PDF issues: viewpoint from the experiments

LHC EW WG General Meeting

June 11th, 2024

Simone Amoroso (DESY)

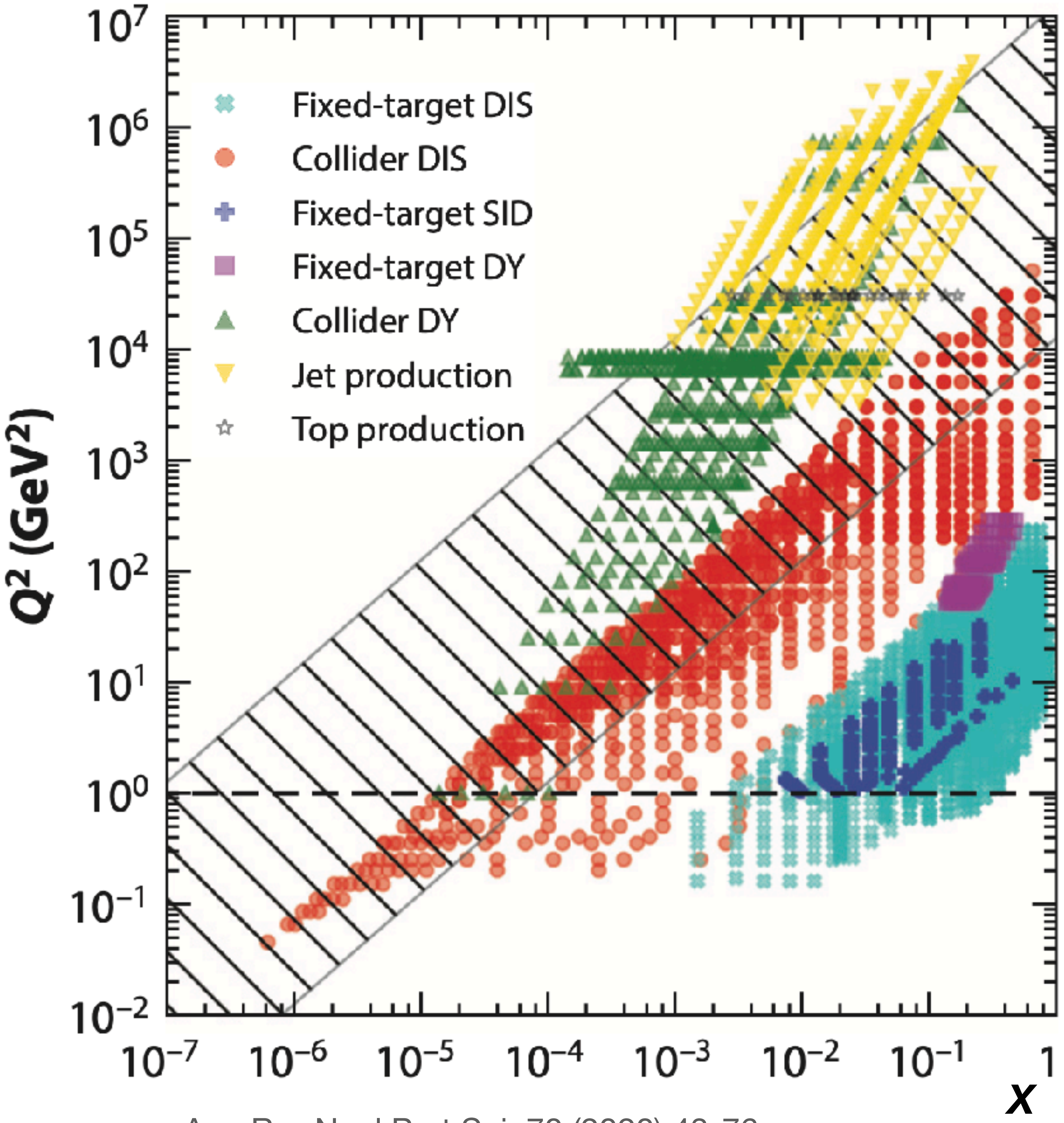
Parton Distribution Functions

- Predictions at a hadron collider require knowledge of the proton structure

$$\sigma = \sum_{i,j} \int dx_1 dx_2\, f_{i/p}(x_1) f_{j/p}(x_2)\, \hat{\sigma}(x_1 x_2 s)$$

- Cross-sections calculated as convolution of short-distance cross-sections with Parton Distribution Functions (PDFs)

- A universal quantity, PDFs are inferred from a given set of measurements and can be used to predict any cross-section

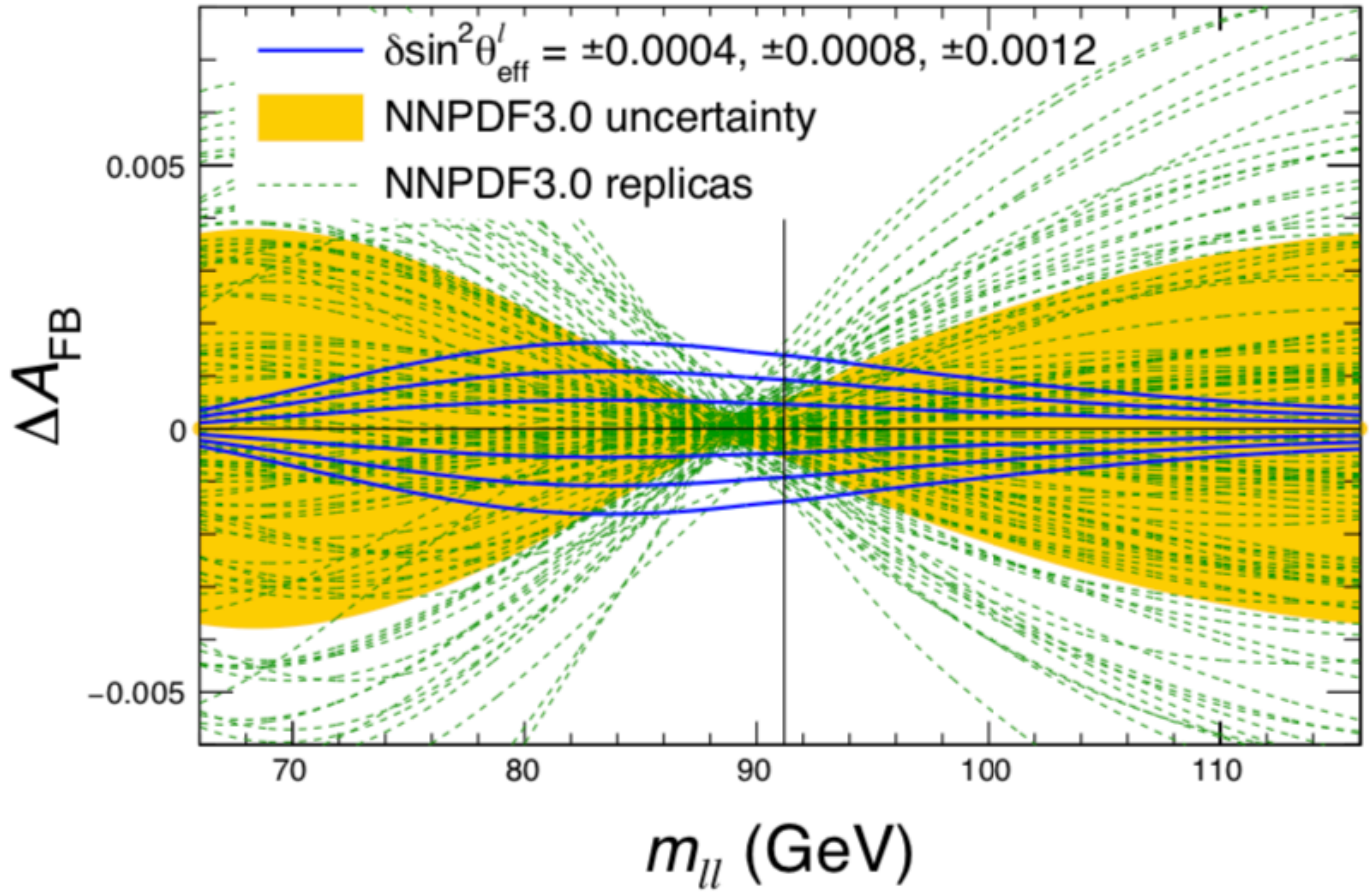


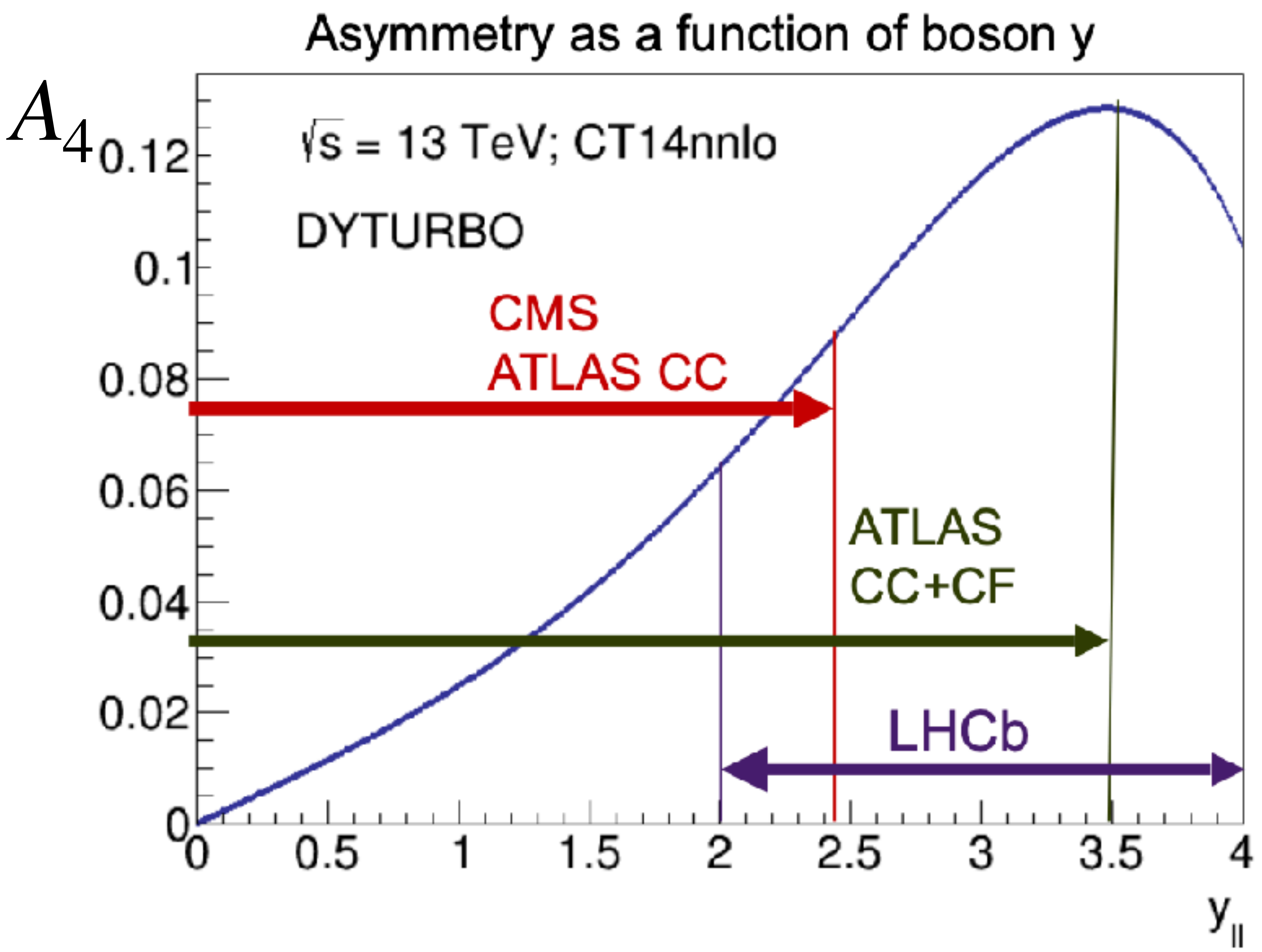
Ann.Rev.Nucl.Part.Sci. 70 (2020) 43-76

$\sin^2\theta^l_{\text{eff}}$ AT THE LHC AND PDFs

- Weak mixing angle extracted from polarization asymmetries in Drell-Yan

- Parton-level effect measured at particle level -> dependence on PDFs
  - Direction of incoming quark/anti-quark inferred from Z rapidity sign

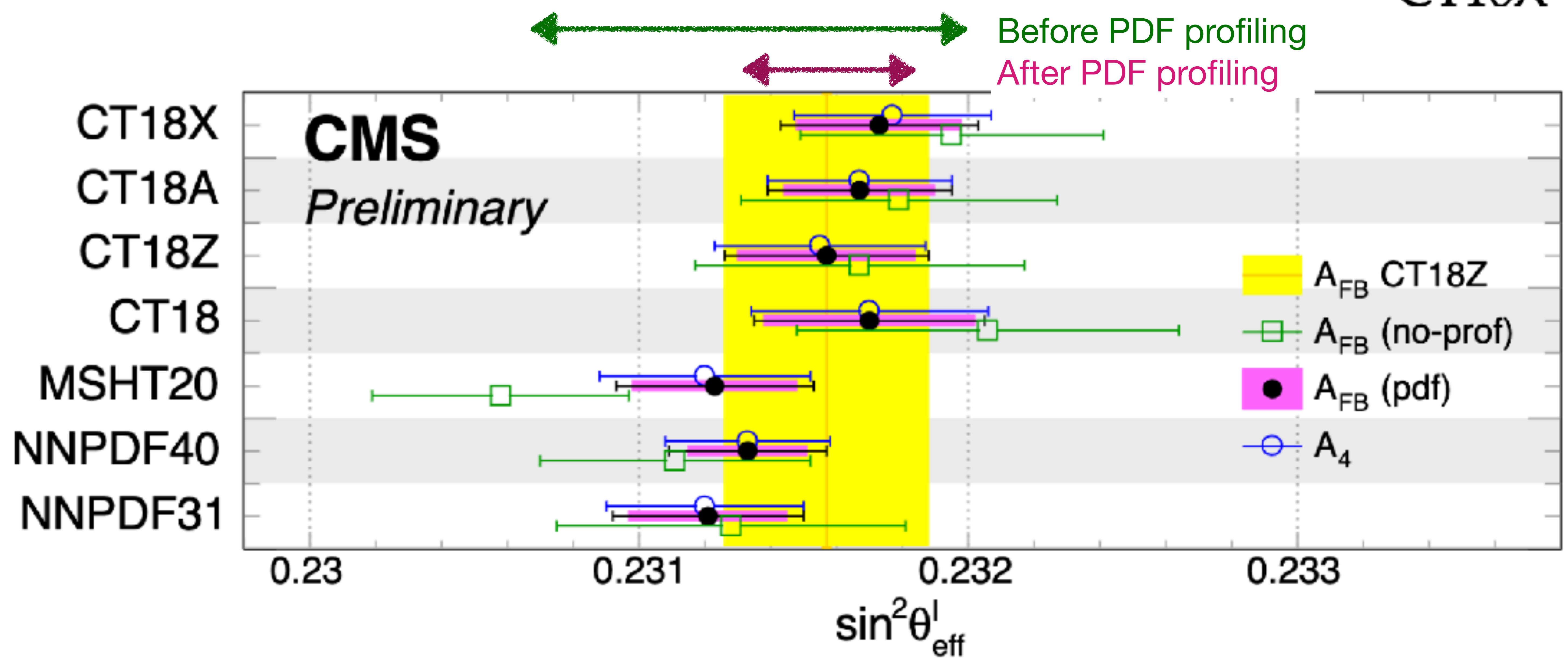




PDF uncertainties on $\sin^2\theta^l_{\mathrm{eff}}$ - CMS 13 TeV

- All PDF sets provide an equally good description of the data
- PDF spread and uncertainties reduced in the fit

PDF	A_{FB} (816 bins)		A_4 (63 bins)	
	χ^2_{min}	$\sin^2\theta^\ell_{\mathrm{eff}}$	χ^2_{min}	$\sin^2\theta^\ell_{\mathrm{eff}}$
NNPDF31	724.7	23121 ± 29	58.5	23120 ± 30
NNPDF40	730.5	23133 ± 24	62.6	23133 ± 25
MSHT20	735.8	23123 ± 30	71.0	23120 ± 32
CT18	728.4	23170 ± 35	62.2	23170 ± 36
CT18Z	730.7	23157 ± 31	61.3	23155 ± 32
CT18A	730.3	23167 ± 28	63.6	23167 ± 28
CT18X	728.5	23173 ± 30	61.8	23177 ± 30

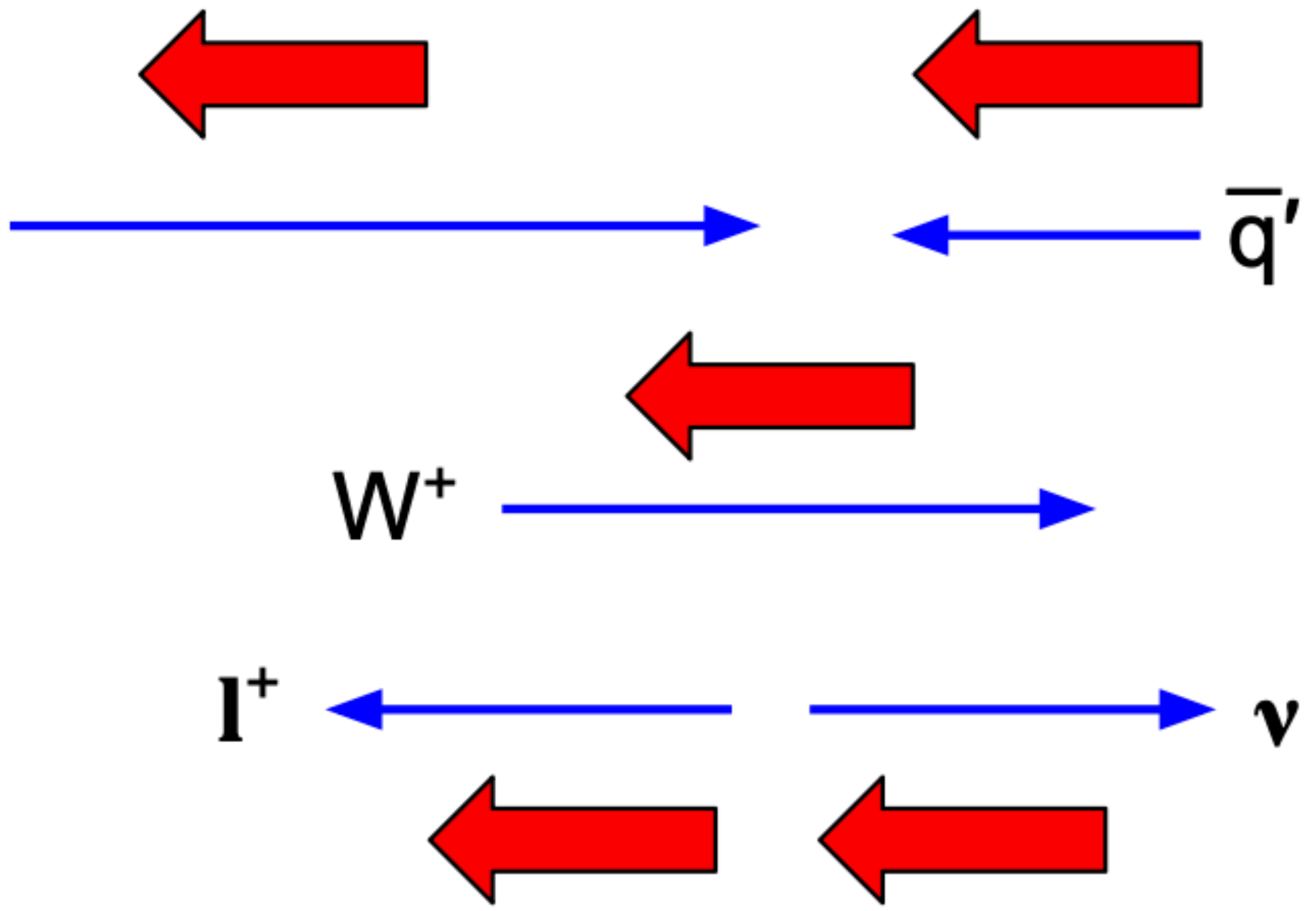


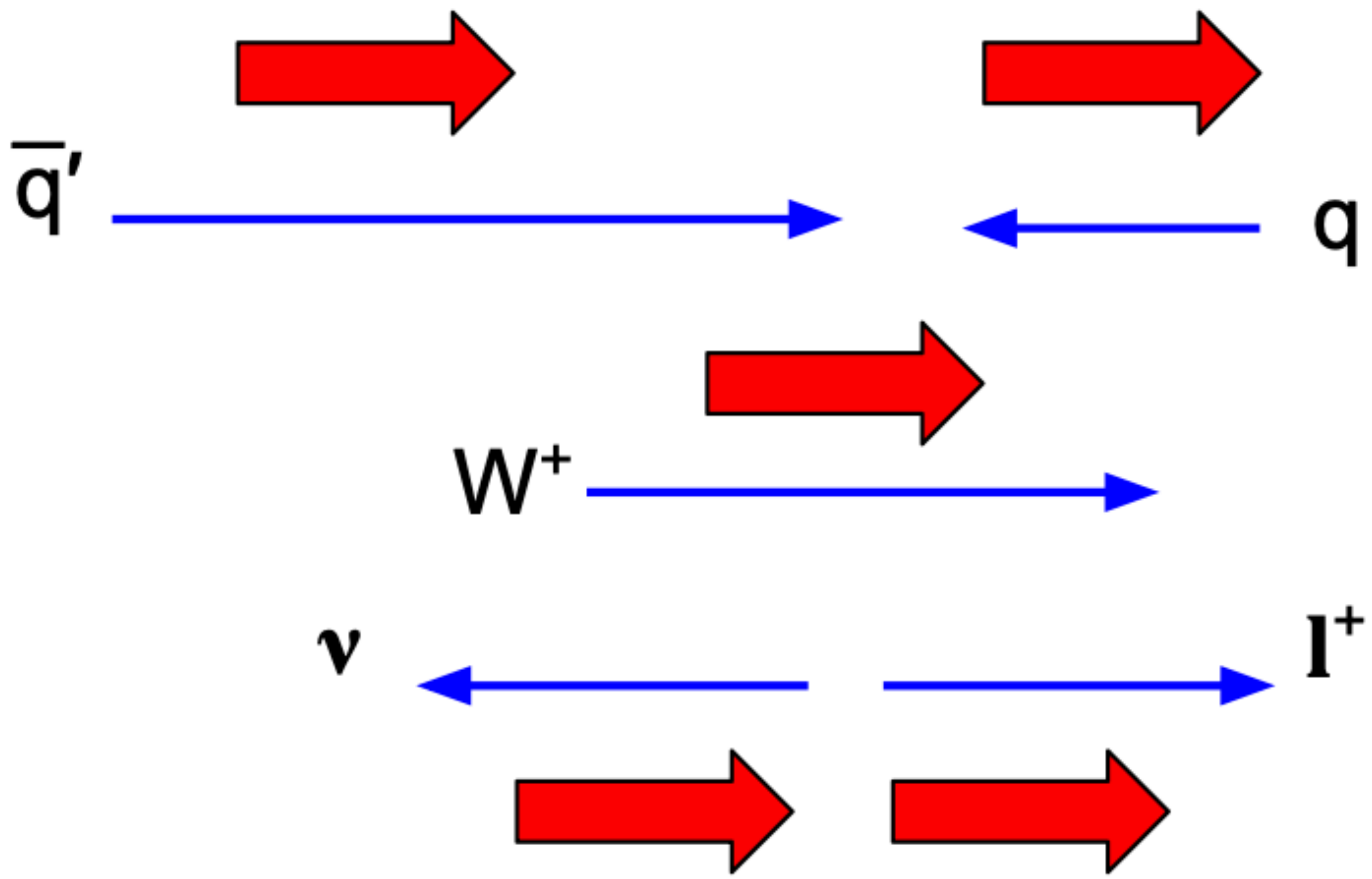
- $\sin^2\theta^l_{\mathrm{eff}}$ values with different PDFs consistent at the ~1 sigma level
- Use CT18Z as covering the central values obtained with the other sets

SMP-22-010

PDFs and W-helicities

- Lepton direction in W-boson decays retains information on the boson polarization
- Left-handed couplings of the W correlate polarization and rapidity of the boson with the direction of the quark/anti-quark, and hence the direction of the outgoing lepton



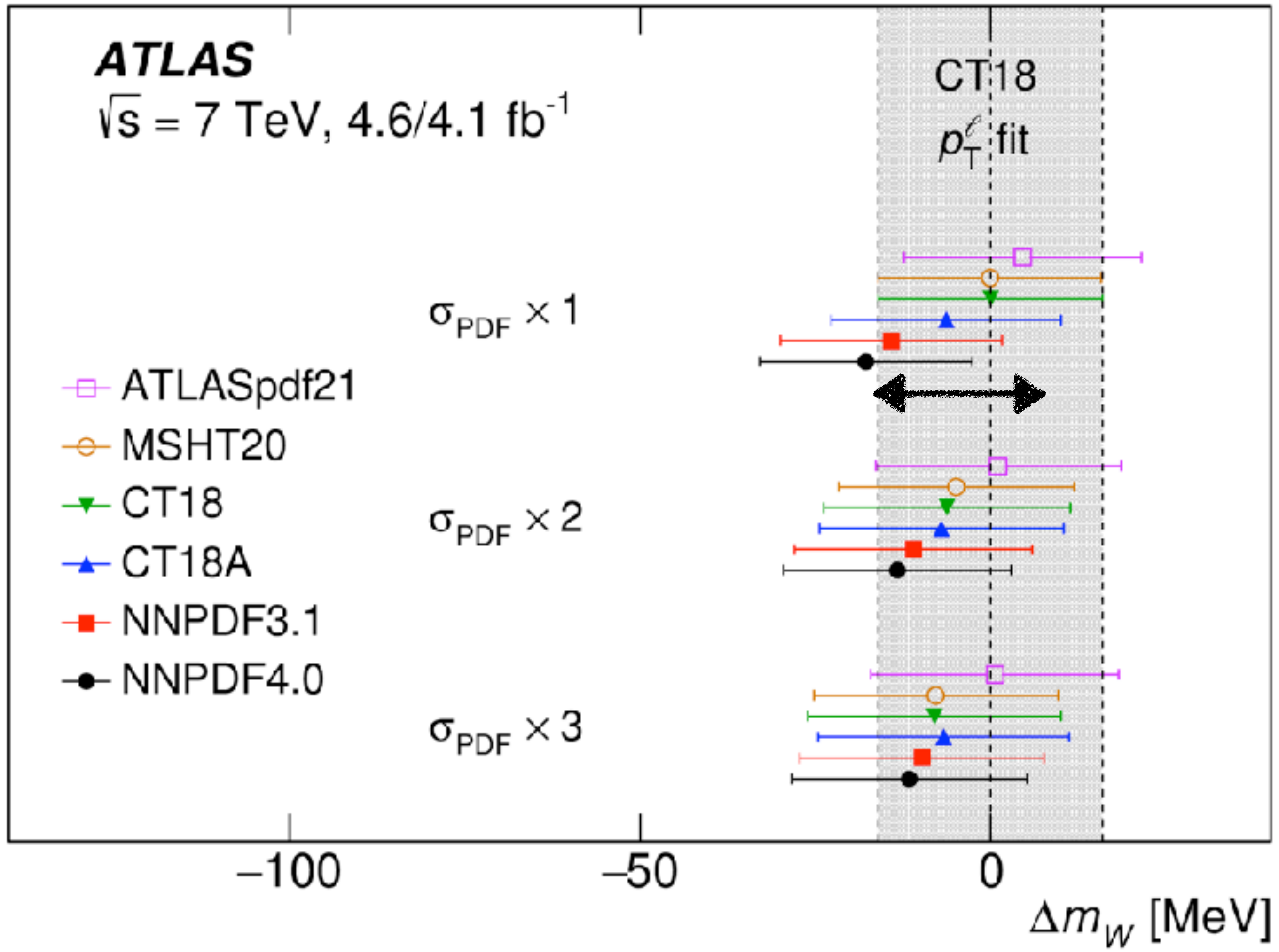


- Effect induced by PDFs, important uncertainty in W-mass determinations

PDFs in the ATLAS 7 TeV m_W

- Large PDF dependence: NNPDF4.0 and CT18NNLO differ by 21 MeV

- PDF uncertainties smaller and at the 3-9 MeV level

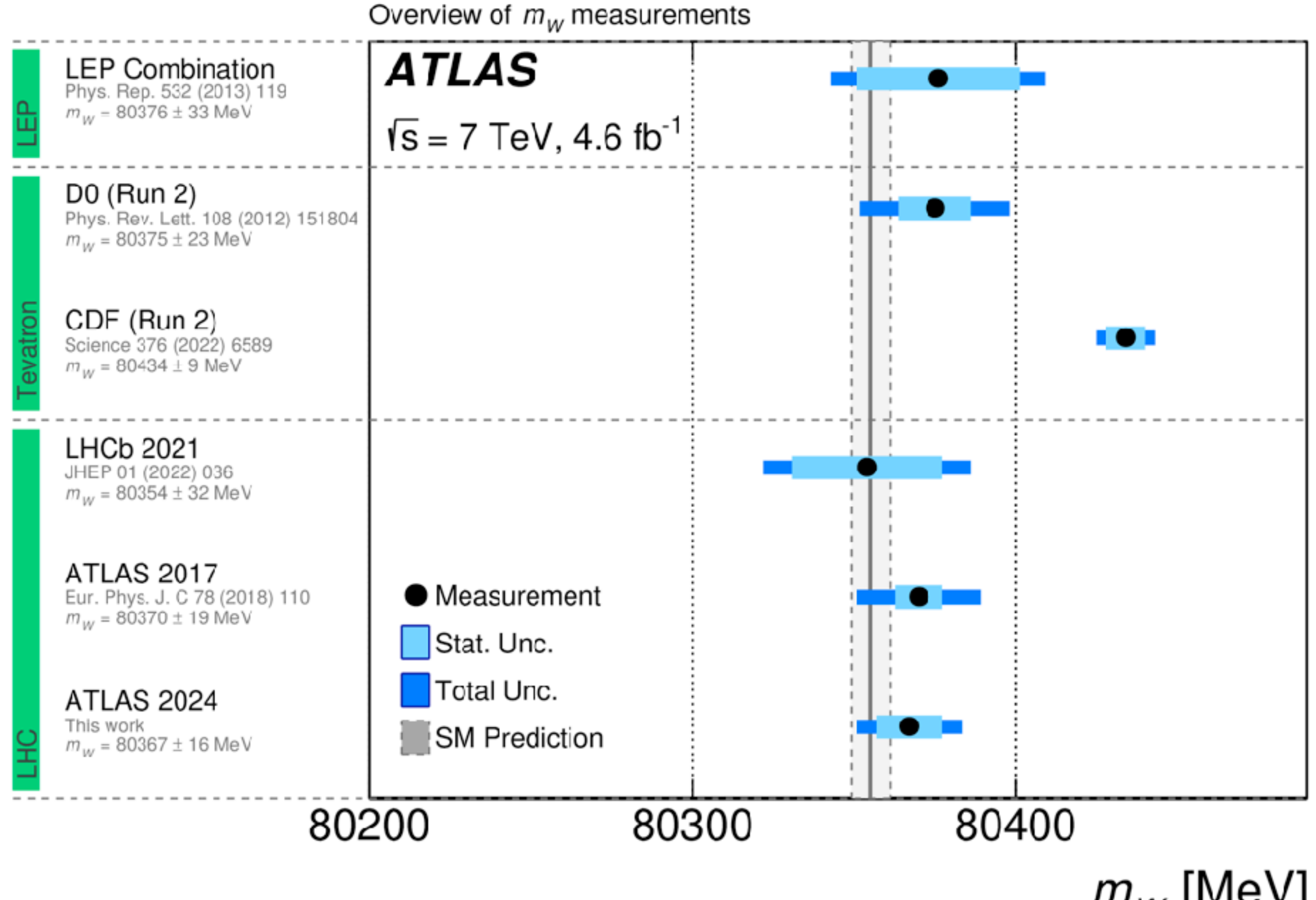


STDM-2019-24

PDF set	Combined m_W [MeV]
CT14	80363.6 ± 15.9
CT18	80366.5 ± 15.9
CT18A	80357.2 ± 15.6
MMHT2014	80366.2 ± 15.8
MSHT20	80359.3 ± 14.6
ATLASpdf21	80367.6 ± 16.6
NNPDF3.1	80349.6 ± 15.3
NNPDF4.0	80345.6 ± 14.9

PDFs in m_W combinations

- Different hadron collider measurements are correlated through the PDFs
- How much? Measurements all made with very different sets



STDM-2019-24

Mw average from LEP and LHC

Poor-man's combination until Tevatron Mw is understood

Average of LEP and LHC:

- LEP combination: 80 376 ± 25 stat ± 22 syst MeV
- ATLAS: 80 370 ± 7 stat ± 11 exp syst ± 14 model ± 8 PDF MeV
- LHCb: 80 354 ± 23 stat ± 10 exp syst ± 17 model ± 9 PDF MeV
- Assume correlations:
  - ATLAS/LHCb: model between 0 and 1, PDF between 0 and -0.5
  - LEP/LHC: none
- Mw(LHC) = 80366 ± 19 MeV
- Combine with LEP (fully uncorrelated):

  Mw(LEP+LHC) = 80369 ± 16 MeV

  (Same result if all three measurements combined in one step, with χ^2/ndf = 0.28/2)
- Previous TEV+LEP+ATLAS combination: 80379 ± 13 MeV

Roman Kogler 5 The Global Electroweak Fit

PDFs in the Tevatron/LHC m_W combination

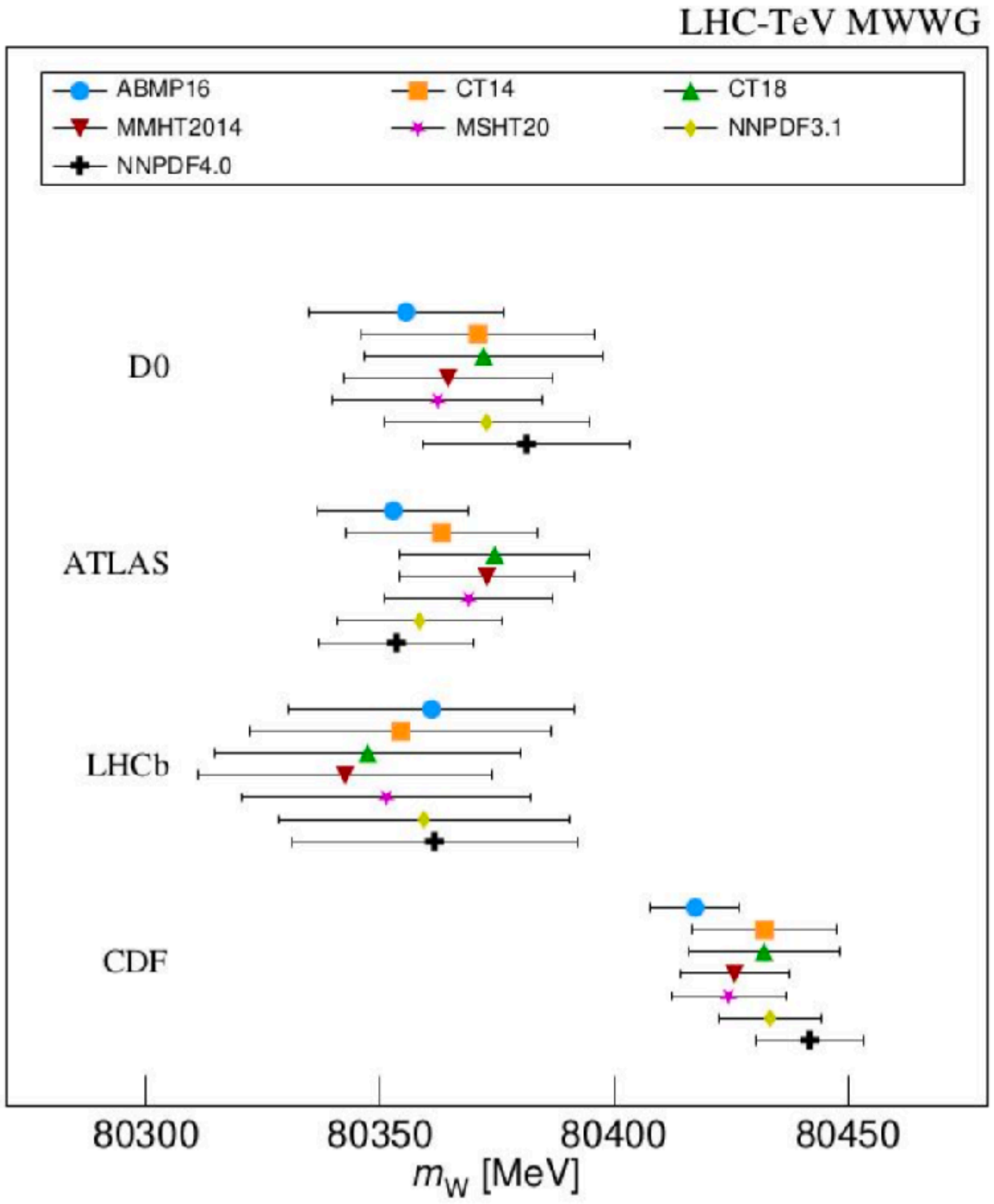
- Combination of hadron collider W-boson mass measurements

- Measurements corrected to the same PDF before averaging

- Different PDF choices evaluated

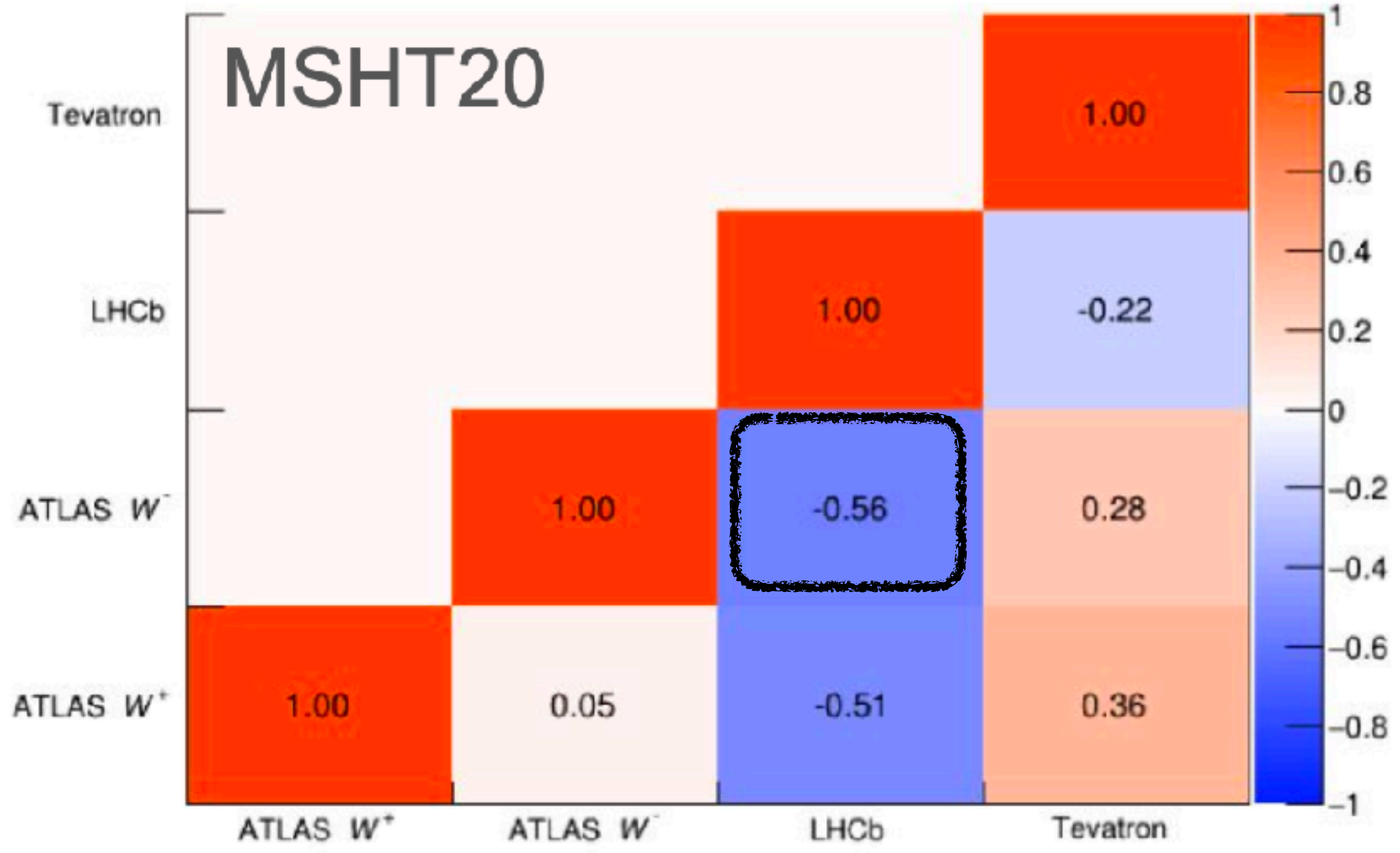
- Variations larger than the PDF uncertainties and often as large as the total uncertainty

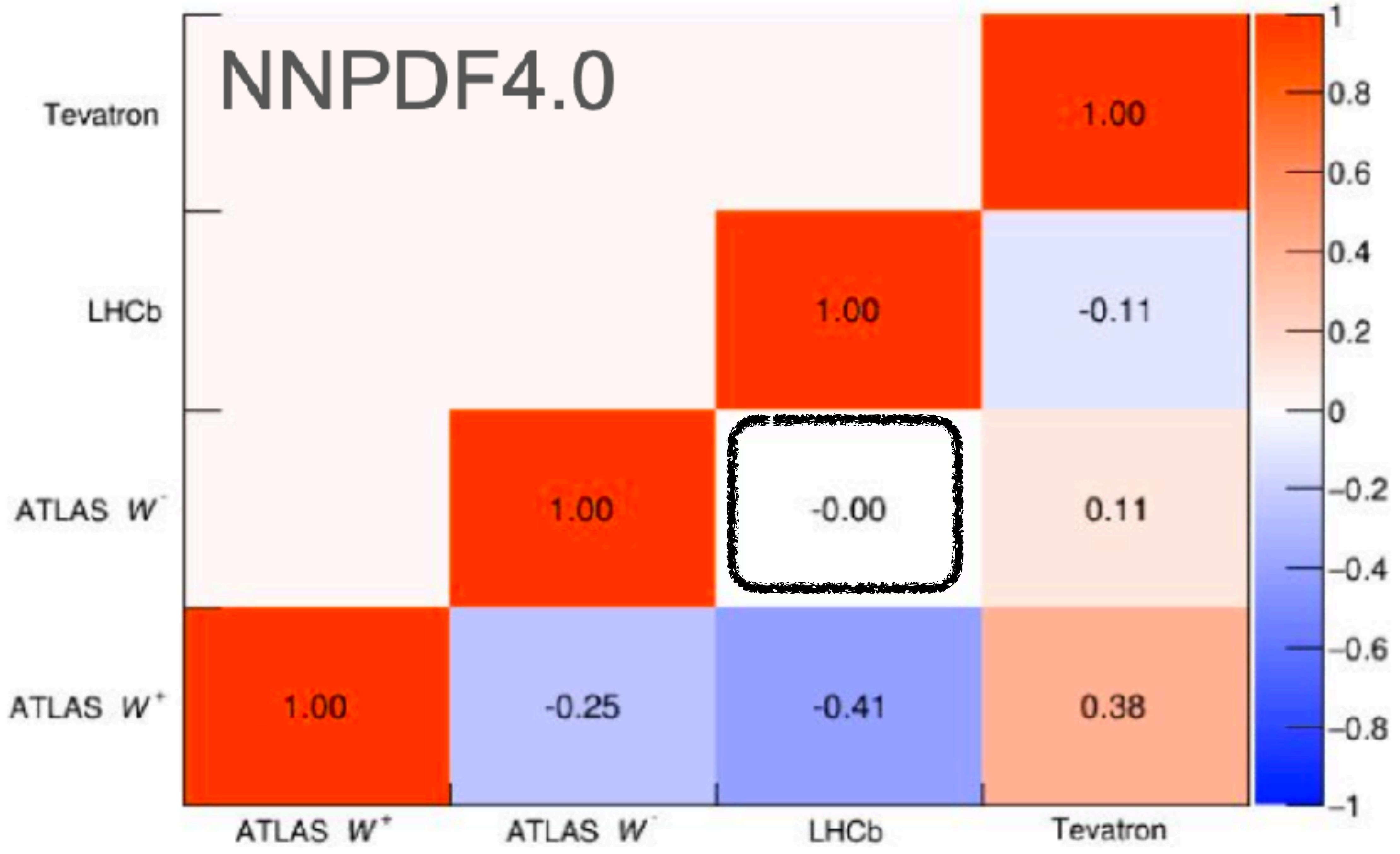
Eur.Phys.J.C 84 (2024) 5, 451



PDFs in the Tevatron/LHC m_W combination

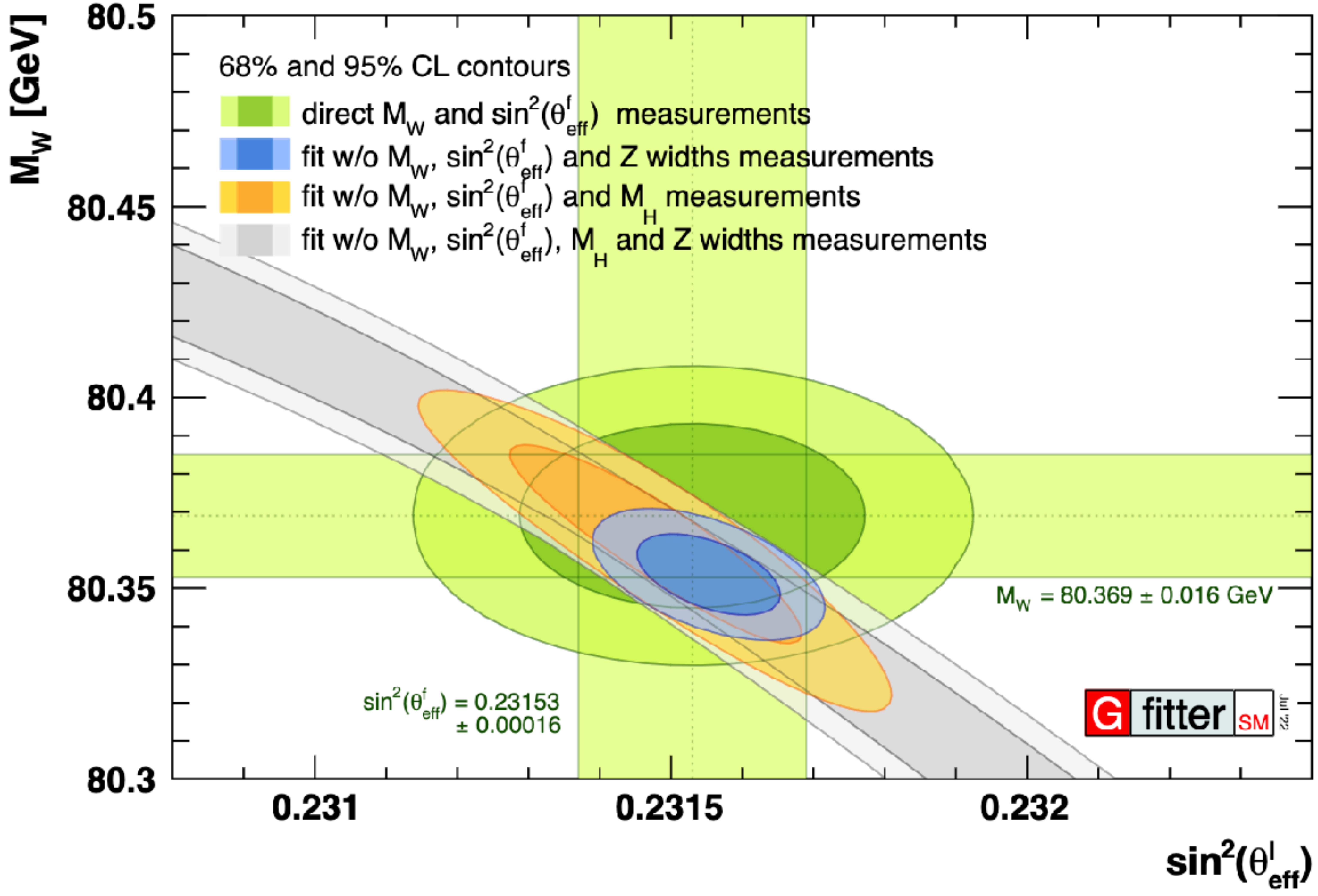
- Different predicted patter of PDF correlations for the different experiments





Eur.Phys.J.C 84 (2024) 5, 451

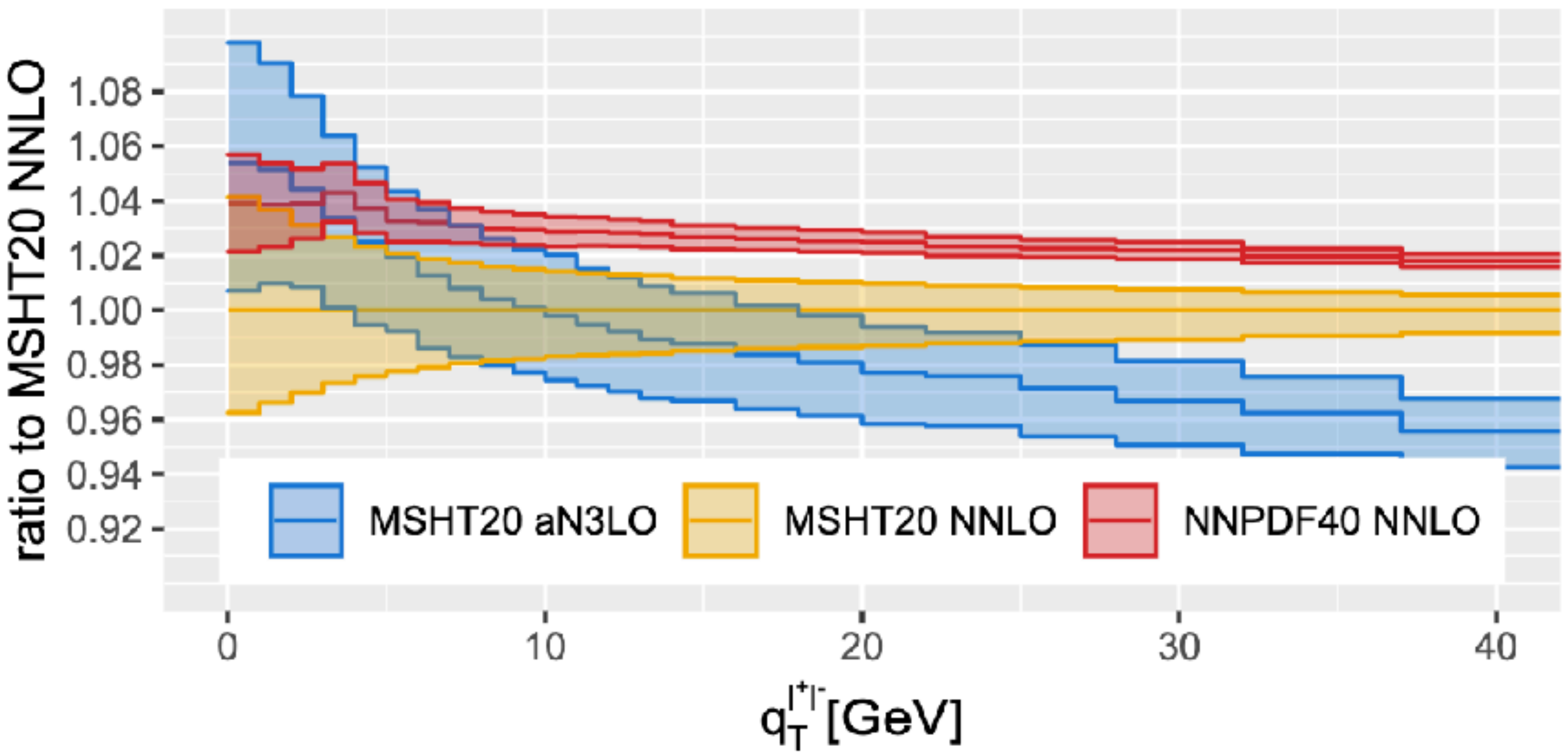
Hadron collider measurements in the EW fit

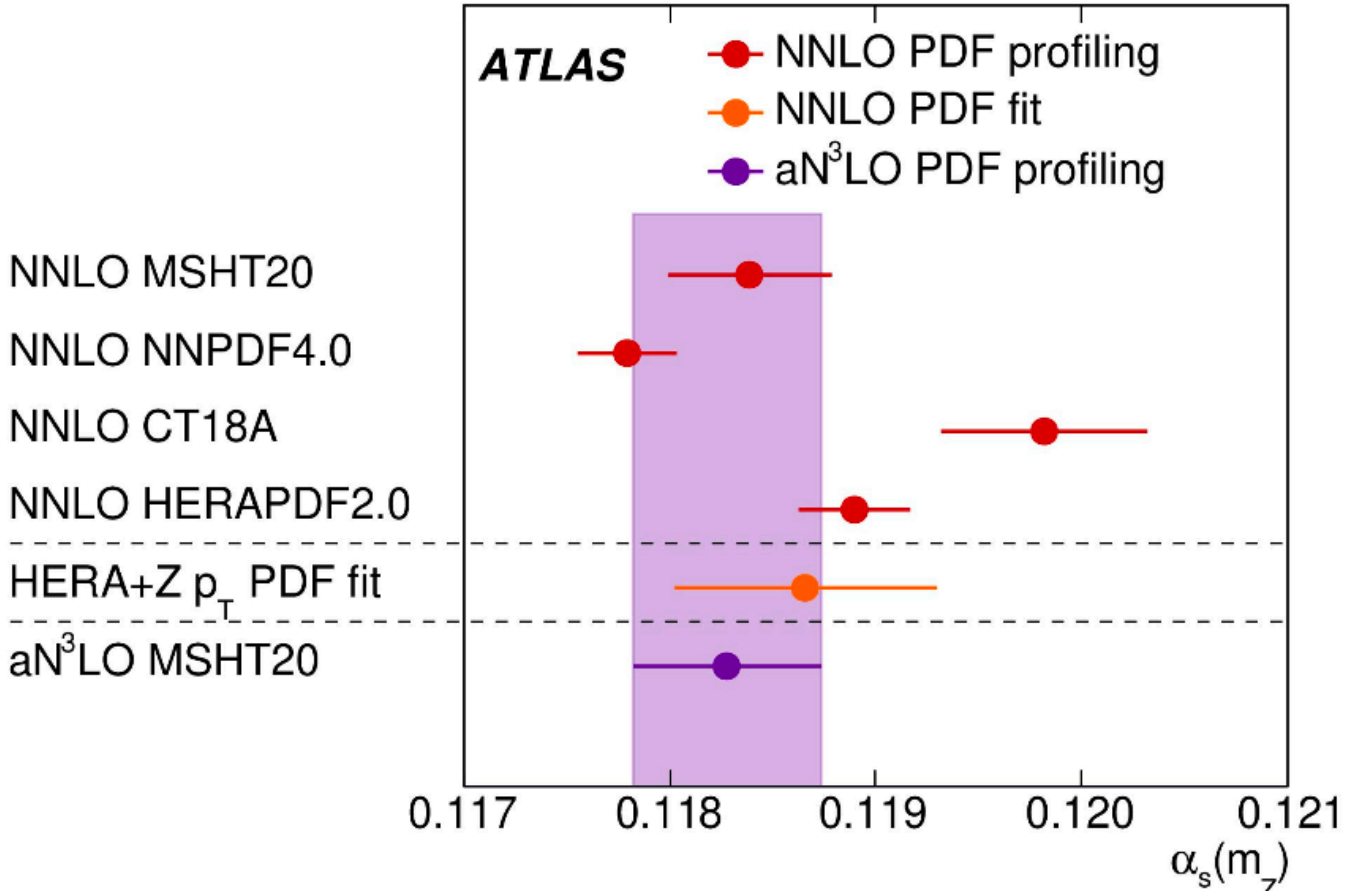


- Different measurements taken to be uncorrelated
- Not true anymore if PDFs become the largest uncertainty
- How much correlated? How much does it depend on the PDF chosen?

PDFs in ATLAS $\alpha_S(m_Z)$ from Z pT

T. Neumann, MCFM NNLO Z+jet



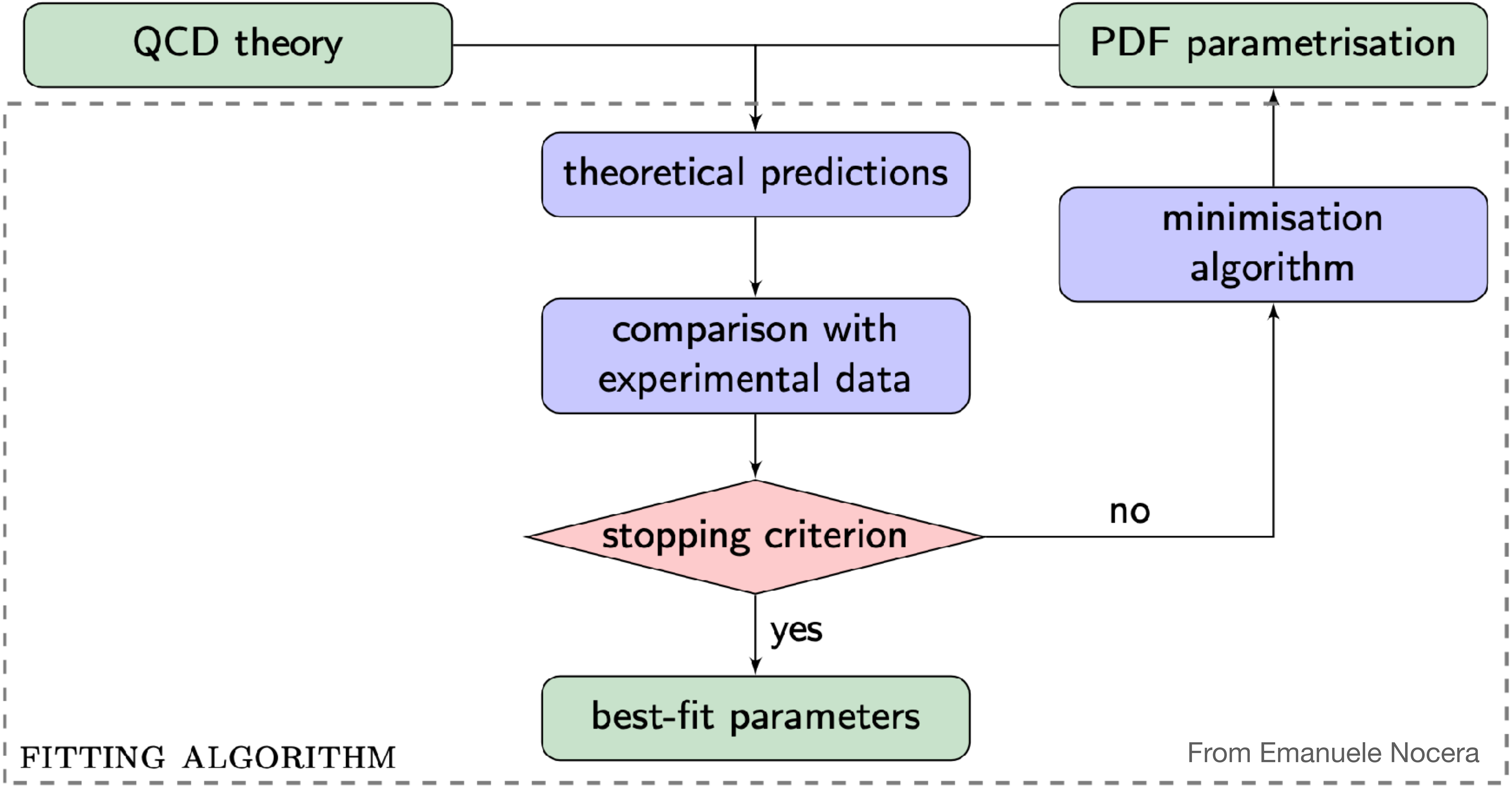


- Spread in NNLO PDFs ~0.00102, driven by NNPDF4.0 vs CT18A
- CT18 not compatible with other set within PDF uncertainties
- Often as large as the total measurement uncertainty

PDF set	$\alpha_s(m_Z)$	PDF uncertainty
MSHT20 [37]	0.11839	0.00040
NNPDF4.0 [84]	0.11779	0.00024
CT18A [29]	0.11982	0.00050
HERAPDF2.0 [65]	0.11890	0.00027

STDM-2023-01

PDFs fitting 101



From Emanuele Nocera

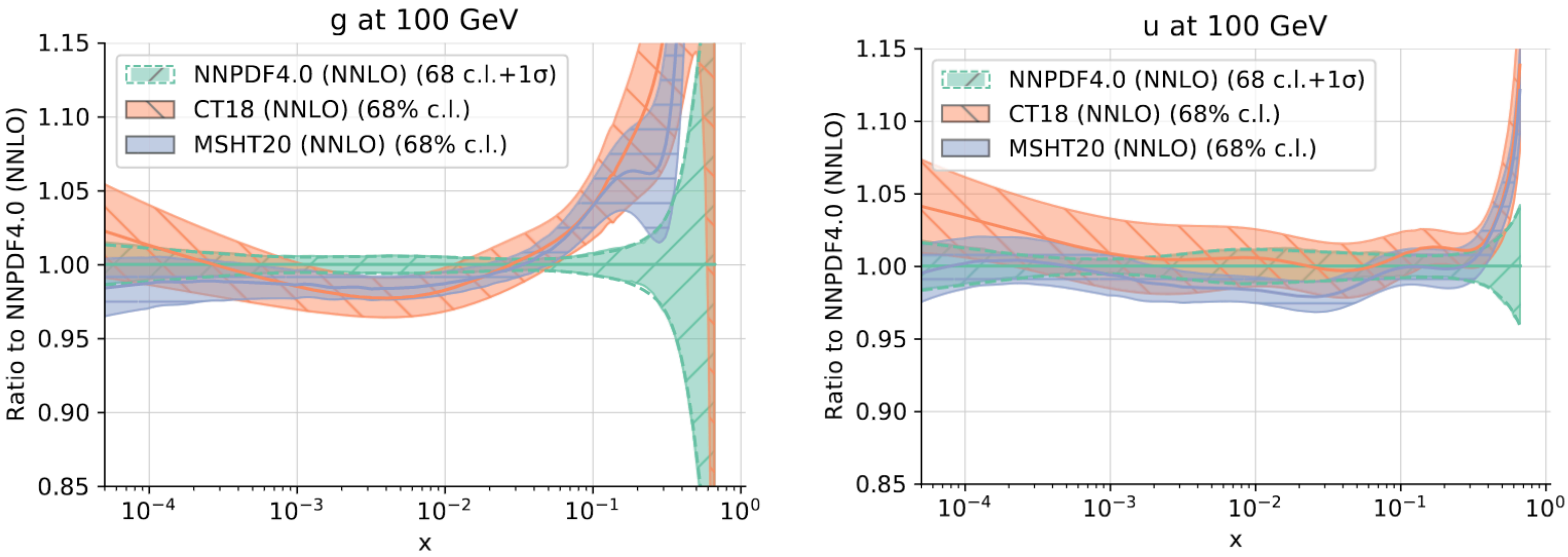
History of PDF determinations

- Global PDFs extracted by several groups making different choices in their input data, non-perturbative parametrization, heavy-flavor scheme and fit methodology: ABMP, CTEQ-TEA, MSHT, NNPDF

From Jun Gao

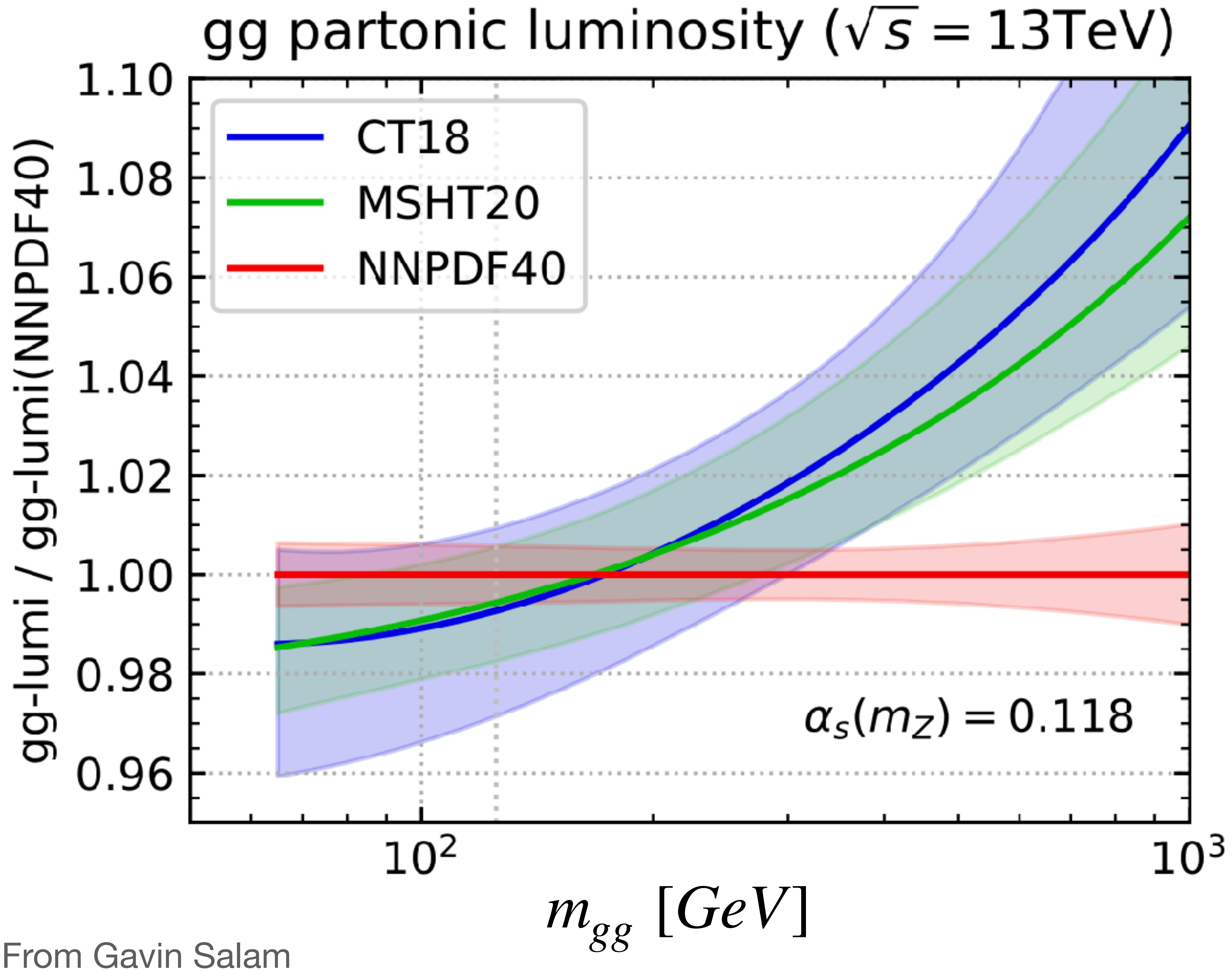
The Status of global PDF fits

- A complex problem that leads to a variety of solutions …



2109.02653

The status of global PDF fits



gg-lumi, ratio to PDF4LHC15 @ m_H

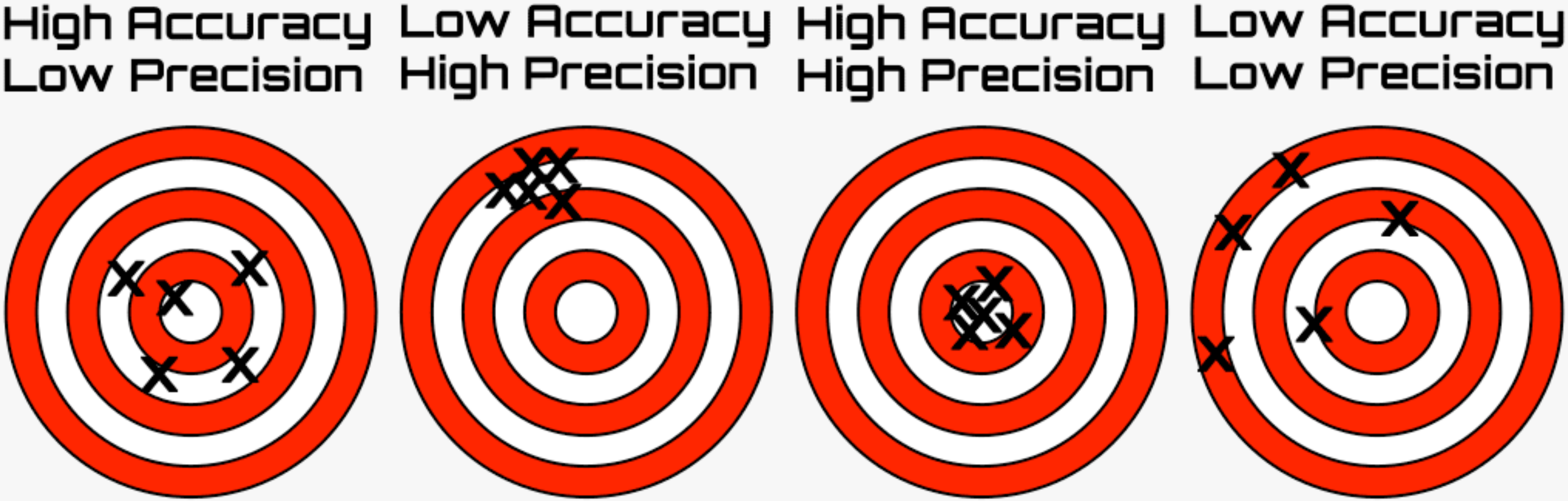
PDF4LHC15	1.0000	±	0.0184
PDF4LHC21	0.9930	±	0.0155
CT18	0.9914	±	0.0180
MSHT20	0.9930	±	0.0108
NNPDF40	0.9986	±	0.0058

$\times 3$

- NNPDF and MSHT now at %-level precision
- Yet in significant tension with each other

From Gavin Salam

Precision vs Accuracy



PDFs and their uncertainties

- Most of the measurements and theory inputs are common among PDF groups
- Expect PDF uncertainties to be largely correlated (but how much exactly?)

$$\text{Data} = \text{PDF} \otimes \sigma_H$$

Experimental:

- Dataset choice
- Experimental uncertainties
- Statistical + systematics
- Correlations
- χ^2 definition
- Outliers treatment

Methodological:

- Parametrization choice
- Regularization (what is a good PDF?)
- Uncertainties prescription (i.e. tolerances)

Theoretical:

- Perturbative QCD/EW order (NLO, NNLO, …)
- Heavy-flavor schemes (FONLL, RT, ACOT, …)
- Perturbative scales
- Theory calculations (q_T-subtr, antenna, …)
- Stat./Grid accuracies

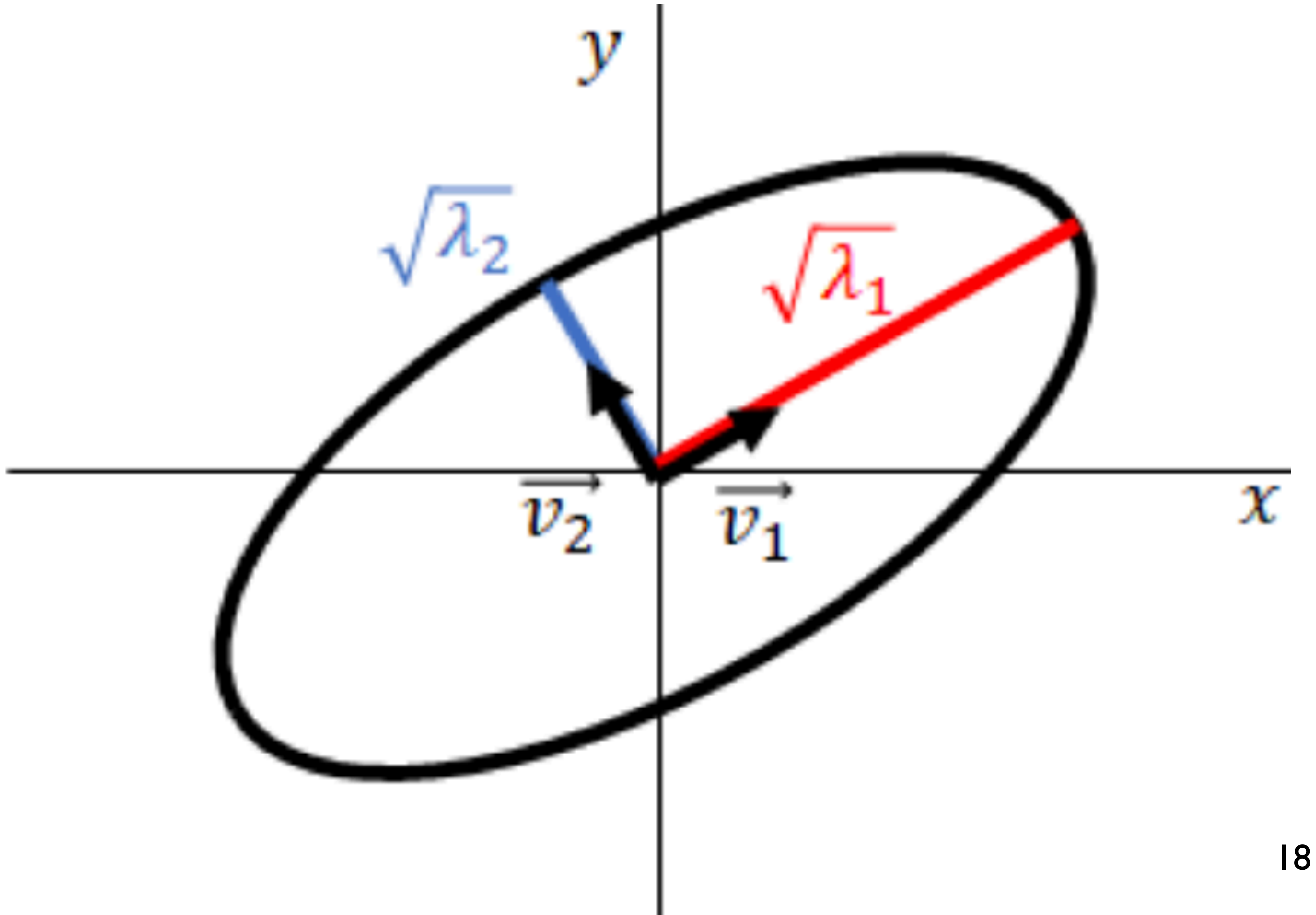
Constraining PDFs using new data

- PDF groups typically provide uncertainties as Hessian eigenvector
  - Or as replicas, but the following considerations would equally apply

- Eigenvectors provide a suitable representation of the PDF likelihood near the fit minimum suitable for propagating them in statistical analyses

- Our data carries information on the PDFs but not enough to fully determine them

- Combine the likelihood of our measurement with the (approximated) likelihood of the PDF fit to extract our parameter



Hessian Profiling of PDFs

- Include the PDF eigenvectors in the χ^2/likelihood as covariance/nuisances

Data · Theory

$$\chi^2(\mathbf{b}_{\rm exp},\mathbf{b}_{\rm th}) = \sum_{i=1}^{N_{\rm data}} \frac{\left[D_i - T_i\left(1-\sum_k \gamma_{ik}^{\rm th} b_{k,\rm th} - \sum_j \gamma_{ij}^{\rm exp} b_{j,\rm exp}\right)\right]^2}{\delta_{i,\rm uncor}^2 T_i^2 + \delta_{i,\rm stat}^2 D_i T_i}$$

Nuisance parameter impacts

$$+ \sum_i \log \frac{\delta_{i,\rm uncor}^2 T_i^2 + \delta_{i,\rm stat}^2 D_i T_i}{\delta_{i,\rm uncor}^2 D_i^2 + \delta_{i,\rm stat}^2 D_i^2}$$

Uncorrelated and statistical uncertainties

Experimental nuisances

$$+ \sum_{j=1}^{N_{\rm exp.sys}} b_{j,\rm exp}^2 + \sum_{k=1}^{N_{\rm th.sys}} b_{k,\rm th}^2 .$$

Theory nuisances

- The data pulls and constrains (a linear combination of) the PDF nuisances
- The values of the nuisance parameters at the minimum define a new *profiled PDF* with (generally) smaller uncertainties

$$f_0' = f_0 + \sum_k b_{k,\rm th}^{\rm min}\left(\frac{f_k^+ - f_k^-}{2} + b_{k,\rm th}^{\rm min}\frac{f_k^+ + f_k^- - 2f_0}{2}\right)$$

Defines “profiled” PDFs

- This reduction in PDF uncertainties happens as long as their covariance is included in the fit, even if the nuisance parameters are not explicitly used

PDF uncertainties and tolerances

- PDF uncertainties are derived using ***"Tolerance factors"*** $\Delta\chi^2 = T^2$
  - Introduced to avoid underestimated uncertainties due to bad goodness-of-fit
  - Corresponds to ***scaling the errors*** on ALL input measurements by a factor ***T^2***
- Crude approach to deal with model deficiencies, and analogous to PDG scaling

- Different groups use different approaches:
  - **CTEQ-TEA**:
    Global T^2 ~ 30 designed to also cover for different PDF parameterizations
  - **MSHT**:
    Dynamical T^2 ~ 10 different for each eigenvector
  - **NNPDF/ABMP**:
    No tolerance applied, T^2=1 (does this mean their GoF is good?)

Profiling and tolerances in LHC measurements

- When we propagate PDF eigenvectors we ignore tolerances (***T***^{2} =1)
  - The impact of our data on PDFs is overestimated
  - For CT (MSHT) PDFs we effectively assign a weight of ~30 (3) to the new data
  - Equivalent to taking the PDG average of some quantity, but removing the scaling factor on its uncertainty

- Our result is not equivalent to including our data in the original PDF fit
  - Consistent approach requires rescaling of PDF eigenvectors by ***T***^{2} and using $\Delta\chi^2 = T^2$ when deriving the constraints on the parameters

- At the same time, we extract a parameter from our measurement, not the PDFs and we do not make an average of different measurements
- What sense does it make in scaling up our measurement uncertainties due to (not-understood) tensions among some other data or with theory?

Profiling and tolerances - III

- Need to construct a procedure that allows us to:
  - Estimate the impact of our data to the PDFs consistently with how the original PDF uncertainties have been derived (with tolerance)
  - Evaluate the measurement uncertainties on the parameter of interest without inflating them (without tolerances)

- How to have different $\Delta\chi^2$ for different uncertainties ?
  - Perform a profiling of PDF with the new measurement including the proper eigenvectors rescaling and tolerance factors on $\Delta\chi^2$
  - Use the PDFs such obtained to extract the POI for fixed PDFs using $\Delta\chi^2 = 1$
  - Determine the PDF uncertainty on the POI through fixed error propagation of the reduced/rotated eigenvectors obtained in step 1

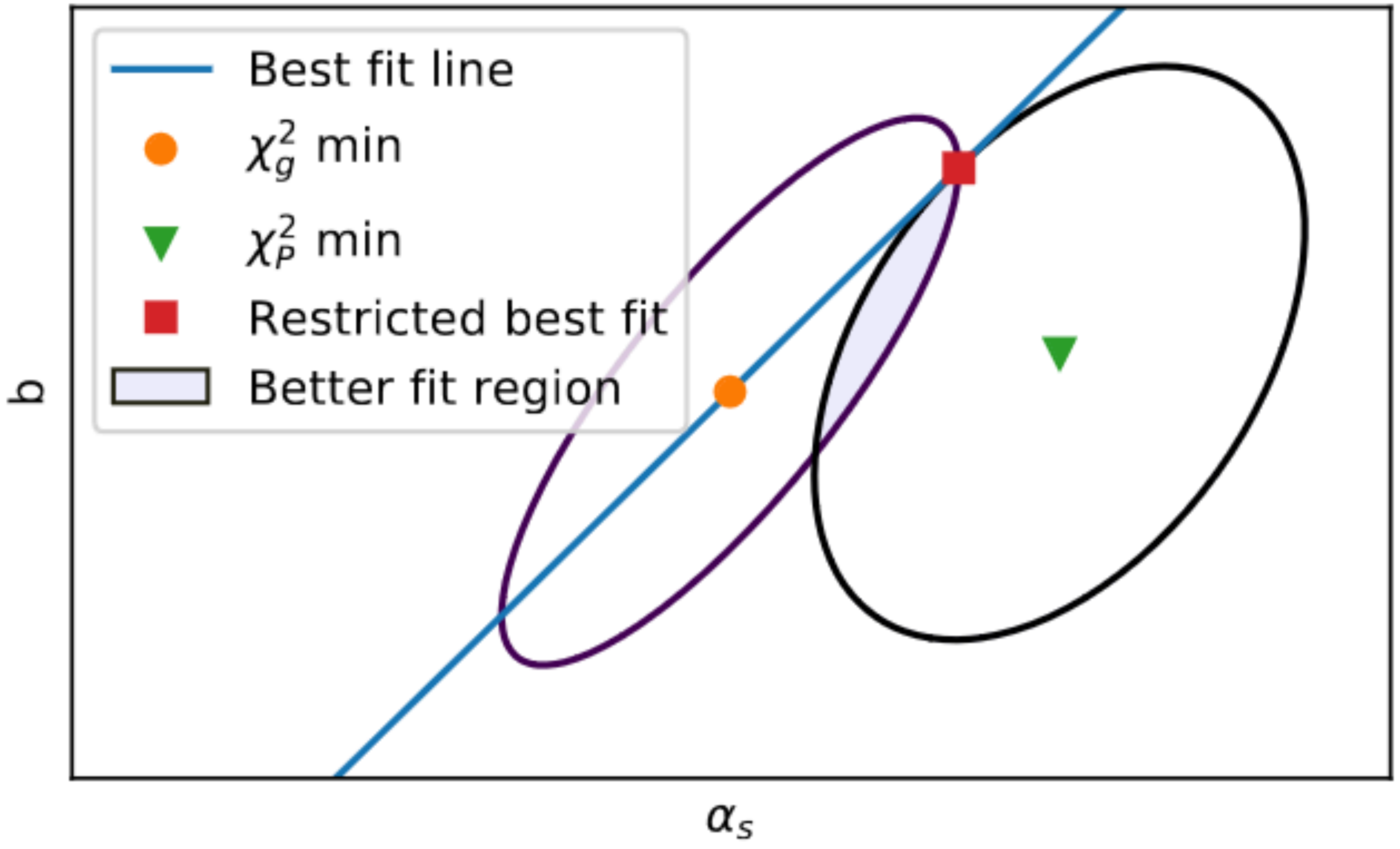
(from T. Cridge)

PDF PROFILING AND $\alpha_S(m_Z)$

- The situation when we extract the strong coupling is more delicate
  - $\alpha_S(m_Z)$ is obtained performing a χ^2 scan using PDFs obtained at fixed $\alpha_S(m_Z)$
  - PDF uncertainty evaluated at the nominal value of $\alpha_S(m_Z)$

2001.04986

  - Ignores correlations between $\alpha_S(m_Z)$ and PDFs
  - The result of a full PDF+$\alpha_S(m_Z)$ fit may well be in a direction not probed by the factorised approximation



PDFs for precise EW measurements

- Global PDF fits aim to provide the best average description using the largest possible set of input measurements
  - Requires compromises in the level of understanding one can achieve
  - For certain measurements we lack detailed experimental informations, or there are known theoretical issues in the interpretation, …

- What we need are accurate PDFs with reliable and complete uncertainties (not necessarily the smallest possible) for specific data/processes
  - Can we foresee **"reduced data PDF fits" using** a subset of **accurate and self-consistent** measurements with state-of-the-art theory ?
  - Can we have an estimate of the impact of each individual choice entering a PDF fit (input data, theory, tolerances, …) ?

The EW precision legacy of the LHC

- Result of LHC EW precision measurements will stay with us for a long time
  - Essential to ensure future-proof measurements that can be updated to newer PDFs and theoretical models and combined across experiments
  - At LEP, this was done through the definition of pseudo-observables

- Closest analogue at the LHC: **simultaneous PDF+EW** parameter extractions
  - Full control over input choices and assumptions entering the PDFs
  - Better defined (and possibly smaller) PDF uncertainties
  - Combinations and updates become (in principle) trivial

- Measurements of unfolded (NC/CC) DY cross-sections, asymmetries (A_{FB}, A_i) and boson/lepton p_T would encompass the full program of EW precision measurements (m_W, $\sin^2\theta^l_{\text{eff}}$, $\alpha_S(m_Z)$)

Summary

- Uncertainties propagation for EW measurements at the LHC ***almost broken*** by PDFs

- A **deeper understanding of PDFs** and their uncertainties is a must
  - We are not after the best average description of all measurements, but for the **most accurate PDFs using the most reliable measurements**

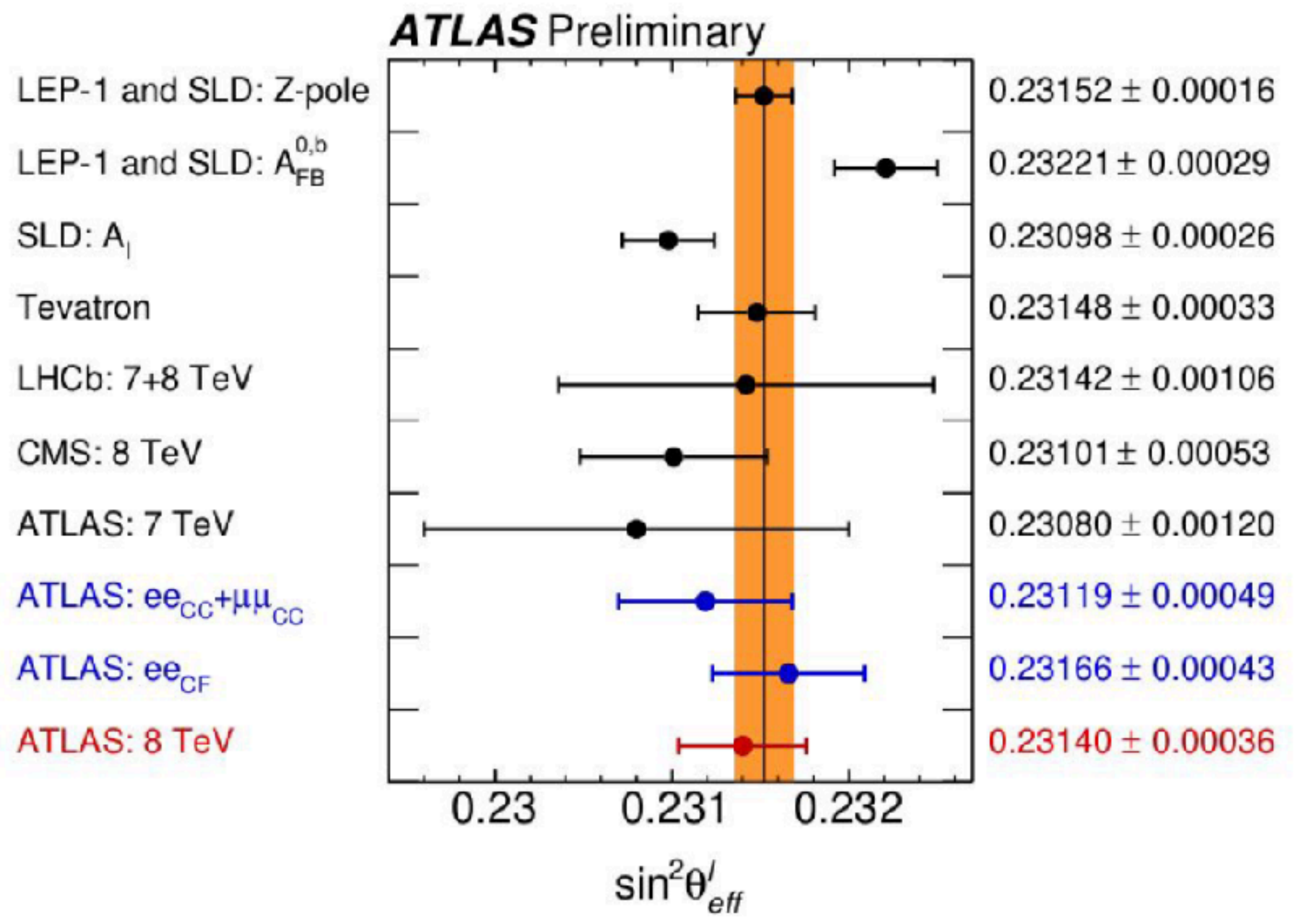
- **PDF uncertainties** on (most) hadron collider measurements **are underestimated**
  - Tolerances should be accounted for when including PDFs in fits (but how exactly?)
  - Spread of central values often larger than PDF uncertainties, how to quote it?

- Ultimately, **simultaneous PDF+EW fits** are probably the way to go to allow control over PDFs and their uncertainties and would provide a framework for preservation and reinterpretation of LHC EW precision measurements
  - But opens up new questions related to the definition, choice and consistency of the input data, and the many other prescriptions related to PDFs, …

BACKUP

PDFs in $\sin^2\theta^l_{\rm eff}$ - ATLAS 8 TeV

$$0.23140 \pm 0.00021\ (\text{stat.}) \pm 0.00024\ (\text{PDF}) \pm 0.00016\ (\text{syst.})$$



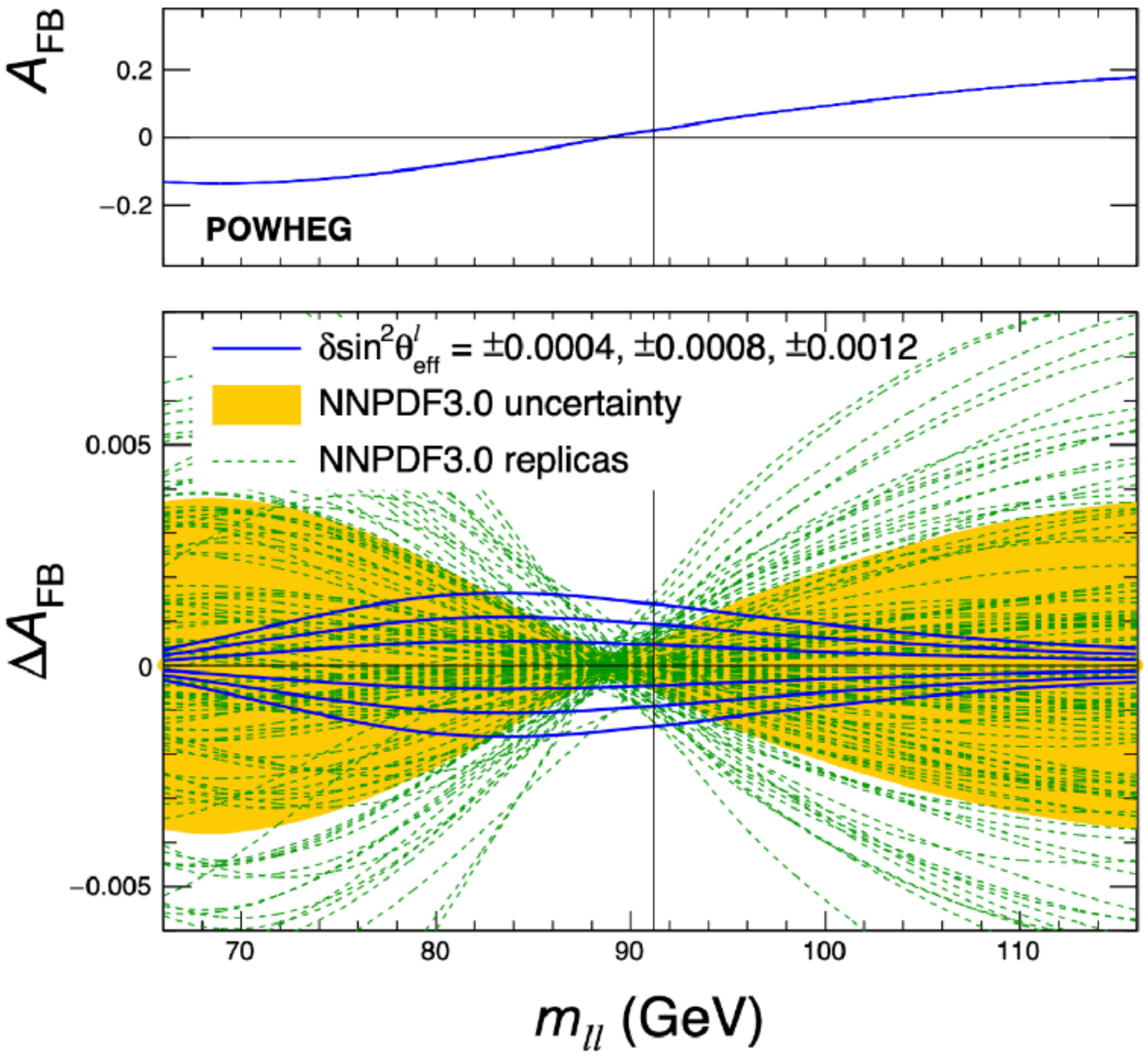
	CT10	CT14	MMHT14	NNPDF31
$\sin^2\theta^\ell_{\rm eff}$	0.23118	0.23141	0.23140	0.23146
	Uncertainties in measurements			
Total	39	37	36	38
Stat.	21	21	21	21
Syst.	32	31	29	31

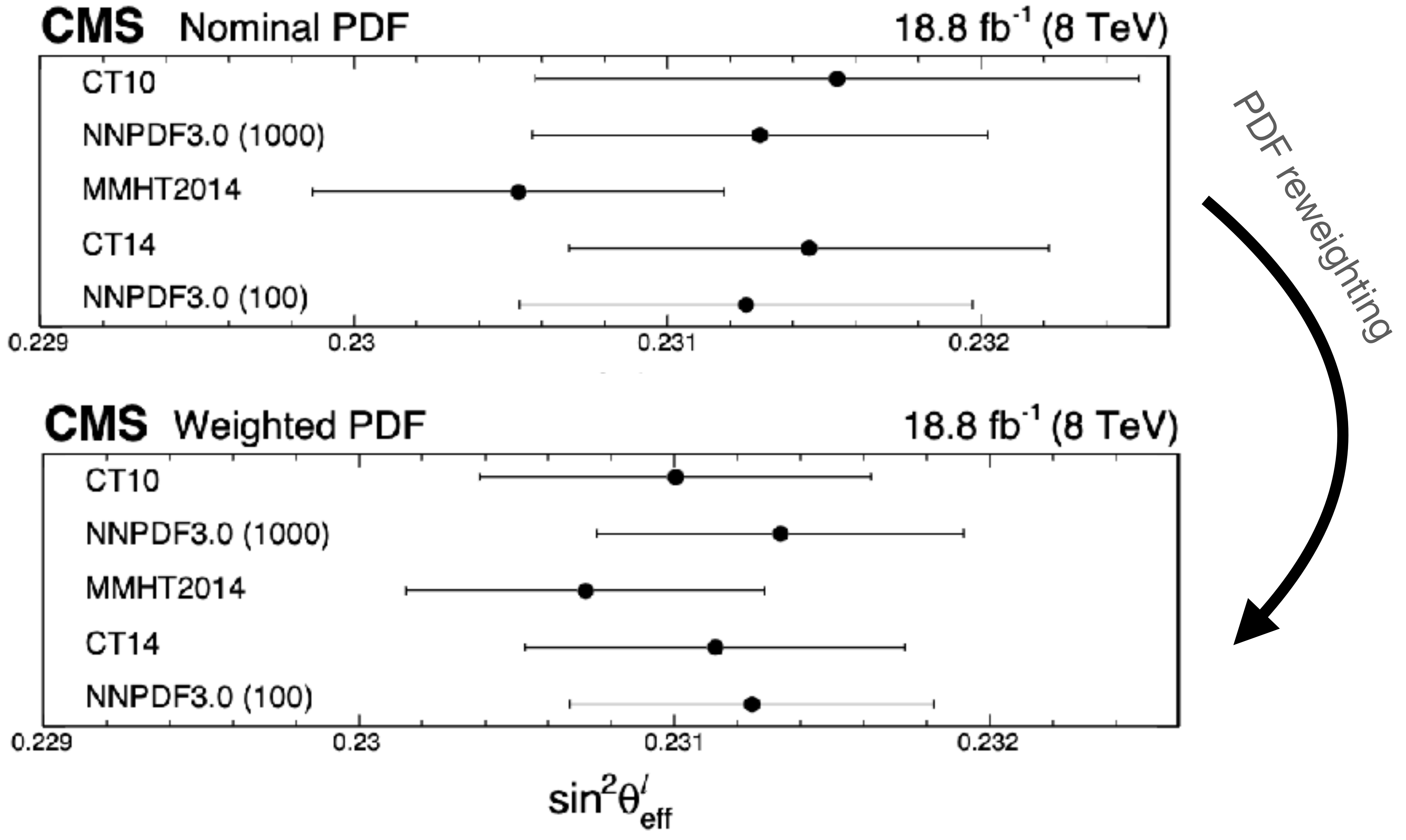
- Large uncertainty from envelope of PDFs, $3 \cdot 10^{-3}$, but using old PDF sets

PDFs in $\sin^2\theta^l_{\mathrm{eff}}$ - CMS 8 TeV

- LHC measurements rely on the correlation pattern in the PDFs to reduce their impact on the weak mixing angle

$$\sin^2\theta^\ell_{\mathrm{eff}} = 0.23101 \pm 0.00036\,(\mathrm{stat}) \pm 0.00018\,(\mathrm{syst}) \pm 0.00016\,(\mathrm{theo}) \pm 0.00031\,(\mathrm{PDF})$$





- PDF uncertainty of $3 \cdot 10^{-4}$ vs MSHT14/NNPDF30 spread of $6 \cdot 10^{-4}$

Constraining PDFs using new data

- Most of the times our data carries useful information about the PDFs that we would want to exploit, but not enough to fully determine them

- Ideally would include the new data in a simultaneous fit of the parameter of interest and the PDF (requires more data and is complicated, see later)

- Approximations allow to propagate PDF uncertainties in statistical analyses in a way that, under certain assumptions, would reproduce the result of including it in a new fit

Forced error propagation

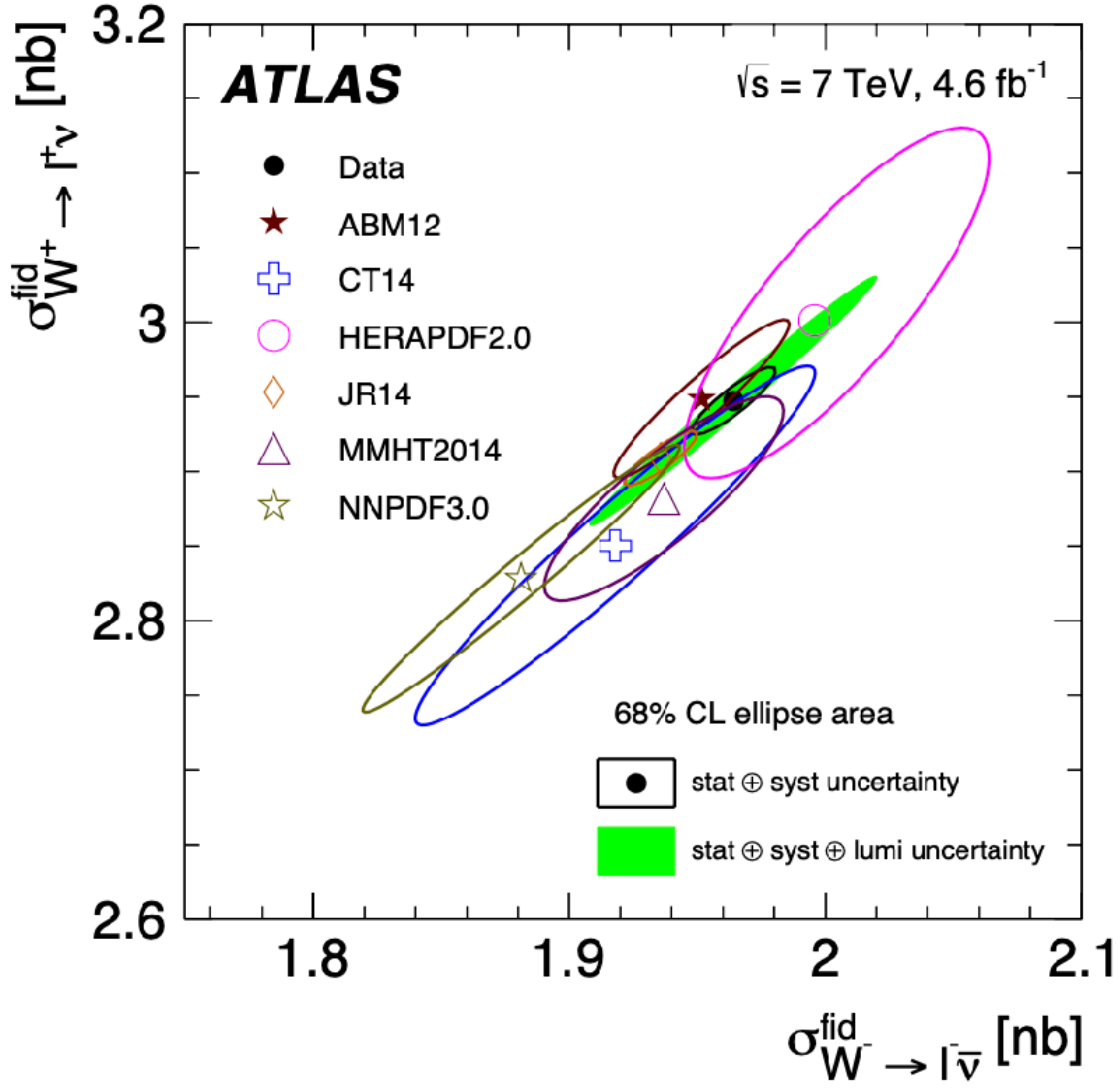
- Sometimes our data is not very sensitive to PDF or we simply want to report the original PDF uncertainty on the measured parameter

- In this case we would only like to propagate the input PDF uncertainties to the parameter of interest

- Forced error propagation (also called externalized or offset uncertainties)
  - Repeat the measurement for each eigenvector/replica
  - Evaluate the difference in the fitted parameter

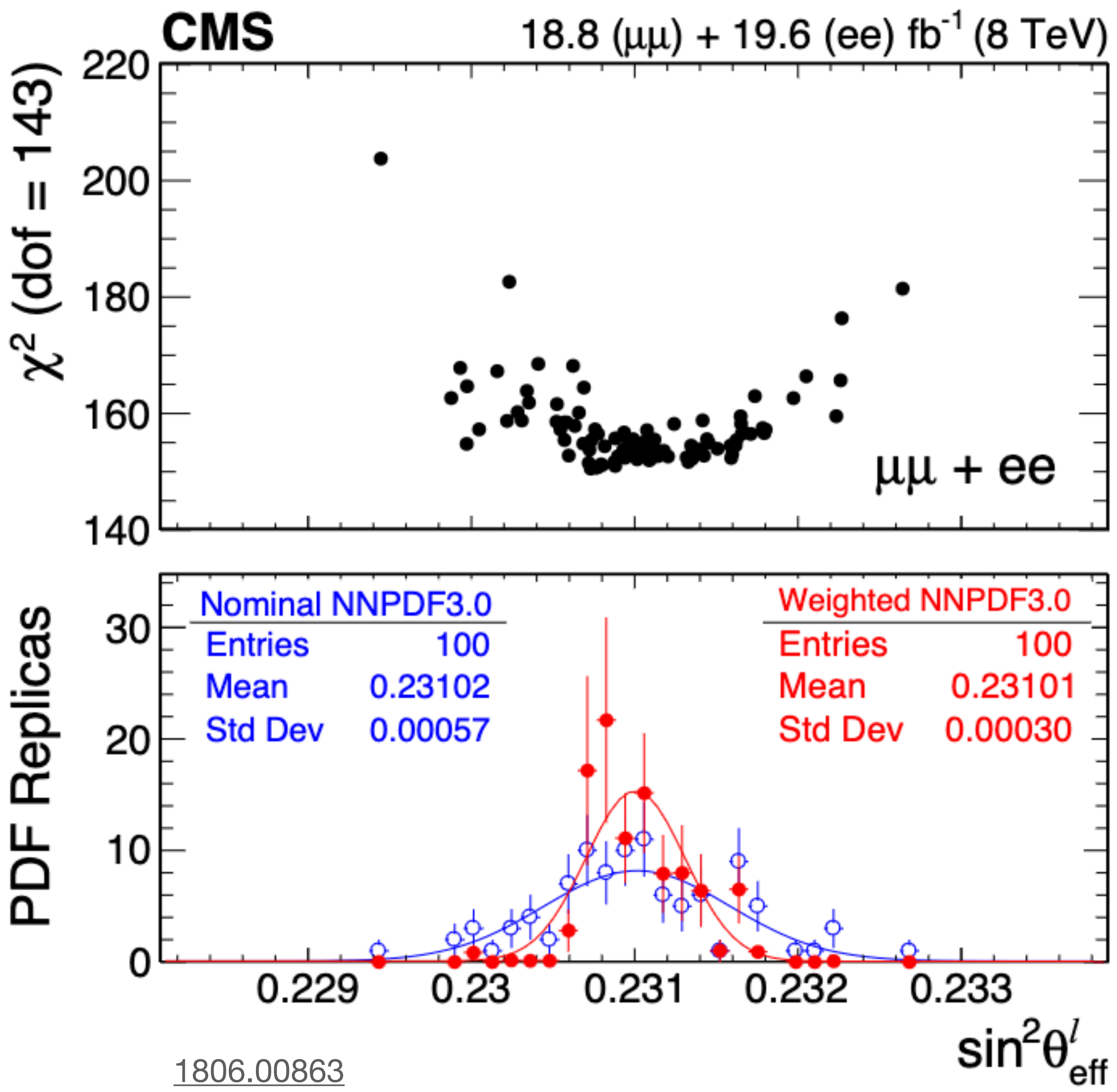
$$\delta_\theta = \hat{\mu}(\hat{\theta} + \sigma_\theta) - \hat{\mu}(\hat{\theta})$$

1612.03016



Bayesian reweighing

- A completely analogous procedure can be defined for replica uncertainties



1806.00863

- Downweight replicas far from the data according to:

$$w_k = \frac{(\chi^2_k)^{\frac{1}{2}(N_{\rm data}-1)}\,\exp^{-\frac{1}{2}\chi^2_k}}{\frac{1}{N_{\rm rep}}\sum_{k=1}^{N_{\rm rep}}(\chi^2_k)^{\frac{1}{2}(N_{\rm data}-1)}\,\exp^{-\frac{1}{2}\chi^2_k}}$$

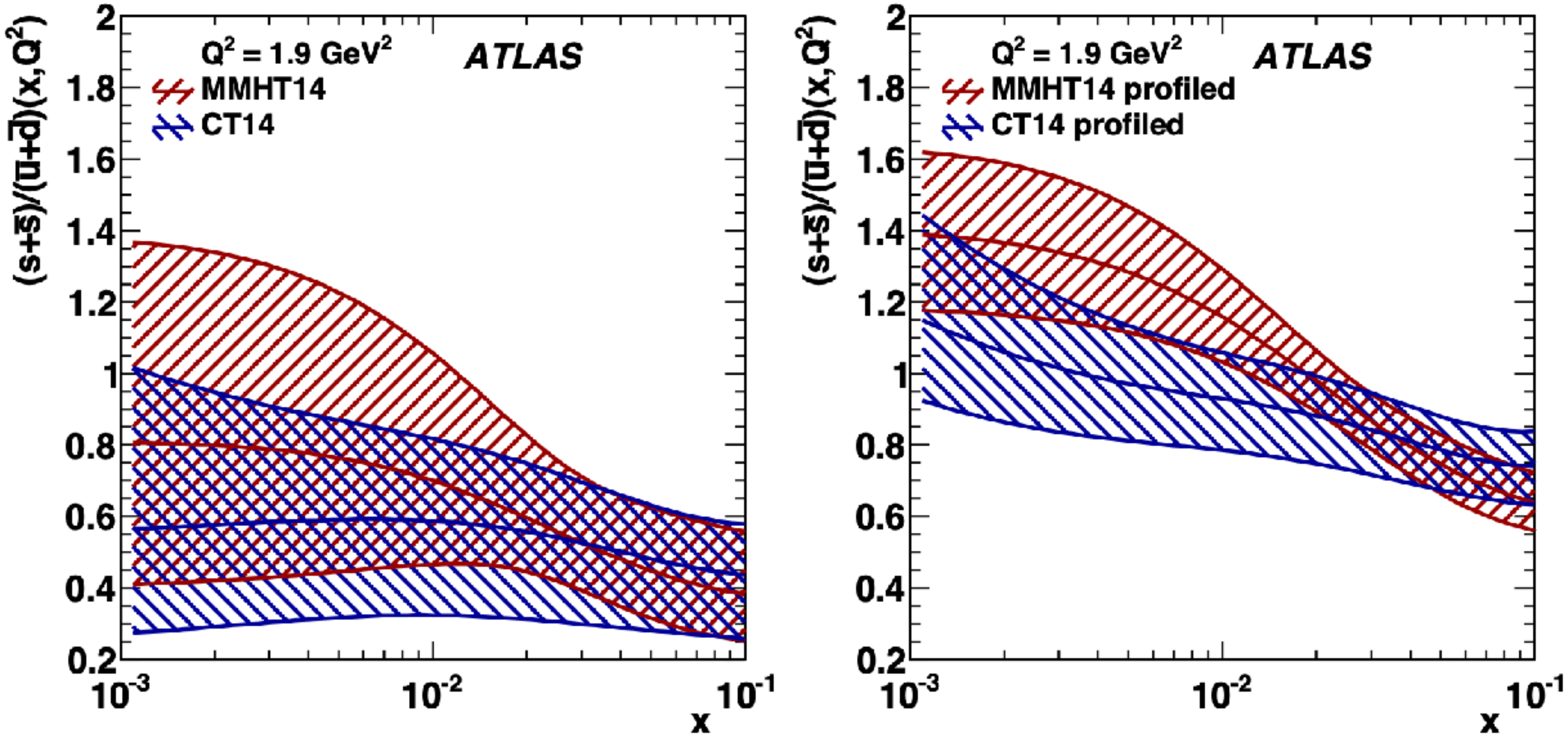
- Recompute the uncertainty on your observable with the reweighted replicas

$$\langle \mathcal{O}^{\rm new}({\rm PDF})\rangle = \frac{1}{N_{\rm rep}}\sum_{k=1}^{N_{\rm rep}} w_k \mathcal{O}({\rm PDF}_k)$$

Profiling and tolerances - I

- Hessian approximation is only valid around the minimum
- In most cases including the new data strongly constraints and modifies the original PDFs
- When are the profiling results reliable?

1612.03016



Augmented likelihood for PDFs with global tolerance

1. Start by defining the correspondence between $\Delta\chi^2$ and cumulative probability level: 68% c.l. ⇔ $\Delta\chi^2 = T^2$.
2. Write the **augmented** likelihood density for this definition:

$$P(D_i|T_i) \propto e^{-\chi^2/(2T^2)}$$

3. When profiling 1 new experiment with the prior imposed on PDF nuisance parameters $\lambda_{\alpha,th}$:

$$\chi^2(\vec{\lambda}_{\rm exp}, \vec{\lambda}_{\rm th}) = \sum_{i=1}^{N_{pt}} \frac{\left[D_i + \sum_\alpha \beta_{i,\alpha}^{\rm exp}\lambda_{\alpha,\rm exp} - T_i - \sum_\alpha \beta_{i,\alpha}^{\rm th}\lambda_{\alpha,\rm th}\right]^2}{s_i^2} + \sum_\alpha \lambda_{\alpha,\rm exp}^2 + \sum_\alpha T^2\lambda_{\alpha,\rm th}^2. \qquad \beta_{i,\alpha}^{\rm th} = \frac{T_i(f_\alpha^+) - T_i(f_\alpha^-)}{2},$$

new experiment

priors on expt. systematics and PDF params

4. Alternatively, we can reparametrize $\chi^{2\prime} \equiv \chi^2/T^2$, so that 68% c.l. ⇔ $\Delta\chi^{2\prime} = 1$. We have

$$P(D_i|T_i) \propto e^{-\chi^{2\prime}/2}$$

$$\chi^{2\prime}(\vec{\lambda}_{\rm exp}, \vec{\lambda}_{\rm th}) = \sum_{i=1}^{N_{pt}} \frac{\left[D_i + \sum_\alpha \beta_{i,\alpha}^{\rm exp}\lambda_{\alpha,\rm exp} - T_i - \sum_\alpha \beta_{i,\alpha}^{\rm th}\lambda_{\alpha,\rm th}\right]^2}{s_i^2\ T^2} + \sum_\alpha \frac{\lambda_{\alpha,\rm exp}^2}{T^2} + \sum_\alpha \lambda_{\alpha,\rm th}^2.$$

consistent redefinition

5. Inconsistent redefinitions:

$$\chi^{2\prime}(\vec{\lambda}_{\rm exp}, \vec{\lambda}_{\rm th}) = \sum_{i=1}^{N_{pt}} \frac{\left[D_i + \sum_\alpha \beta_{i,\alpha}^{\rm exp}\lambda_{\alpha,\rm exp} - T_i - \sum_\alpha \beta_{i,\alpha}^{\rm th}\lambda_{\alpha,\rm th}\right]^2}{s_i^2} + \sum_\alpha \lambda_{\alpha,\rm exp}^2 + \sum_\alpha \lambda_{\alpha,\rm th}^2.$$

and $P(D_i|T_i) \propto e^{-\chi^{2\prime}/2}$

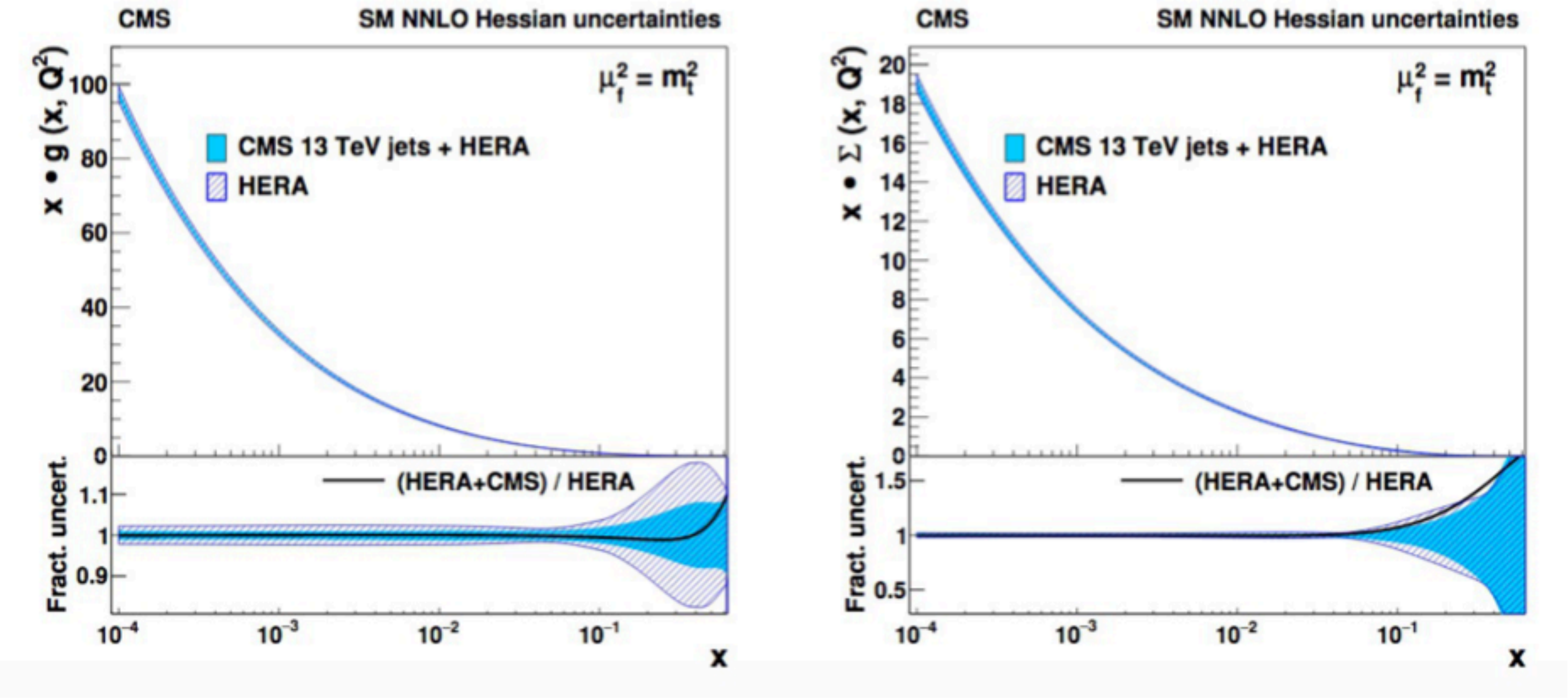
or $P(D_i|T_i) \propto e^{-\chi^{2\prime}/(2T^2)}$

[equivalent to $s_i \to s_i/T$ or $\lambda_{\alpha,th} \to \lambda_{\alpha,th}T$ without $\beta_{i,\alpha,th} \to \beta_{i,\alpha,th}/T$]

Refitting PDFs

- Sometimes data is so sensitive to PDFs that we know profiling would not work
  - Or our measurement is already included in PDF sets, using it to extract other SM parameters would give biased results
- In those cases **perform a new PDF fit** including minimal relevant data

2111.10431



- Often done in the context of alphas extractions from jet measurements
- CMS 13 teV inclusive jets fitted with HERA DIS data to obtain:

$$\alpha_S(m_Z) = 0.1166 \pm 0.0014\,(\text{fit}) \pm 0.0007\,(\text{model}) \pm 0.0004\,(\text{scale}) \pm 0.0001\,(\text{param.})$$

LHC measurements and PDFs

- PDFs most precisely determined from DIS data, but not all combinations probed
  - d_v is less precisely determined than u_v, no flavour decomposition of the light sea

- LHC data cannot replace DIS, but can provide complementary information and help resolve tensions and disagreements which happen in global fits (or among them)

Drell-Yan	Flavour decomposition of the sea, u_v, d_v, γ PDF
W+charm	Strange PDF
Jets	High-x gluon PDF
Photon	Medium-x gluon PDF
Top pair	Medium- and high-x gluon PDF

LHC data in PDF fits

Data set	Ref.	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
ATLAS W, Z 7 TeV ($\mathcal{L} = 35$ pb^{-1})	[51]	✓	✓	✓	✓	✓
ATLAS W, Z 7 TeV ($\mathcal{L} = 4.6$ fb^{-1})	[52]	✓	✓	✗	(✓)	✓
ATLAS low-mass DY 7 TeV	[53]	✓	✓	✗	(✓)	✗
ATLAS high-mass DY 7 TeV	[54]	✓	✓	✗	(✓)	✓
ATLAS W 8 TeV	[79]	✗	(✓)	✗	✗	✓
ATLAS DY 2D 8 TeV	[78]	✗	✓	✗	✗	✓
ATLAS high-mass DY 2D 8 TeV	[77]	✗	✓	✗	(✓)	✓
ATLAS $\sigma_{W,Z}$ 13 TeV	[81]	✗	✓	✓	✗	✗
ATLAS W+jet 8 TeV	[93]	✗	✓	✗	✗	✓
ATLAS Z p_T 7 TeV	[259]	(✓)	✗	✗	(✓)	✗
ATLAS Z p_T 8 TeV	[63]	✓	✓	✗	✓	✓
ATLAS $W + c$ 7 TeV	[83]	✗	✓	✗	(✓)	✗
ATLAS $\sigma_{t\bar{t}}^{\rm tot}$ 7, 8 TeV	[65]	✓	✓	✓	✗	✗
ATLAS $\sigma_{t\bar{t}}^{\rm tot}$ 7, 8 TeV	[260–265]	✗	✗	✓	✗	✗
ATLAS $\sigma_{t\bar{t}}^{\rm tot}$ 13 TeV ($\mathcal{L} = 3.2$ fb^{-1})	[66]	✓	✗	✓	✗	✗
ATLAS $\sigma_{t\bar{t}}^{\rm tot}$ 13 TeV ($\mathcal{L} = 139$ fb^{-1})	[134]	✗	✓	✗	✗	✗
ATLAS $\sigma_{t\bar{t}}^{\rm tot}$ and Z ratios	[266]	✗	✗	✗	✗	(✓)
ATLAS $t\bar{t}$ lepton+jets 8 TeV	[67]	✓	✓	✗	✓	✓
ATLAS $t\bar{t}$ dilepton 8 TeV	[89]	✗	✓	✗	✗	✓
ATLAS single-inclusive jets 7 TeV, R=0.6	[73]	✓	(✓)	✗	✓	✓
ATLAS single-inclusive jets 8 TeV, R=0.6	[86]	✗	✓	✗	✗	✗
ATLAS dijets 7 TeV, R=0.6	[148]	✗	✓	✗	✗	✗
ATLAS direct photon production 8 TeV	[100]	✗	(✓)	✗	✗	✗
ATLAS direct photon production 13 TeV	[101]	✗	✓	✗	✗	✗
ATLAS single top R_t 7, 8, 13 TeV	[94, 96, 98]	✗	✓	✓	✗	✗
ATLAS single top diff. 7 TeV	[94]	✗	✓	✗	✗	✗
ATLAS single top diff. 8 TeV	[96]	✗	✓	✗	✗	✗

Data set	Ref.	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
CMS W asym. 7 TeV ($\mathcal{L} = 36$ pb^{-1})	[267]	✗	✗	✗	✗	✓
CMS Z 7 TeV ($\mathcal{L} = 36$ pb^{-1})	[268]	✗	✗	✗	✗	✓
CMS W electron asymmetry 7 TeV	[55]	✓	✓	✗	✓	✓
CMS W muon asymmetry 7 TeV	[56]	✓	✓	✓	✓	✗
CMS Drell-Yan 2D 7 TeV	[57]	✓	✓	✗	(✓)	✓
CMS Drell-Yan 2D 8 TeV	[269]	(✓)	✗	✗	✗	✗
CMS W rapidity 8 TeV	[58]	✓	✓	✓	✓	✓
CMS W, Z p_T 8 TeV ($\mathcal{L} = 18.4$ fb^{-1})	[270]	✗	✗	✗	(✓)	✗
CMS Z p_T 8 TeV	[64]	✓	✓	✗	(✓)	✗
CMS $W + c$ 7 TeV	[76]	✓	✓	✗	(✓)	✓
CMS $W + c$ 13 TeV	[84]	✗	✓	✗	✗	(✓)
CMS single-inclusive jets 2.76 TeV	[75]	✓	✗	✗	✗	✓
CMS single-inclusive jets 7 TeV	[147]	✓	(✓)	✗	✓	✓
CMS dijets 7 TeV	[74]	✗	✓	✗	✗	✗
CMS single-inclusive jets 8 TeV	[87]	✗	✓	✗	✓	✓
CMS 3D dijets 8 TeV	[149]	✗	(✓)	✗	✗	✗
CMS $\sigma_{t\bar{t}}^{\rm tot}$ 5 TeV	[88]	✗	✓	✗	✗	✗
CMS $\sigma_{t\bar{t}}^{\rm tot}$ 7, 8 TeV	[146]	✓	✓	✗	✗	✗
CMS $\sigma_{t\bar{t}}^{\rm tot}$ 8 TeV	[271]	✗	✗	✗	✗	✓
CMS $\sigma_{t\bar{t}}^{\rm tot}$ 5, 7, 8, 13 TeV	[68, 272–280]	✗	✗	✓	✗	✗
CMS $\sigma_{t\bar{t}}^{\rm tot}$ 13 TeV	[69]	✓	✓	✓	✗	✗
CMS $t\bar{t}$ lepton+jets 8 TeV	[70]	✓	✓	✗	✗	✓
CMS $t\bar{t}$ 2D dilepton 8 TeV	[90]	✗	✓	✗	✓	✓
CMS $t\bar{t}$ lepton+jet 13 TeV	[91]	✗	✓	✗	✗	✗
CMS $t\bar{t}$ dilepton 13 TeV	[92]	✗	✓	✗	✗	✗
CMS single top $\sigma_t + \sigma_{\bar{t}}$ 7 TeV	[95]	✗	✓	✓	✗	✗
CMS single top R_t 8, 13 TeV	[97, 99]	✗	✓	✓	✗	✗
CMS single top 13 TeV	[281, 282]	✗	✗	✗	✗	(✓)

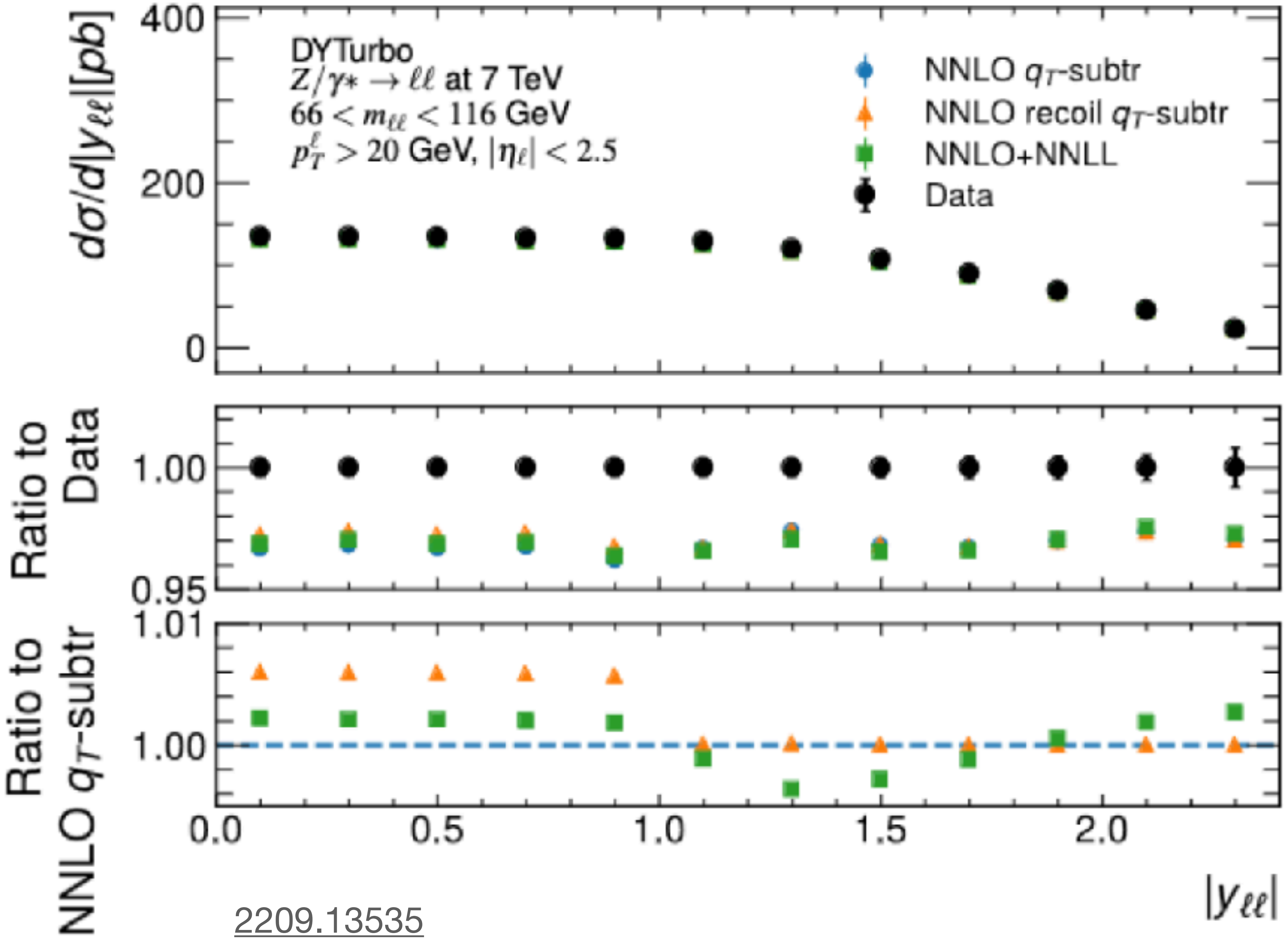
Data set	Ref.	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
LHCb Z 7 TeV ($\mathcal{L} = 940$ pb^{-1})	[59]	✓	✓	✗	✗	✓
LHCb $Z \to ee$ 8 TeV ($\mathcal{L} = 2$ fb^{-1})	[61]	✓	✓	✓	✓	✓
LHCb W 7 TeV ($\mathcal{L} = 37$ pb^{-1})	[283]	✗	✗	✗	✗	✓
LHCb $W, Z \to \mu$ 7 TeV	[60]	✓	✓	✓	✓	✓
LHCb $W, Z \to \mu$ 8 TeV	[62]	✓	✓	✓	✓	✓
LHCb $W \to e$ 8 TeV	[80]	✗	(✓)	✗	✗	✗
LHCb $Z \to \mu\mu, ee$ 13 TeV	[82]	✗	✓	✗	✗	✗

2109.02653

✓	in baseline dataset
✗	not considered
(✓)	impact assessed but excluded from baseline

Theory deficiencies

- A bad description of the data can also come from deficiencies in the theory predictions
- Precise ATLAS 7TeV W,Z cross-sections notoriously give a bad χ^2 in PDF fits
  - CT18Z special set including it, NNPDF modifies the experimental uncertainties



2209.13535

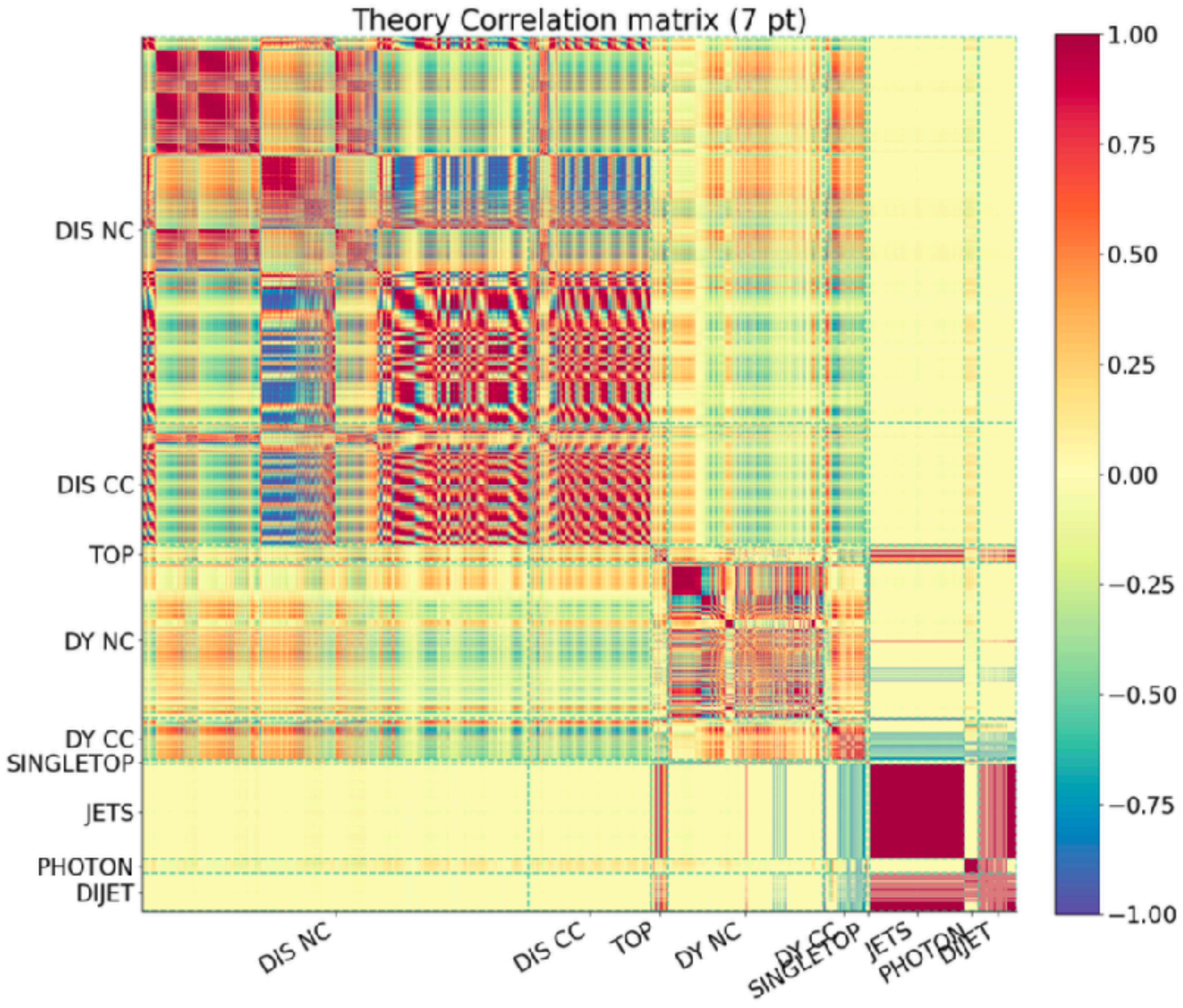
- Impact of fiducial cuts on theory neglected until a few years ago
- Large χ^2 improvement for all PDFs

PDF set	Total χ^2 (ndf=61) NNLO q_T subtr.	NNLO recoil q_T-subtr	NNLO+NLL
CT18ANNLO68	96	84	74
MSHT20nnlo	111	87	79
NNPDF31	91	84	71
NNPDF40nnlo	89	83	69

- In how many other cases are we blaming data for theory issues?

PDFs and theory uncertainties

- PDF fits historically performed without any theoretical uncertainty
- The precision of our measurements is clearly challenging this choice



1906.10698

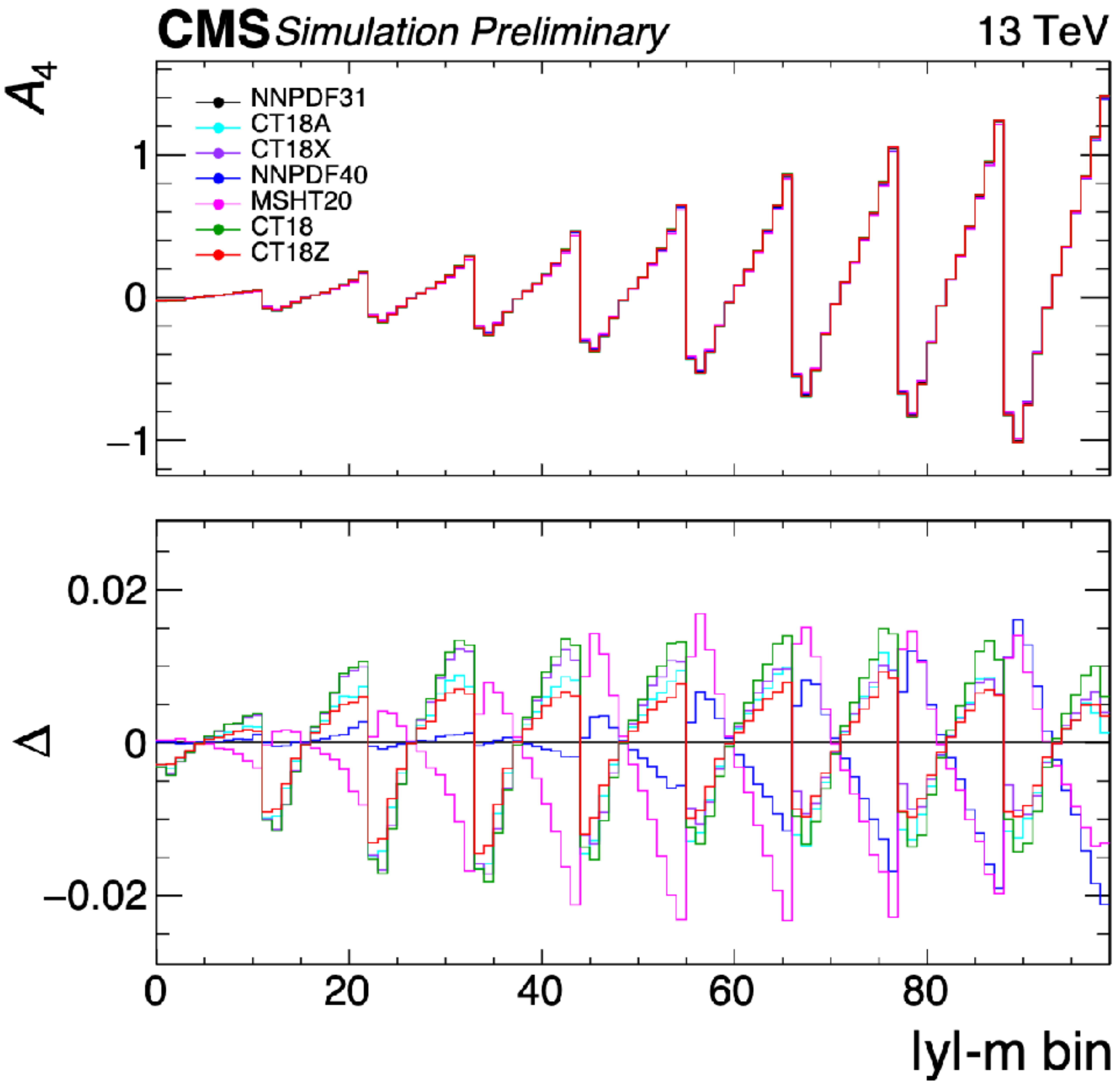
- Attempts by NNPDF and MSHT (at n3lo) to include uncertainties for QCD missing orders
- Shown to give more flexibility to the PDFs and slightly alleviate tensions
- However move the problem to converting scale variations into nuisance parameters (see F.Tackmann, A.Huss, …)

Not all measurements are born equal

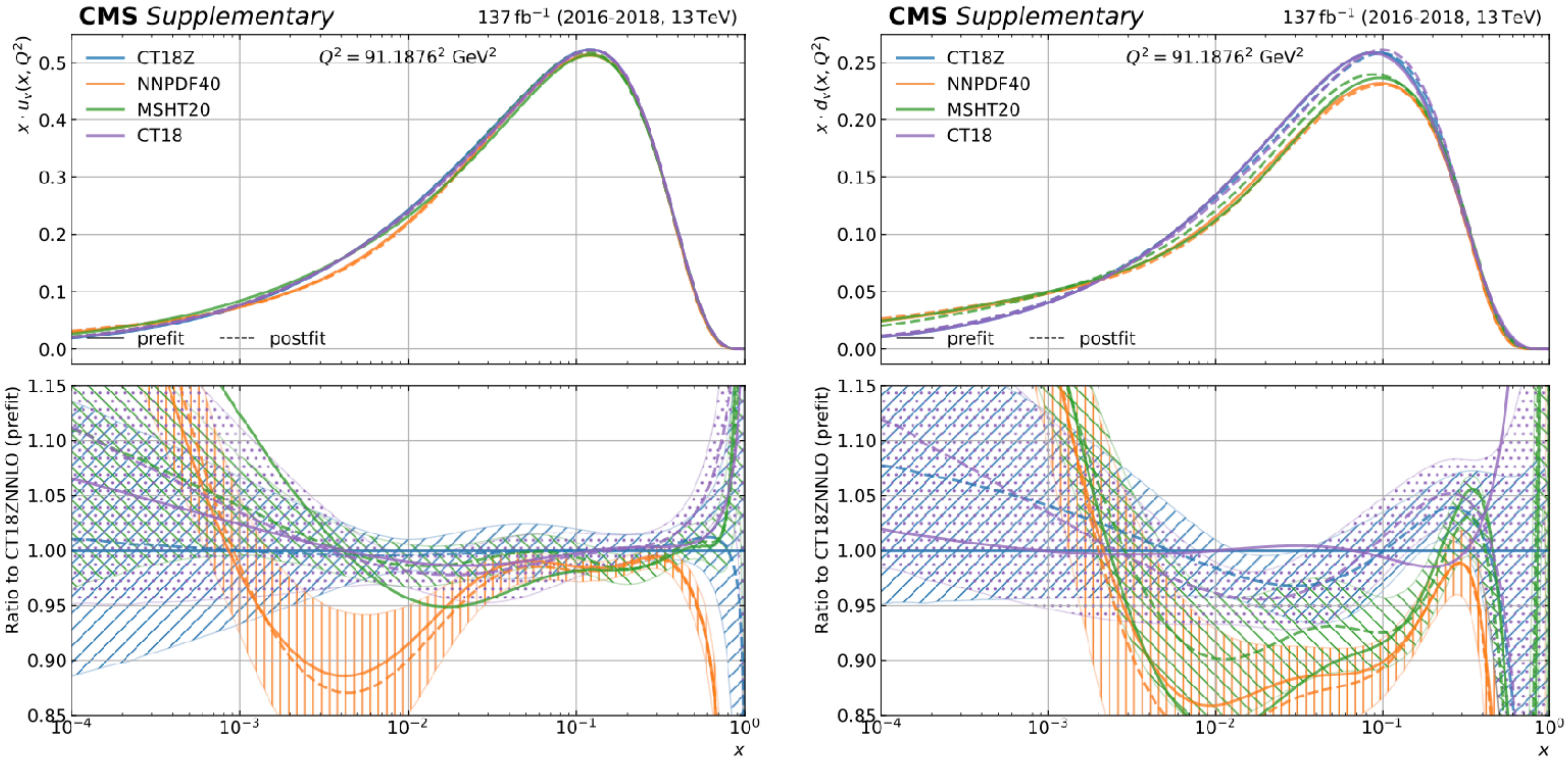
- Certain measurements provide enough of information for reinterpretation, others do not
- Certain measurements have been cross-checked across multiple channels, others not
- Certain measurements can be shown to agree with theory, others not
- Certain observables are direct measurements, others extrapolations using theory (stable tops, parton-level jets, Born-leptons, …)

- Suggest to identify a subset of precise and self-consistent measurements which we believe to be well described by theory to be used for **“reduced data PDF fits”**
- Involving both PDF fitting groups and experimental collaborations
- Similar to PDF4LHC benchmarking, but aimed at a deeper understanding of differences in PDFs and alleviate the need for tolerances
- Could consider a “PDF challenge” in which we provide you with pseudo data generated under a known probability distribution (including tensions) and we compare the PDF+uncertainties returned by the various PDF fitting approaches

Parton distribution functions dependence of A_4



Pre- and Post-fit PDFs comparison



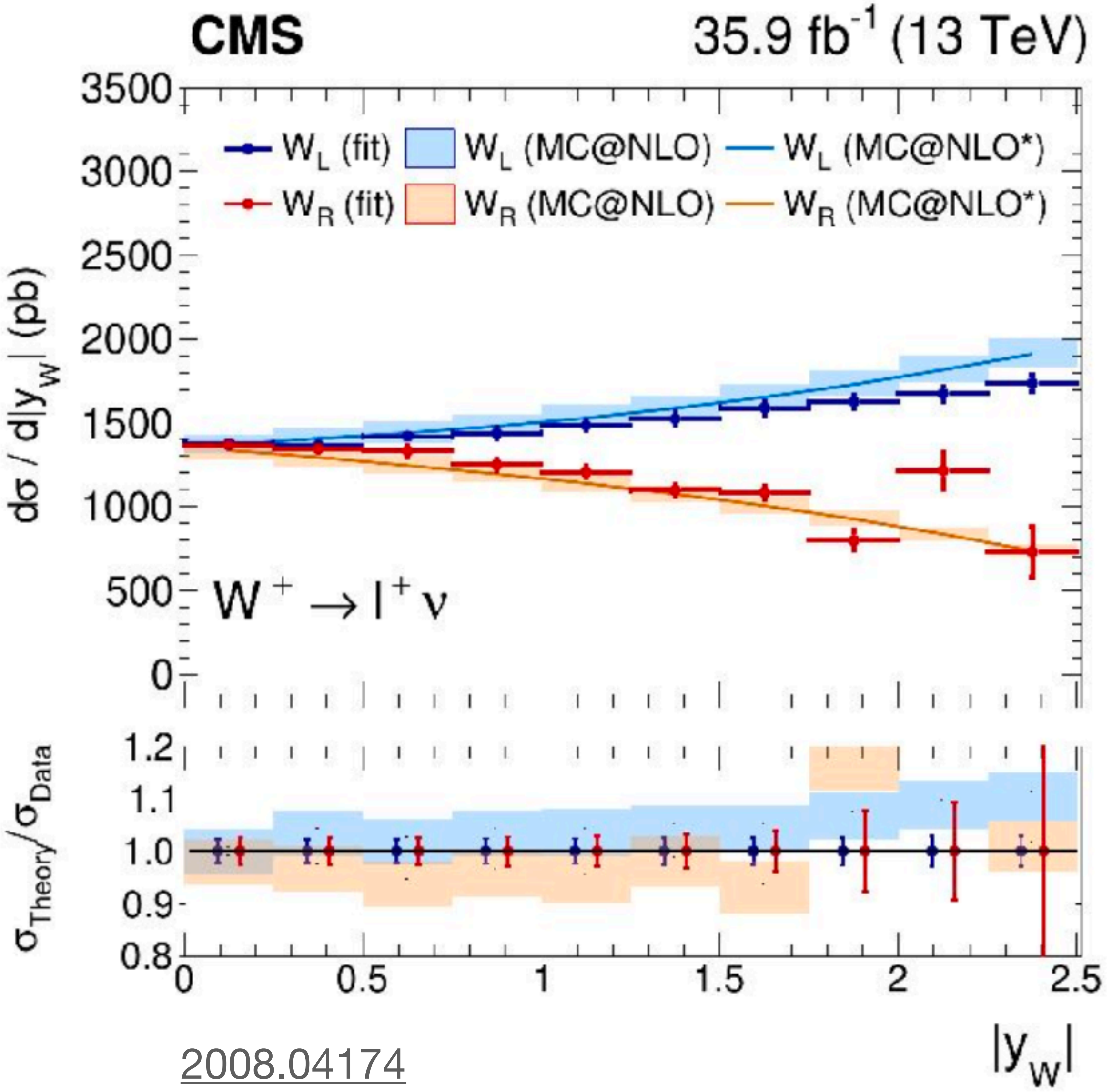
- The profiled PDFs are pulled by less than one sigma wrt the original ones
- Important validation of the results of PDF profiling

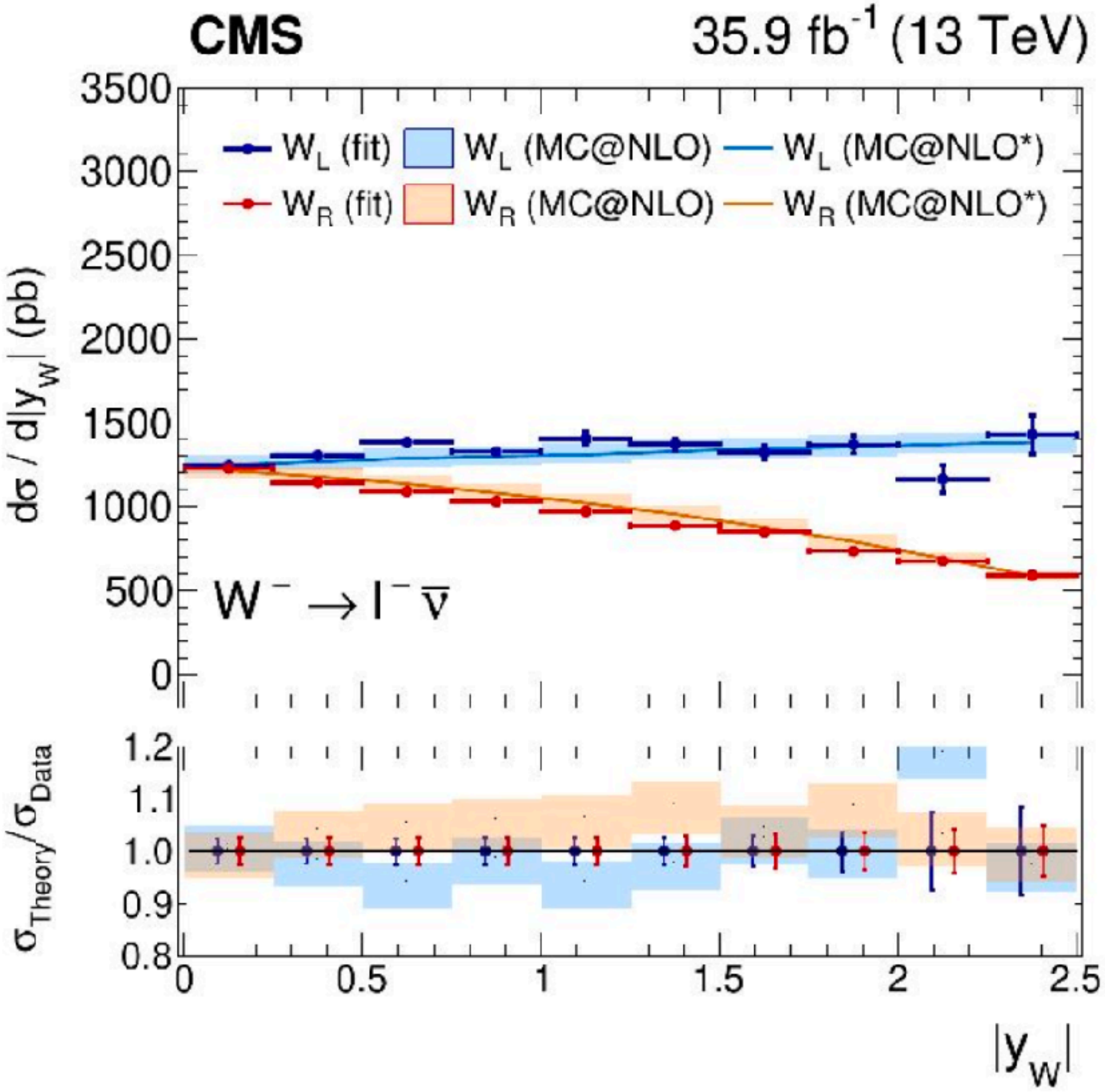
Lepton pT/eta cross-sections

- W-boson rapidity and helicity can be inferred statistically from $p_T^{\rm lep}$- $\eta^{\rm lep}$
- Need predictions with q_T-resummation to describe $p_T^{\rm lep}$

2008.04174

Polarized W cross-sections

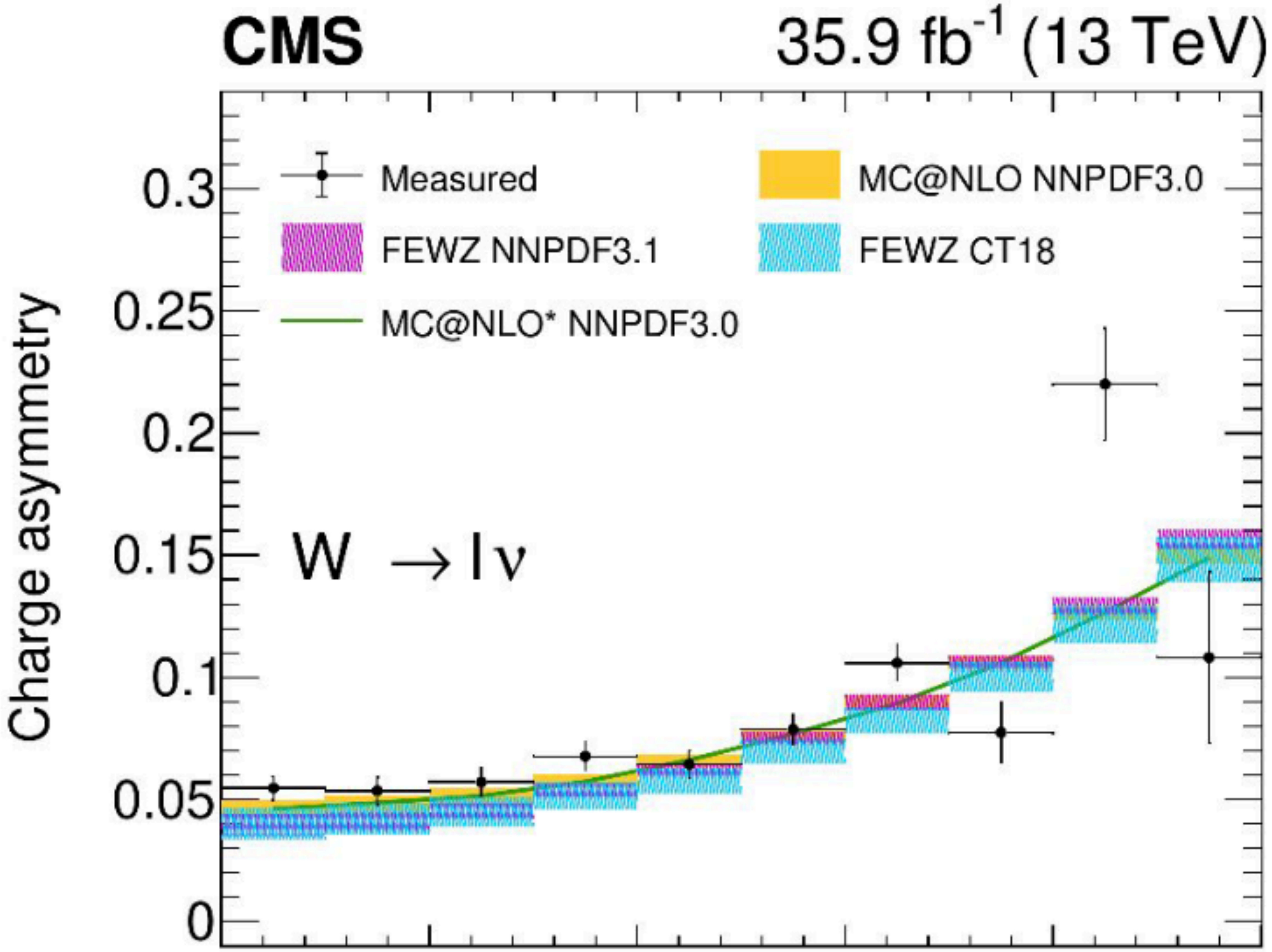


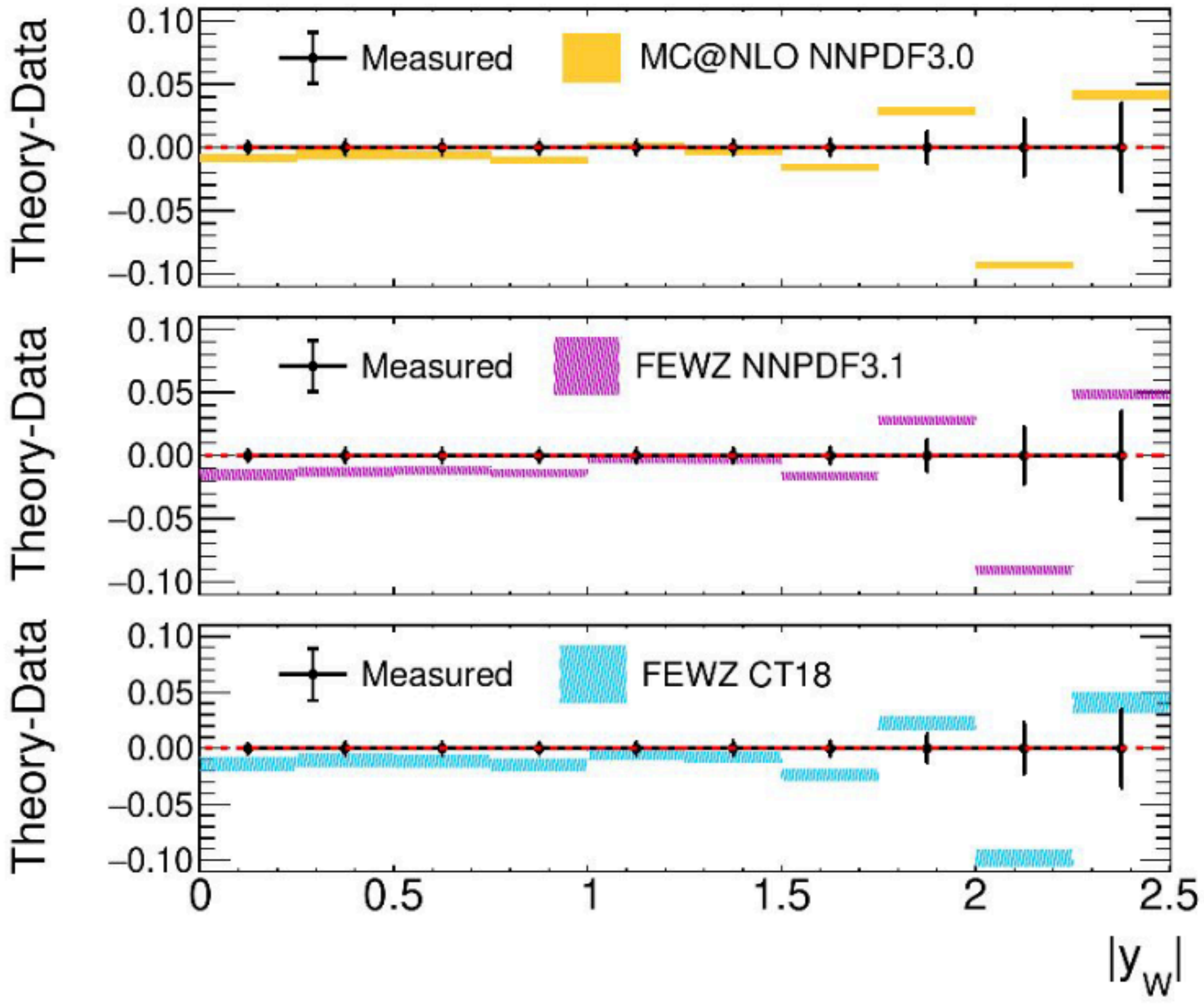


2008.04174

W charge asymmetry measurement

- Helicities in W integrated results measured without assumptions on underlying polarization
- Avoids circularity in PDF uncertainties in e.g. Tevatron W-asymmetry measurements

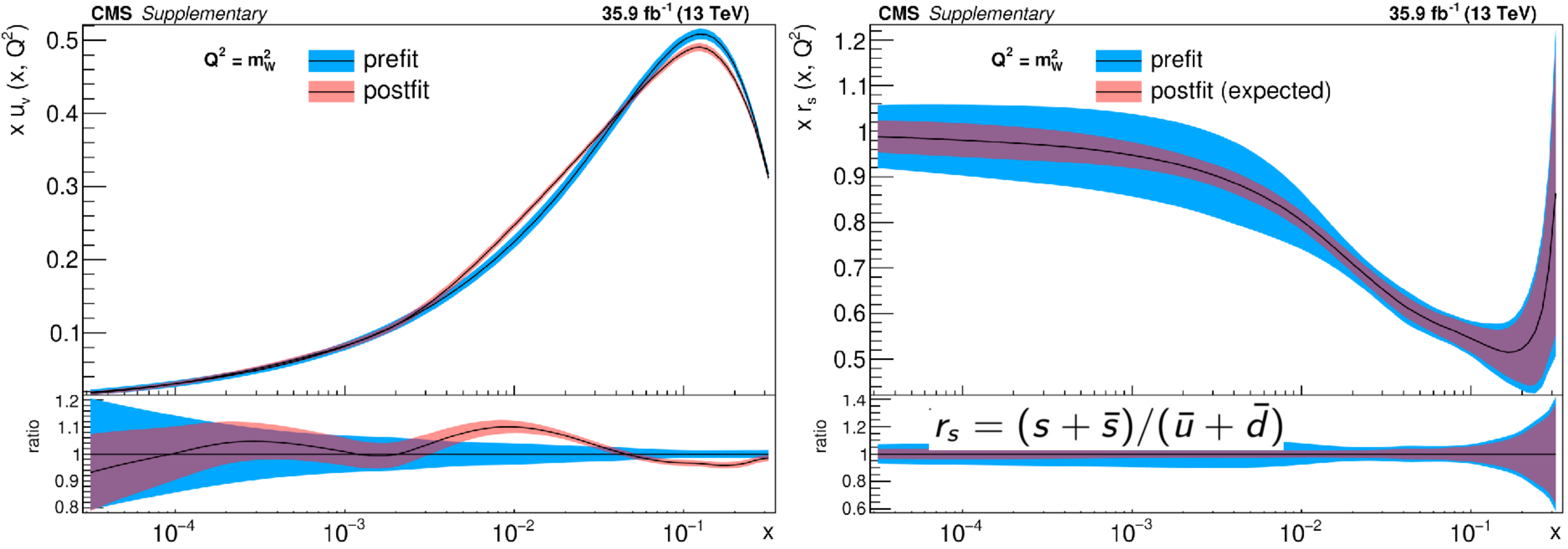




2008.04174

PDF constraints

- Sensitivity to PDFs evaluated using aMC@NLO+Pythia predictions and NNPDF30
- Large reduction in uncertainties for valence and strange PDFs



2008.04174

Goodness-of-fit measure

Consider usual PDF fit probability:

Theory — Data — Hessian matrix - contains uncorrelated (s_k) and correlated uncertainties (β_k)

$$P(T|D) \propto \exp(-\chi^2) \propto \exp\left(-\frac{1}{2}(T-D)^T H_0 (T-D)\right)$$

$$\propto \exp\left(-\frac{1}{2}\sum_{k=1}^{N_{pt}} \frac{1}{s_k^2}\left(D_k - T_k - \sum_{\alpha=1}^{N_{corr}} \beta_{k,\alpha}\lambda_\alpha\right)^2 + \sum_{\alpha=1}^{N_{corr}} {\lambda_\alpha}^2\right)$$

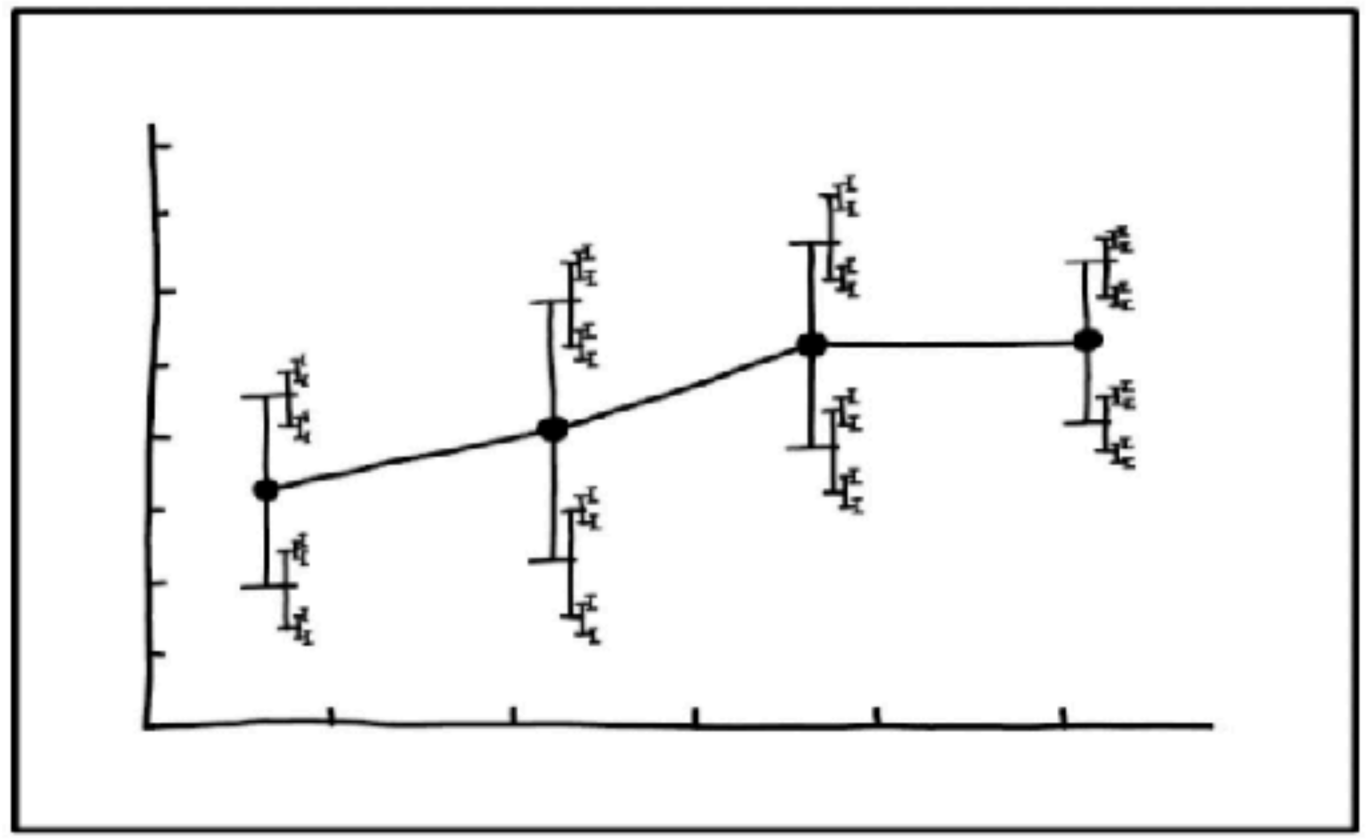
Experimental Nuisance parameters

Errors on Errors

- Include an uncertainty on nuisance parameters variances

$$\frac{(u_i - \theta_i)^2}{\sigma_{u_i}^2} \rightarrow \left(1 + \frac{1}{2r_i^2}\right) \ln\left[1 + 2r_i^2 \frac{(u_i - \theta_i)^2}{v_i}\right]$$

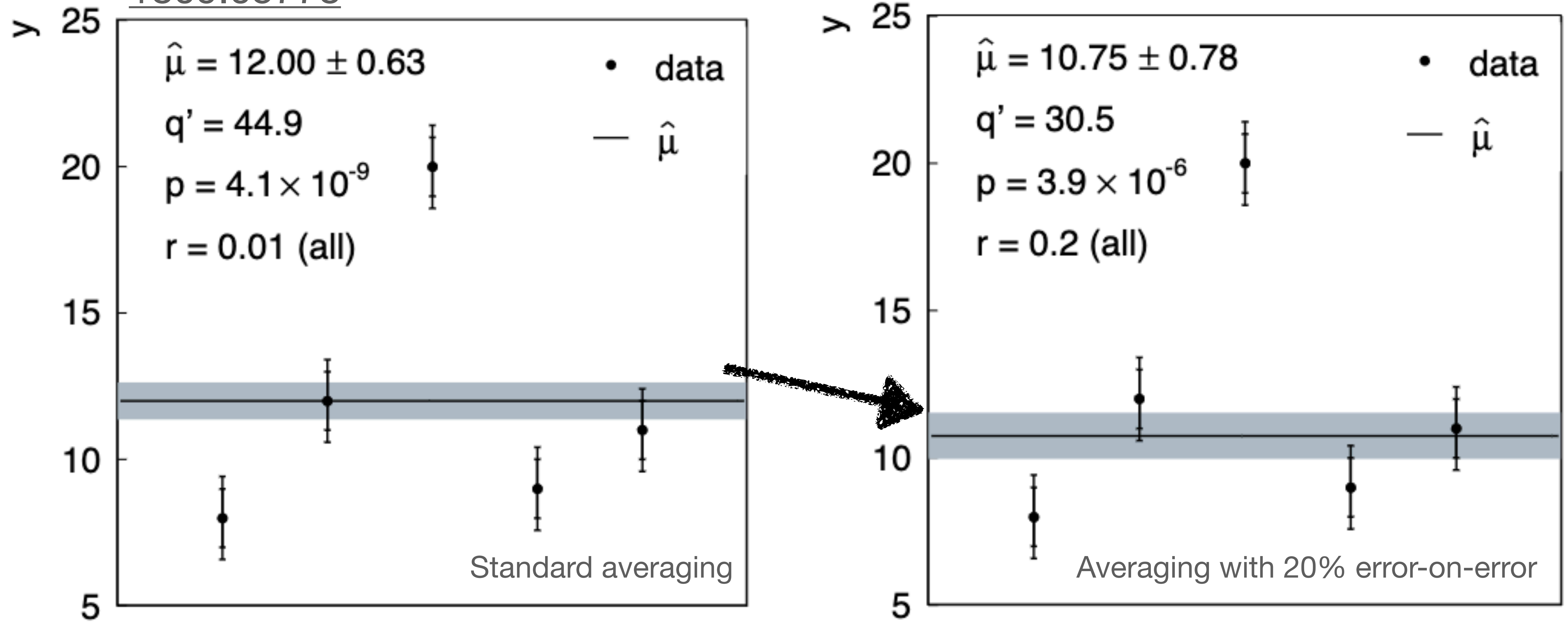
error-on-error



https://xkcd.com/2110/

- Replaces Gaussians with Student t distributions with fatter tails

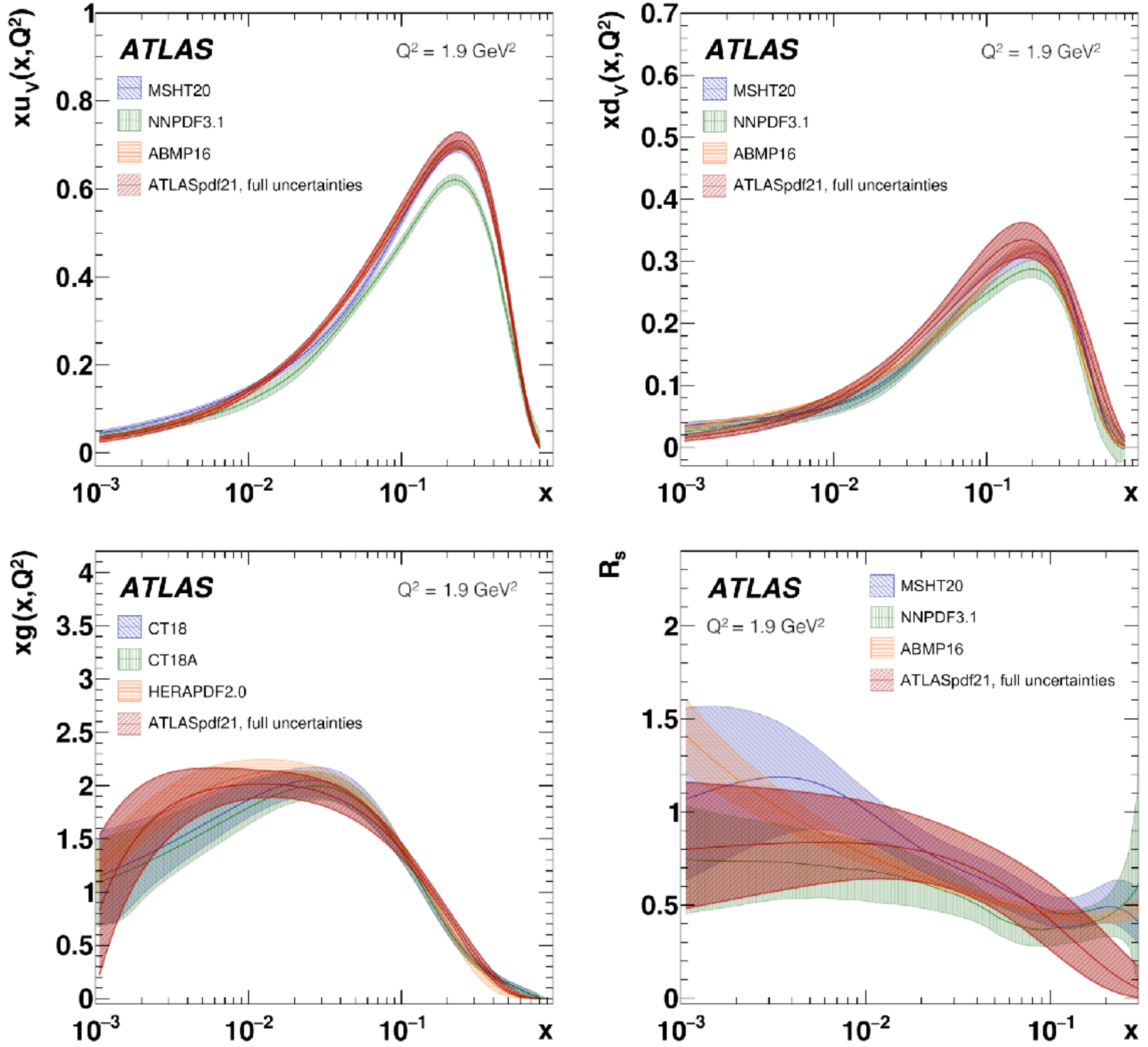
1809.05778



Small error-on-error (r=20%) sufficient to improve consistency

And reduces sensitivity to outlier measurements

How easily could this be tested in a real PDF fit?



ATLASpdf21 results

- Good description of the fitted data

ATLASpdf21	CT18	CT18A
2010/1620 (1.24)	2135/1641 (1.30)	2133/1641 (1.30)
MSHT20	**HERAPDF2.0**	**NNPDF3.1**
2218/1641 (1.35)	2262/1641 (1.37)	2109/1641 (1.29)

- Significant impact of the ATLAS data on the valence distributions
- PDFs in agreement with global fits