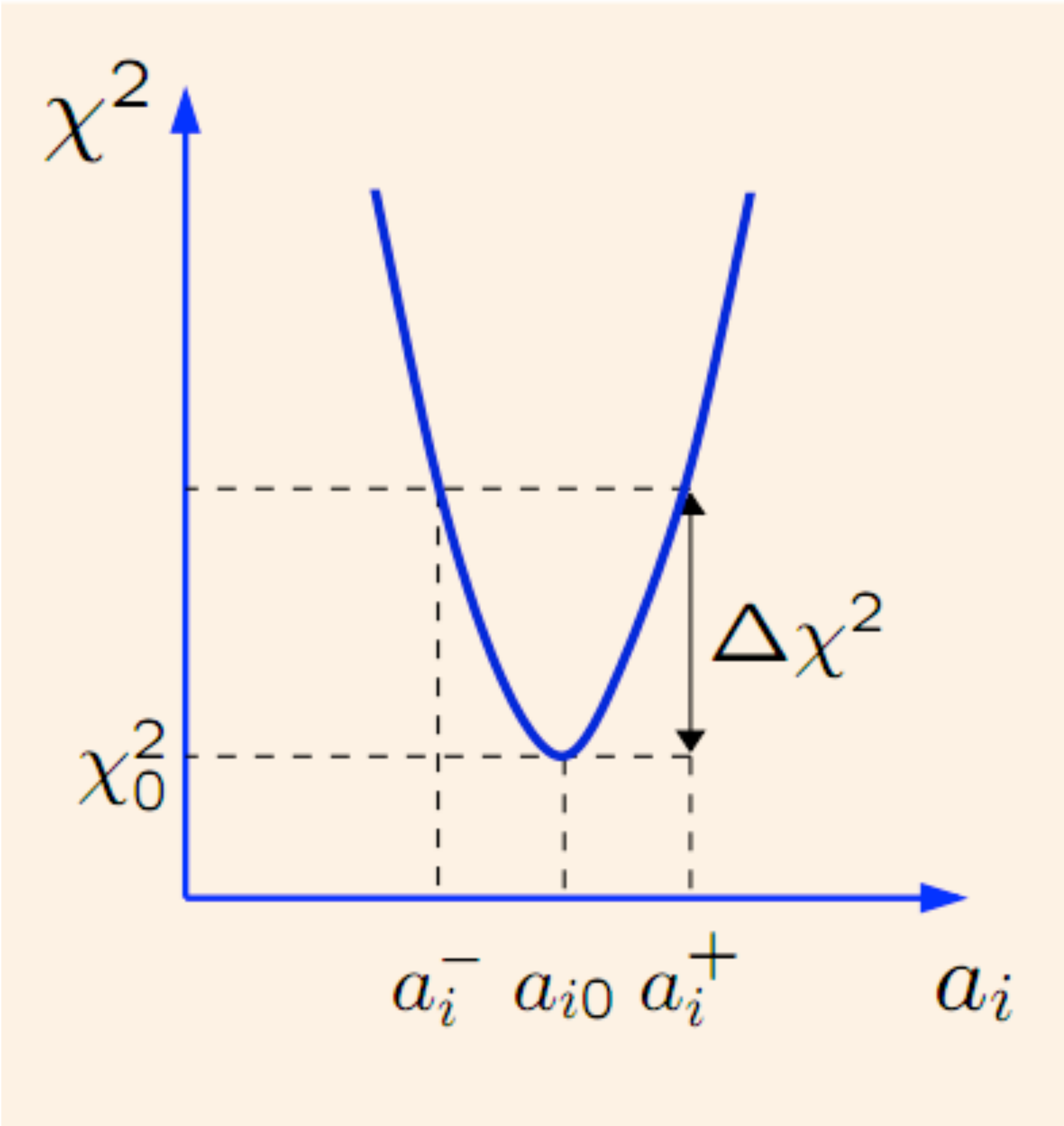
$$\sigma_X^2 = (H)_{ij}\, \partial_i X\, \partial_j X \xrightarrow{\text{diagonalisation}} \sigma_X^2 = |\vec{\nabla} X|^2$$

z_i eigenvectors of H with unit eigenvalues

➡ The total uncertainty is the sum in quadrature of the uncertainties due to each parameter

➡ $\Delta\chi^2 = \sum z_i^2$ the surfaces of constant χ^2 are spheres in the z space of radius $\sqrt{\Delta\chi^2}$

Hessian method



- According to textbook statistics, the 1σ contour in parameter space is given by

$$\Delta\chi^2 = 1$$

- Projection of the radius one sphere would give the uncertainty on parameters and on the PDFs, observables…

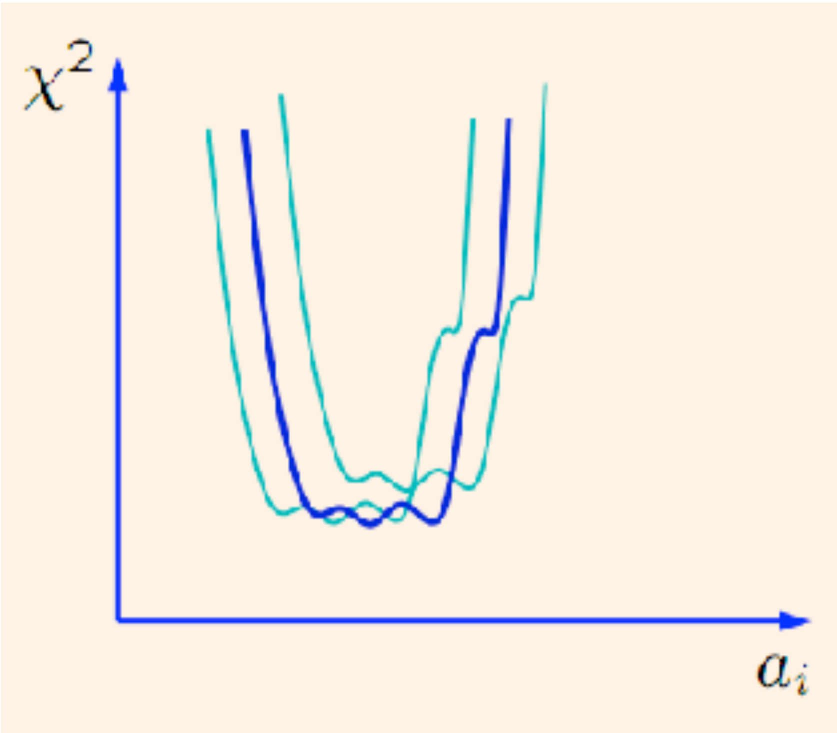
- The textbook statistics should work in case of perfectly compatible Gaussian errors

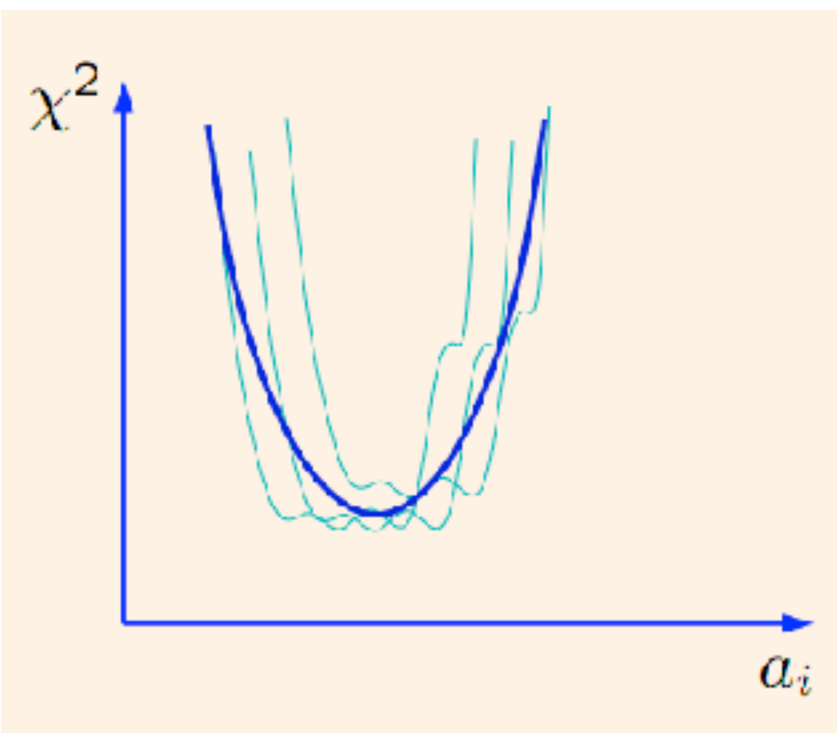
- But in practice, for global fits a tolerance is introduced

- NB: introducing a tolerance corresponds to blow up uncertainties by a factor $\sqrt{\Delta\chi^2}$

P. Nadolsky, CTEQ summer school 2009

Hessian method





The actual χ^2 function displays

- A well pronounced global minimum $\chi_0{}^2$
- Some tensions between datasets in the vicinity of the minimum
- Some dependence on assumptions about flat directions (= unconstrained combinations of PDF parameters)

The likelihood is approximately described by a quadratic χ^2 with a revised tolerance condition

$$\Delta\chi^2 \leq T^2$$

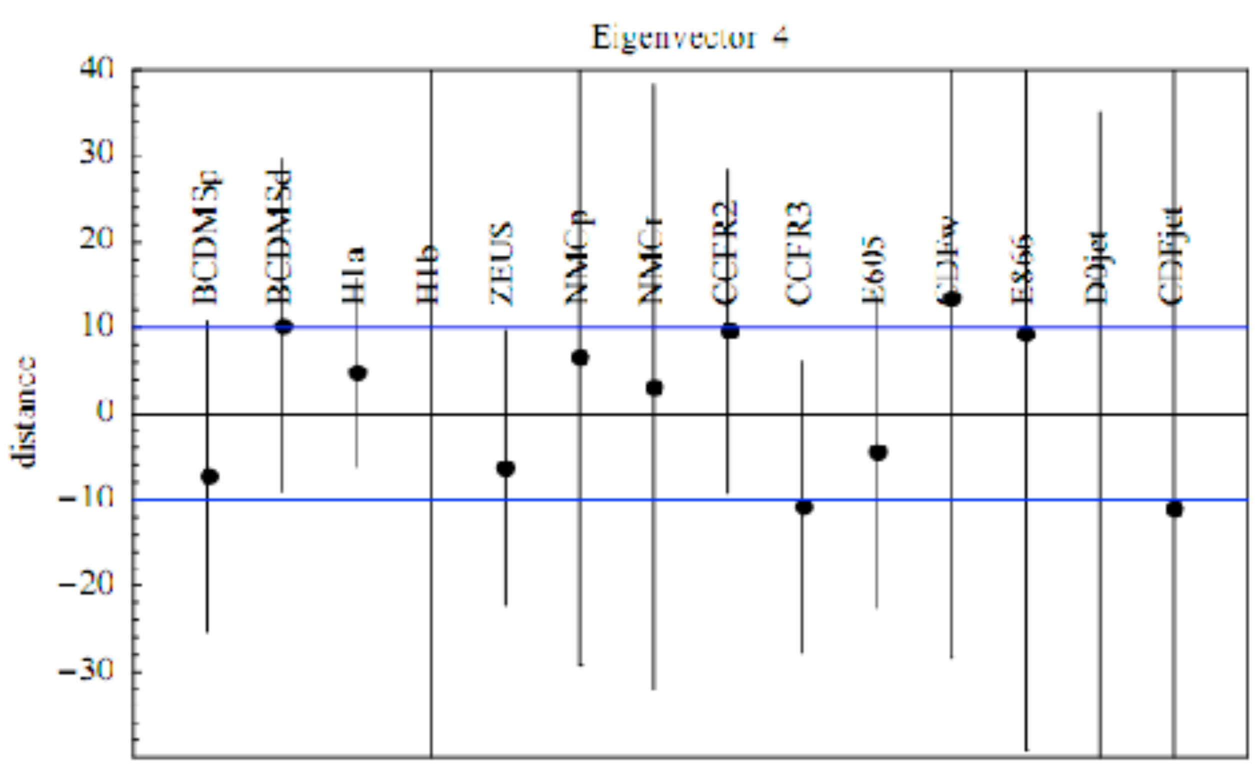
P. Nadolsky, CTEQ summer school 2009

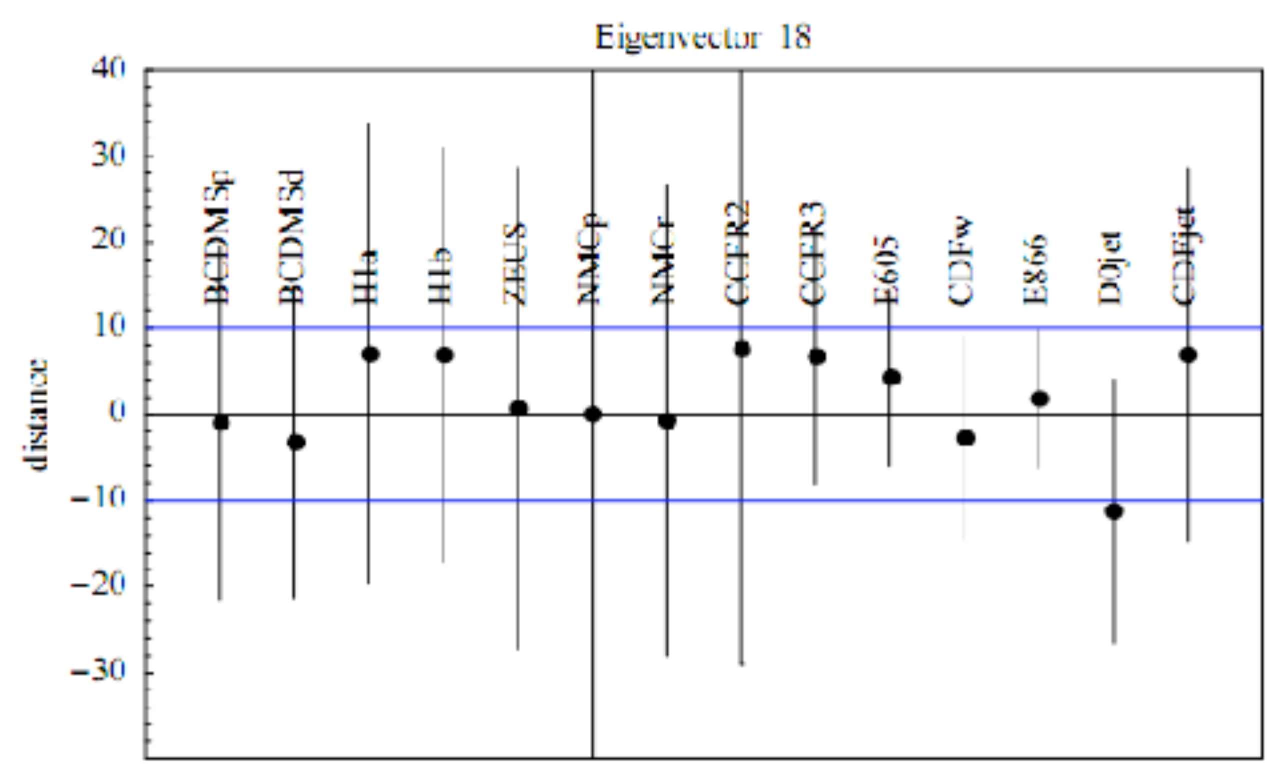
Hessian method

CTEQ6 tolerance criterion

- Acceptable values of PDF parameters must agree at ~ 90% C.L. with all experiments included in the fit, for a plausible range of assumptions about the PDF parametrisation, scale dependence, systematic uncertainty

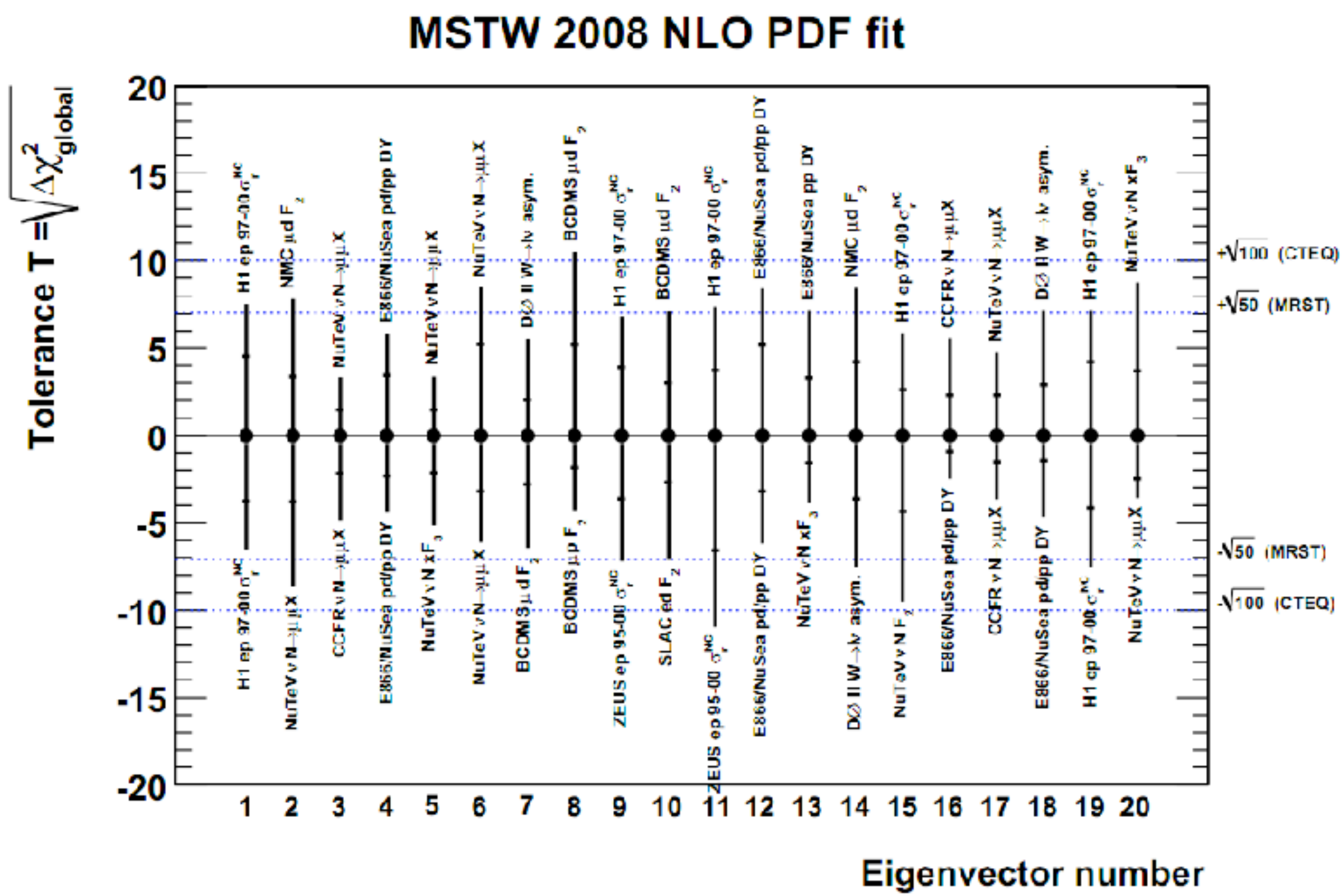
- Can be crudely approximated by assuming T ~ 10 for all PDF parameters





Hessian method

MSTW08 tolerance criterion



A dynamical tolerance, which varies according to the considered parameter

Monte Carlo method

- First idea by Giele Keller Kosover (hep-ph/0104052)
- Monte Carlo in parameter space

$$X(\vec{a})$$

MC sampling

$$\langle X \rangle = \int d\vec{a} X[\vec{a}] \mathcal{P}[\vec{a}]$$

P probability of parameter values

MC sampling in **parameter** space

$$\langle X \rangle \sim \frac{1}{N_{\text{rep}}} \sum_{i=1}^{N_{\text{rep}}} X(\vec{a}_i)$$

$$\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$$

Problem

How many replicas are needed?
Three bins per parameter ⇒ $3^{N_{par}}$ bins

E.g. for 23 parameters need more than 10^{11} replicas!!!

Monte Carlo method

- Forte, J. I Latorre, Piccione (hep-ph/0701127)
- First applied to structure functions then to PDFs

$$X(\vec{a})$$

MC sampling

$$\langle X \rangle = \int d\vec{a} X[\vec{a}] \mathcal{P}[\vec{a}]$$

P probability of parameter values

MC sampling in **data** space

$$\langle X \rangle \sim \frac{1}{N_{\mathrm{rep}}} \sum_{i=1}^{N_{\mathrm{rep}}} X(\vec{a}_i)$$

$$\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$$

Idea

Choose parameters along ∇X ⟺
Choose replicas of the data, i.e. work in the space of data and project back into PDF space

How many replicas does one need? 1-dim average of N_{rep} converges to true average with standard deviation $\sigma/\sqrt{N_{rep}}$

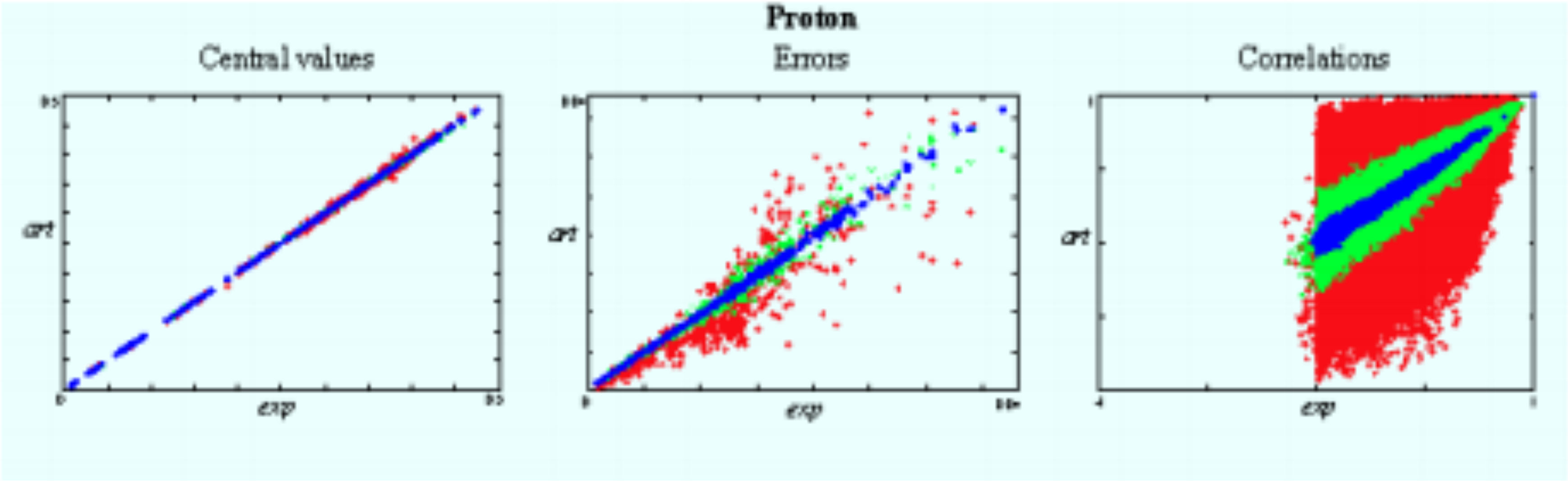
E.g. 10 replicas are enough for getting "true" central value with σ/3 accuracy

Monte Carlo method

- Generate artificial data according to distribution

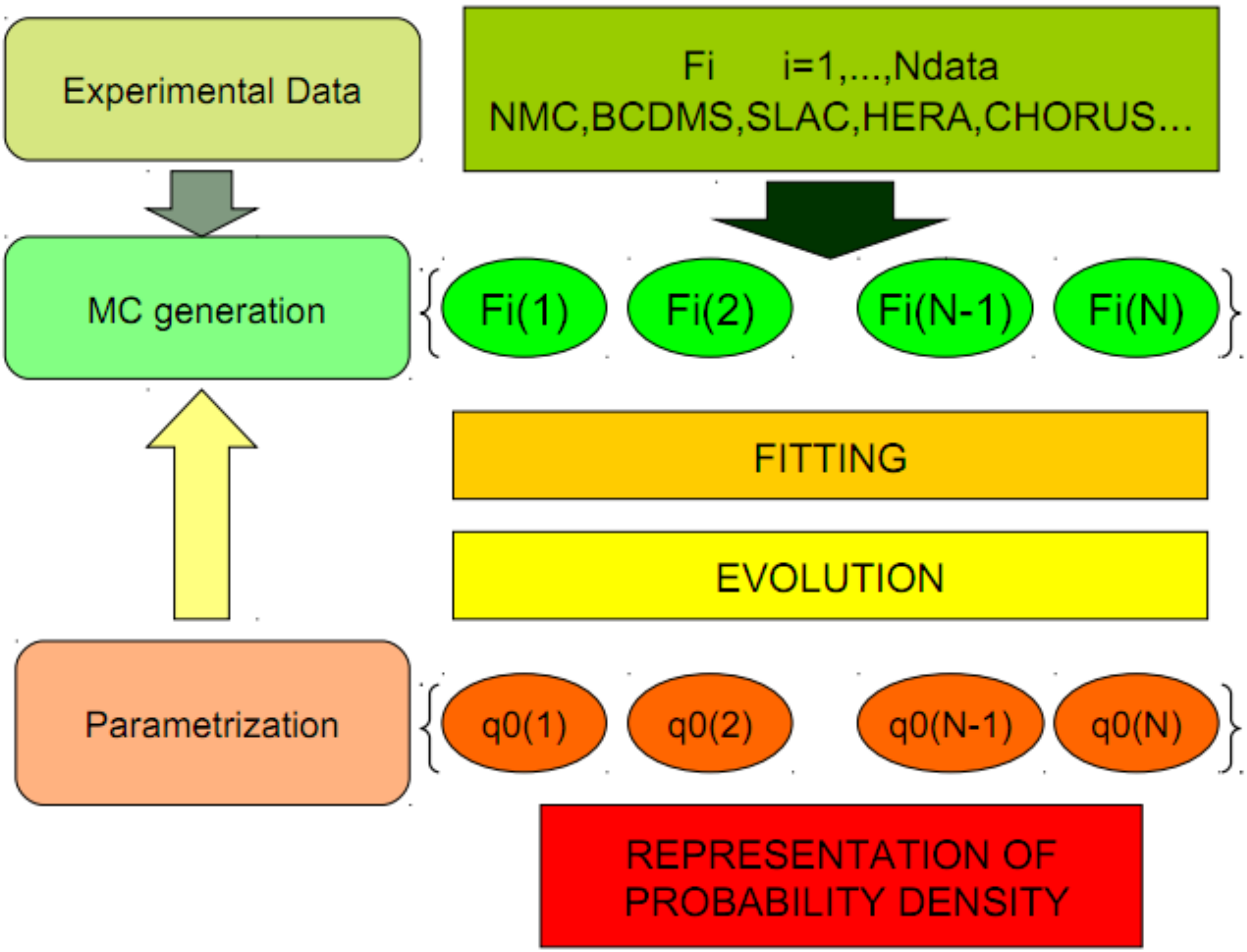
$$F_p^{(\mathrm{art})(k)} = S_{p,N}^{(k)} F_p^{(exp)} \left(1 + \sum_{l=1}^{N_c} r_{p,l}^{(k)} \sigma_{p,l} + r_p^{(k)} \sigma_{p,s}\right)$$

- r_i are univariate Gaussian random numbers such that if two points have correlated systematic uncertainties, they oscillate in the same directions
- S normalisation factors
- Validate Monte Carlo replicas against experimental data

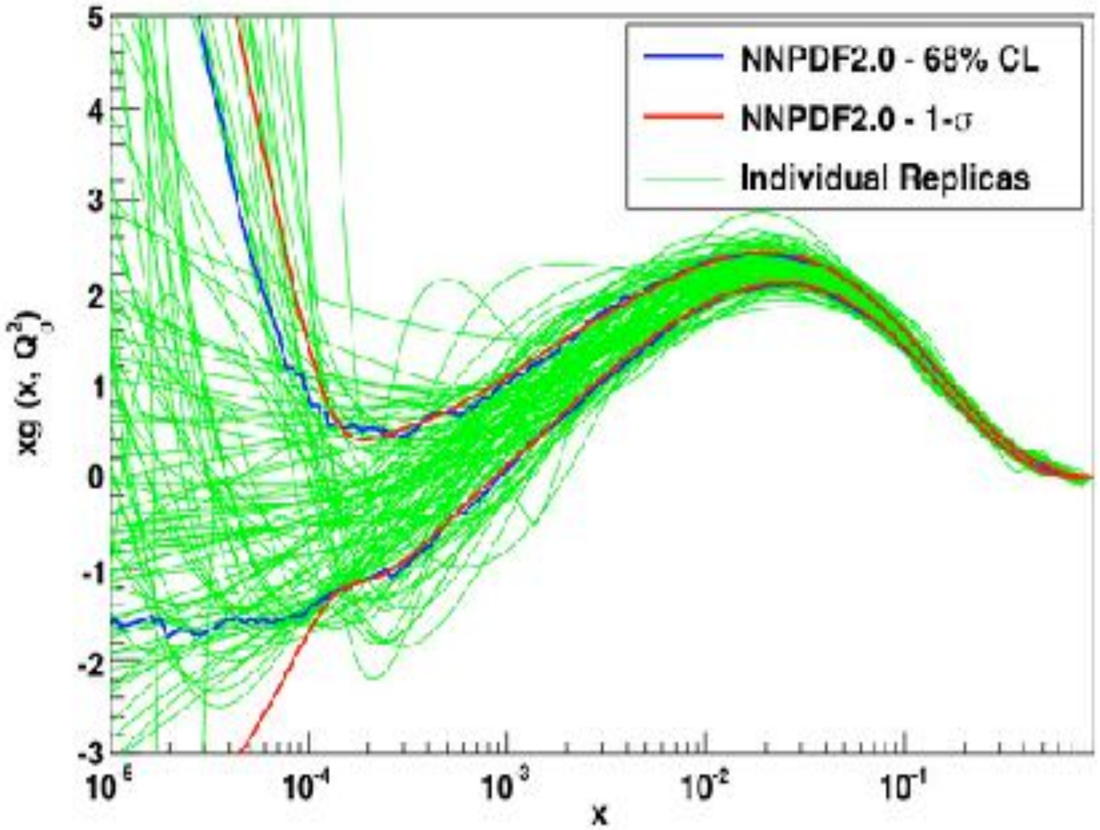


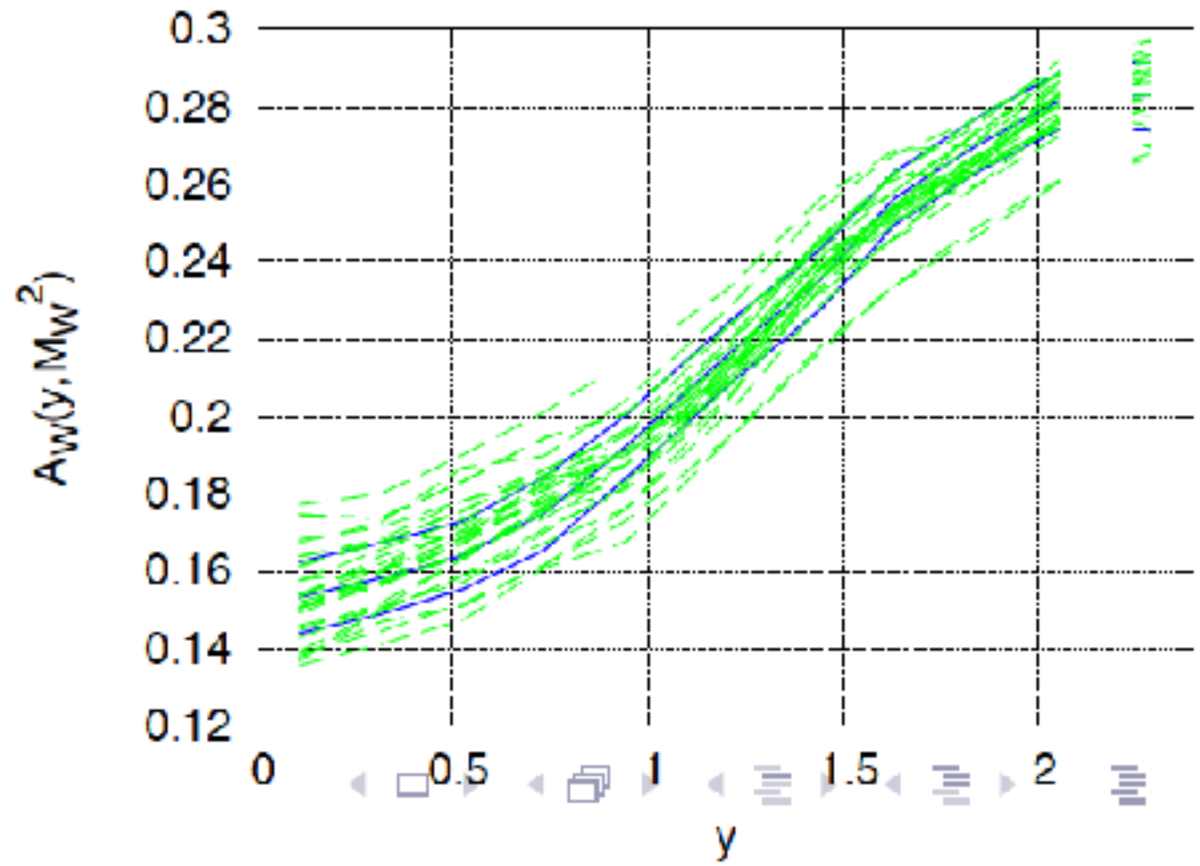
- Convergence rate increases with N_{rep}
- Correlations reproduced to % accuracy with 1000 reps

Monte Carlo method



Monte Carlo method





$$\langle f_J \rangle = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} f_J^{(k)}$$

$$\langle A(\{f\}) \rangle = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} A(\{f^{(k)}\})$$

$$\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$$

Individual replicas may fluctuate significantly, average quantities such as central values and 1σ error bands are smooth inasmuch as stability is reached due to the dimension of the ensemble increasing

The NNPDF solution

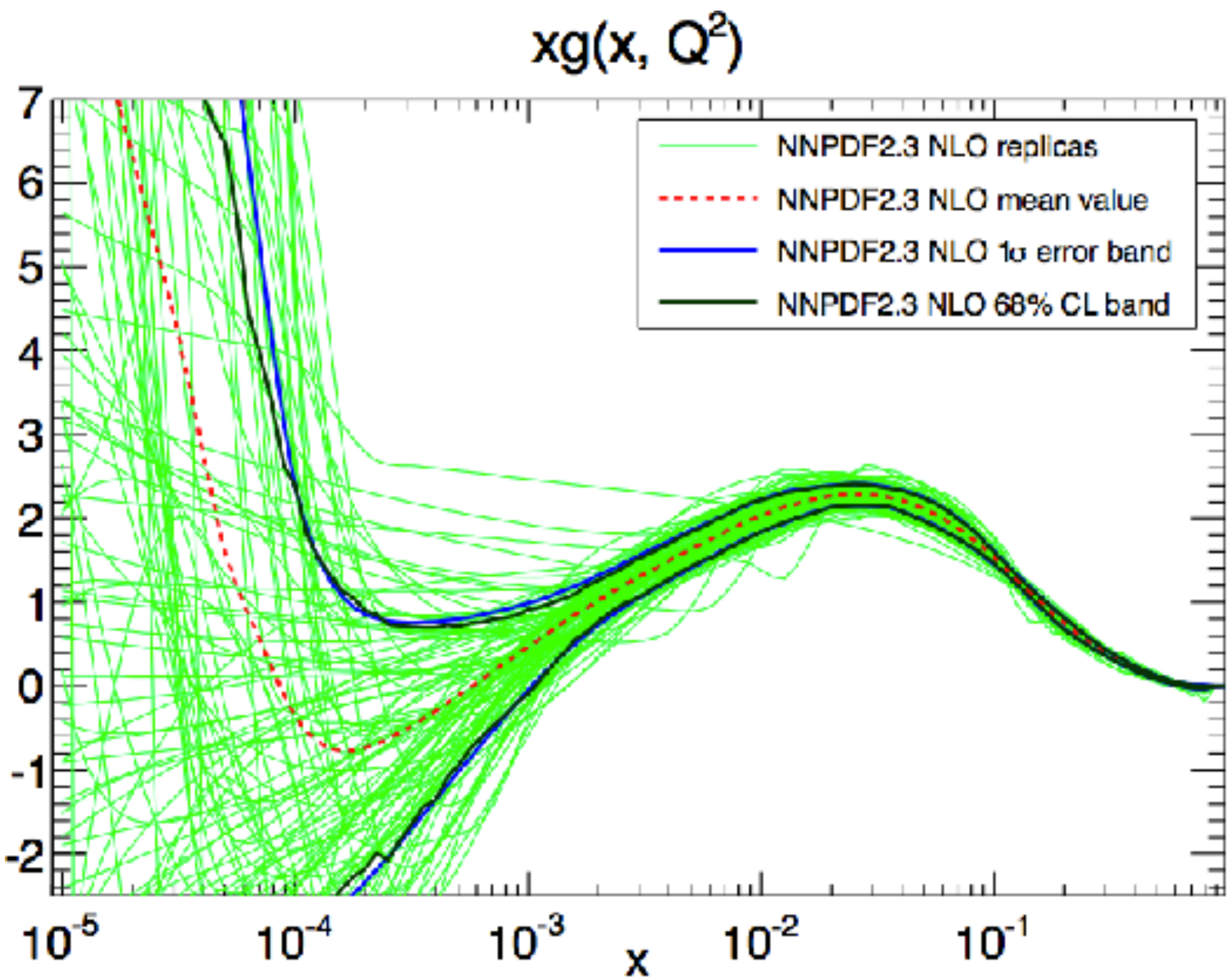
Monte Carlo sampling

Neural Network

http://nnpdf.mi.infn.it

- Fit of structure function **(2005)**
- DIS-only fit of PDFs **(2008)**
- First NNPDF global fit **(2010)**
- First fit including LHC data **(2013)**
- Closure test **(2016)**
- Fitted charm **(2018)**
- …

The NNPDF solution



The N(eural)N(etwork)PDFs:

- Monte Carlo techniques: sampling the probability measure in PDF functional space
- Neural Networks: all independent PDFs are associated to an unbiased and flexible parametrization: O(300) parameters versus O(30) in polynomial parametrization
- Genetic algorithm and cross-validation methods

✓ Precise error estimate not driven by theoretical prejudice
✓ No need to add new parameters when new data are included
✓ Statistical interpretation of uncertainty bands
✓ Possibility to include data via re-weighting: no need to refit

Summary for the user

Hessian method (CTEQ/TEA, MSTW, ABKM, HERAPDF)

$$\langle \mathcal{F} \rangle = \mathcal{F}[q^{(0)}]$$

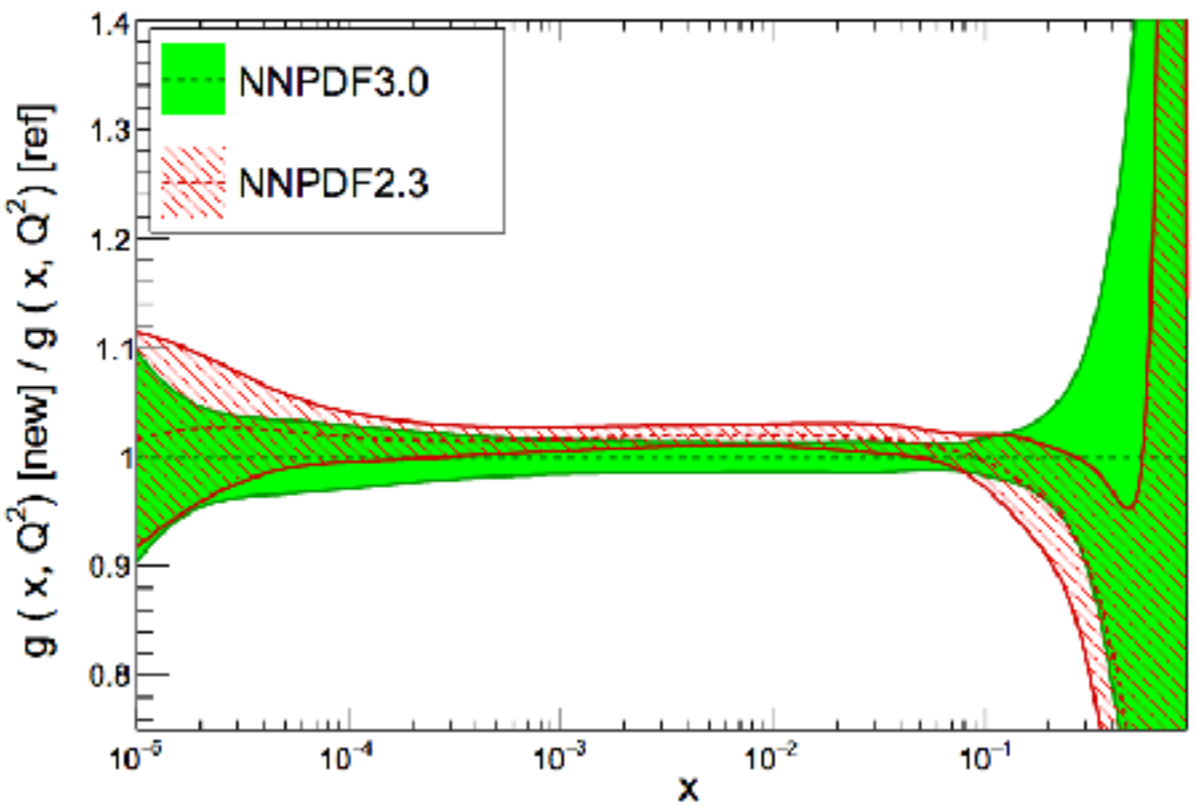
$$\sigma_{\mathcal{F}}^{\mathrm{Hess}} = \frac{1}{2} \left(\sum_{k=1}^{N_{\mathrm{set}}/2} \left(\mathcal{F}[\{q^{(2k-1)}\}] - \mathcal{F}[\{q^{(2k)}\}] \right)^2 \right)^{1/2}$$

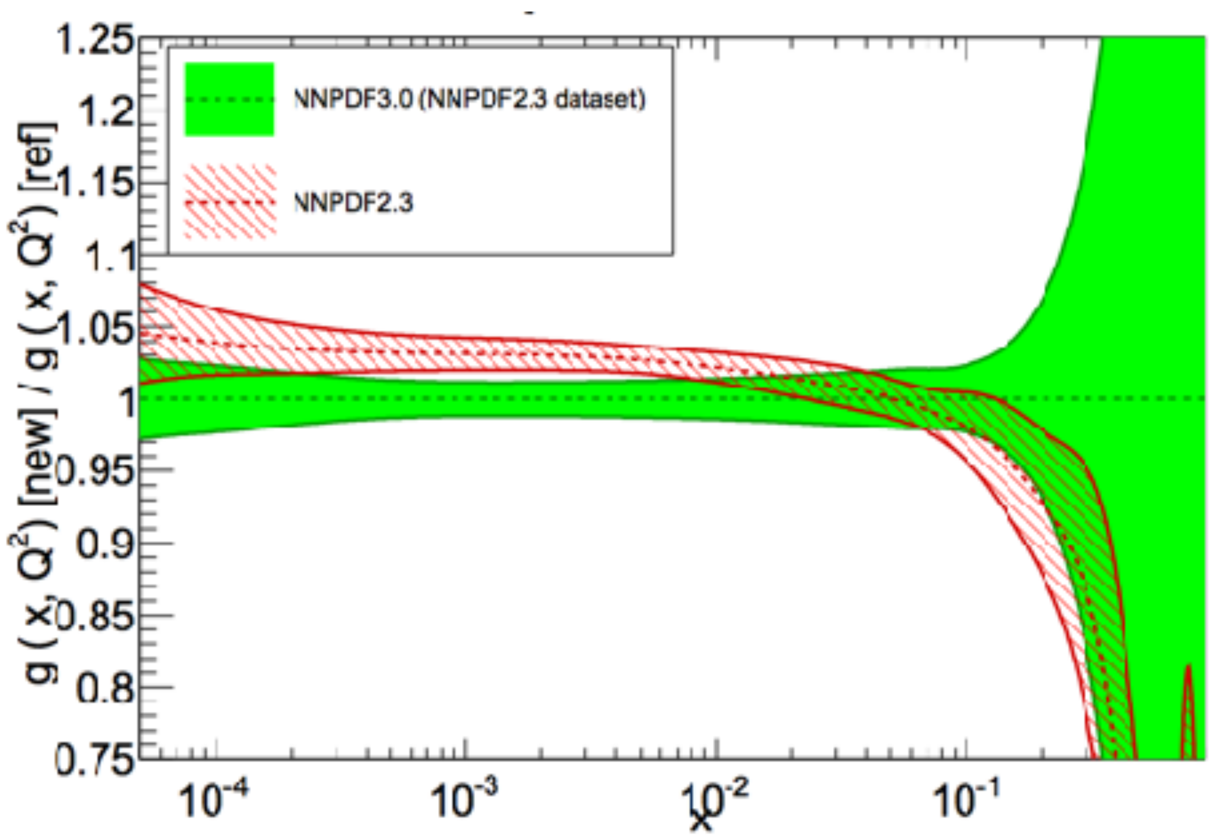
Monte Carlo method (NNPDF)

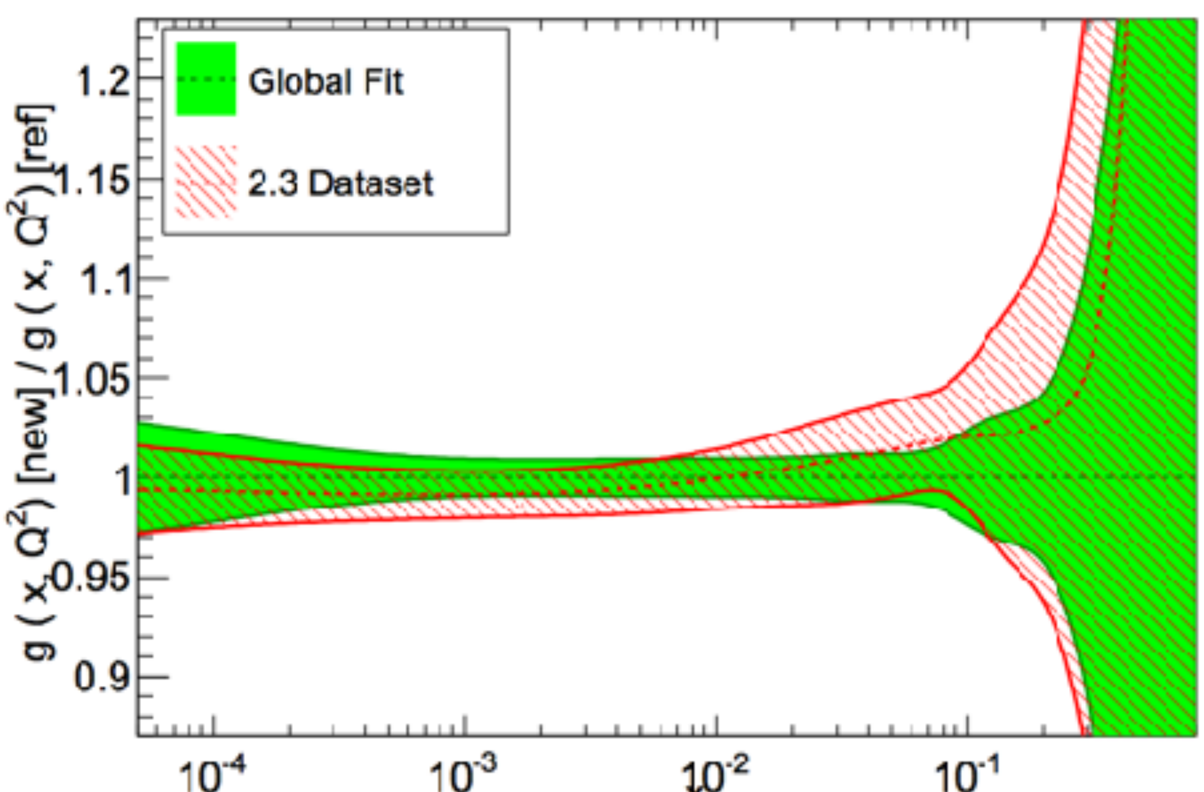
$$\langle \mathcal{F} \rangle = \frac{1}{N_{\mathrm{set}}} \sum_{i=1}^{N_{\mathrm{set}}} \mathcal{F}[q^{(i)}]$$

$$\sigma_{\mathcal{F}}^{\mathrm{MC}} = \left(\frac{1}{N_{\mathrm{N_{set}}}} \sum_{k=1}^{N_{\mathrm{set}}} \left(\mathcal{F}[\{q^{(k)}\}] - \langle \mathcal{F}[\{q\}] \rangle \right)^2 \right)^{1/2}$$

Key issue: methodology





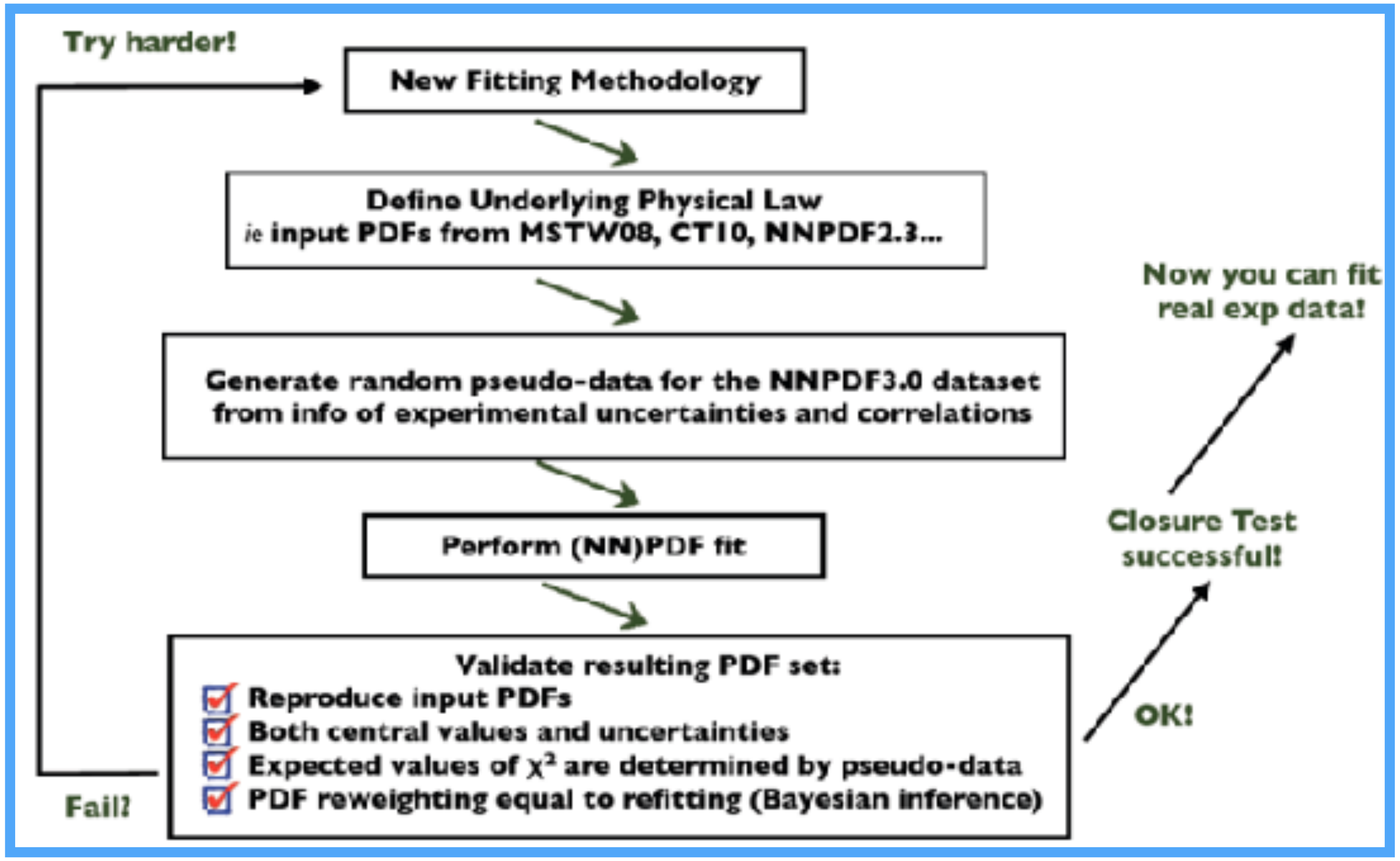


- NNPDF2.3 -> NNPDF3.0: included many new data (LHC and combined HERA) & change in fitting methodology (genetic algorithm and stopping criterion)
- Main changes in the gluon are due to the change in methodology
- How to make sure that we have a "perfect" methodology?

Statistical validation

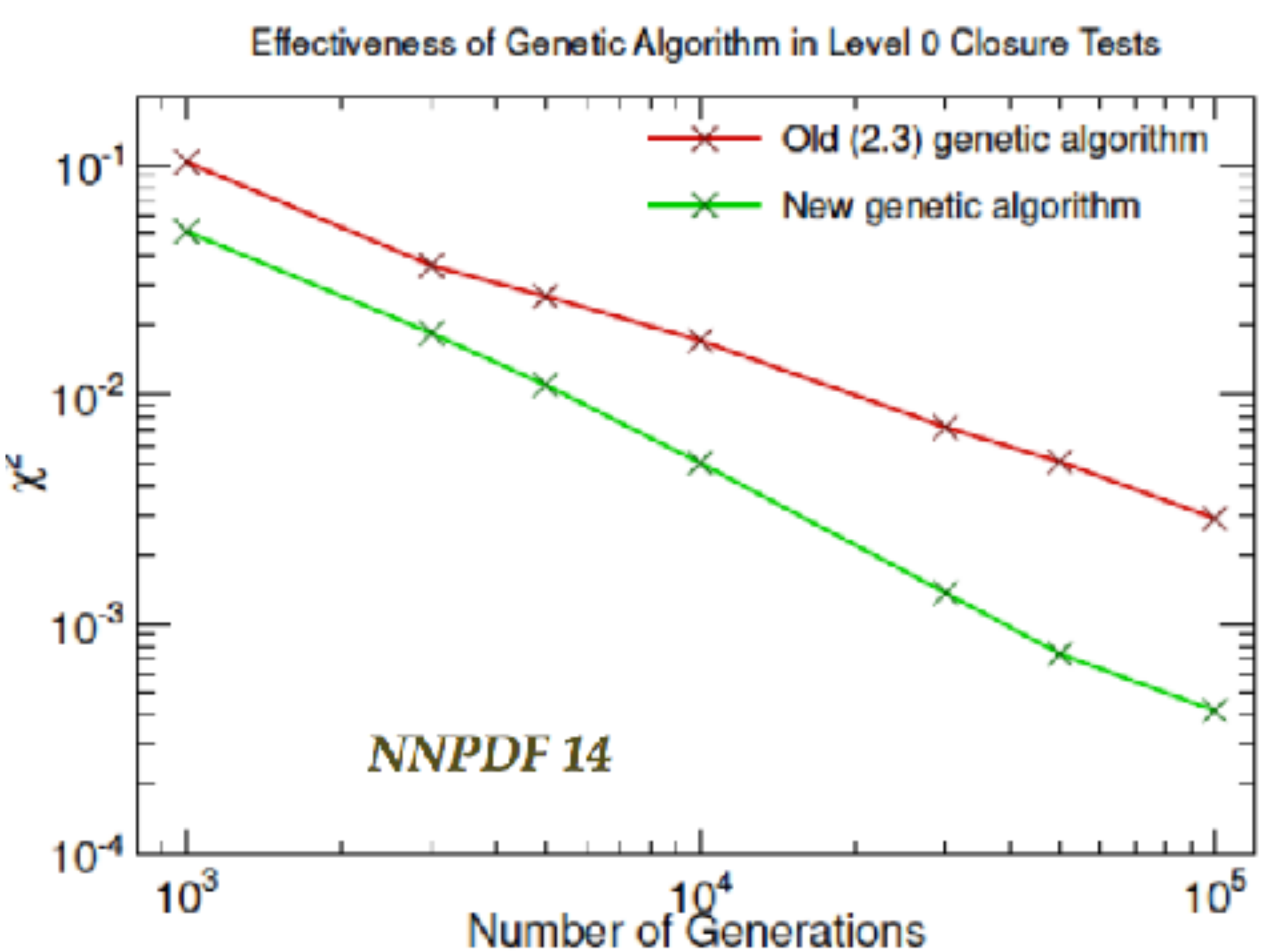
Closure test: the ultimate check of PDF fitting

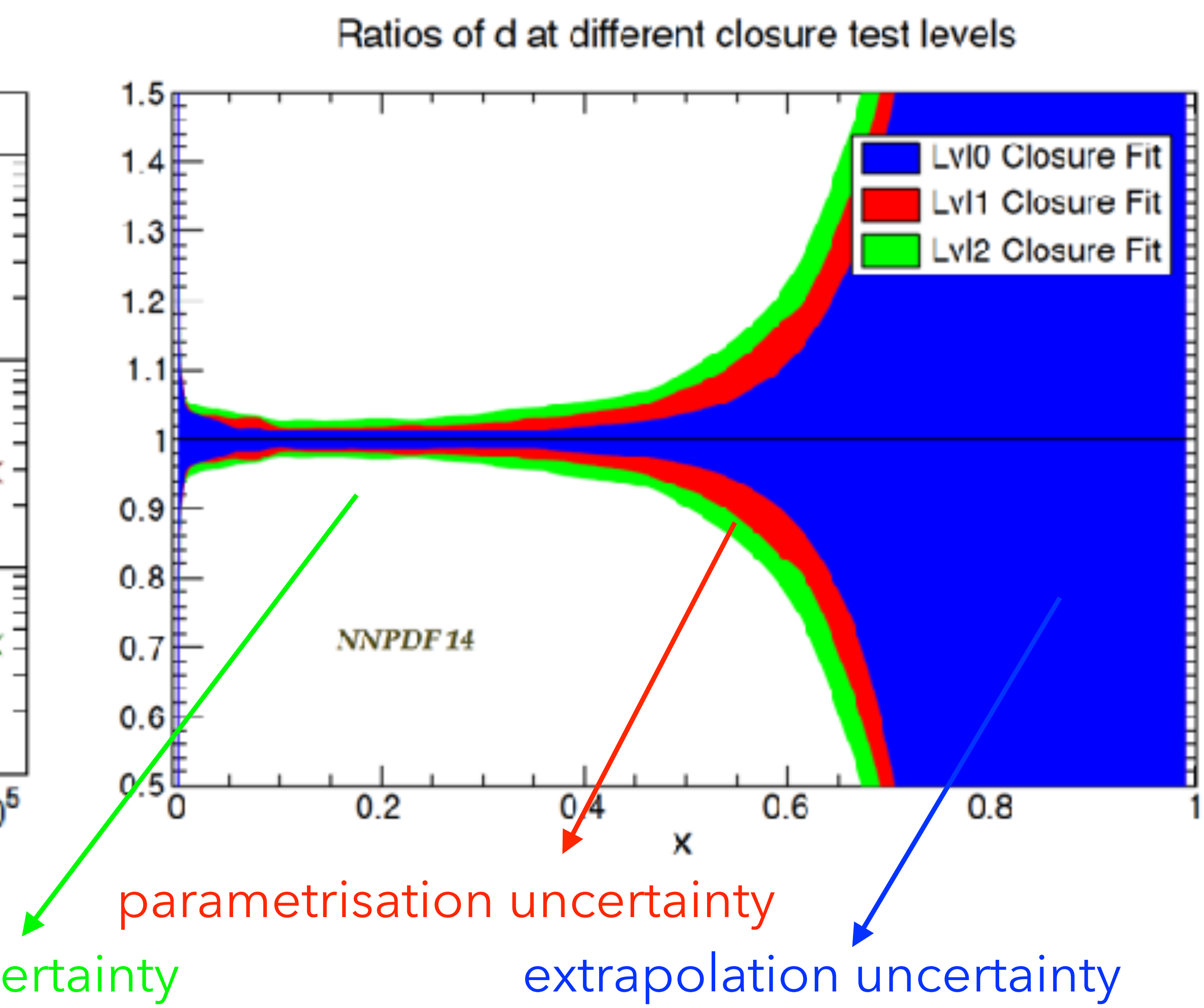
- Assume PDFs known: generate fake experimental data with them and th predictions
- Can decide data uncertainty (zero uncertainty - level 1-2, or as in real data - level 3)
- Fit PDFs to fake data
- Check whether fit reproduces the underlying "truth"
  - ➡ Check whether true values are gaussianly distributed about the fit
  - ➡ Check whether uncertainties are faithful
  - ➡ Trace different sources of uncertainty



Statistical validation

- **Level-0:** if pseudo-data are identical to the input theory, then agreement with theory should be arbitrarily good, i.e. $\chi^2 \to 0$ but PDF uncertainty $\to 0$ only in the region where there are enough data
- **Level-1:** add uncertainty to pseudo data equal to actually experimental uncertainties: replicas fit same data over and over again, then $\chi^2 \to 1$ and test equivalent minima (parametrisation Δ)
- **Level-2:** generate Monte Carlo replicas of pseudo-data with fluctuations, then $\chi^2 \to 2$ (data Δ)

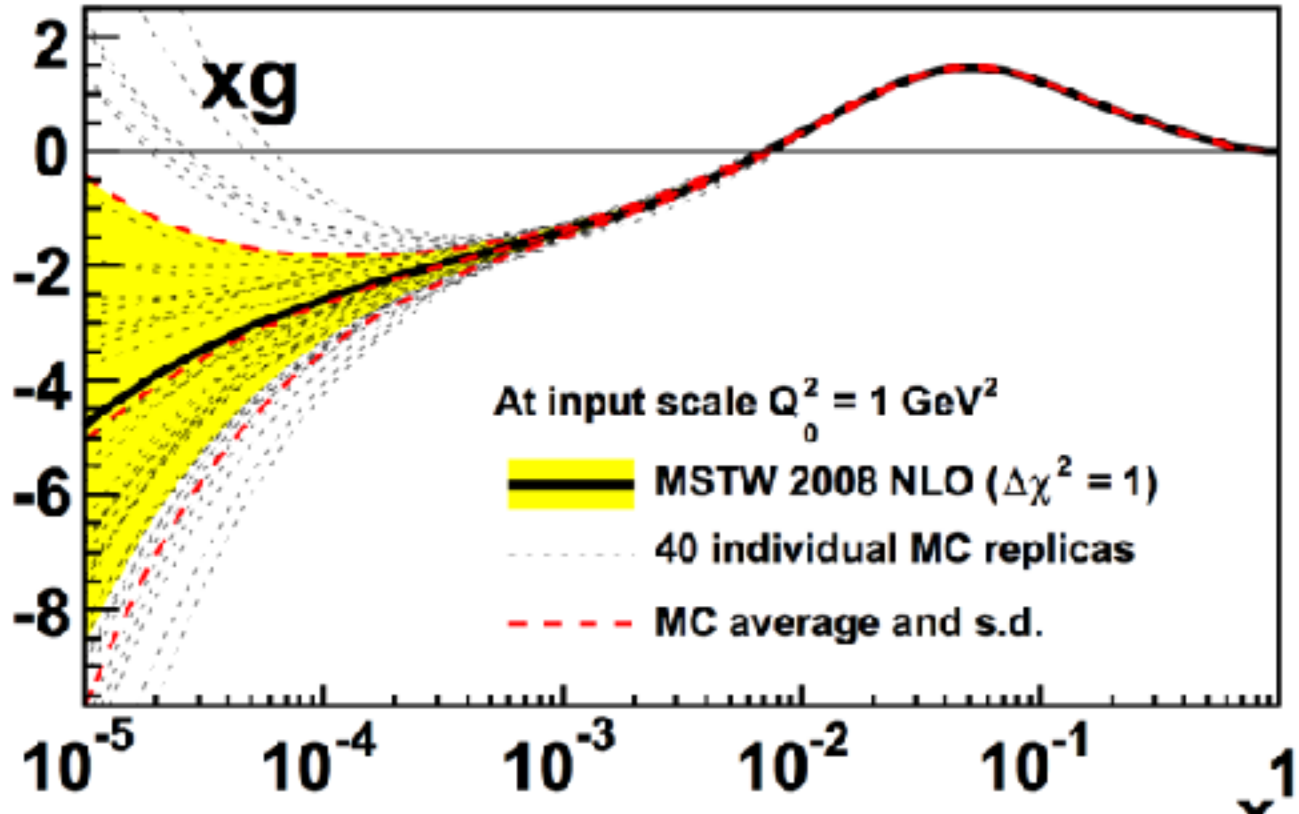




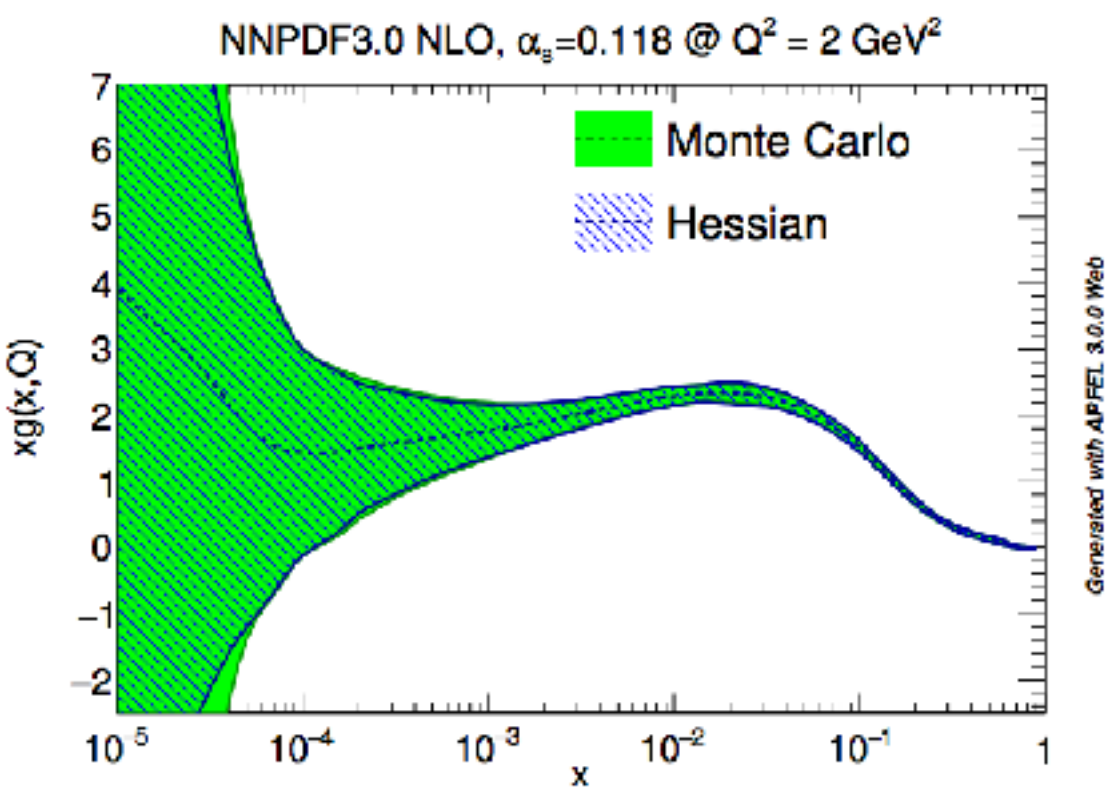
3 Δs comparable in data region

Hessian ⟺ Monte Carlo

- To convert Hessian into Monte Carlo, generate multi-gaussian replicas in the fitted parameters space
- Accurate when the number of replicas similar to that that reproduces the data



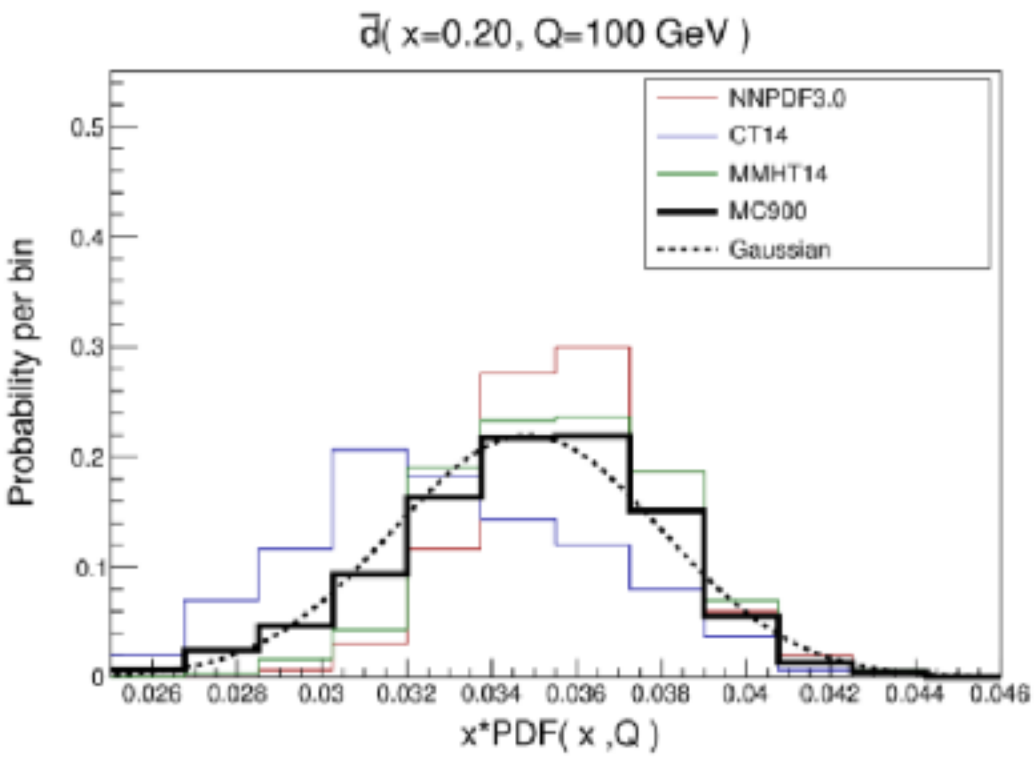
Thorne, Watt, 1205.4024



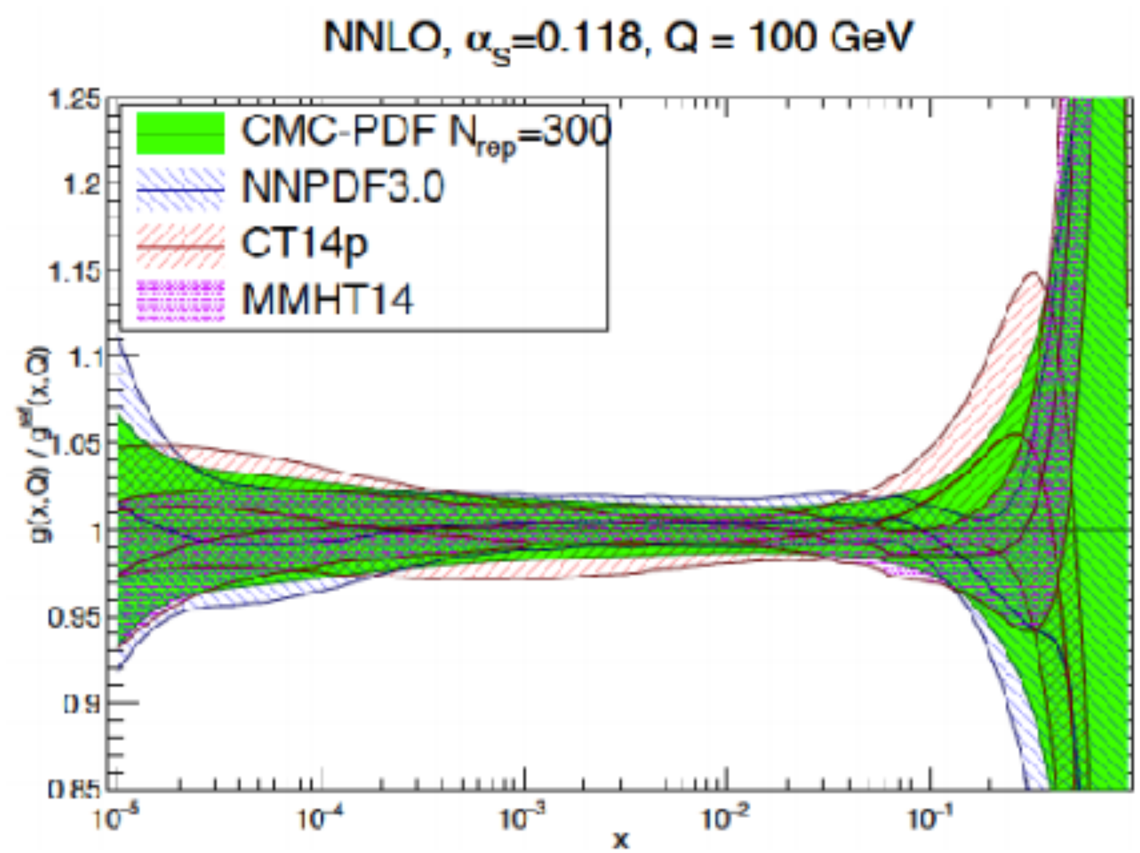
Carrazza et al 1505.06736

- To convert Monte Carlo into Hessian, sample the replicas f(x) at discrete set of points and construct the ensuing covariance matrix
- Eigenvectors of the covariance matrix as a basis in the vector space spanned by the replicas by the singular-value decomposition
- Number of dominant eigenvectors similar to numbers of replicas for accurate representation

Hessian ⟺ Monte Carlo



- Using Monte Carlo conversion of Hessian sets, can combine different PDF sets, combining MC replicas into a single set
- Useful for conservative estimate
- Combined set approximatively Gaussian



PDF4LHC15 recipe

- Monte Carlo combination of most recent global PDF sets **[Forte, Watt]**
- Each replica receives the same weight: uncertainty smaller than in the envelope, as in the latter outliers are given a larger weight
- New compression studies: N=40 replicas are virtually identical to the original 300 replicas from the point of view of correlation, standard deviation, observables **[Carrazza et al.]**

Statistics and methodology summary

➡ PDF determination: Hessian Method
- Simple linear error propagation
- Tolerance required for realistic uncertainties
- Parametrisation bias possible

➡ PDF determination: Monte Carlo method
- Two-step procedure: data MC -> PDF MC
- Very general parametrisation allowed
- Need optimal fit determination method (cross-validation)

➡ PDF representation: Hessian vs Monte Carlo
- Conversion possible either way
- Compression method available either way
- MC very flexible, Hessian very efficient

➡ PDF validation: the closure test
- Performed in the MC approach (so far)
- Interpolation and functional uncertainties significant