PDFs at the LHC

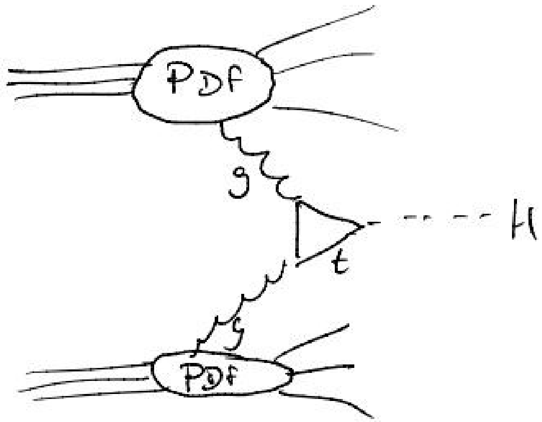
- Methods
- Challenges
- Results

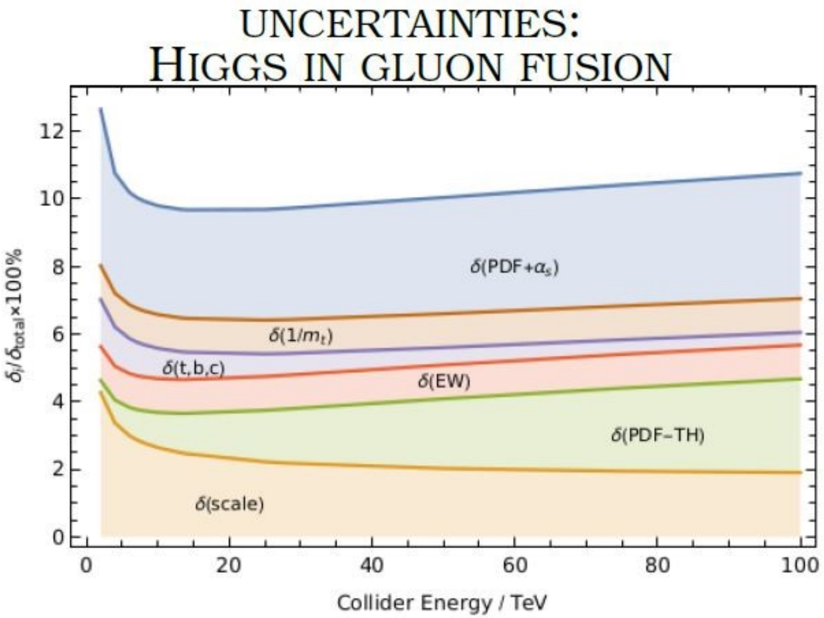
Richard Ball
Edinburgh

Christmas meeting 2022
Durham

Why PDFs?

Need accurate and precise PDFs to compute SM (and BSM) processes at LHC





- Compare SM predictions to LHC data e.g. *W, Z, H,* $t\bar{t}$, etc
- Extract physical parameters e.g. α_s, m_q, m_W, $sin^2\theta_W$ etc
- Search for new physics e.g. SUSY, SMEFT

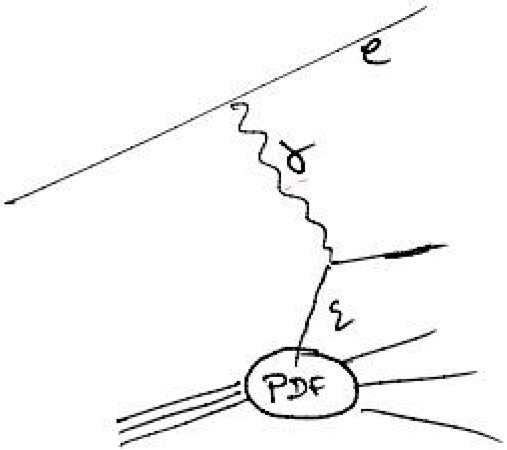
Currently: dominant uncertainty often PDF + α_s (few %)

Ultimate aim: 1% PDF uncertainties, to make the most of (HL)-LHC

PDF Determination

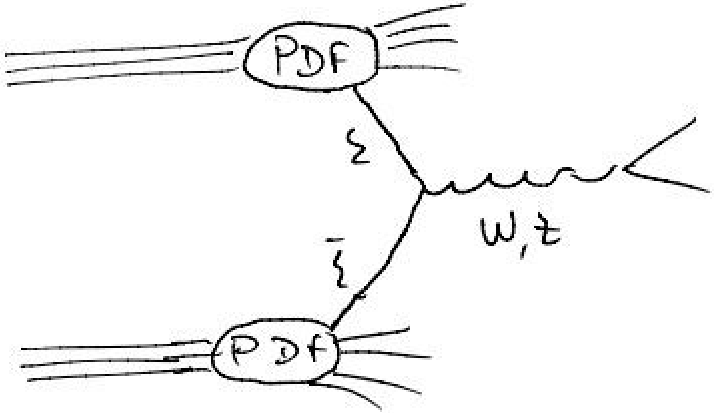
Factorization and Universality

DIS:



$$\sigma \sim C \otimes f$$

Hadronic:



$$\sigma \sim H \otimes f \otimes f$$

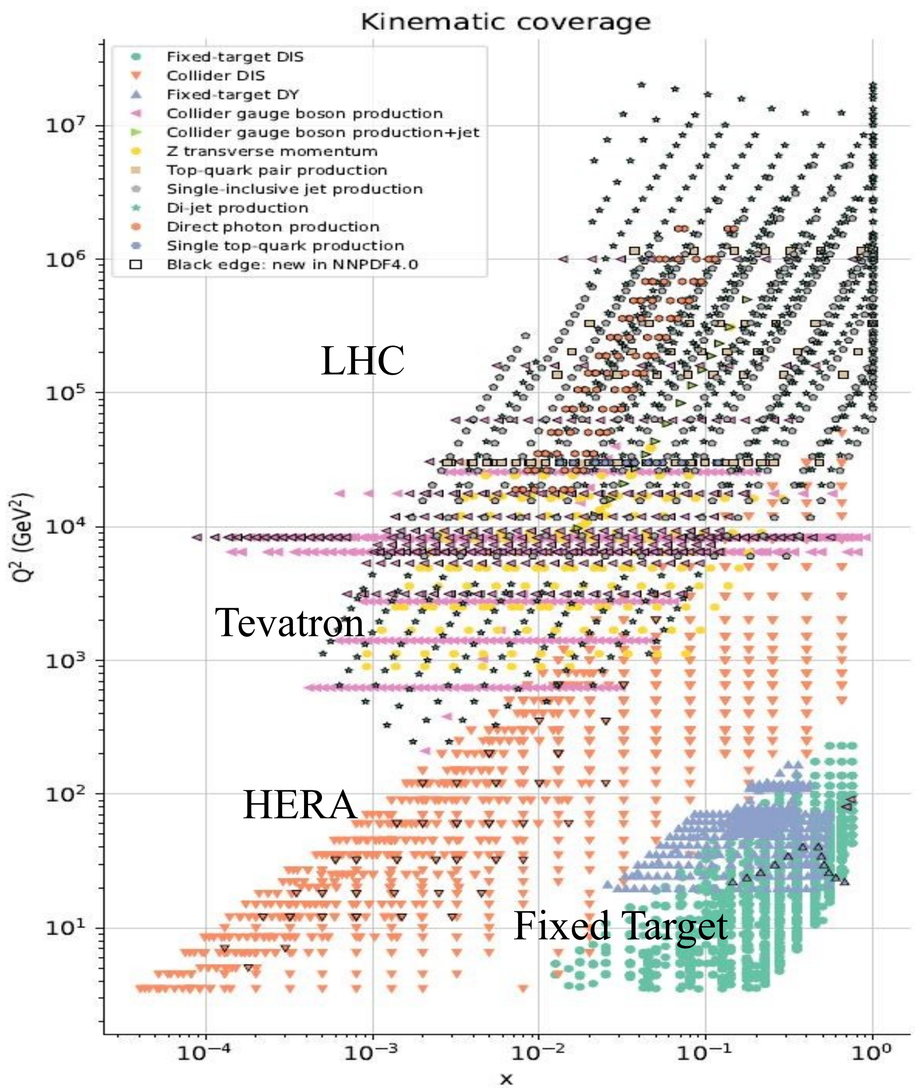
- C and H hard partonic cross-sections: process dependent: perturbative
- $f = \{g, u, \bar{u}, d, \bar{d}, s, \bar{s}, \ldots\}$ PDFs: process independent: nonperturbative
- Scale dependence: $\frac{\partial f}{\partial \ln Q^2} = P \otimes f$: P process independent: perturbative

So PDFs $f(x, Q_0^2)$ **nonperturbative, but universal:**

extract from global experimental datasets (DIS + Hadronic)

Global Datasets

	Process	Subprocess	Partons
Fixed Target	$\ell^{\pm}\{p,n\} \to \ell^{\pm} + X$	$\gamma^* q \to q$	$q, \bar{q}, g$
	$\ell^{\pm} n/p \to \ell^{\pm} + X$	$\gamma^* d/u \to d/u$	d/u
	$pp \to \mu^+\mu^- + X$	$u\bar{u}, d\bar{d} \to \gamma^*$	$\bar{q}$
	$pn/pp \to \mu^+\mu^- + X$	$(u\bar{d})/(u\bar{u}) \to \gamma^*$	$\bar{d}/\bar{u}$
	$\nu(\bar{\nu})N \to \mu^-(\mu^+) + X$	$W^* q \to q'$	$q, \bar{q}$
	$\nu N \to \mu^-\mu^+ + X$	$W^* s \to c$	s
	$\bar{\nu} N \to \mu^+\mu^- + X$	$W^* \bar{s} \to \bar{c}$	$\bar{s}$
Collider DIS	$e^{\pm} p \to e^{\pm} + X$	$\gamma^* q \to q$	$g, q, \bar{q}$
	$e^+ p \to \bar{\nu} + X$	$W^+\{d,s\} \to \{u,c\}$	d, s
	$e^{\pm} p \to e^{\pm} c\bar{c} + X$	$\gamma^* c \to c, \gamma^* g \to c\bar{c}$	c, g
	$e^{\pm} p \to e^{\pm} b\bar{b} + X$	$\gamma^* b \to b, \gamma^* g \to b\bar{b}$	b, g
	$e^{\pm} p \to \text{jet} + X$	$\gamma^* g \to q\bar{q}$	g
Tevatron	$p\bar{p} \to \text{jet} + X$	$gg, qg, qq \to 2j$	g, q
	$p\bar{p} \to (W^{\pm} \to \ell^{\pm}\nu) + X$	$ud \to W^+, \bar{u}\bar{d} \to W^-$	$u, d, \bar{u}, \bar{d}$
	$p\bar{p} \to (Z \to \ell^+\ell^-) + X$	$uu, dd \to Z$	u, d
	$p\bar{p} \to t\bar{t} + X$	$qq \to t\bar{t}$	q
LHC	$pp \to \text{jet} + X$	$gg, qg, q\bar{q} \to 2j$	g, q
	$pp \to (W^{\pm} \to \ell^{\pm}\nu) + X$	$u\bar{d} \to W^+, d\bar{u} \to W^-$	$u, d, \bar{u}, \bar{d}, g$
	$pp \to (Z \to \ell^+\ell^-) + X$	$q\bar{q} \to Z$	$q, \bar{q}, g$
	$pp \to (Z \to \ell^+\ell^-) + X, p_{\perp}$	$gq(\bar{q}) \to Zq(\bar{q})$	$g, q, \bar{q}$
	$pp \to (\gamma^* \to \ell^+\ell^-) + X$, Low mass	$q\bar{q} \to \gamma^*$	$q, \bar{q}, g$
	$pp \to (\gamma^* \to \ell^+\ell^-) + X$, High mass	$q\bar{q} \to \gamma^*$	$\bar{q}$
	$pp \to W^+\bar{c}, W^-c$	$sg \to W^+c, \bar{s}g \to W^-\bar{c}$	$s, \bar{s}$
	$pp \to t\bar{t} + X$	$gg \to t\bar{t}$	g
	$pp \to D, B + X$	$gg \to c\bar{c}, b\bar{b}$	g
	$pp \to J/\psi, \Upsilon + pp$	$\gamma^*(gg) \to c\bar{c}, b\bar{b}$	g
	$pp \to \gamma + X$	$gq(\bar{q}) \to \gamma q(\bar{q})$	g



Wide range of SM processes:

from $e^{\pm}p$, νN, pp, $p\bar{p}$ collisions

Kinematics: wide range of x and Q^2

Likelihood

$f(x) \to T[f]$ **Easy!** - just compute (at LO, NLO, NNLO, etc)

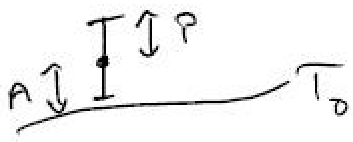
Then compare theory $T[f]$ to experimental data D:

$$P(D|f) \sim \exp(-\chi^2[f]/2)$$ ‘Likelihood’

$$\chi^2[f] = (D - T[f])^T C^{-1} (D - T[f])$$ C is experimental covmat

Note: if f_0 is ‘true’ PDF:

- $\|D - T[f_0]\| \sim$ ‘accuracy’ of the data, A
- $\|C\|^{1/2} \sim$ ‘precision’ of the data, P
- data are ‘faithful’ if $A \sim P$, i.e. $\chi^2[f_0] \sim N_{\rm dat}$



But we can never **know** f_0 : the best we can hope for is $P(f|D)$

Moreover mapping $D \to f(x)$ is ill-defined: D are discrete, $f(x)$ is a function

Bayes Theorem

Bayes, c.1760

How can we determine $P(f|D)$?

$$P(f|D) \propto P(D|f)P(f)$$

Likelihood: we can compute this

'Prior': how can we possibly find this???

Two fundamentally distinct approaches to finding the prior:

- 'Modelling': e.g. assume $f(x) \sim x^a(1-x)^b$

  and that $P[f] = 1$ in this space of functions, zero outside

  fit parameters $\{a, b\}$ by maximising the likelihood

- 'Machine Learning': NN + CV + MC (+ CT) + HO

  no model: use data to also infer probable 'smoothness' of $f(x)$

  and thus infer $P[f]$ throughout the space of functions

Modelling PDFs

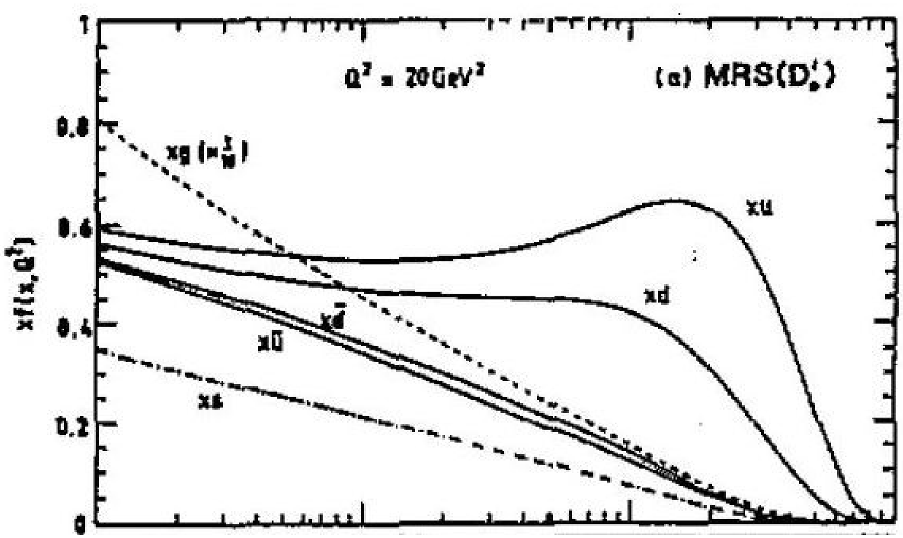
Choose the prior by hand

Parametrising PDFs

Typically: $$f(x) \sim x^a (1-x)^b Poly(\sqrt{x})$$ '*Poly*' = quadratic, Chebyschev, Bernstein

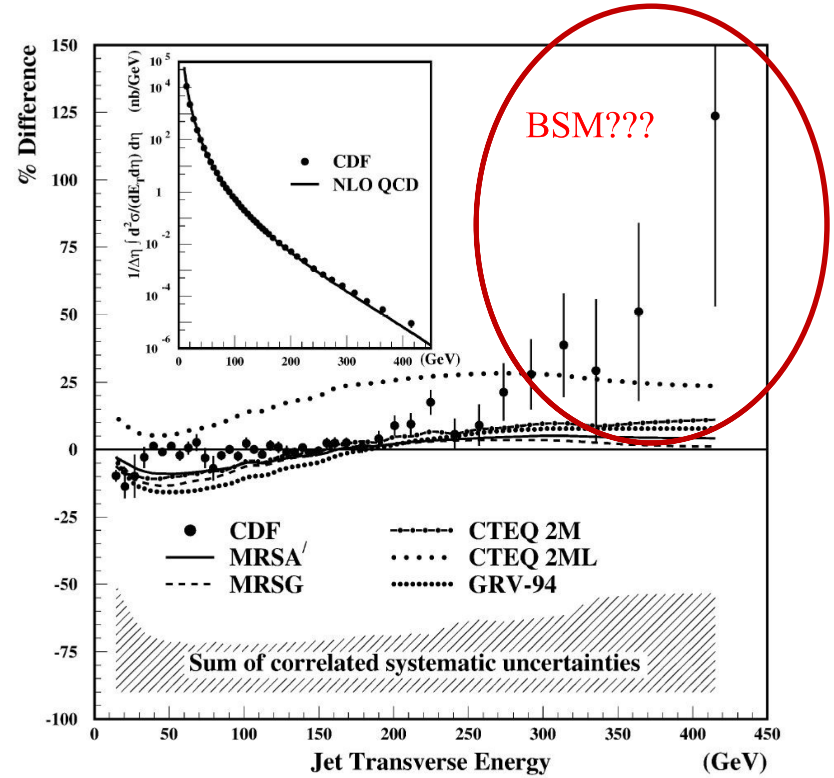
'Art' : choose suitable functions: not too many parameters, not too few

- Assume $P[f]$ has uniform support in this space of functions
- Maximise $P(f|D)$ by maximising $P(D|f)$: Maximum Likelihood
- Gives 'best fit' PDF : minimises $\chi^2[f]$ in parameter space $\{a, b, \ldots\}$



Martin, Roberts and Stirling (MRS), c.1993

NLO, VFNS, ~ 20 free parameters



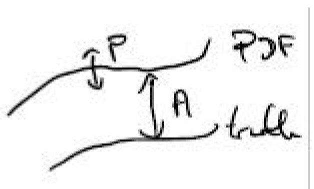
Need PDF Uncertainties!

CDF incl jets c.1995

Hessian Uncertainties

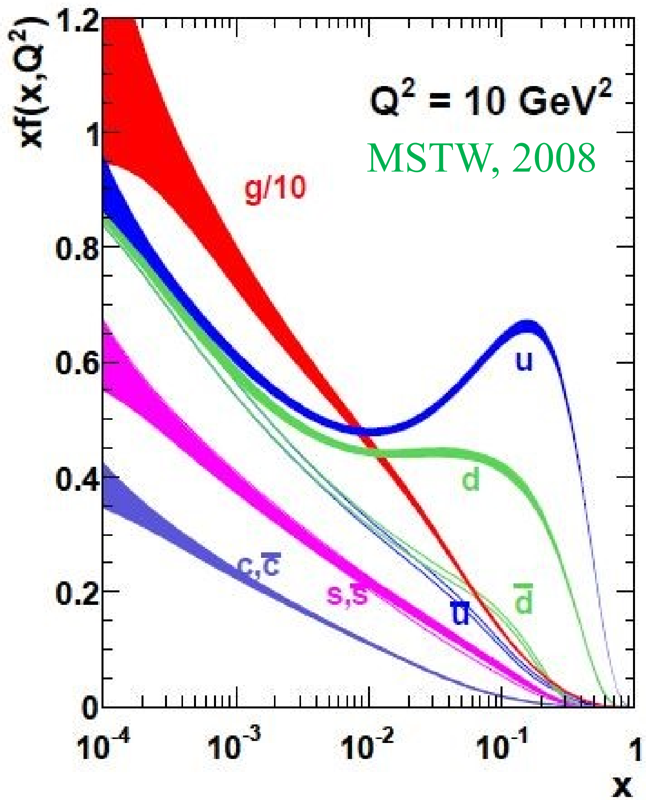
Propagate data uncertainties into PDF parameters: $\Delta\chi^2 = 1$
- gives Gaussian $P(f|D)$

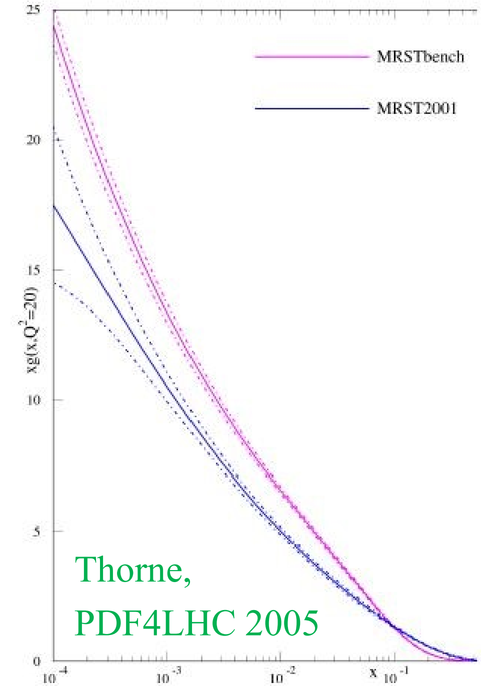
Problem: precision ≪ accuracy : PDF unc. not faithful



Reasons:

- data inconsistent (exp unc too small)?
- modelling overconstrains $P[f]$:
  reduces PDF uncertainty





Solution: inflate exp unc by a factor T: $\Delta\chi^2 = T^2$
so that precision ~ accuracy   CTEQ, 2002

Call T ‘Tolerance’. Typically $T \sim 5 - 10$

More sophisticated: ‘Dynamical Tolerance’:
tune each evec separately

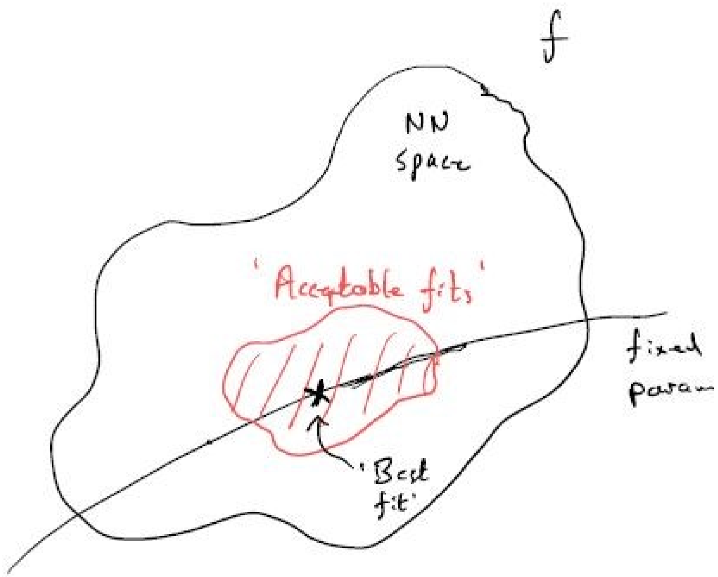
NLO, NNLO: GMVFNS:
~ 30 – 40 free parameters

Machine Learning PDFs

Let the data choose the prior

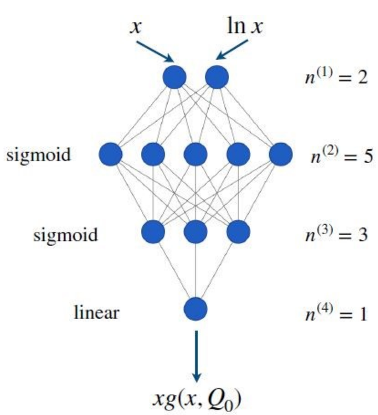
Forte, Latorre et al, 2002

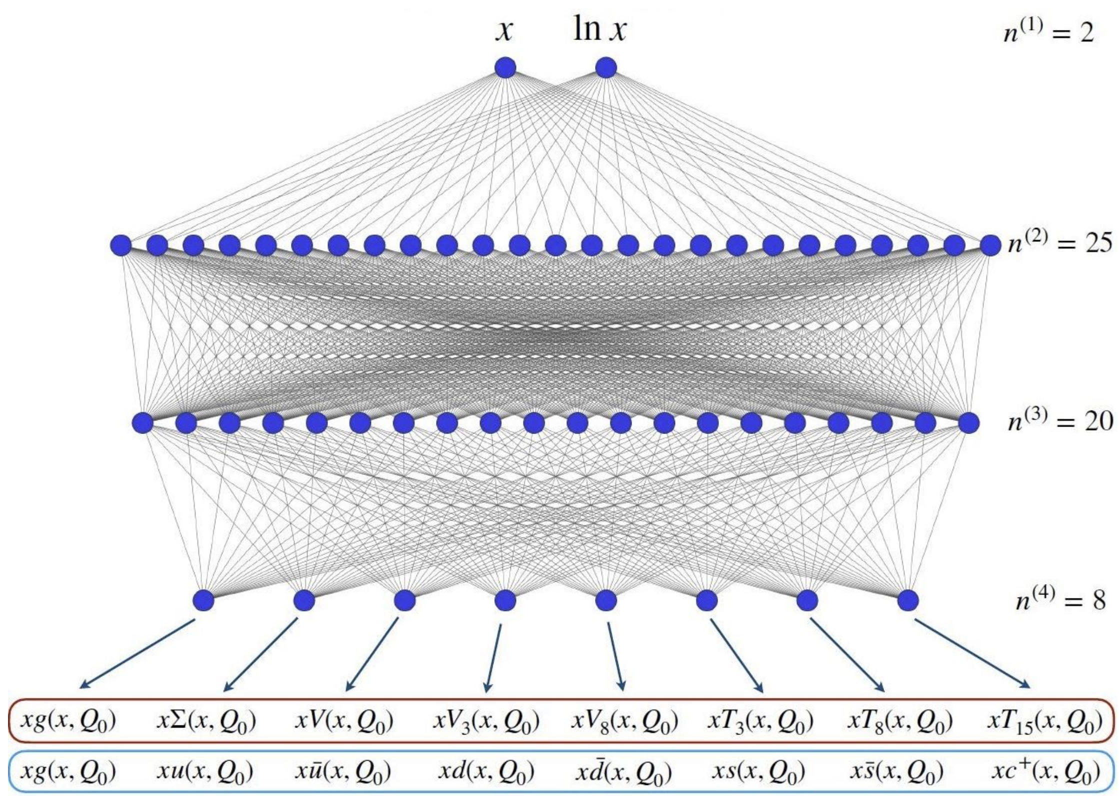
Step 1: Neural Networks



Basic idea: choose a parametrization so large that it can fit any conceivable PDF

Eliminates bias: if see any sign parametrization too small, just make the network even bigger!

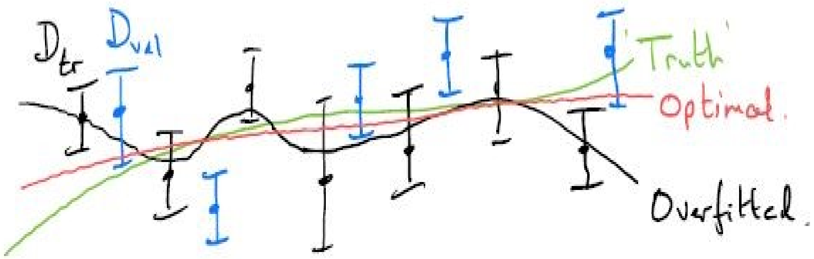




Old NN architecture (up to NNPDF3.1): 296 free parameters (37 for each PDF)

New NN architecture (NNPDF4.0): 763 free parameters

Step 2: Cross-Validation



Fit NN to experimental data using $\chi^2[f]$

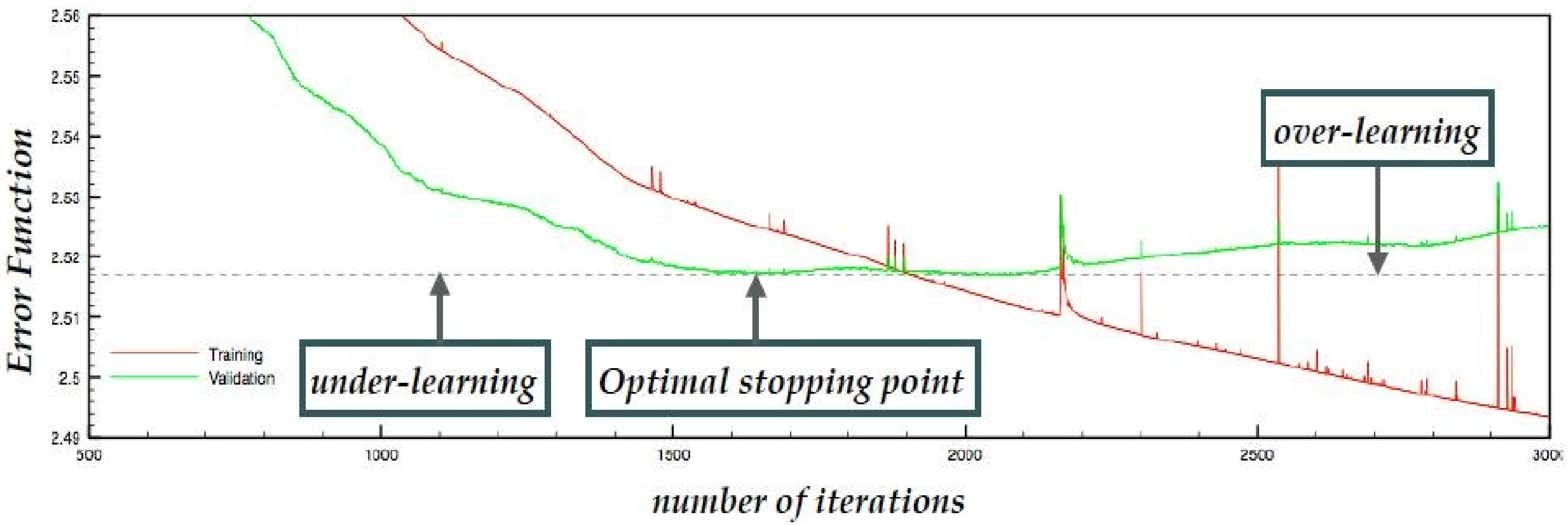
Problem: NN can fit anything!

– large number of redundant parameters

– must avoid fitting random data fluctuations

Solution: 'Cross-validation': $\{D\} \to \{D_{\rm tr}, D_{\rm val}\}$

: 'training' set and 'validation' set



- 'Optimal' fit is not the 'best fit': stop before $\chi^2_{\rm tr}$ is too low, to avoid overfitting
- 'Optimal' fits are not unique: space of NN very big. 'Functional Uncertainty'
- 'Optimal' fit is smoother (and thus closer to 'truth') than overfits

Step 3: Monte Carlo Replicas

Giele & Kosower, 1998

Propagate data unc into PDF unc

Hessian not much use (redundant parameters): instead

- Generate random 'data replicas' $\{D^r\}$:

$$\langle D^r \rangle = D \qquad \langle (D^r - D)(D^r - D)^T \rangle = C$$

- Fit NN to each data replica: $D^r \to f^r$ 'PDF replicas'

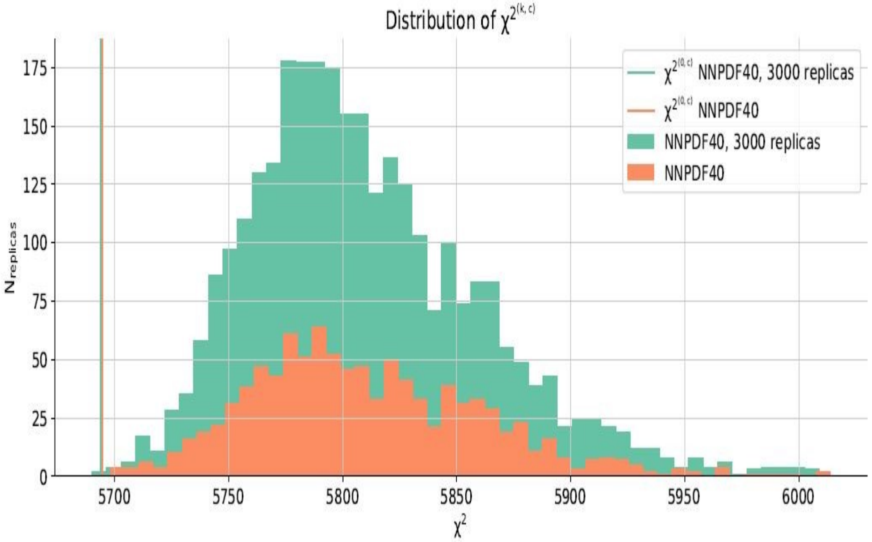
    using cross-validation: random tr/val split, random initial seed

- Each data replica equally likely, so each PDF replica equally likely:

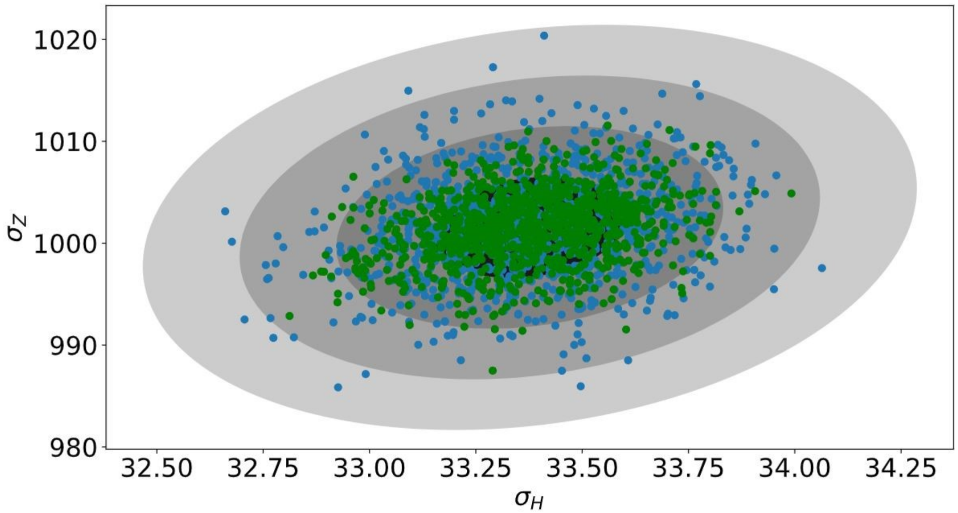
$$T_0 = \langle T[f^r] \rangle \qquad \mathrm{Cov}T = \langle (T[f^r] - T_0)(T[f^r] - T_0)^T \rangle$$

The PDF replicas $\{f^r\}$ give importance sampling of $P(f|D)$

Importance sampling of χ^2



Importance sampling in plane of H and Z tot xsecs



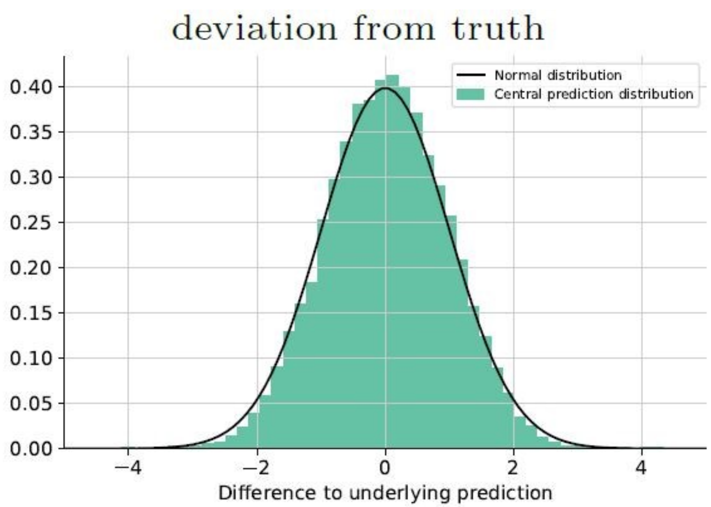
Total Uncertainty vs Functional Uncertainty

Uncertainties ~ 100 replicas: Correlations ~ 1000 replicas

Closure Testing: Trust but Check!

- Choose a 'prior PDF' f_0 : anything you like, within reason: assumed truth
- Generate 'perfect data' D_0 from f_0 : $D_0 = T[f_0]$: no theory inconsistencies
- Generate perfect data replicas $\{D_0^r\}$ using experimental covmat C: no data inconsistencies
- Fit NN in usual way, with CV: $D_0^r \to f_0^r$
- Check PDFs faithful, i.e. that accuracy ~ precision:

$$\|\langle f_0^r \rangle - f_0\| \sim \|\langle (f_0^r - \langle f_0^r \rangle)^2 \rangle\|^{1/2}$$

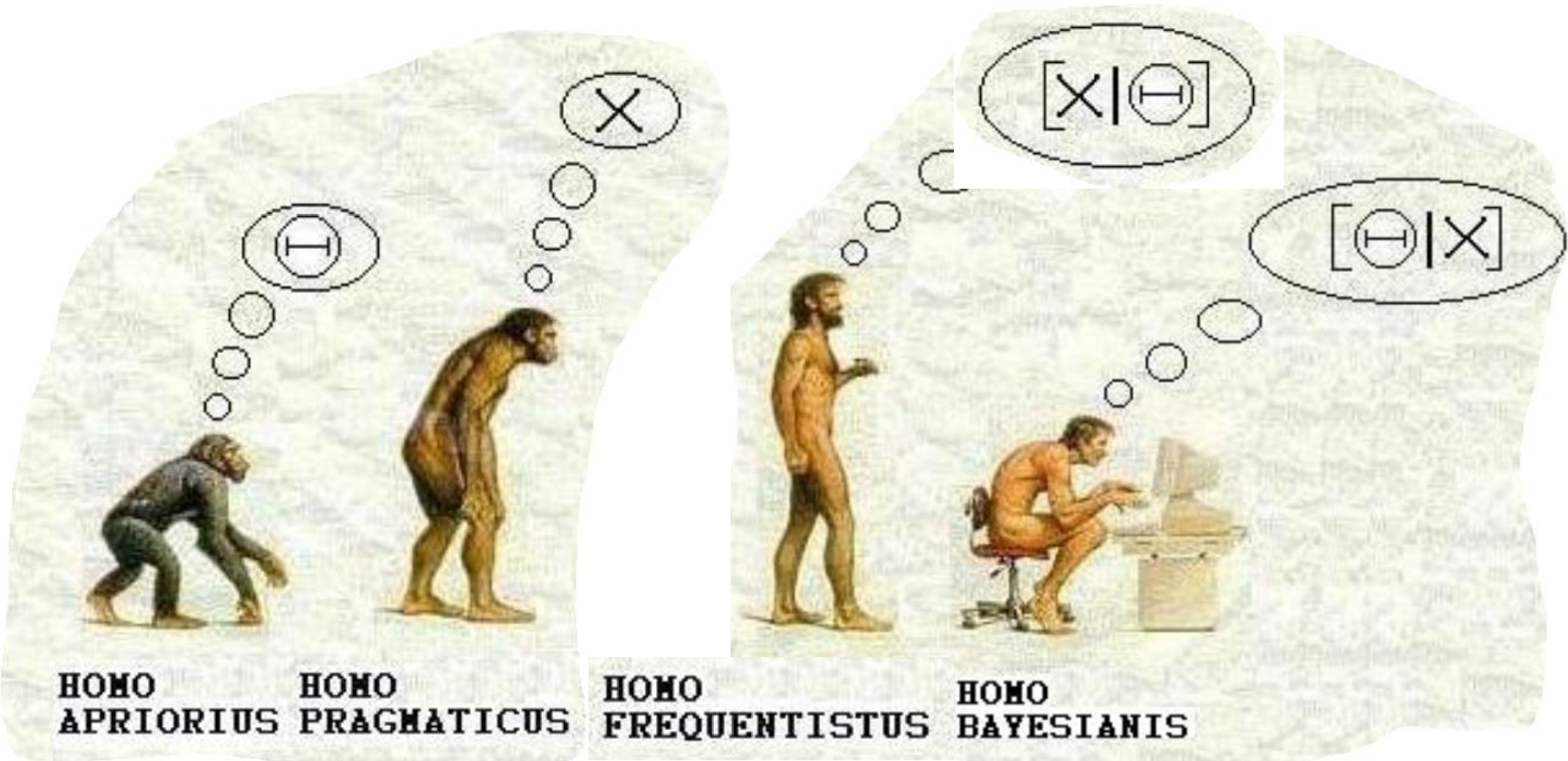


$\sqrt{\text{bias/variance}}$ i.e. A/P	$\xi_{1\sigma}^{(\text{data})}$ i.e. fraction within 1σ
1.03 ± 0.05	0.68 ± 0.02

Can also do 'future tests'
– test uncertainties in extrapolation

Note: closure test does not determine whether the precision is optimal! – only that it is faithful
Better methodology can always give more precise PDFs…

Evolution?



Middle ages: 'Religion'

18/19th century: 'Phenomenology'

20th century: 'Science'

21st century: ML/AI ???

Current Global PDFs

Global Data Sets

Increasing use of LHC datasets, from Run I and now Run II: DY, W, Z, top, incl jets, dijets,…

- Fixed Target DIS: SLAC/BCDMS/NMC
- Neutrino DIS: CCFR/CHORUS/NuTeV
- HERA: H1/ZEUS (NC,CC,c,b)
- Fixed Target DY: E605/E866/E906
- Tevatron: CDF/D0 (W, Z, incl jets, top)

ATLAS

Data set	NNPDF4.0	NNPDF3.1	ABMP16	CT18	MSHT20
ATLAS W, Z 7 TeV (2010)	✓	✓	✓	✓	✓
ATLAS W, Z 7 TeV (2011)	✓	✓	✗	✓	✓
ATLAS low-mass DY 7 TeV	✓	✓	✗	✗	✗
ATLAS high-mass DY 7 TeV	✓	✓	✗	✗	✓
ATLAS W 8 TeV	✓	✗	✗	✗	✓
ATLAS DY 2D 8 TeV	✓	✗	✗	✗	✓
ATLAS high-mass DY 2D 8 TeV	✓	✗	✗	✗	✓
ATLAS $\sigma_{W,Z}$ 13 TeV	✓	✗	✓	✗	✗
ATLAS W^++jet 8 TeV	✓	✗	✗	✗	✓
ATLAS Z p_T 8 TeV	✓	✓	✗	✓	✓
ATLAS $\sigma_{t\bar{t}}^{\rm tot}$ 7, 8 TeV	✓	✓	✓	✗	✗
ATLAS $\sigma_{t\bar{t}}^{\rm tot}$ 13 TeV	✓	✓	✓	✗	✗
ATLAS $t\bar{t}$ lepton+jets 8 TeV	✓	✓	✗	✓	✓
ATLAS $t\bar{t}$ dilepton 8 TeV	✓	✗	✗	✗	✓
ATLAS single-inclusive jets 7 TeV, R=0.6	✗	✓	✗	✓	✓
ATLAS single-inclusive jets 8 TeV, R=0.6	✓	✗	✗	✗	✗
ATLAS dijets 7 TeV, R=0.6	✓	✗	✗	✗	✗
ATLAS direct photon production 13 TeV	✓	✗	✗	✗	✗
ATLAS single top R_t 7, 8, 13 TeV	✓	✗	✓	✗	✗
ATLAS single top diff. 7, 8 TeV	✓	✗	✗	✗	✗
ATLAS single top diff. 8 TeV	✓	✗	✗	✗	✗

CMS

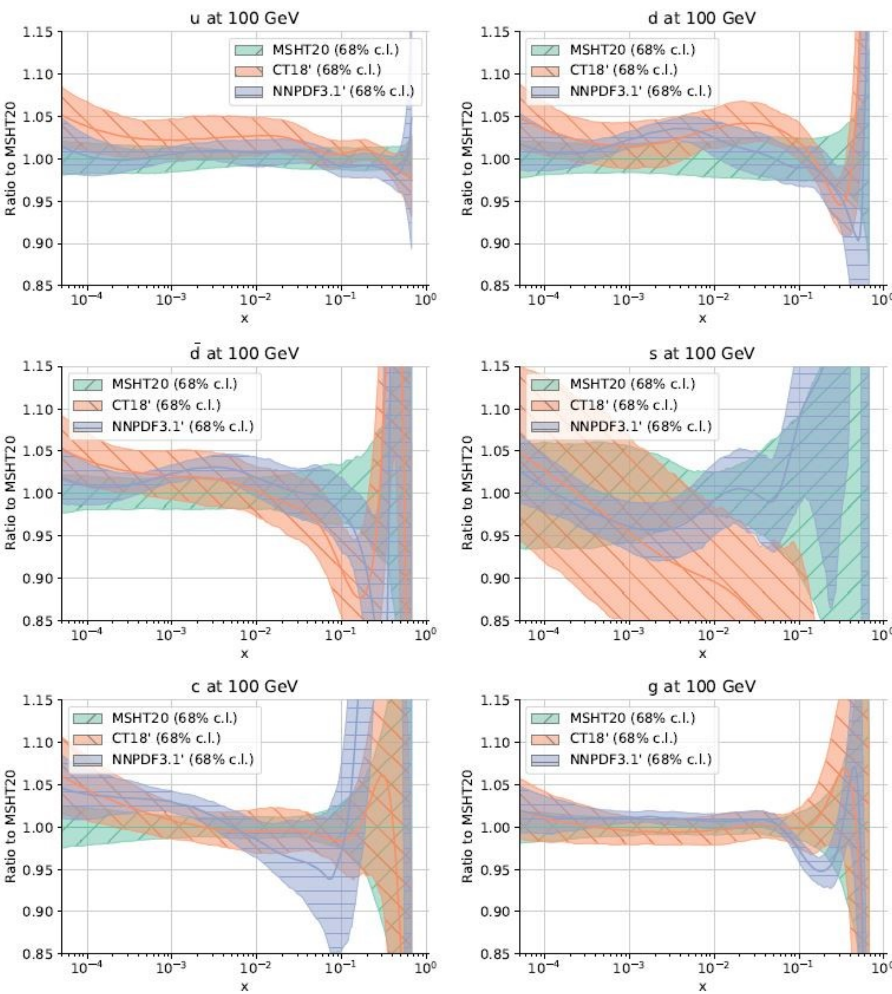
Data set	NNPDF4.0	NNPDF3.1	ABMP16	CT18	MSHT20
CMS W electron asymmetry 7 TeV	✓	✓	✗	✓	✓
CMS W muon asymmetry 7 TeV	✓	✓	✓	✓	✗
CMS Drell-Yan 2D 7 TeV	✓	✓	✗	✗	✓
CMS W rapidity 8 TeV	✓	✓	✓	✓	✓
CMS Z p_T 8 TeV	✓	✓	✗	✓	✗
CMS $W+c$ 7 TeV	✓	✓	✗	✗	✓
CMS $W+c$ 13 TeV	✓	✗	✗	✗	✗
CMS single-inclusive jets 2.76 TeV	✗	✓	✗	✗	✓
CMS single-inclusive jets 7 TeV	✗	✓	✗	✓	✓
CMS dijets 7 TeV	✓	✗	✗	✗	✗
CMS single-inclusive jets 8 TeV	✓	✗	✗	✓	✓
CMS 3D dijets 8 TeV	✗	✗	✗	✗	✗
CMS $\sigma_{t\bar{t}}^{\rm tot}$ 5 TeV	✓	✗	✓	✗	✗
CMS $\sigma_{t\bar{t}}^{\rm tot}$ 7, 8 TeV	✓	✓	✓	✗	✓
CMS $\sigma_{t\bar{t}}^{\rm tot}$ 13 TeV	✓	✓	✓	✗	✗
CMS $t\bar{t}$ lepton+jets 8 TeV	✓	✓	✗	✗	✓
CMS $t\bar{t}$ 2D dilepton 8 TeV	✓	✗	✗	✓	✓
CMS $t\bar{t}$ lepton+jet 13 TeV	✓	✗	✗	✗	✗
CMS $t\bar{t}$ dilepton 13 TeV	✓	✗	✗	✗	✗
CMS single top $\sigma_t + \sigma_{\bar{t}}$ 7 TeV	✓	✗	✓	✗	✗
CMS single top R_t 8, 13 TeV	✓	✗	✓	✗	✗

LHCb

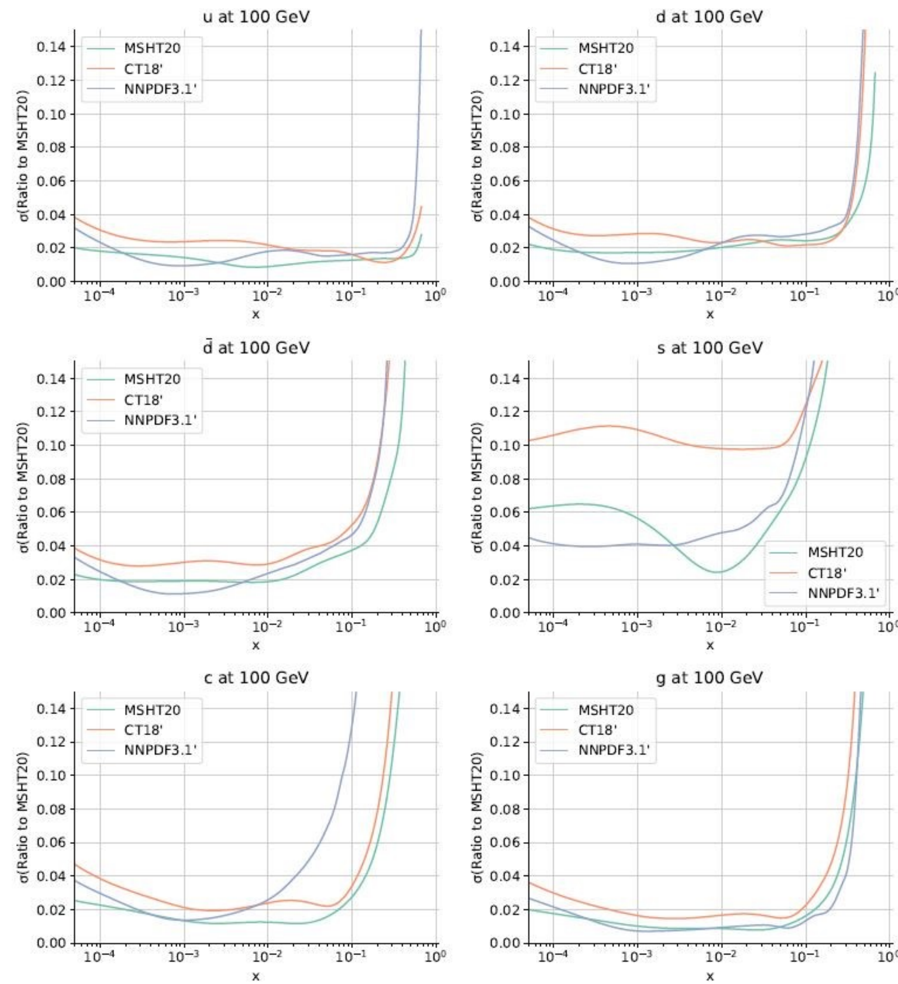
Data set	NNPDF4.0	NNPDF3.1	ABMP16	CT18	MSHT20
LHCb Z 940 pb	✓	✓	✗	✗	✓
LHCb $Z \to ee$ 2 fb	✓	✓	✓	✓	✓
LHCb $W, Z \to \mu$ 7 TeV	✓	✓	✓	✓	✓
LHCb $W, Z \to \mu$ 8 TeV	✓	✓	✓	✓	✓
LHCb $Z \to \mu\mu, ee$ 13 TeV	✓	✗	✗	✗	✗

Global PDFs

- MRS/MRST/MMHT/MSHT: Hessian + dynamical tolerance : 'MSHT20'
- MT/CTEQ/CT: Hessian + tolerance : 'CT18'
- NNPDF: NN+CV+MC (+CT) : 'NNPDF3.1'



PDFs (normalised to MSHT) ~ consistent



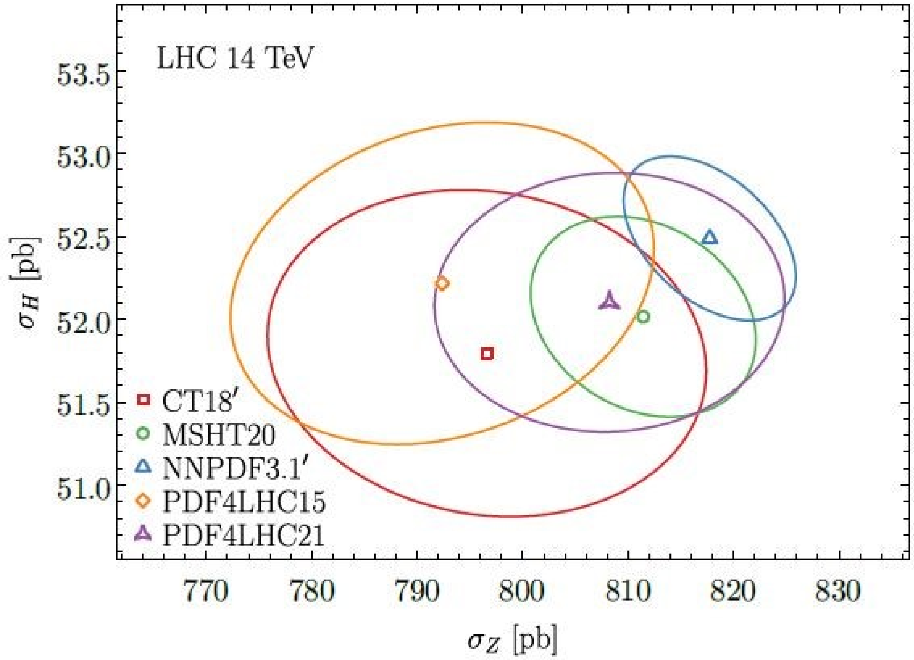
PDF uncertainties ~ few %

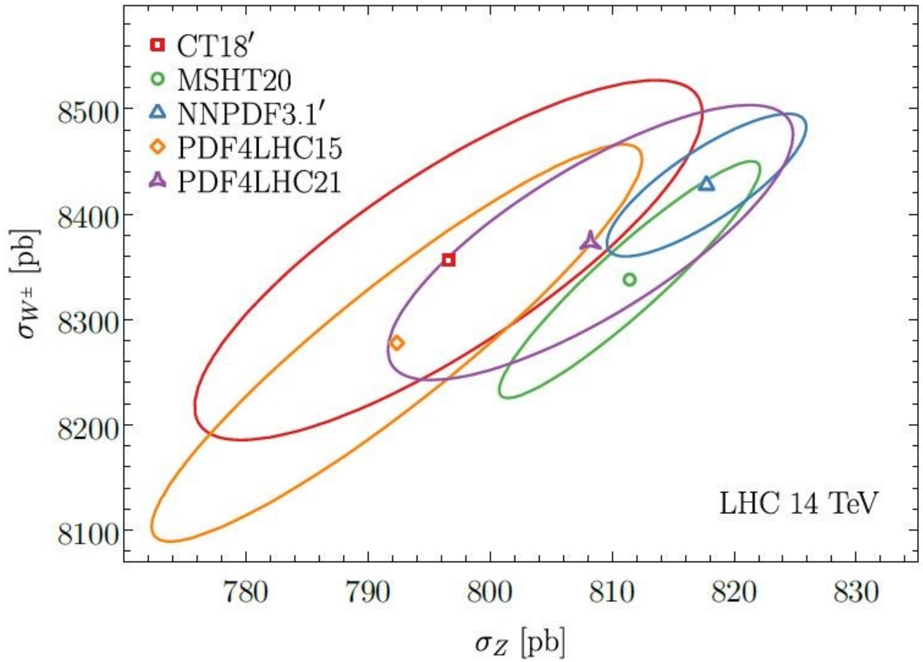
Consistency, Precision, Combination

Compare e.g. predictions for total cross-sections: $W^{\pm}, Z, H$

All consistent, but $\sigma_{\{CT18\}} > \sigma_{\{MSHT20\}} > \sigma_{\{NNPDF3.1\}}$

PDF4LHC 2021





Combination: PDF4LHC21 = CT18 ⊕ MSHT20 ⊕ NNPDF3.1 300 replicas each

Gives conservative estimate of overall PDF uncertainty ~ few %

Performed using various tools: Hessian → Monte Carlo, MC2HESSIAN, META-PDF, CMC

Thorne & Watt, 2012

Gao & Nadolsky, 2014

Carrazza et al, 2015

Towards 1% Precision

Step 4: Hyperoptimization

N3fit: Carrazza, Cruz-Martinez, 2019

Hyperparameters: not data, or theory, or the PDF: rather the ‘technical’ parameters:

- NN architecture (number & size of layers), activation functions, intitialization,etc
- Fitting parameters: optimizer, learning rate, stopping parameters, etc

Traditionally hyperparameters chosen by hand (fiddled).

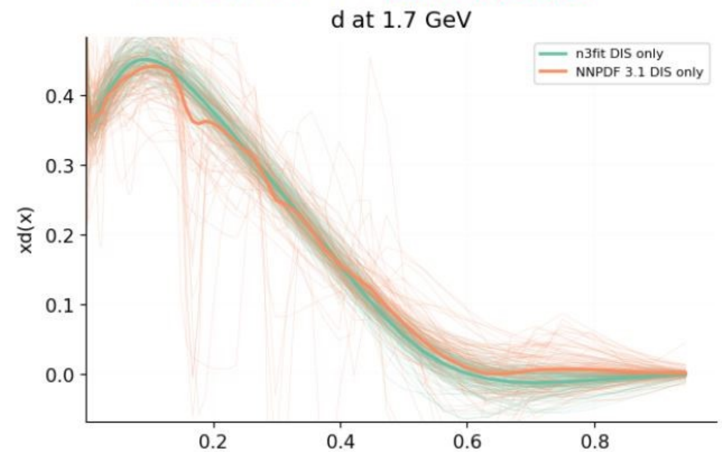
Better to choose objectively, optimising χ^2_{val} (hyperopt)

Hyperparameters highly correlated: need thousands of fits to explore full space

Test set: choose independent data set (not fitted)

Result: smoother, more precise PDFs

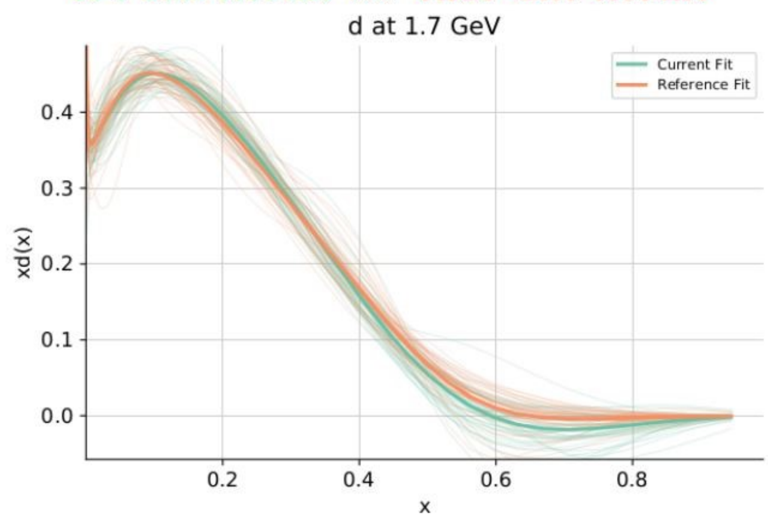
HYPEROPT VS HANDPICKED



K-folding: divide data into many independent data sets: test on random subsets: assures generalisability

Result: even smoother, more precise PDFs!

K-FOLD HYPER VS. TEST-SET HYPER



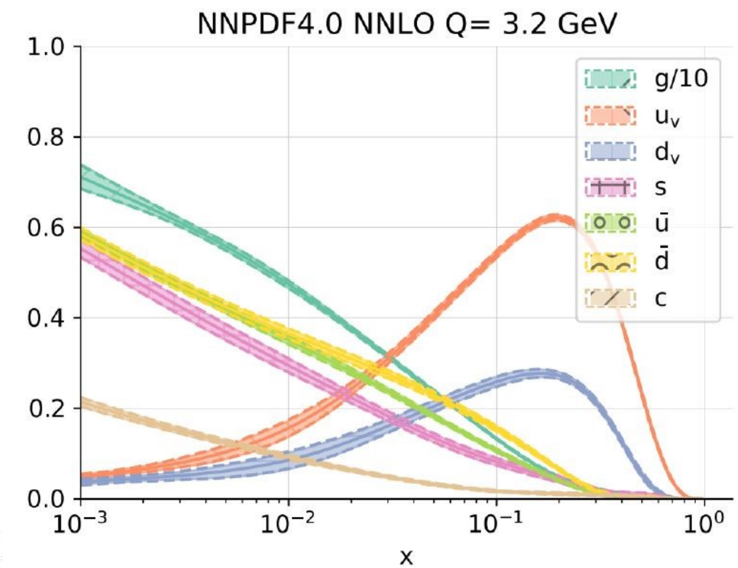
Also much faster (more efficient):

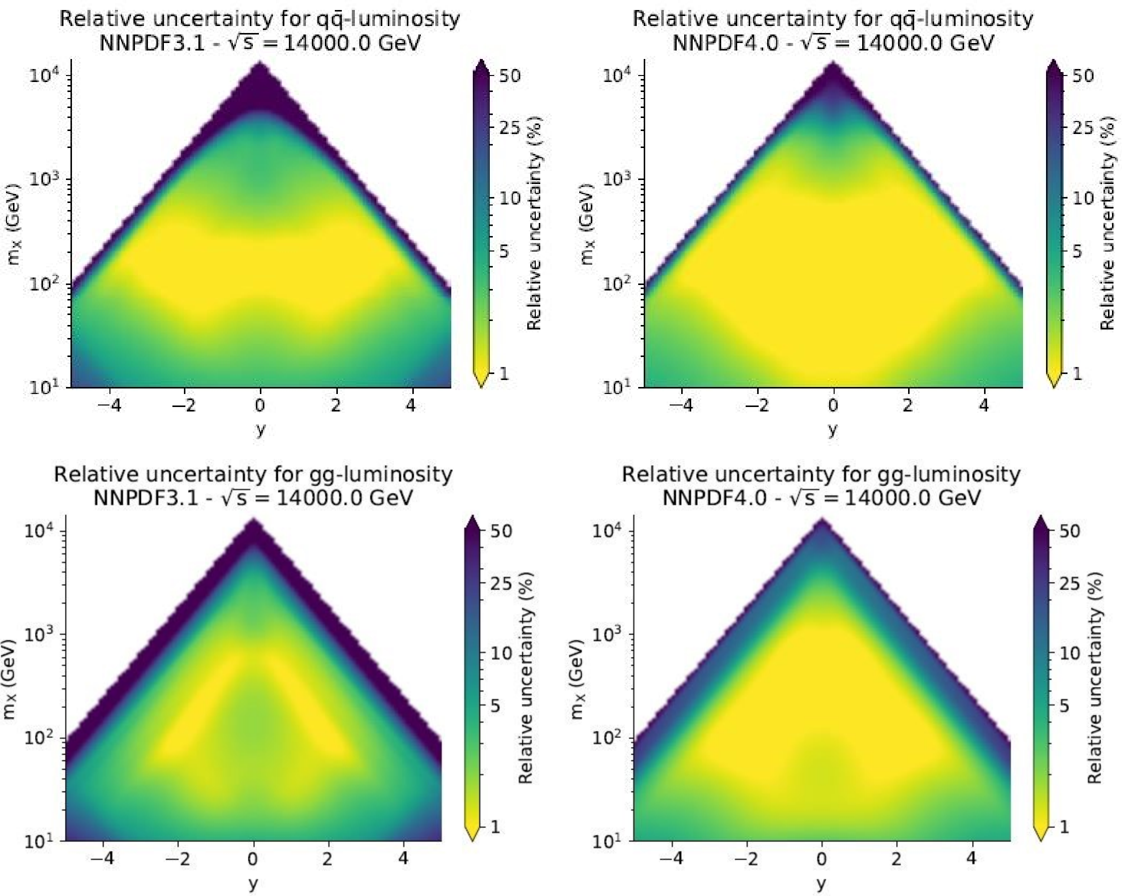
- NNFit (NNPDF3.1): 18hr/replica
- N3fit (NNPDF4.0): 38min/replica

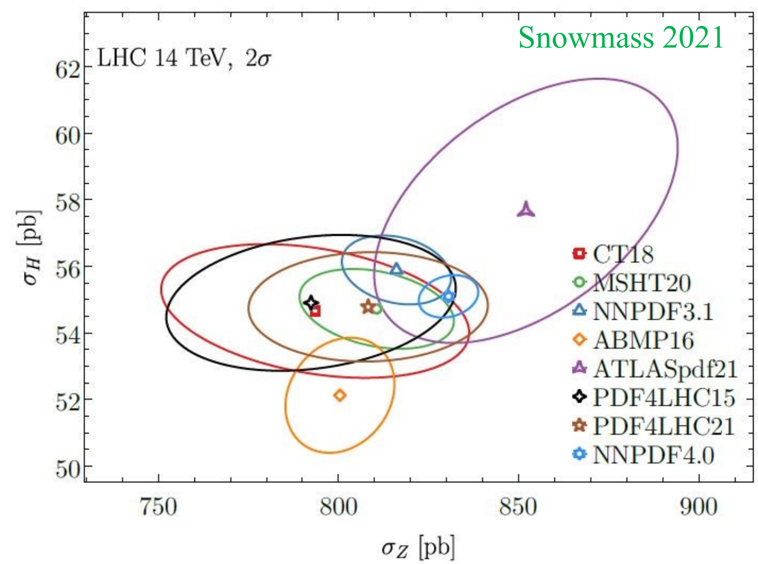
NNPDF4.0

NNPDF3.1 (2017) and NNPDF4.0 (2021) consistent, both faithful (closure test), but 4.0 much more precise:

- better methodology (hyperopt)
- more LHC data, new processes
- better theory (positivity, sum rules, nucl unc)







NNPDF4.0 most precise set to date

A ML open-source QCD fitting framework

Search docs

- Getting started
- Fitting code: `n3fit`
- Code for data: validphys
- Handling experimental data: Buildmaster
- Storage of data and theory predictions
- Theory
- Continuous integration and deployment
- Servers
- External codes
- Tutorials
- Adding to the Documentation

» The NNPDF collaboration

View page source

The NNPDF collaboration

The NNPDF collaboration performs research in the field of high-energy physics. The NNPDF collaboration determines the structure of the proton using contemporary methods of artificial intelligence. A precise knowledge of the so-called **Parton Distribution Functions (PDFs)** of the proton, which describe their structure in terms of their quark and gluon constituents, is a crucial ingredient of the physics program of the Large Hadron Collider of CERN.

The NNPDF code

The scientific output of the collaboration is freely available to the publi through the arXiv, journal repositories, and software repositories. Along with this online documentation, we release the NNPDF code used to produce the latest family of PDFs from NNPDF, NNPDF4.0. The code is made available as an open-source package together with the user-friendly examples and an extensive documentation presented here.

The code can be used to produce the ingredients needed for PDF fits, to run the fits themselves, and to analyse the results. This is the first framework used to produce a global PDF fit made publicly available, enabling for a detailed external validation and reproducibility of the NNPDF4.0 analysis. Moreover, the code enables the user to explore a number of phenomenological applications, such as the assessment of the impact of new experimental data on PDFs, the effect of changes in theory settings on the resulting PDFs and a fast quantitative comparison between theoretical predictions and experimental data over a broad range of observables.

If you are a new user head along to Getting started and check out the Tutorials.

NNPDF is not a black box!

DIY global fitting is at last possible…. and encouraged

Missing Higher Order Uncertainties (MHOU)

Theory Uncertainties: PDFs extracted from experimental data using theory (eg NNLO QCD)

So PDFs have theoretical uncertainties as well as experimental uncertainties

Problem: how do we incorporate theory uncertainties in PDF determination?

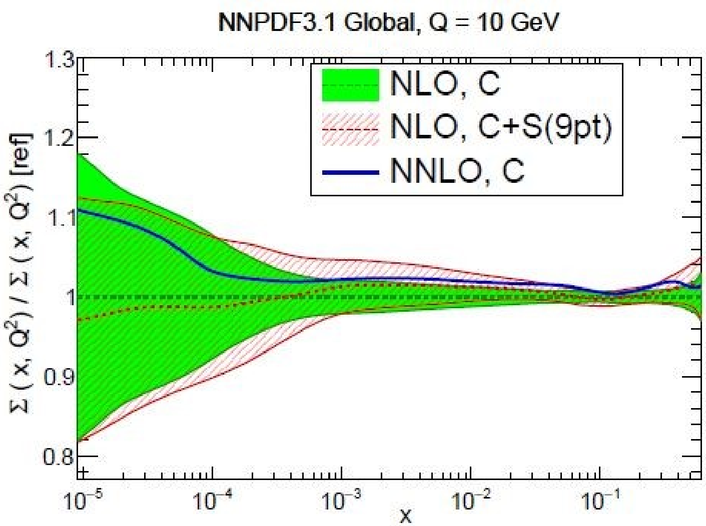
Solution: the 'theory covariance matrix' S

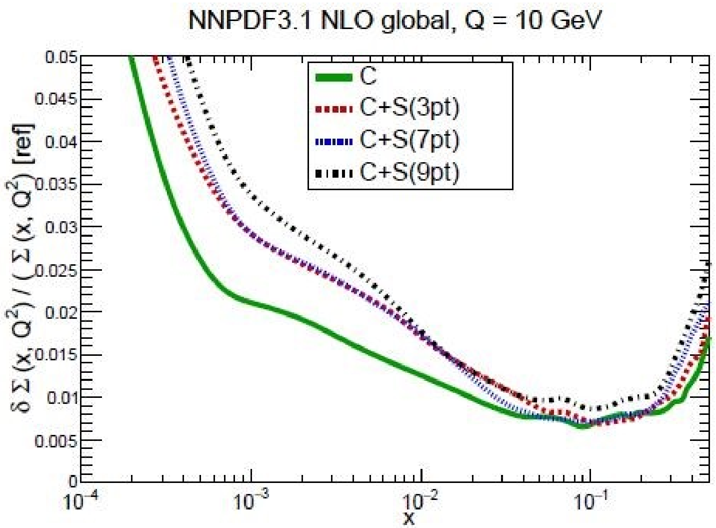
Assumes experimental and theoretical unc are Gaussian and independent, so $C \to C + S$

Tested in NNPDF4.0 on nuclear uncertainties: data with nucl unc deweighted in the fit

MHOU: estimate for every data point in the fit by scale variation:

- μ_F variation - estimates MHOU in parton evolution (correlated across all processes)
- μ_R variation - estimates MHOU in hard processes (correlated only within process)





Data with $S \gg C$ deweighted in fit: shift towards NNLO: modest increase in PDF unc

NNPDF4.0 + NNLO MHOU coming soon!

EWK corrections

For 1% precision, need EWK corrections:

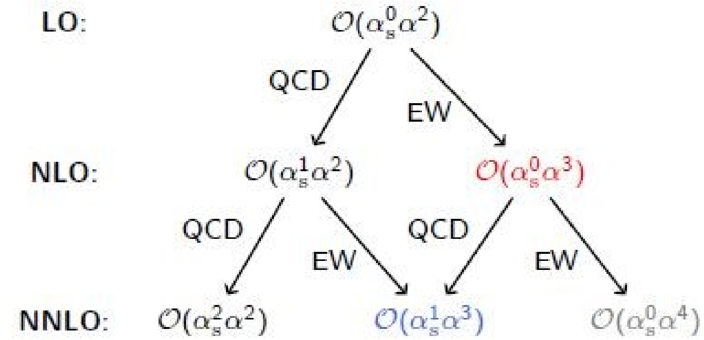
$\alpha_s^2 \sim \alpha$ so NNLO QCD ~ NLO EWK

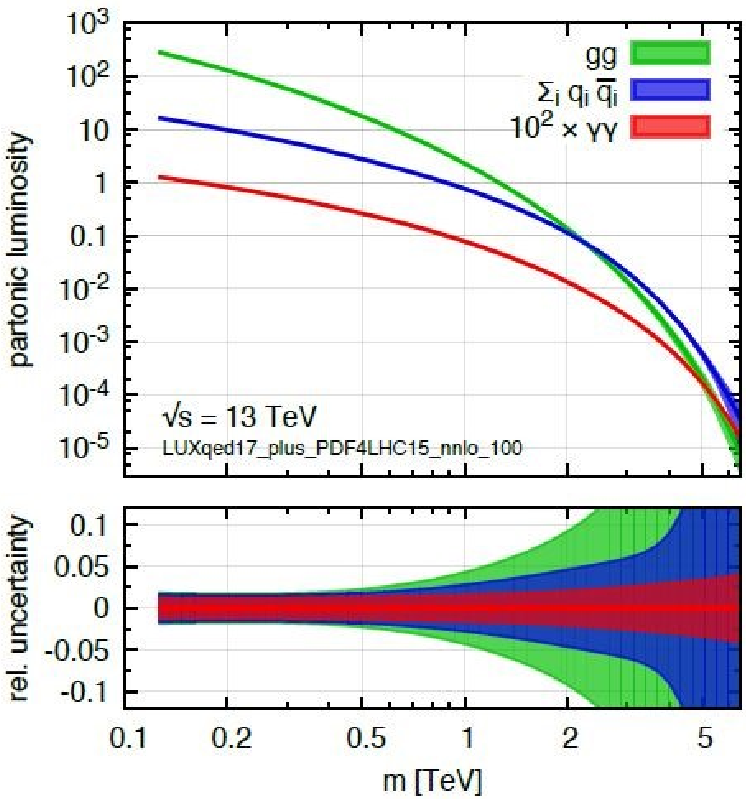
Currently ad hoc (corrns vs cuts): need systematic treatment

Double counting Problem: sometimes experimentalists subtract ISR, add FSR

Subtracting ISR problematic: can't be unfolded

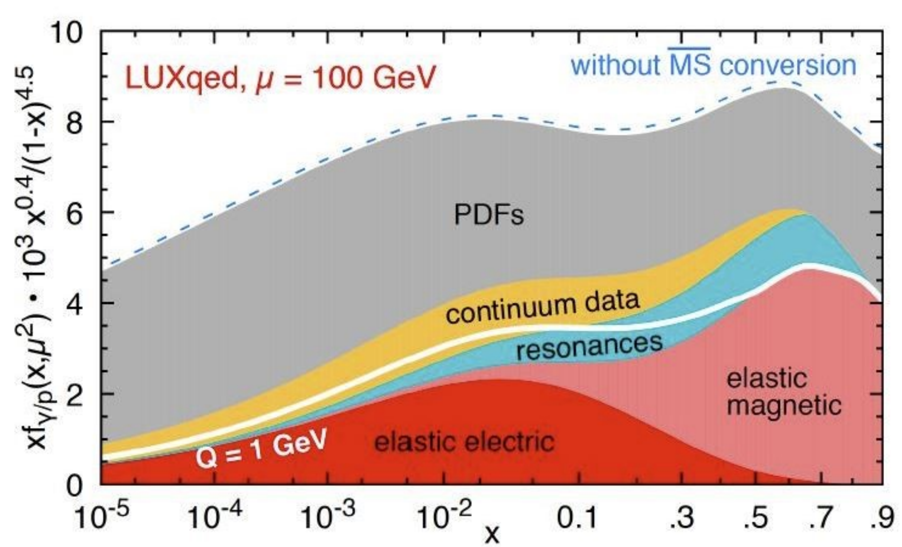
e.g. DY:





Photon PDF $\gamma(x, Q^2)$ can be computed in terms of elastic FF and inelastic SF data, and PDFs : LUXqed

Manohar, Nason, Salam, Zanderighi (2016)

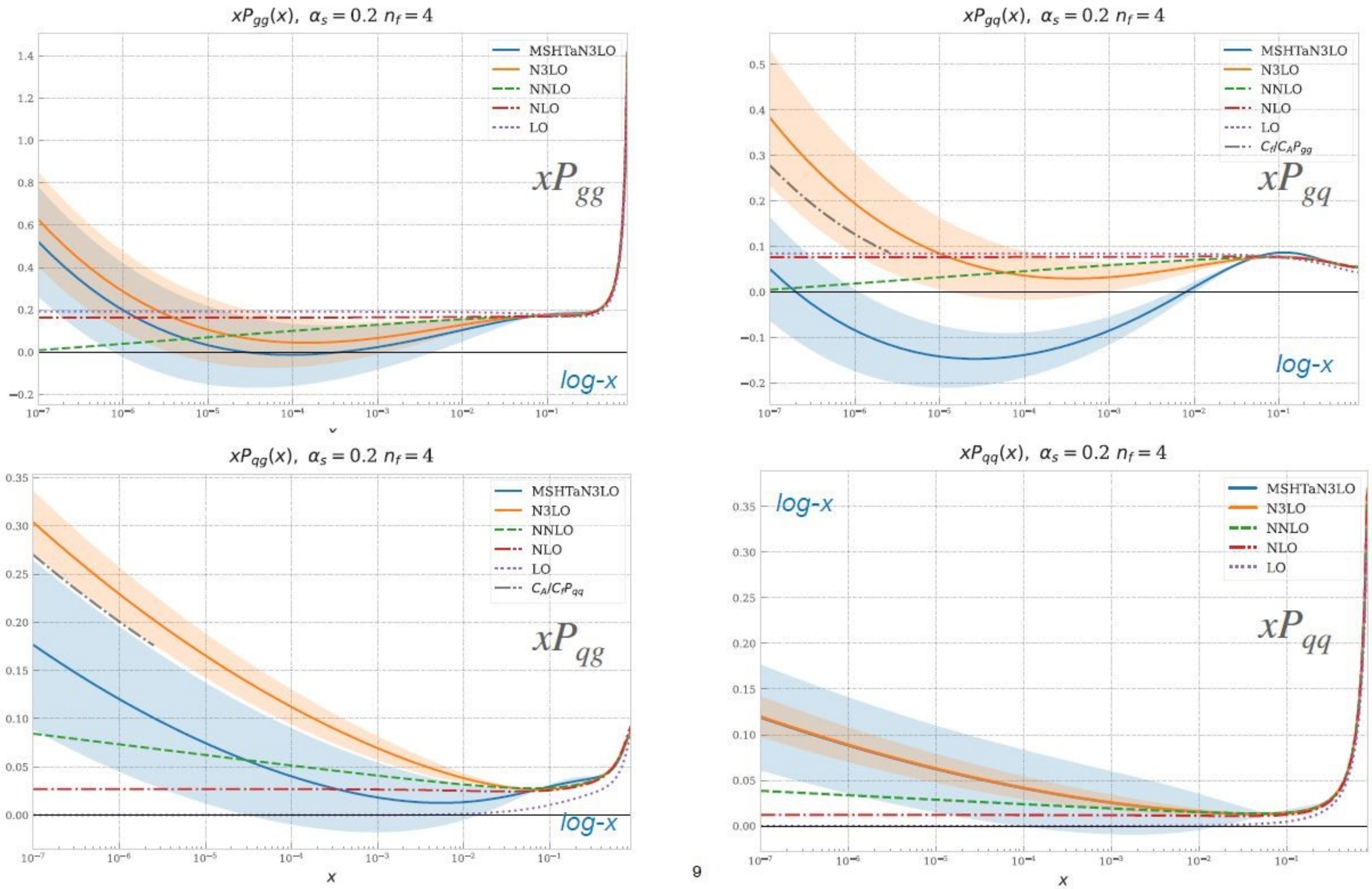


NNPDF4.0 + NLO EWK coming soonish!

N3LO corrections

Exact hard xsecs for DIS, DY, Higgs

Partial results for splitting functions: large n_f, large x, small x, moments



MSHT (2022),
Magni (NNPDF), 2022

Large uncertainties
at small x, as expected

Notes: use theory covmat for N3LO unc

if have estimates for MHOU, can combine NNLO processes and N3LO processes in global fit

NNPDF4.0 @ N3LO coming soon!

Parametric Uncertainties

'Physical parameters': not data, or theory, or the PDF, or hyper: rather

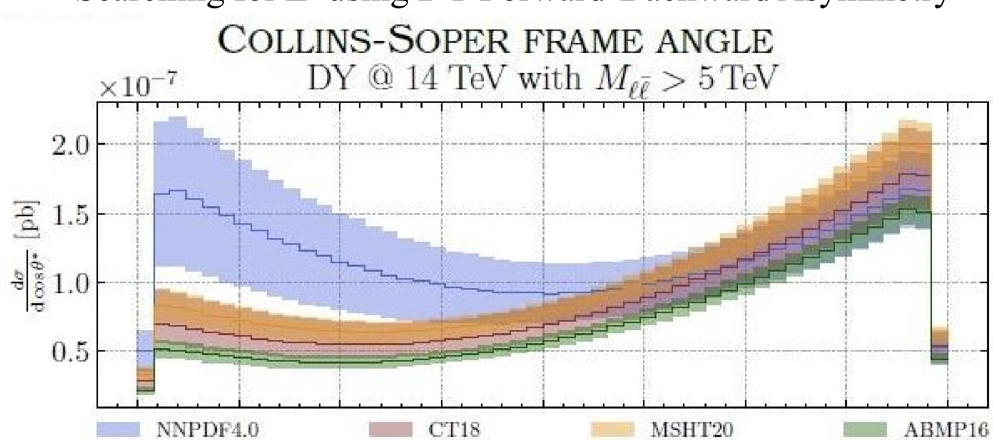
$$\alpha_s,\ m_c,\ m_b, m_t, m_W, \sin^2\theta_W, m_H,\ SMEFT\ \text{parameters}, \text{etc}$$

Determinations of these will generally be correlated: with each other, and with the PDFs

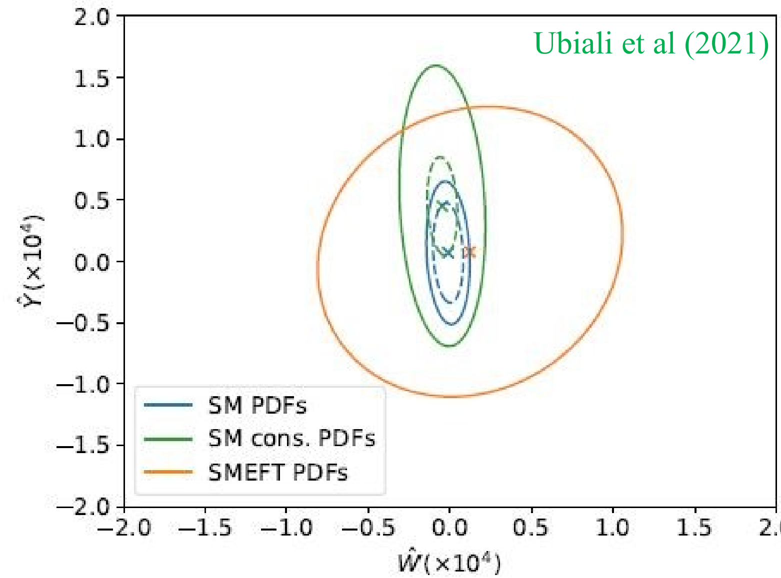
Several methods of dealing with this:

- PDF 'profiling' (Hessian only): e.g. ATLAS m_W
- Simultaneous Hessian fits: e.g. MSHT, CT α_s, H1 EWK
- Correlated replicas (NN only): e.g. NNPDF α_s
- Simultaneous NN fits (SIMUnet): e.g. SMEFT
- Theory covariance matrix: extract parameters from uncertainty in fit

Searching for Z' using DY Forward-Backward Asymmetry



PDF sets disagree at very large x



BSM is not BSM if it can be absorbed in PDFs!

Intrinsic Charm

Article | Open Access | Published: 17 August 2022

Evidence for intrinsic charm quarks in the proton

The NNPDF Collaboration

Nature **608**, 483–487 (2022) | Cite this article

37k Accesses | **3** Citations | **359** Altmetric | Metrics

Abstract

The theory of the strong force, quantum chromodynamics, describes the proton in terms of quarks and gluons. The proton is a state of two up quarks and one down quark bound by gluons, but quantum theory predicts that in addition there is an infinite number of quark–antiquark pairs. Both light and heavy quarks, whose mass is respectively smaller or bigger than the mass of the proton, are revealed inside the proton in high-energy collisions. However, it is unclear whether heavy quarks also exist as a part of the proton wavefunction, which is determined by non-perturbative dynamics and accordingly unknown: so-called intrinsic heavy quarks[1]. It has been argued for a long time that the proton could have a sizable intrinsic component of the lightest heavy quark, the charm quark. Innumerable efforts to establish intrinsic

ScienceNews

NEWS PARTICLE PHYSICS

Protons contain intrinsic charm quarks, a new study suggests

Understanding a proton's charm could refine intel from particle colliders

Protons are commonly thought to contain only three quarks — two up quarks and one down quark (illustrated). But there's new evidence that protons may also contain intrinsic charm quarks and antiquarks.

SEFA KAR/ISTOCK/GETTY IMAGES PLUS

NewScientist

CYBER MONDAY SALE

Physicists surprised to discover the proton contains a charm quark

The textbook description of a proton says it contains three smaller particles - two up quarks and a down quark - but a new analysis has found strong evidence that it also holds a charm quark

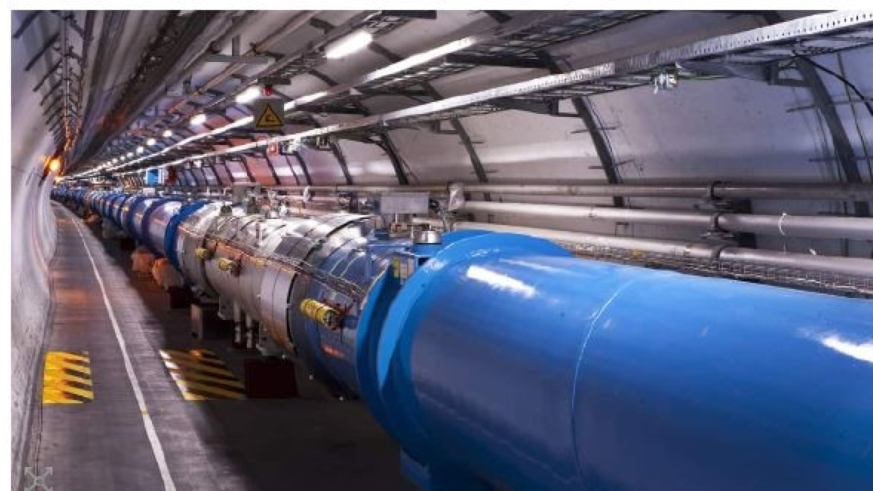
physicsworld | Magazine | Latest | People | Impact

particles and interactions

PARTICLES AND INTERACTIONS | RESEARCH UPDATE

Protons contain intrinsic charm quarks, machine-learning analysis suggests

23 Aug 2022



The Large Hadron Collider: evidence for intrinsic charm quarks in protons has been found in LHC data.

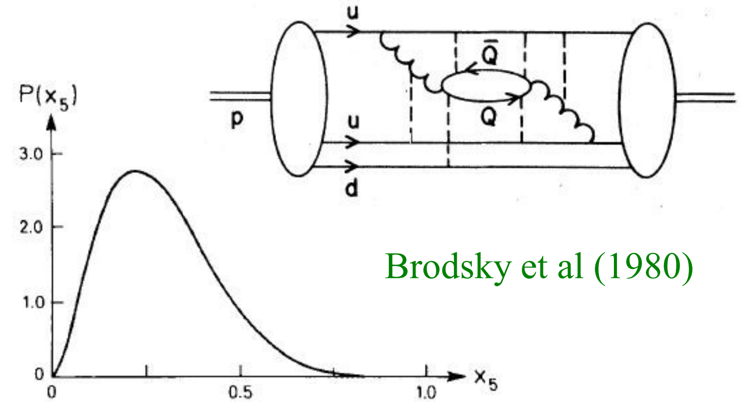
Intrinsic Charm?

Standard PDF Paradigm:

- Light partons: $L = g, u, \bar{u}, d, \bar{d}, s, \bar{s}$: $m_L \ll 1\ GeV$: nonpert: fit PDFs
- Heavy partons: $H = c, \bar{c}, b, \bar{b}, t, \bar{t}$: $m_H \gg 1\ GeV$: use pert QCD

But $m_c \simeq 1.5\ GeV$:

nonperturbative ('intrinsic') charm?



Brodsky et al (1980)

Test empirically: fit the charm PDF!

(in a global PDF fit, e.g. NNPDF4.0)

GM-VFNS: $\begin{cases} Q \simeq m_c\text{: threshold effects, need mass dependence} \\ Q \gg m_c\text{: large } \ln Q^2/m_c^2 \text{ ; need to resum (DGLAP)} \end{cases}$

ACOT/FONLL

In $N_F = 4$ scheme, charm PDF $f_c^4(Q^2)$ evolves perturbatively

In $N_F = 3$ scheme, charm PDF f_c^3 does not evolve: 'intrinsic'

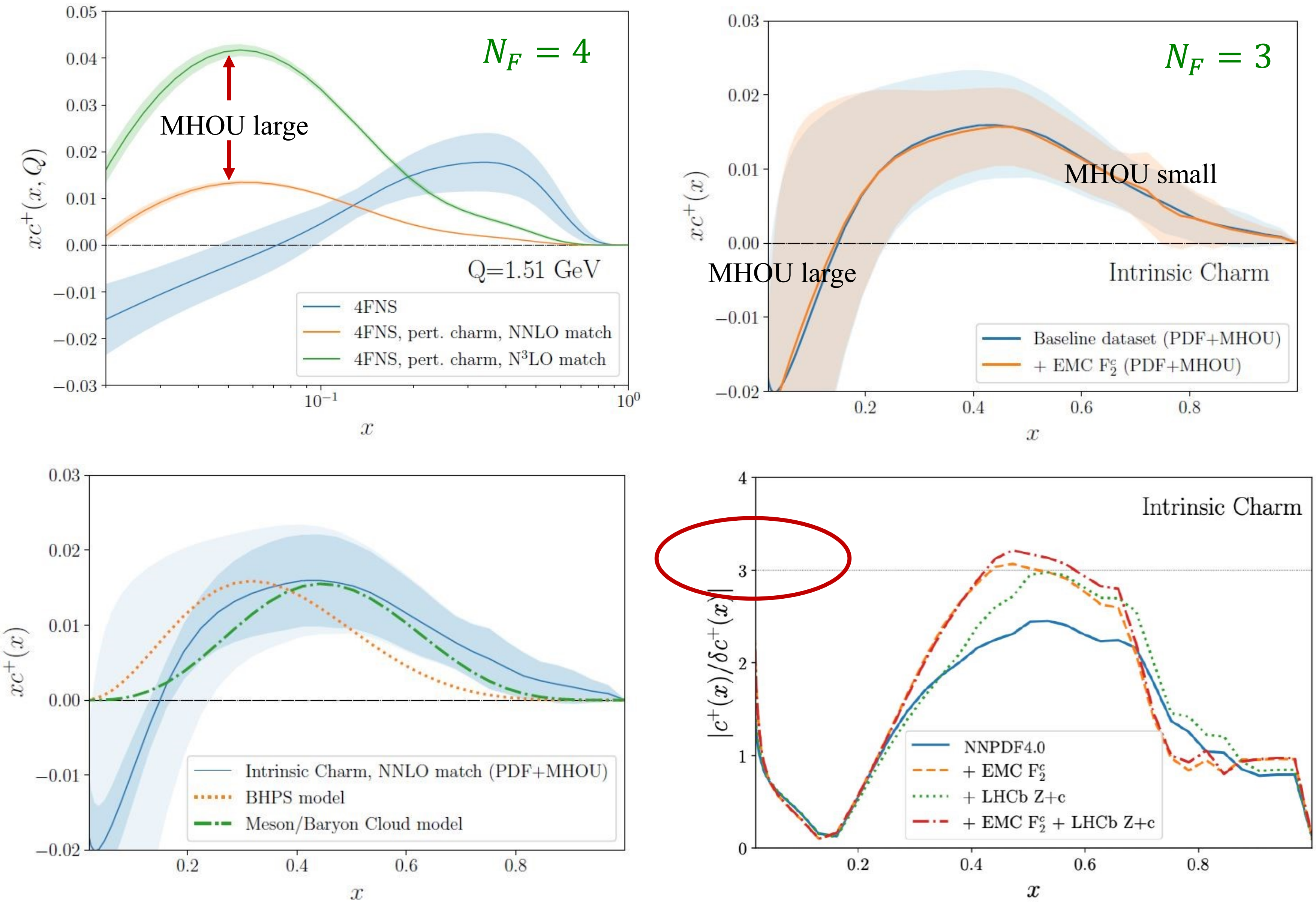
} Matching conditions (N3LO)

Notes: f_c^3 is leading twist

if f_c^3=0 all charm is perturbative: no intrinsic charm

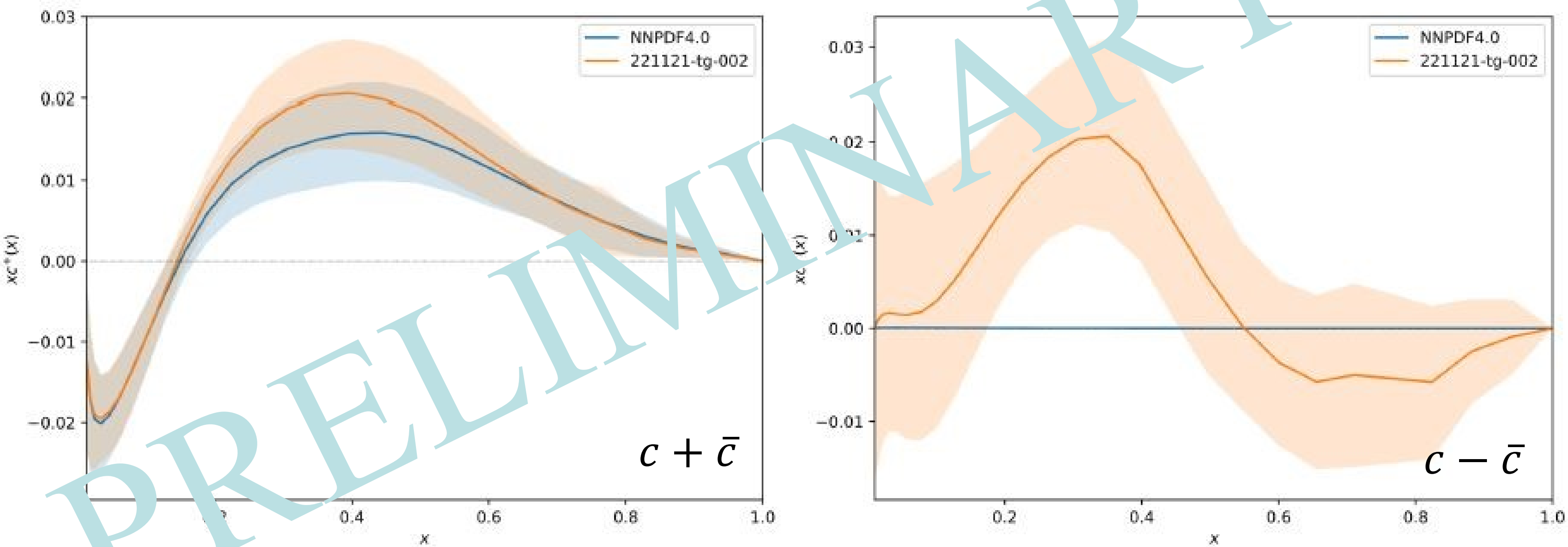
3σ Evidence for Intrinsic Charm

NNPDF (2022)



Evidence for Valence Charm???

In NNPDF4.0, we assume $c = \bar{c}$
What happens if we release this constraint?



Summary & Outlook

- PDFs: very active at LHC… and EIC
- Towards 1% uncertainties:
  - N3fit methodology (NN+CT+MC+HO)
  - MHOU
  - EWK
  - $\alpha_s,\ m_c,\ m_b, m_t, m_W, \ldots$ etc, etc….

  Leading to NNPDF4.1
- Opensource code
  - NNPDF is not a black box!
  - DIY fitting

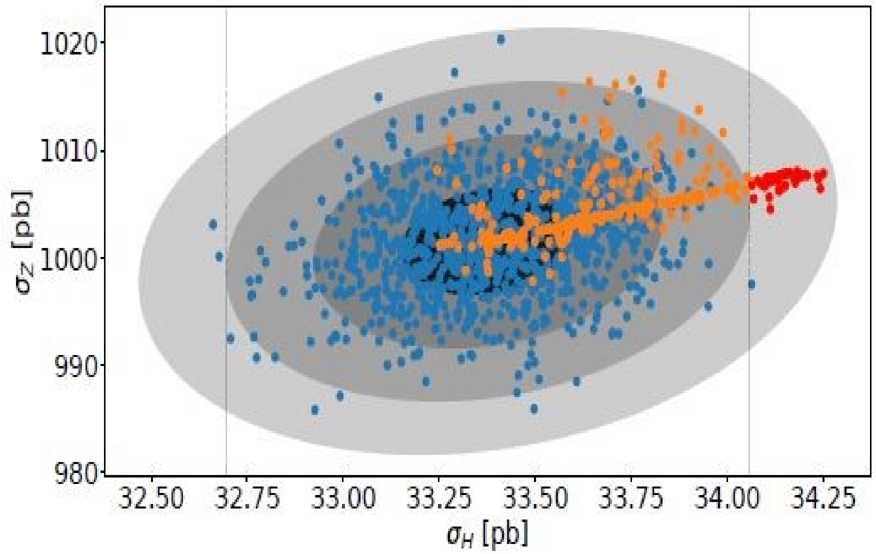
NNPDF

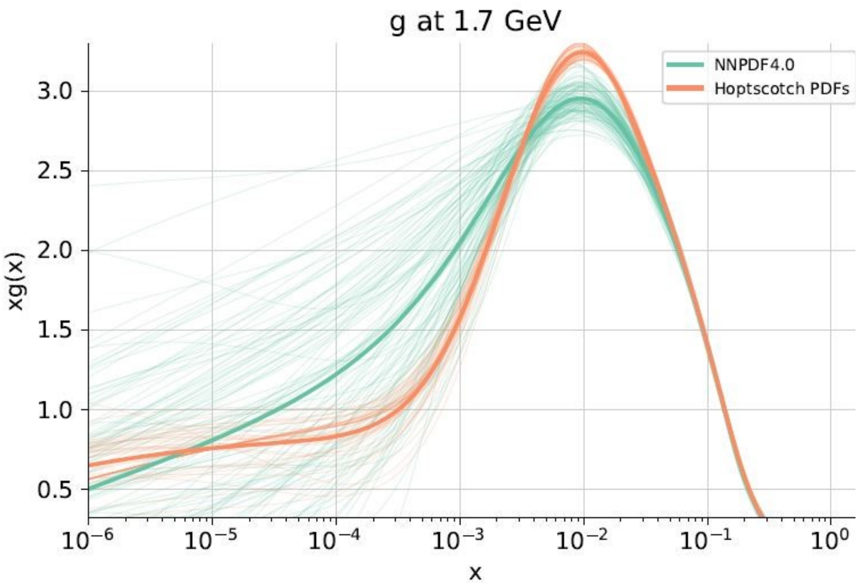
Gargnano (Aug 2022)

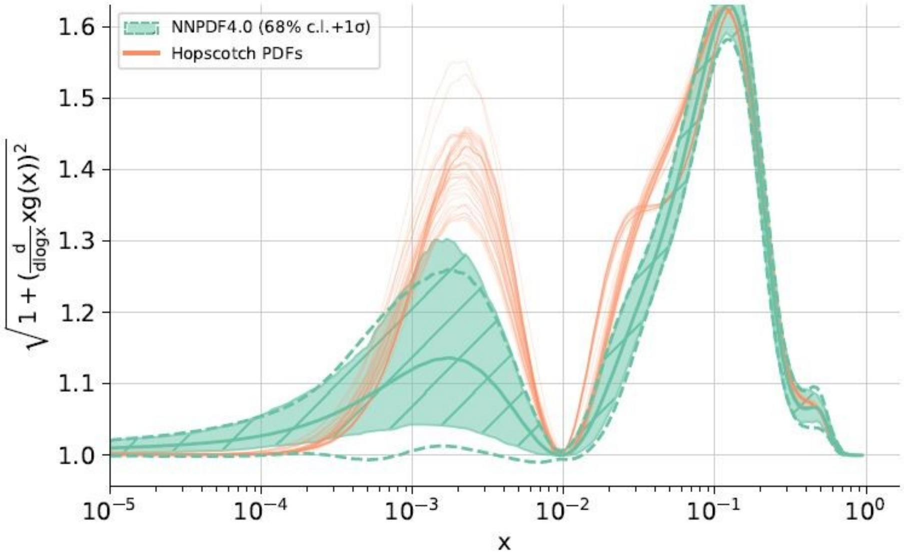
Courtnoy et al (2022)

The Hopscotch Paradox

- Take linear combination of NNPDF replicas : $f_{HS}(x) = \Sigma c_k f_k(x)$
- $f_{HS}(x)$ is a perfectly good PDF
- Minimise $\chi^2[f_{HS}]$ on c_k
- Then $f_{HS}(x)$ has better χ^2 than any NNPDF replica, but can lie outside NNPDF uncertainty!







Hopscotch PDFs too wiggly: overfitted (by construction) But Subtle!