Going Beyond: NNPDF3.0

Next generation PDFs for the LHC Run II

Juan Rojo
STFC Rutherford Fellow, University of Oxford
& CERN, PH Division, TH Unit

PDF4LHC Workshop
CERN, 16/05/2014

Parton Distributions for LHC Run II

- **PDFs** have been an essential ingredient for **Run I** phenomenology, and will be so even more at the upcoming **Run II**
- Many crucial LHC analysis benefit from improved PDFs, from **precision Standard Model measurements**, like the W mass determination, and **Higgs boson characterization** to **BSM searches**
- **NNPDF3.0** is the new forthcoming PDF release from the NNPDF Collaboration, a major upgrade that accounts for recent progress in experimental constraints, theory calculations and methodological improvements
- In this talk: preliminary results on **NNPDF3.0,** including
  - ☑ **New experimental data:** HERA-II structure functions, ATLAS and CMS jets, CMS W+charm, ATLAS and CMS Drell-Yan production, top quark production ....
  - ☑ **Improved theory calculations:** Approximate NNLO K-factors for jets, electroweak effects for Drell-Yan data, **APPLgrid/FastNLO/aMCfast** for all hadronic observables ...
  - ☑ **Fitting methodology:** C++ rewriting of the code, fitting strategy validated on closure tests, extended positivity, optimized Genetic Algorithms minimization ......

DATA

NNPDF3.0: New data

As compared to NNPDF2.3, in NNPDF3.0 we have included **more than 1000 new data points** from recent measurements from HERA and the LHC:

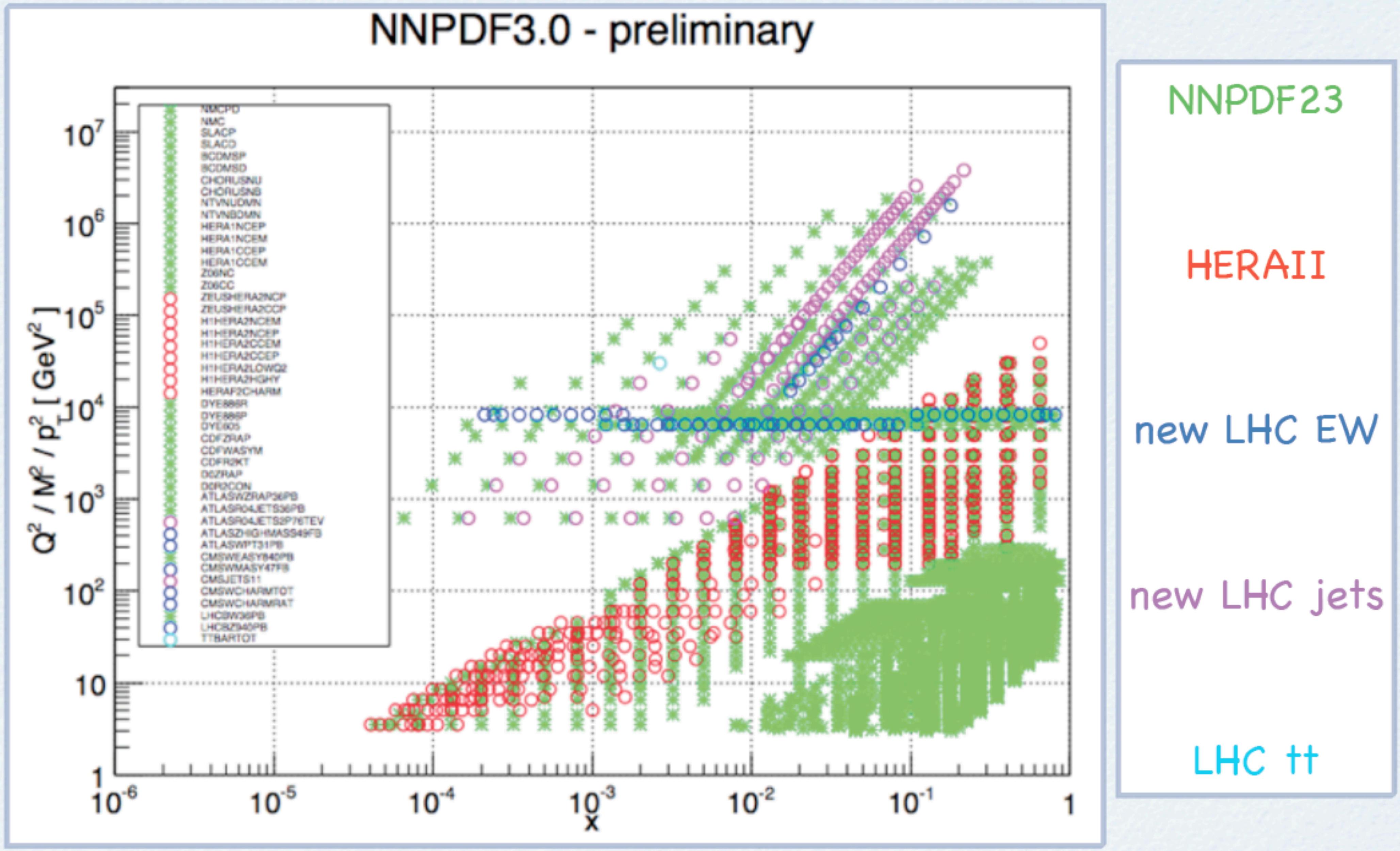
- ☑ **HERA structure function data:** HERA-II structure functions from H1 and ZEUS, combined HERA F_{2c} cross-sections **(to be updated if HERA-II combined data released soon)**
- ☑ **LHC jet data:** CMS **7 TeV inclusive jets** from 2011, ATLAS **7 TeV inclusive jets** from 2010, ATLAS **2.76 TeV jets** including their correlation with 7 TeV data
- ☑ **LHC electroweak data:** ATLAS **W, Z** from 2010, CMS **electron and muon asymmetries** from 2011, LHCb **W distributions** from 2010 and **Z rapidity distributions** from 2011, CMS **W+charm** production data, ATLAS and CMS **Drell-Yan production**, ATLAS **W p_T** distributions
- ☑ **LHC top quark pair production data**

All these data **are already reasonably well described by NNPDF2.3**

Moderate impact in global fit, much more substantial in collider-only fit

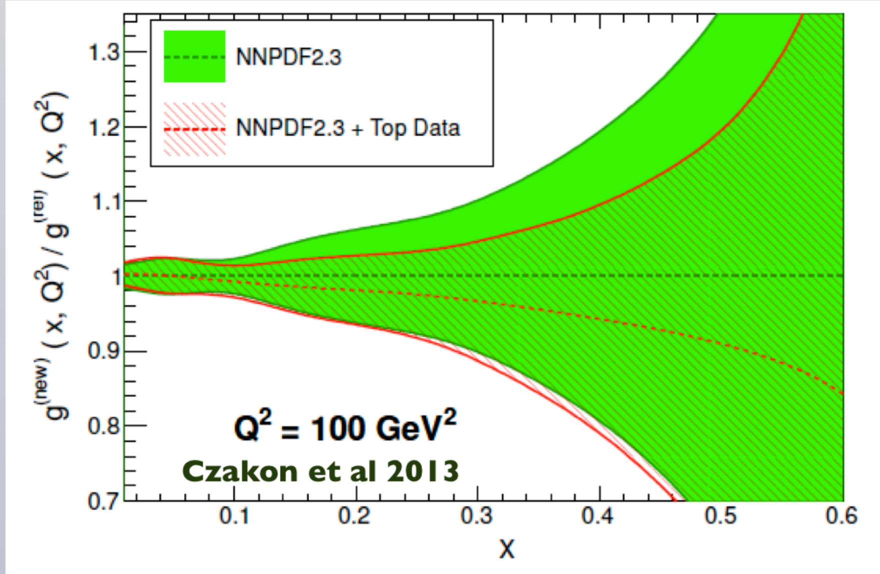
Experiment	Dataset	DOF
NMC		356
	NMCPD	132
	NMC	224
SLAC		74
	SLACP	37
	SLACD	37
BCDMS		581
	BCDMSP	333
	BCDMSD	248
CHORUS		862
	CHORUSNU	431
	CHORUSNB	431
NTVDMN		79
	NTVNUDMN	41
	NTVNBDMN	38
HERA1AV		592
	HERA1NCEP	379
	HERA1NCEM	145
	HERA1CCEP	34
	HERA1CCEM	34
ZEUSHERA2		252
	Z06NC	90
	Z06CC	37
	ZEUSHERA2NCP	90
	ZEUSHERA2CCP	35
H1HERA2		511
	H1HERA2NCEM	139
	H1HERA2NCEP	138
	H1HERA2CCEM	29
	H1HERA2CCEP	29
	H1HERA2LOWQ2	124
	H1HERA2HGHY	52
HERAF2CHARM		47
DYE886		199
	DYE886R	15
	DYE886P	184
DYE605		119
CDF		105
	CDFZRAP	29
	CDFR2KT	76
D0		138
	D0ZRAP	28
	D0R2CON	110
ATLAS		179
	ATLASWZRAP36PB	30
	ATLASR04JETS36PB	90
	ATLASR04JETS2P76TEV	59
CMS		95
	CMSWEASY840PB	11
	CMSWMASY47FB	11
	CMSJETS11	63
	CMSWCHARMTOT	5
	CMSWCHARMRAT	5
	CMSDY2D11	132
LHCB		19
	LHCBW36PB	10
	LHCBZ940PB	9
TOP		6
Total (exps)		4214

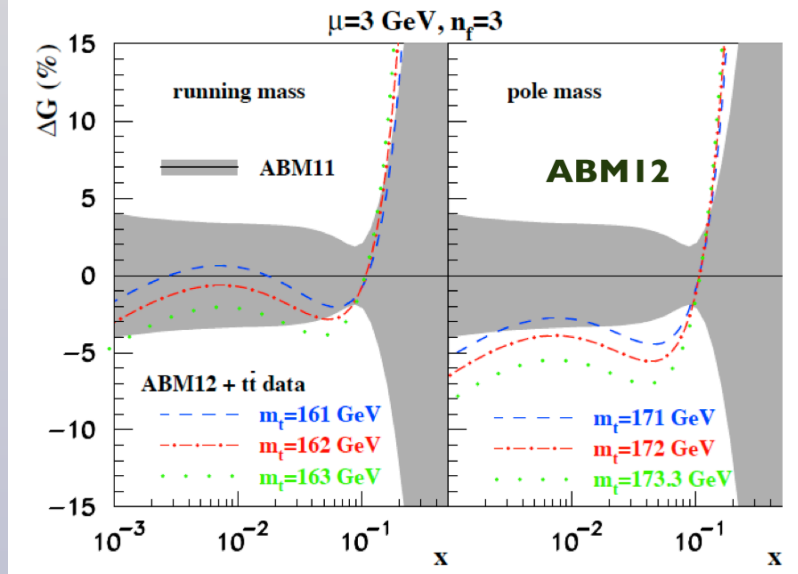
The NNPDF3.0 dataset



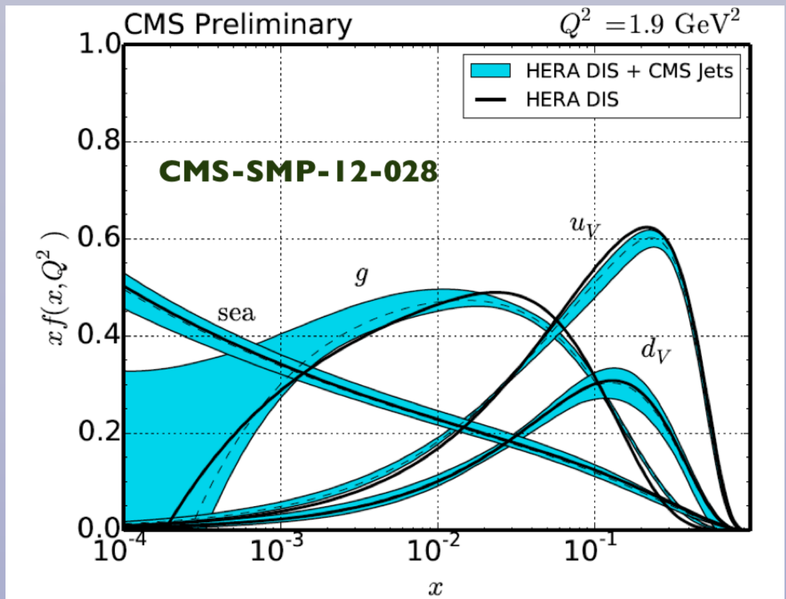
Motivation for new data in NNPDF3.0

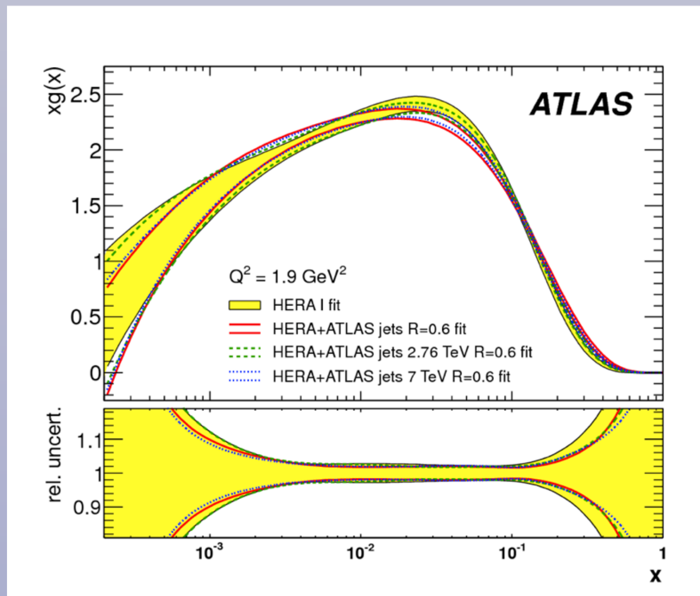
- **Top quark total cross-sections** allow to constrain the **large-x gluon PDF** (NNLO for differential distributions will be available soon, should be able to include as well differential top production measurements)





- **Jet cross-sections** pin down **medium and large-x gluon** and **large-x quarks** (important to include properly information on NNLO corrections, see later)



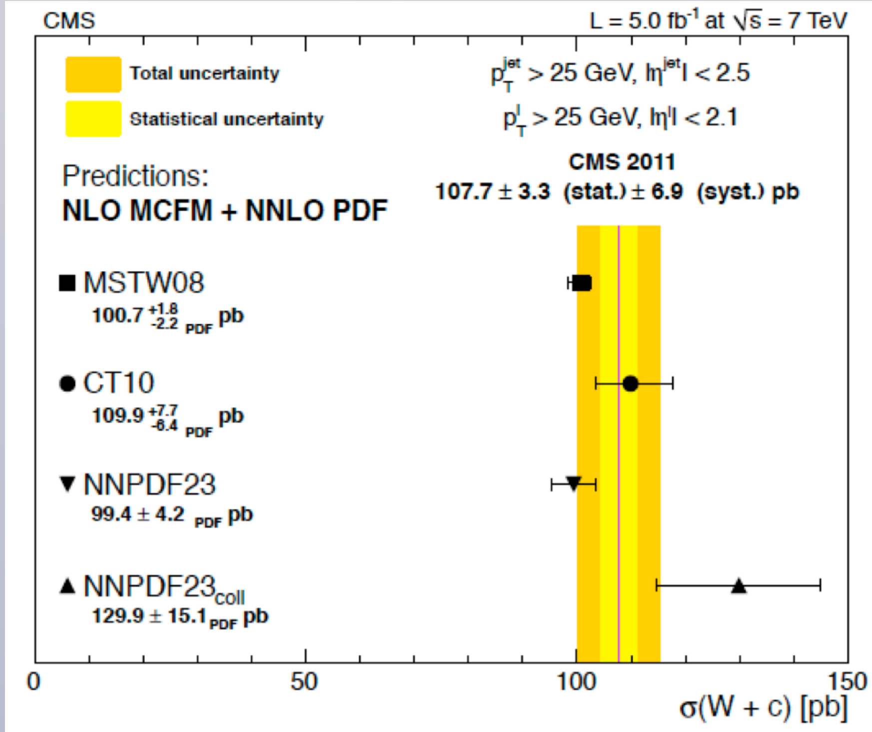


ATLAS data on the **ratio 7 TeV / 2.76 for jet production**, beautiful illustration of the PDF sensitivity of **cross-section ratios between different center-of-mass energies** (Mangano and JR 12)

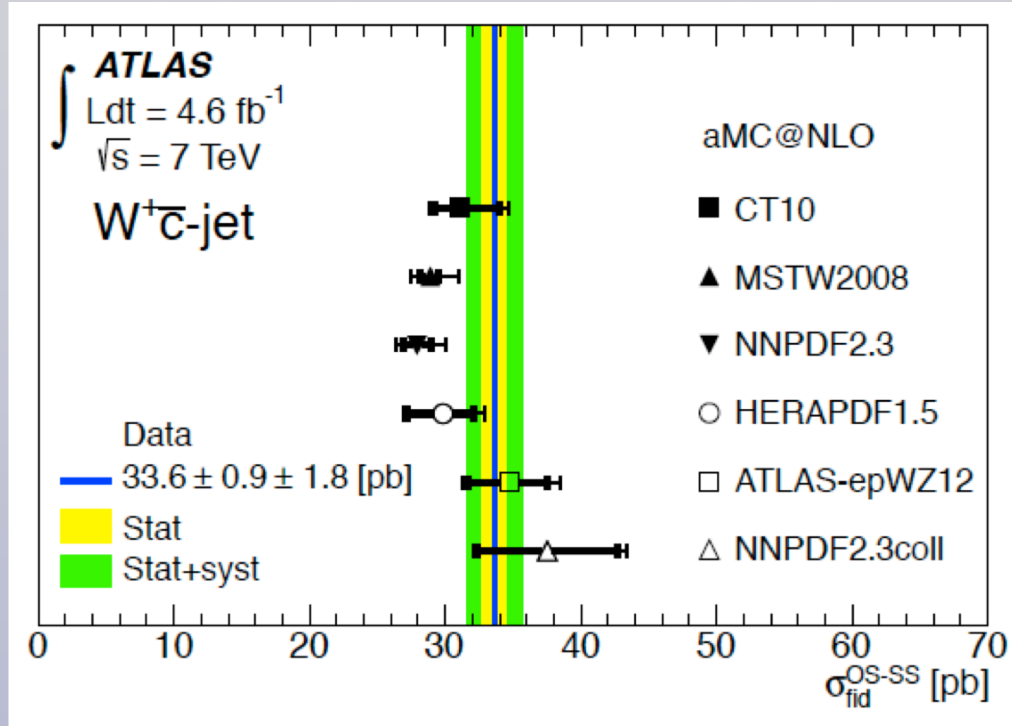
Motivation for new data in NNPDF3.0

- **W+charm** production data directly sensitive to the strange PDF
- Measured by ATLAS (arxiv:1402.6263) and CMS (arxiv:13101138) with somewhat opposite (?) conclusions

CMS: strange suppression in agreement with DIS data



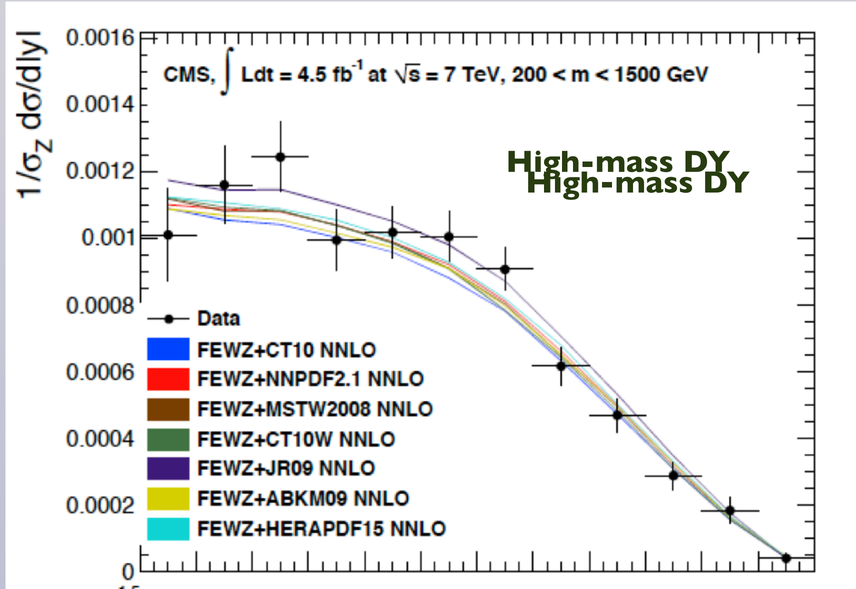
ATLAS: light quark sea symmetric preferred

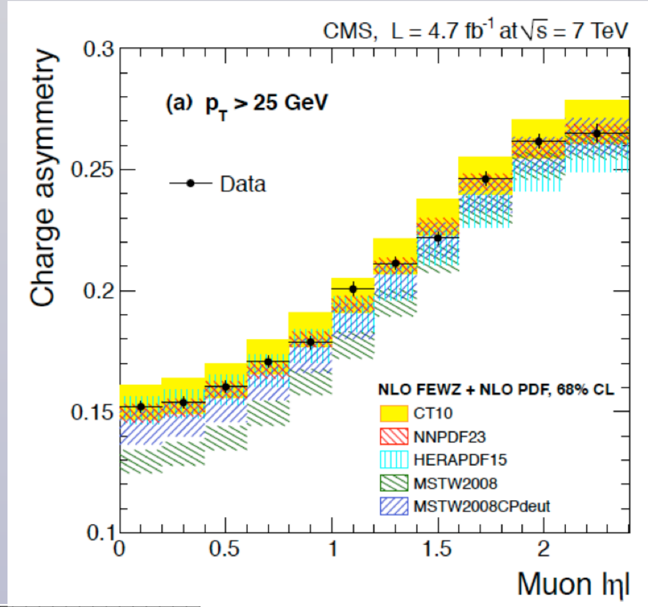


- However, only in the context of a **global fit** the **optimal** value for strangeness can be determined
- A recent analysis in the **ABM framework** (arxiv:1404.6469) suggest that **fits with symmetric strangeness** cannot describe properly fixed target DIS and Drell-Yan data (**see also R. Thorne in past PDF4LHC**) and that one can fit ATLAS data with still a suppressed strangeness (same as found in NNPDF2.3 for **incl W,Z**)
- The **NNPDF3.0** will perform a similar analysis, with the advantage of using a **completely flexible parametrization for s(x,Q)**, which in other analysis uses a very restrictive functional form

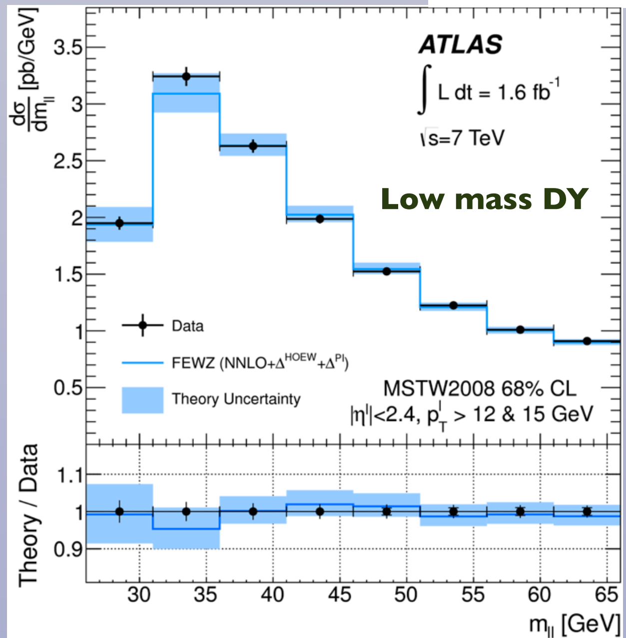
Motivation for new data in NNPDF3.0

- **Electroweak gauge boson production** is an essential measurement for **quark flavor separation**
- The relevance of LHC W,Z production is even greater in **collider-only fits**
- Data on the **Drell-Yan process** at **low and high masses** allow to extend the kinematical coverage in Bjorken-x





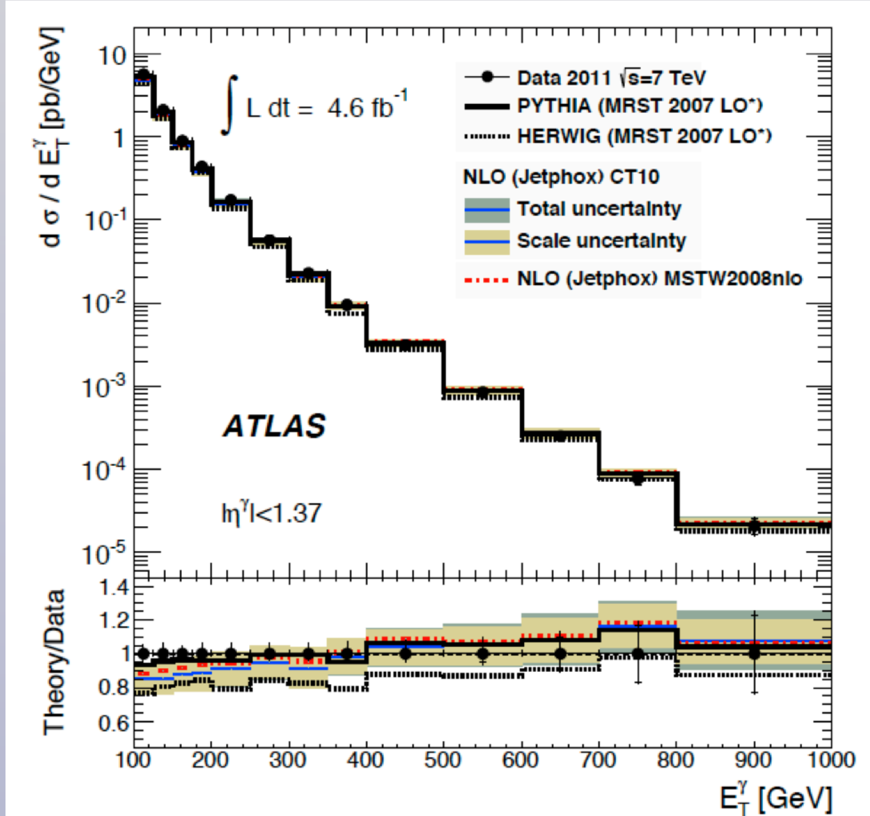
W asymmetry



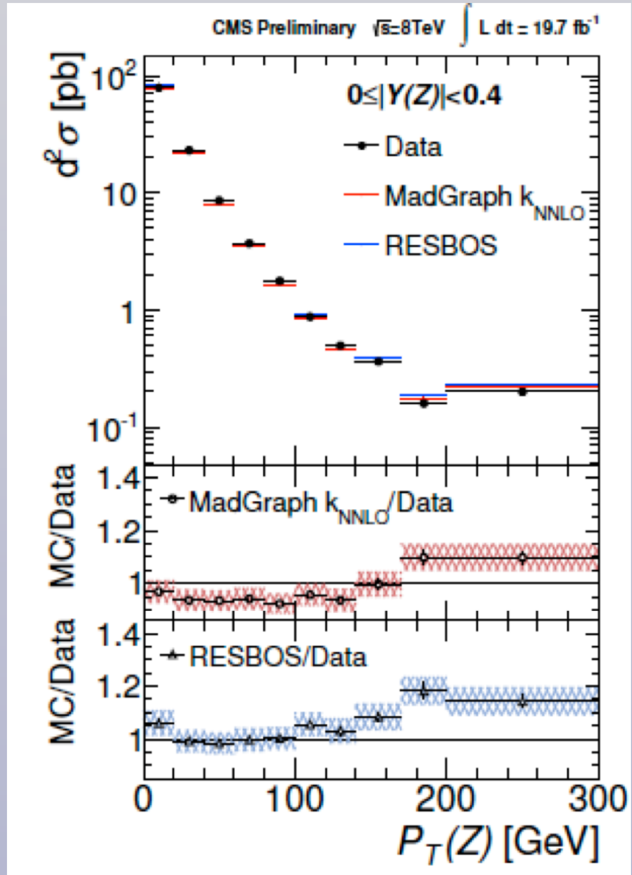
Given the data precision, NNLO corrections are essential, and also **electroweak corrections** at high masses

More data in the pipeline

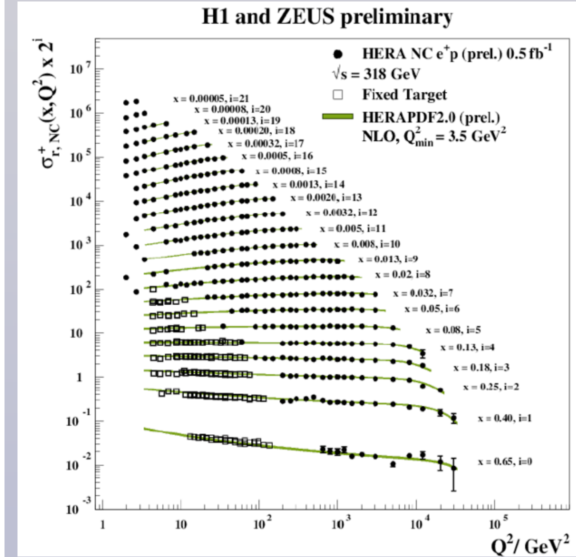
- **The plethora of new LHC data** that is becoming available for PDF fitting makes any PDF fit somewhat outdated shortly after it has been released
- At some point we need to put a **cut-off** about the data to include in NNPDF3.0
- Some **important measurements** that we might try to add in time for NNPDF3.0 include



ATLAS direct photon data from 2011 run
But need APPLgrids and K-factors, already available for the ATLAS analysis



CMS-SMP-13-013
Measurement of the Z pt spectrum at high pt
But data still preliminary
(see Markus's talk)

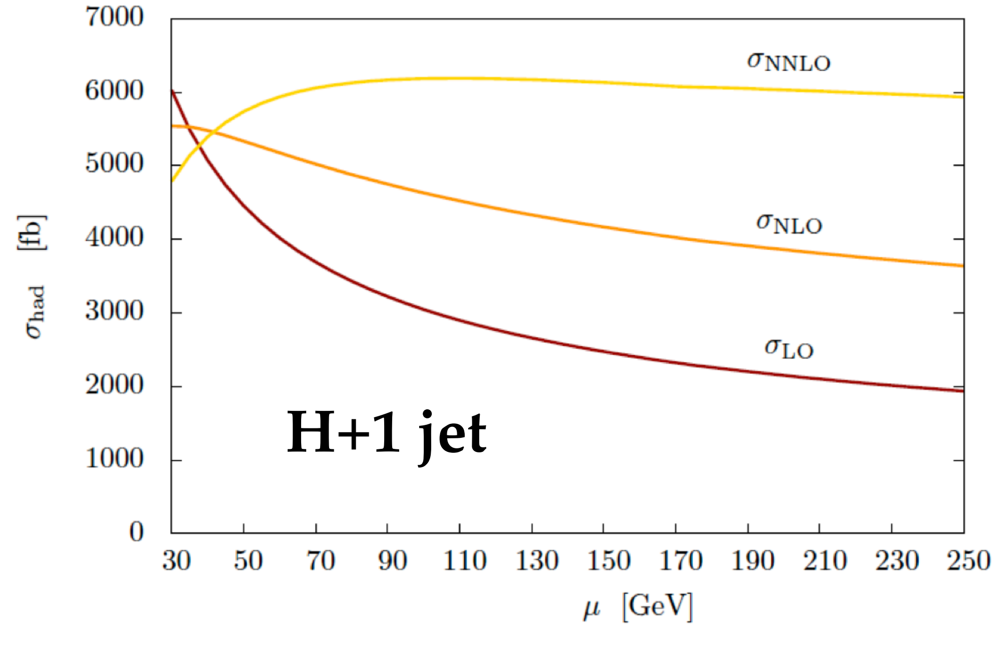


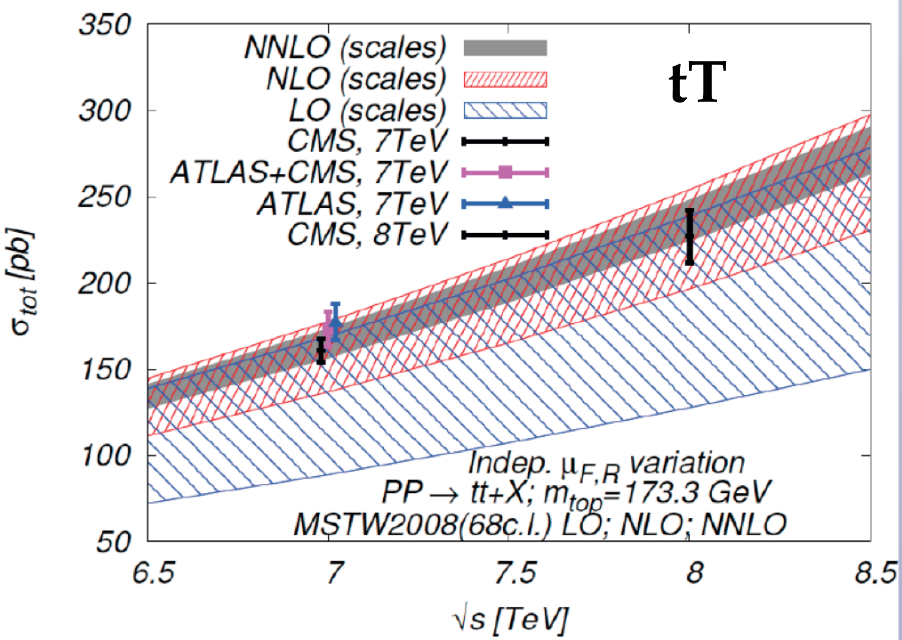
Combined HERA-II data
To supersede the separated H1 and ZEUS HERA-II data already included in 3.0

THEORY

Higher order calculations

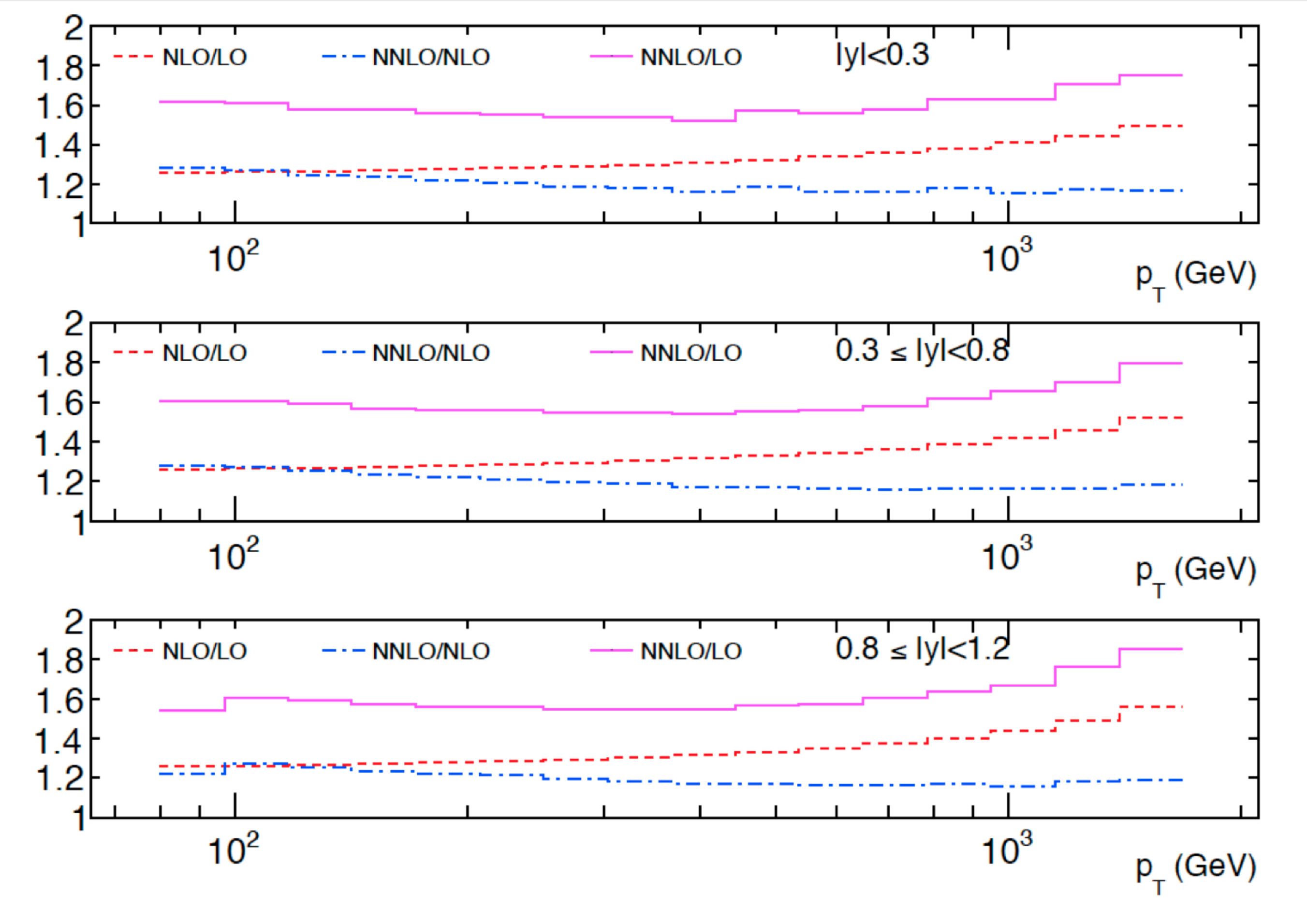
- NNLO calculations are essential to **reduce theoretical uncertainties** in PDF analysis
- Up to last year, only small number of **processes relevant for PDFs** available at **NNLO**
- Recent important progress made on some **key processes for PDF fitting**
  - **NNLO inclusive jet production** in the gluon-gluon channel has been completed (arxiv:1310.3993), jet data essential in PDF fits for **gluons and large-x quarks**
  - The full **NNLO top quark production cross section** is also available (**top++2.0**), **differential distributions** to follow this year (arxiv:1303.6254), allows to pin down **large-x gluon**
  - **Higgs + 1 jet** also available now at **NNLO** (arxiv:1302.6216), important milestone towards the closely related **Z + 1 jet** and **W + 1 jet**, important for **gluon constraints** and **quark flavor separation**

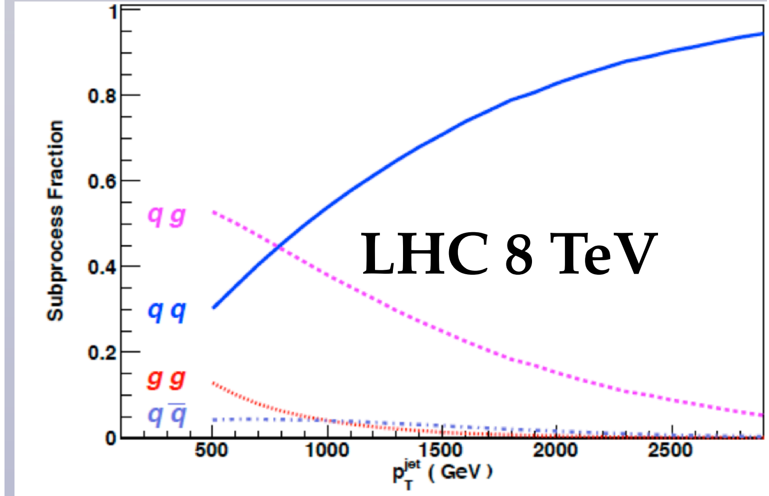




Jets in NNLO global fits

- The recent calculation of the **gluon-gluon channel NNLO jet cross sections** is an important milestone towards the exact inclusion of jet data in NNLO PDF fits: **O(20-25%) enhancements wrt NLO results**
- On the other hand, the **gg channel is small** at medium and large p_T at the LHC energies

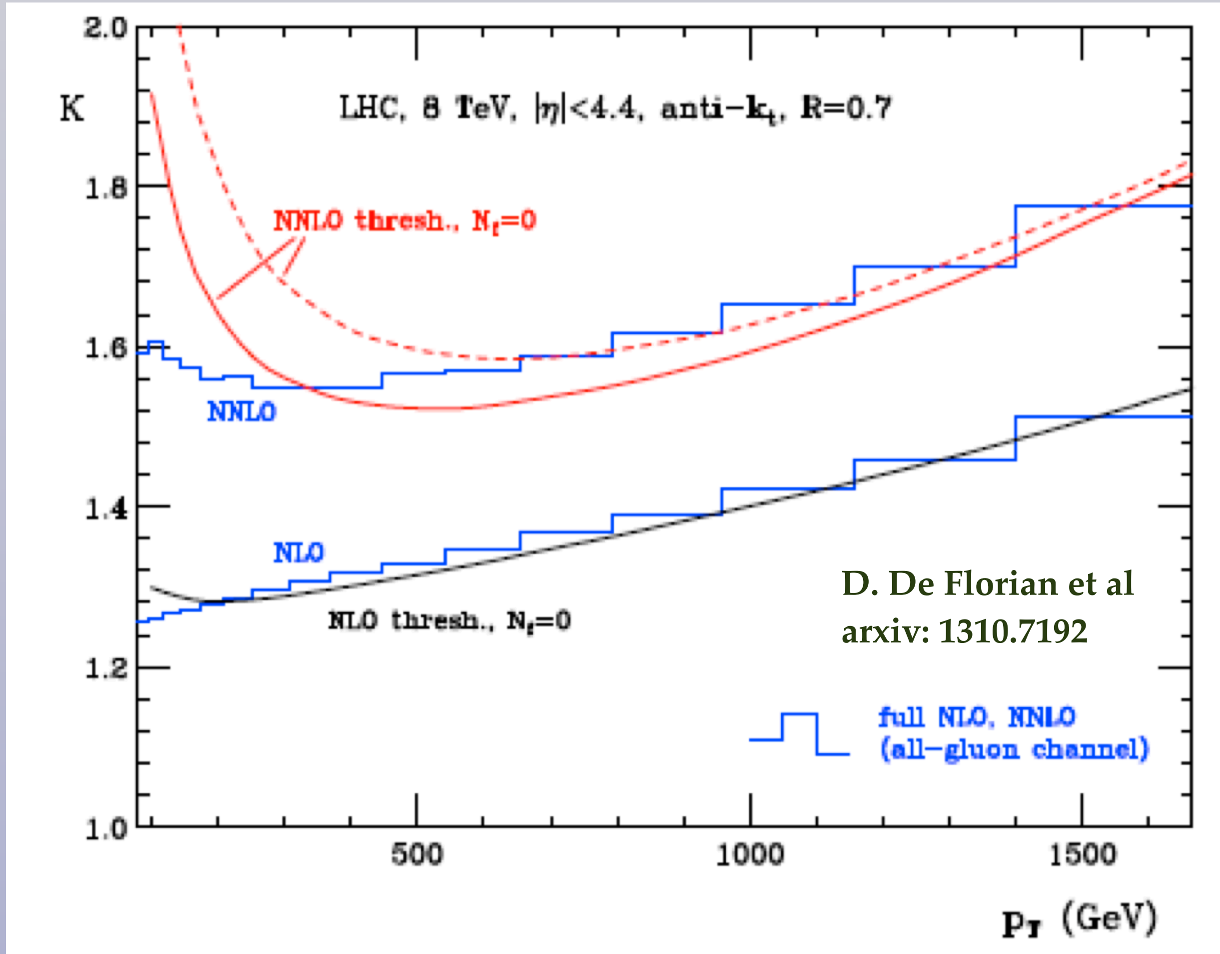




What is the optimal strategy to include **partial NNLO jet results** in the global PDF fit?

Jets in NNLO global fits

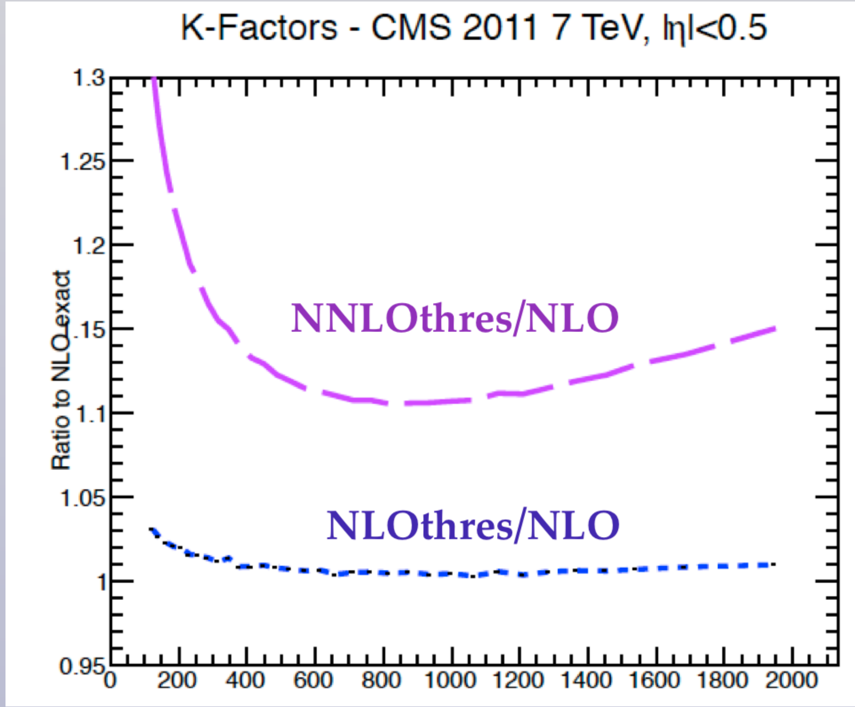
- The recent calculation of the **gluon-gluon channel NNLO jet cross sections** is an important milestone towards the exact inclusion of jet data in NNLO PDF fits: **O(20-25%) enhancements wrt NLO results**
- On the other hand, the **gg channel is small** at medium and large p_T at the LHC energies
- While full NNLO result becomes available, **approximate NNLO** results can be derived from the **improved threshold calculation**: reasonable approximation to exact at large p_T, **breaks down at small p_T**

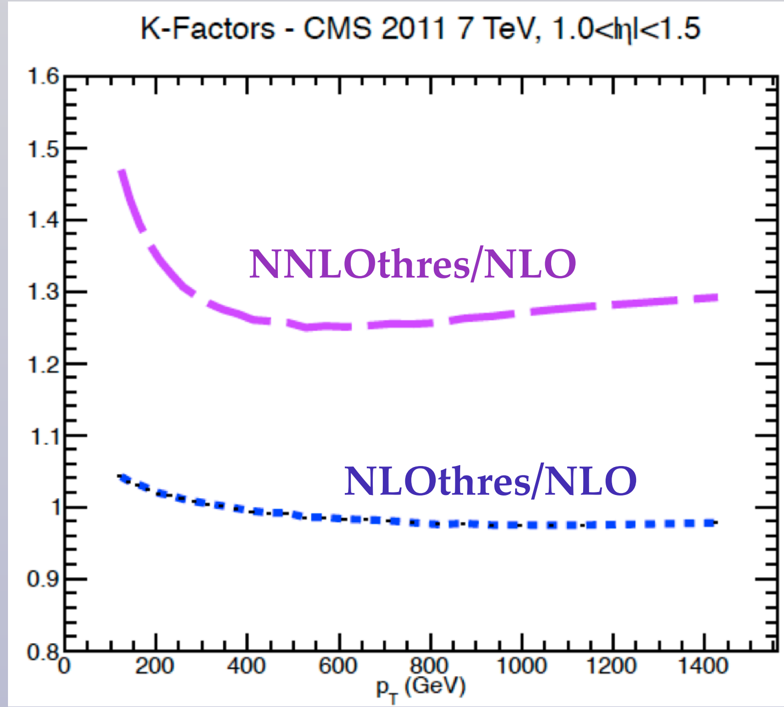


Jets in NNLO global fits

- We can therefore compute **approximate NNLO K-factors** using the **threshold approximation**
- **Comparison with exact gg NNLO** can determine for which values of jet p_T and η the **NNLOthres** calculation can be trusted (assume NNLO K-factor similar in all channels)

Plots by S. Carrazza and J. Pires



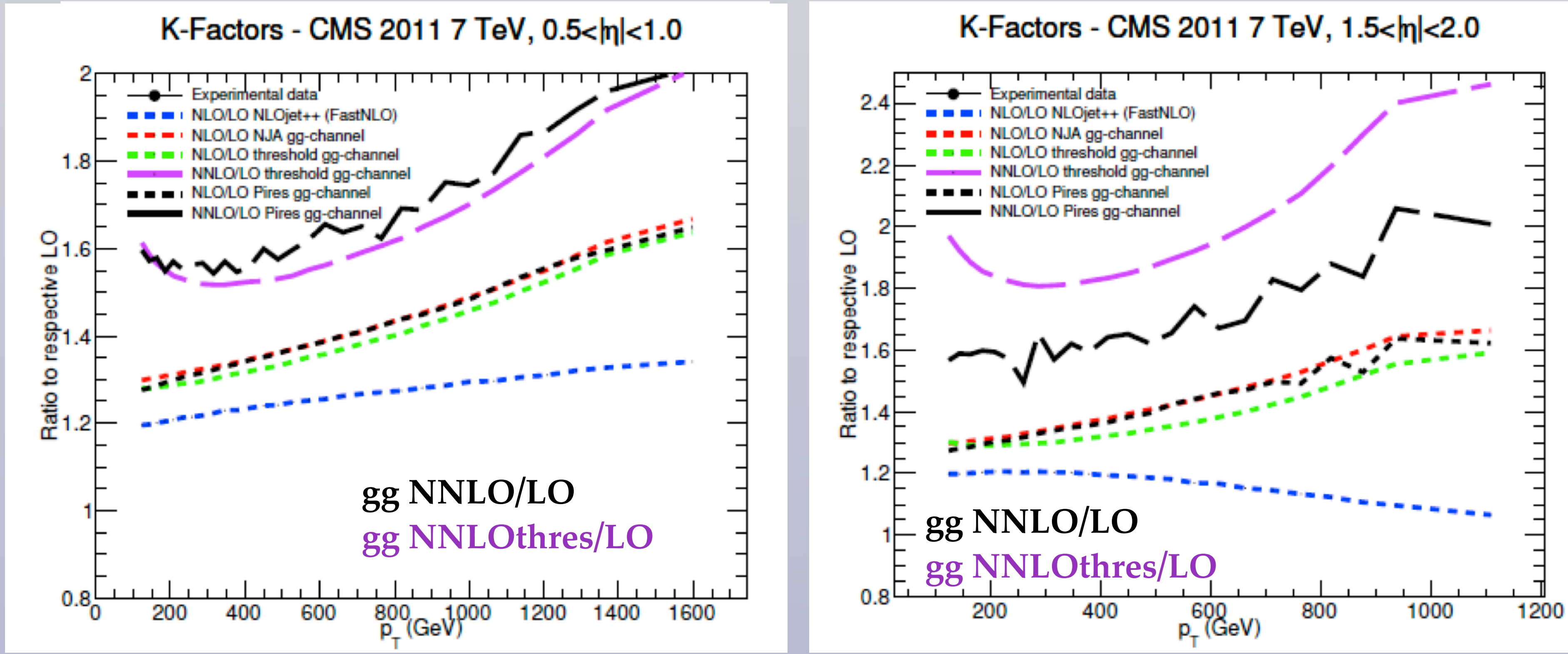


- Until **exact NNLO available**, jet data at small jet p_T and large η **should be excluded from NNLO fit,** since NNLOthres not suitable there
- Small impact for Tevatron jets (where NNLOthres works in a wider range) and ATLAS 2010 jet data (substantial uncertainties), but important for the **ATLAS and CMS jet data from the 2011 and 2012 runs**

Jets in NNLO global fits

- We can therefore compute **approximate NNLO K-factors** using the **threshold approximation**
- **Comparison with exact gg NNLO** can determine for which values of jet p_T and η the **NNLOthres** calculation can be trusted (assume NNLO K-factor similar in all channels)

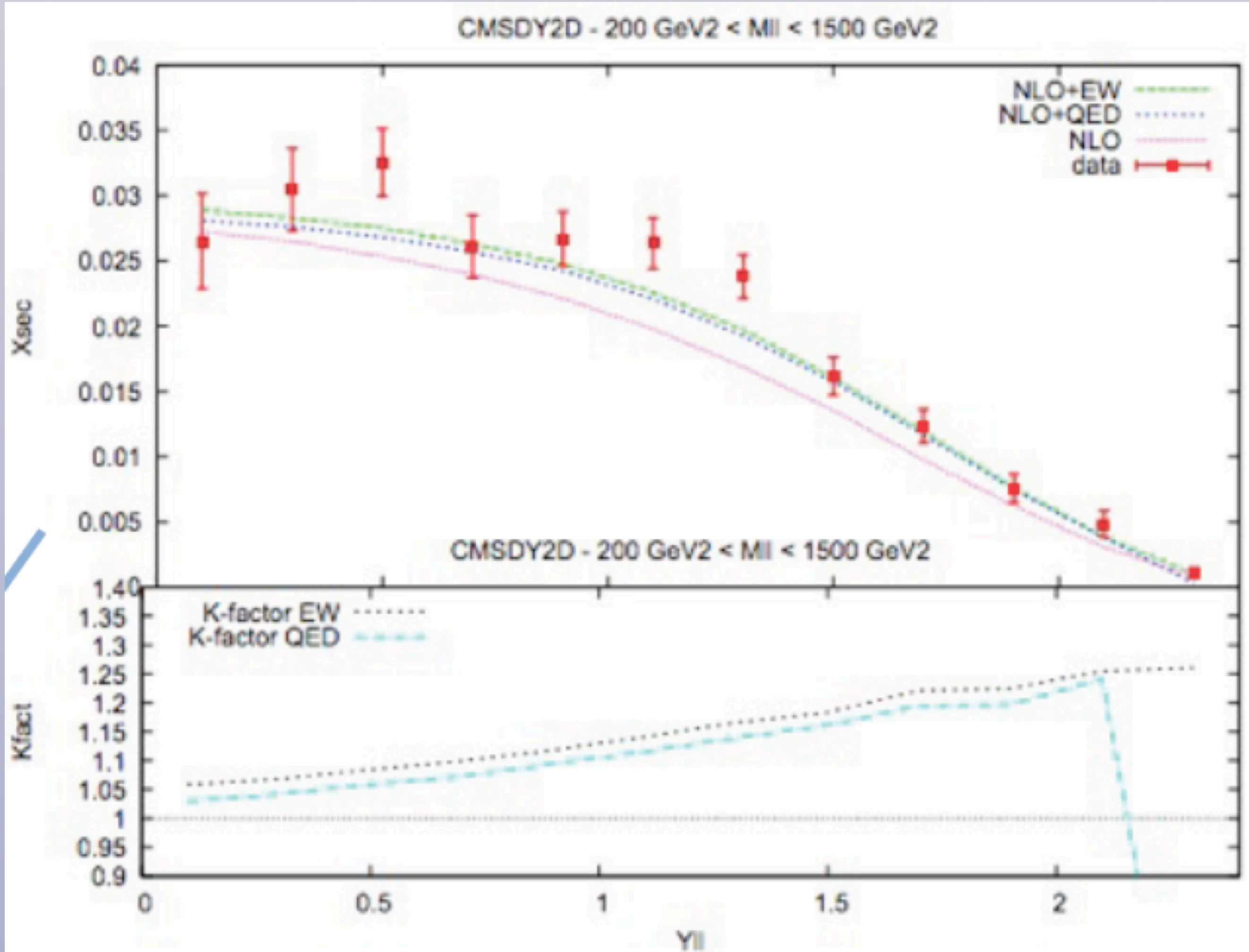
Plots by S. Carrazza and J. Pires



- Until **exact NNLO available**, jet data at small jet p_T and large η **should be excluded from NNLO fit**, since NNLOthres not suitable there
- Small impact for Tevatron jets (where NNLOthres works in a wider range) and ATLAS 2010 jet data (substantial uncertainties , but important for the **ATLAS and CMS jet data from the 2011 and 2012 runs**

Electroweak corrections

- Electroweak corrections are important for **W and Z** production specially at **large invariant masses**
- In NNPDF3.0, full NLO electroweak corrections for all **neutral current Drell-Yan** datasets have been computed with **FEWZ3.1**
- As an illustration, **high-mass Drell-Yan** from CMS 7 TeV



Electroweak corrections up to **25% at the highest dilepton masses**

QED corrections also large, but affected by the large uncertainties of **ɣ(x,Q)**

QED corrections not included in the NNPDF3.0 QCD-only fit (NNPDF3.0 QED to follow)

METHODOLOGY

See also Stefano's talk at the December PDF4LHC meeting

NNPDF3.0 is really NNPDF++

Completion of major software development project: rewriting of most of the **NNPDF fitting framework** from Fortran 77 to C++ and Python

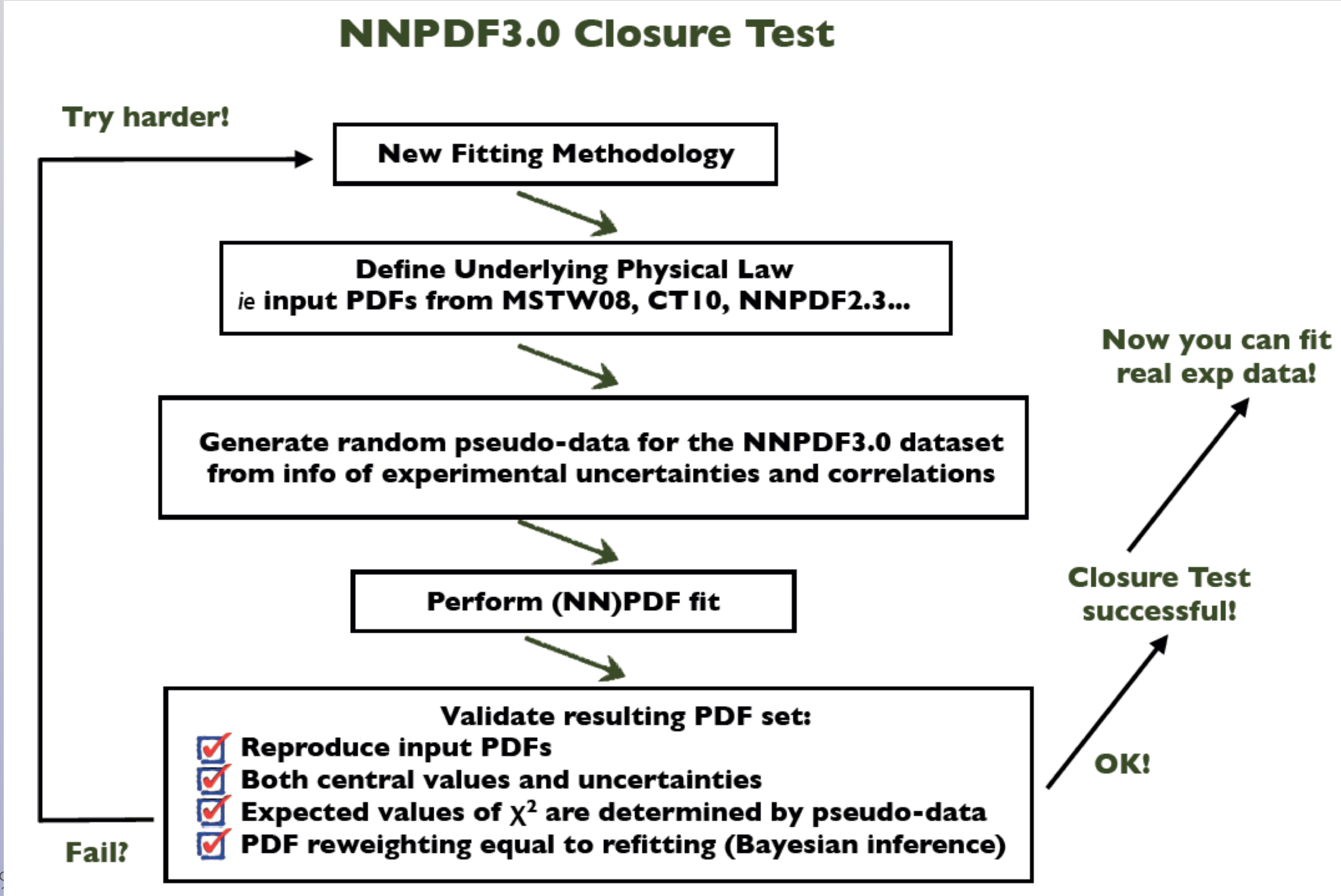
Lines of code by language in NNPDF3.0 fitting framework

Language	files	blank	comment	code
C++	106	6993	6048	26551
Fortran 77	113	115	10161	20872
C/C++ Header	134	1183	857	3920
make	34	792	447	1699
ASP.Net	1	511	0	1390
Bourne Shell	23	261	202	802
Python	8	187	168	565
Fortran 90	1	32	43	117
Bourne Again Shell	3	7	11	34
SUM:	423	10081	17937	55950

- ☑ **Modular structure**: each dataset is an individual object, with the associated theory encapsulated in individual **FK tables:** easy to include new measurements and to upgrade theory for existing ones
- ☑ **Greatly improved fitting efficiency**: main bottleneck for PDF fits is convolution between input PDFs and theory, performed here with assembly-like structure
- ☑ Fits can now be easily parallelized to run in **clusters** and in **Graphical Processing Units**

Closure Testing

Validation and optimization of fitting strategy performed on **closure test with known** underlying physical law



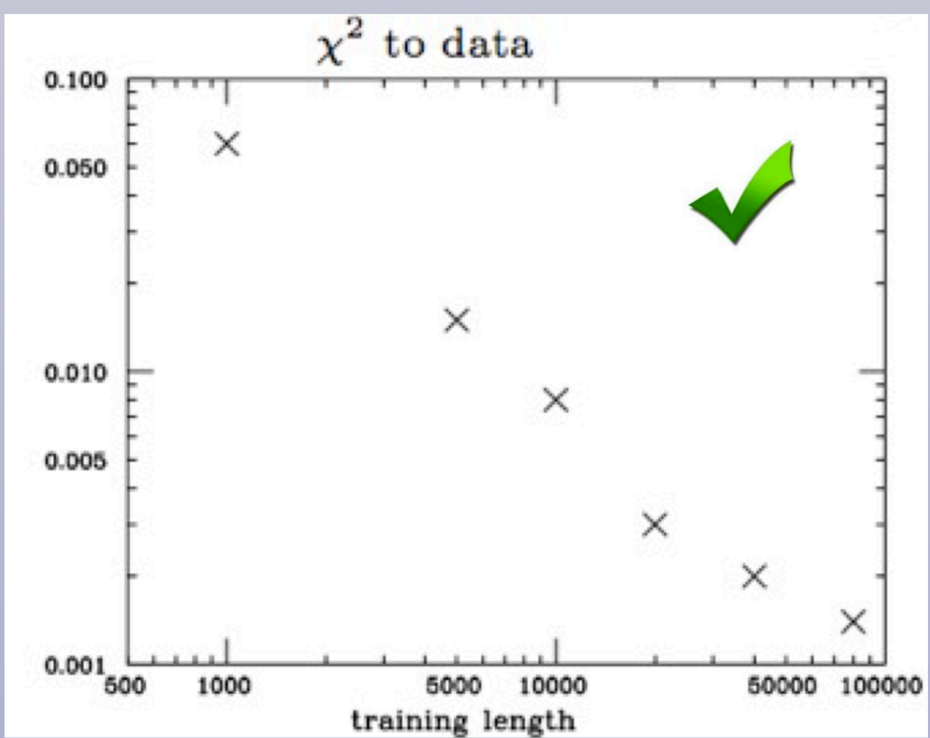
Closure Testing

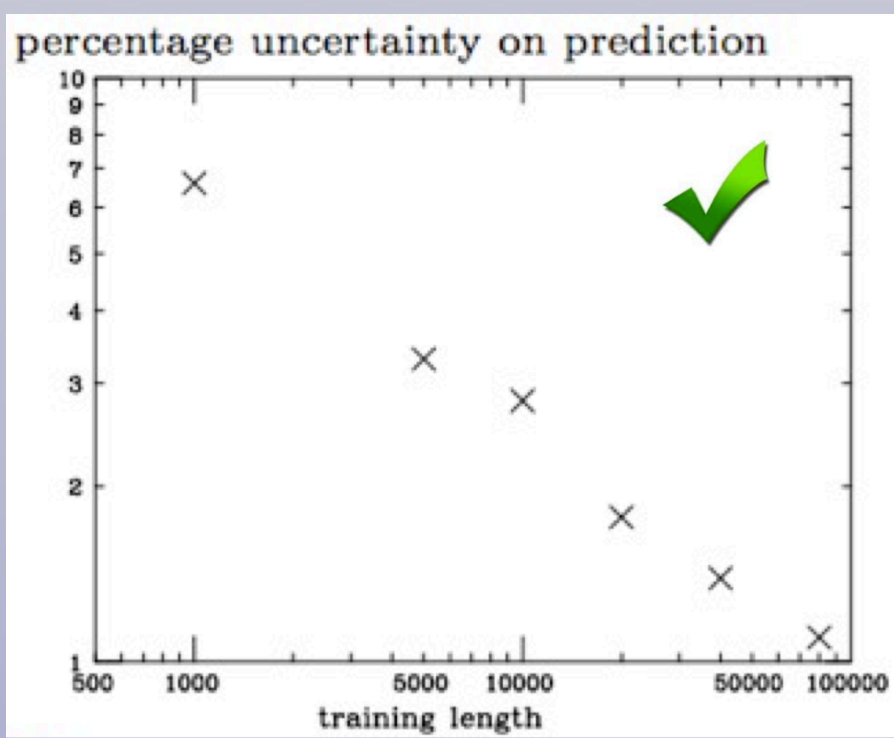
Validation and optimization of fitting strategy performed on **closure test with known** underlying physical law

Three levels of closure tests (**See Stefano's talk in Dec meeting**):

- **Level 0:** no fluctuations on pseudo-data, no Monte Carlo replica generation
- **Level 1**: with fluctuations on pseudo-data, no Monte Carlo replica generation
- **Level 2**: with fluctuations on pseudo-data, with Monte Carlo replica generation

Example: **Level 0 closure tests -** Fit results successfully **converge towards underlying law**: central χ^2 **to pseudo-data tends to zero**, same for PDF uncertainties on predictions (all replicas converge on same underlying law)





Independence of PDF fitting basis

- **Predictions for physical observables** should be independent of the specific choice of PDF fitting basis
- We have explored in **closure tests** that thanks to the **improved NNPDF3.0 methodology**, we achieve almost statistically equivalent fits using **two very different basis**

NNPDF2.3 basis

$$\begin{aligned}
\Sigma(x,Q_0^2) &= \sum_{i=1}^{3}(q_i+\bar{q}_i)(x,Q_0^2) \\
T_3(x,Q_0^2) &= (u+\bar{u}-d-\bar{d})(x,Q_0^2) \\
\Delta_S(x,Q_0^2) &= (\bar{d}-\bar{u})(x,Q_0^2) \\
s^+(x,Q_0^2) &= (s+\bar{s})(x,Q_0^2) \\
V(x,Q_0^2) &= \sum_{i=1}^{3}(q_i-\bar{q}_i)(x,Q_0^2) \\
s^-(x,Q_0^2) &= (s-\bar{s})(x,Q_0^2)
\end{aligned}$$

Natural basis from the point of view of physical observables

NNPDF3.0 basis

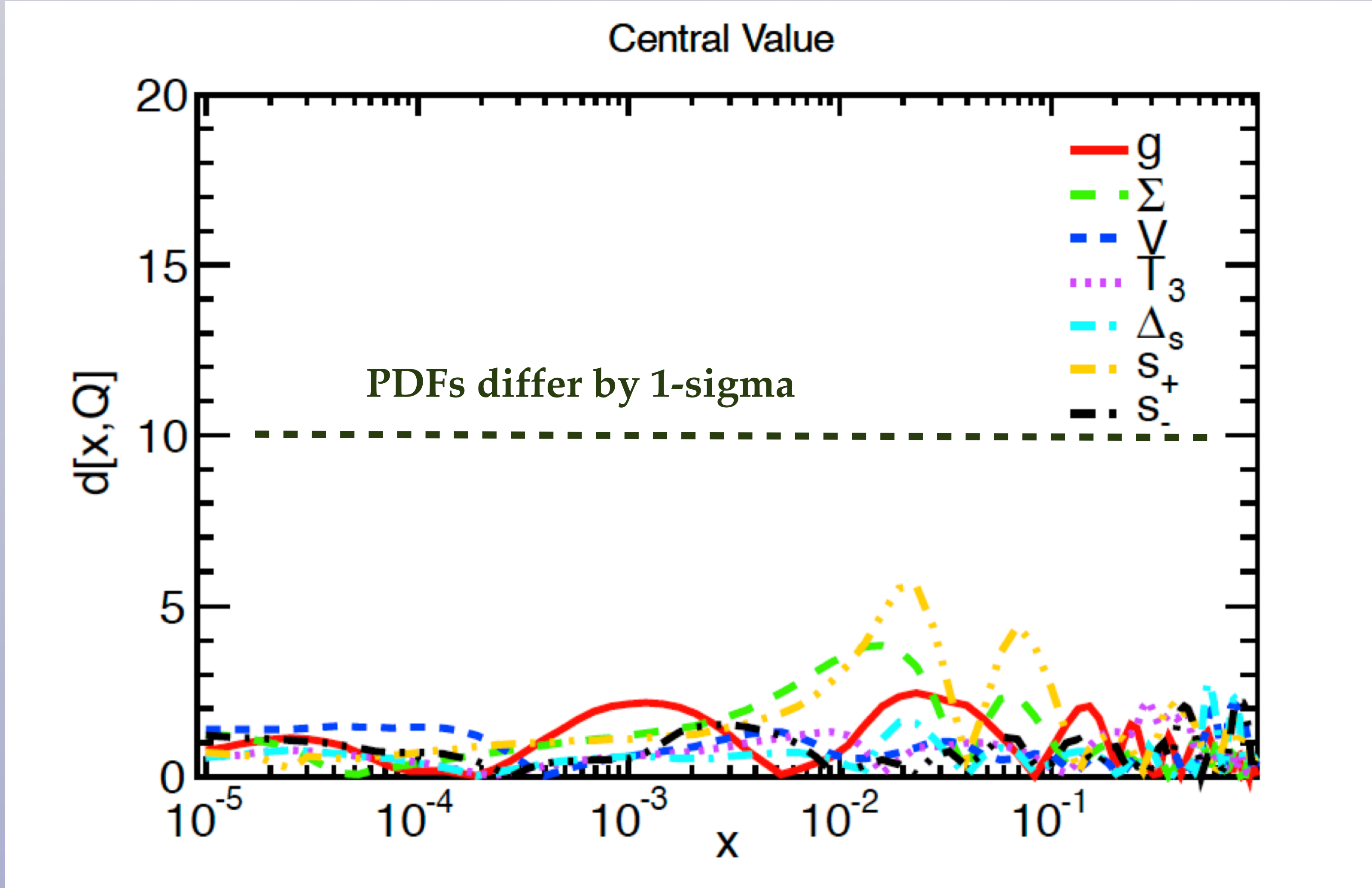
$$\begin{aligned}
\Sigma(x,Q_0^2) &= \sum_{i=1}^{3}(q_i+\bar{q}_i)(x,Q_0^2) \\
T_3(x,Q_0^2) &= (u+\bar{u}-d-\bar{d})(x,Q_0^2) \\
T_8(x,Q_0^2) &= (u+\bar{u}+d+\bar{d}-2s-2\bar{s})(x,Q_0^2) \\
V(x,Q_0^2) &= \sum_{i=1}^{3}(q_i-\bar{q}_i)(x,Q_0^2) \\
V_3(x,Q_0^2) &= (u-\bar{u}-d+\bar{d})(x,Q_0^2) \\
V_8(x,Q_0^2) &= (u-\bar{u}+d-\bar{d}-2s+2\bar{s})(x,Q_0^2)
\end{aligned}$$

Natural basis from point of view of PDF evolution equations

Independence of PDF fitting basis

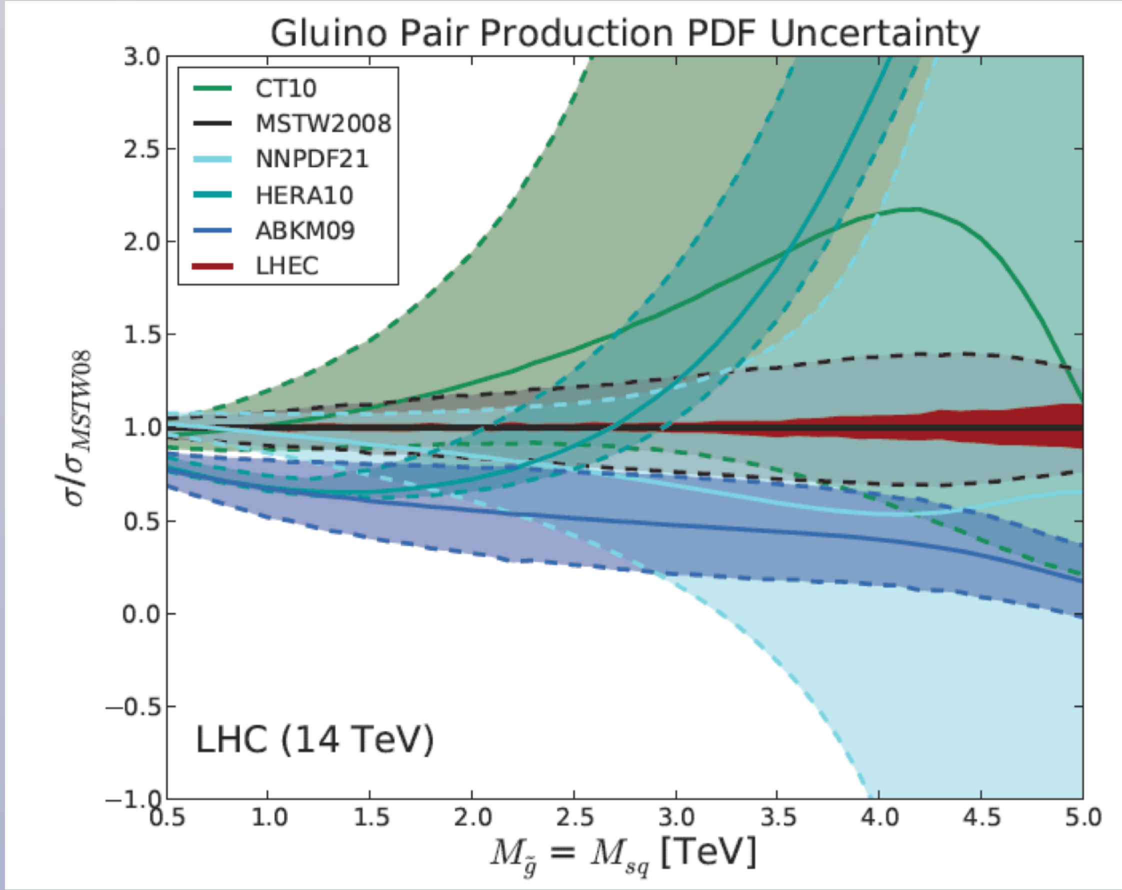
- **Predictions for physical observables** should be independent of the specific choice of PDF fitting basis
- We have explored in **closure tests** that thanks to the **improved NNPDF3.0 methodology**, we achieve almost statistically independent fits using **two very different basis**

Fit in NNPDF2.3 basis vs Fit in NNPDF3.0 basis



Positivity of physical cross-sections

- While PDFs are not **positive definite beyond LO**, **physical cross-sections** should always be positive
- Implementing this condition, **without overconstraining PDFs with a too restrictive parametrization**, is essential for a reliable estimate of PDF uncertainties
- This is particularly crucial in the **large-x region,** production of BSM high-mass particles



See also **Gavin's talk** this afternoon on PDF systematics on high-mass searches using **ColliderReach**

Positivity of physical cross-sections

- While PDFs are not **positive definite beyond LO**, **physical cross-sections** should always be positive
- Implementing this condition, **without overconstraining PDFs with a too restrictive parametrization**, is essential for a reliable estimate of PDF uncertainties
- This is particularly crucial in the **large-x region,** production of BSM high-mass particles

Quark positivity

$$F_2^u(x,Q^2) \propto x\left(u(x,Q^2)+\bar{u}(x,Q^2)\right)+\mathcal{O}\left(\alpha_s\right)$$

$$F_2^d(x,Q^2) \propto x\left(d(x,Q^2)+\bar{d}(x,Q^2)\right)+\mathcal{O}\left(\alpha_s\right)$$

$$F_2^s(x,Q^2) \propto x\left(s(x,Q^2)+\bar{s}(x,Q^2)\right)+\mathcal{O}\left(\alpha_s\right)$$

Anti-Quark positivity

$$\frac{d^2\sigma^{\rm DY,u}}{dM^2\,dy} \propto u(x_1,Q^2)\bar{u}(x_3,Q^2)+\mathcal{O}\left(\alpha_s\right)$$

$$\frac{d^2\sigma^{\rm DY,d}}{dM^2\,dy} \propto d(x_1,Q^2)\bar{d}(x_2,Q^2)+\mathcal{O}\left(\alpha_s\right)$$

$$\frac{d^2\sigma^{\rm DY,s}}{dM^2\,dy} \propto s(x_1,Q^2)\bar{s}(x_2,Q^2)+\mathcal{O}\left(\alpha_s\right)$$

Gluon positivity

$$F_L(x,Q^2) \propto C_g \otimes g(x,Q^2) + C_q \otimes q(x,Q^2)+\mathcal{O}\left(\alpha_s^2\right)$$

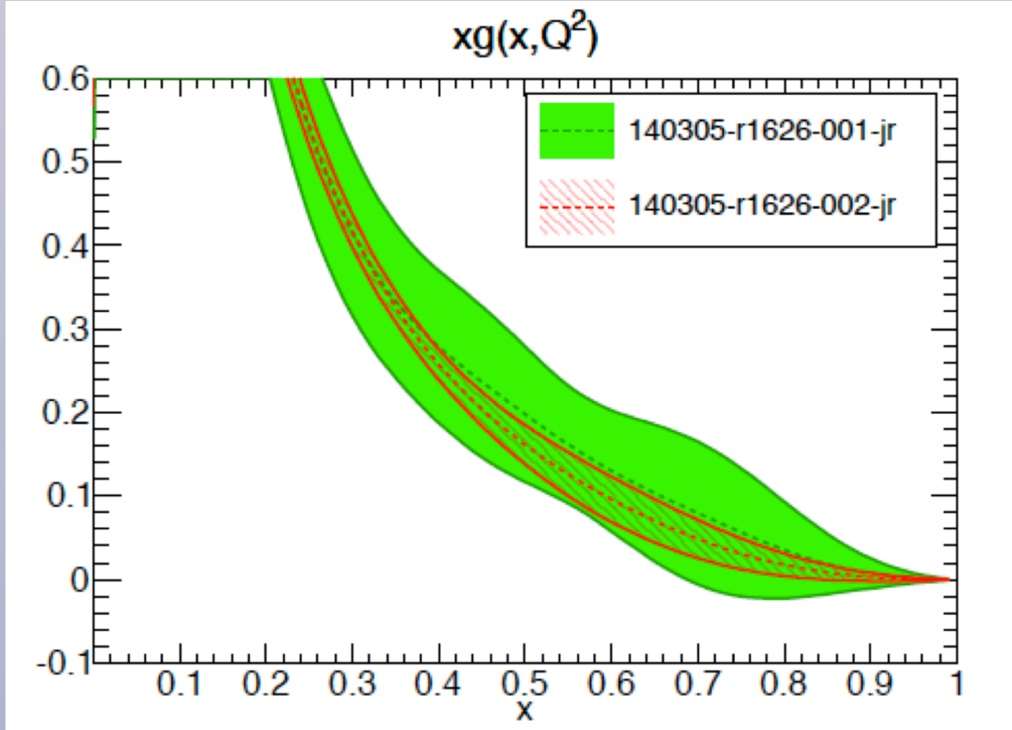
$$\frac{d\sigma^{\rm higgs}}{dy_H} \propto g(x_1,M_H^2)g(x_2,M_H^2)+\mathcal{O}\left(\alpha_s^3\right) \quad M_H \equiv Q_{\rm pos}$$

Positivity of physical cross sections imposed at a **low scale ~ 2 GeV**, then maintained by evolution

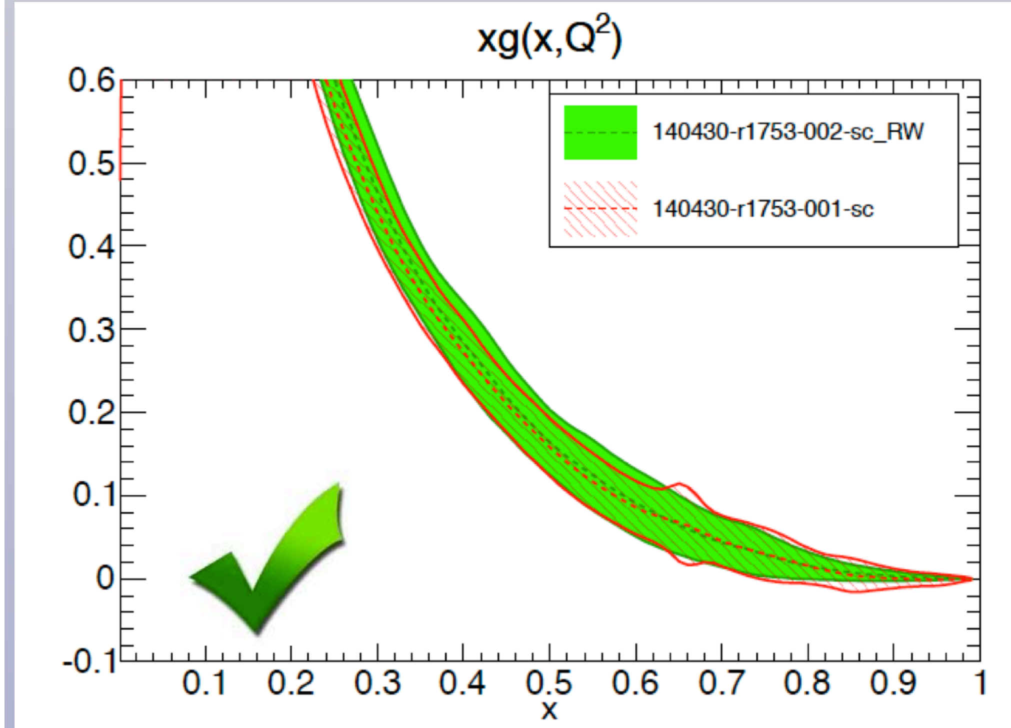
PDF reweighting

- **Statistically identical results** should be obtained when **refitting** or when **reweighting,** even more so in a **closure test**
- Thus **Bayesian inference** can be used as the **ultimate closure test**, sensitive to all the higher moments of the refitted / reweighting PDFs
- As an illustration, compare effects of **CDF, D0 and ATLAS jet data** included by refitting and by reweighting. Compare also **NNPDF** and **Giele-Keller** prescriptions.

Fit wo jet data vs Fit with jet data



Fit with jet data vs RW with jet data

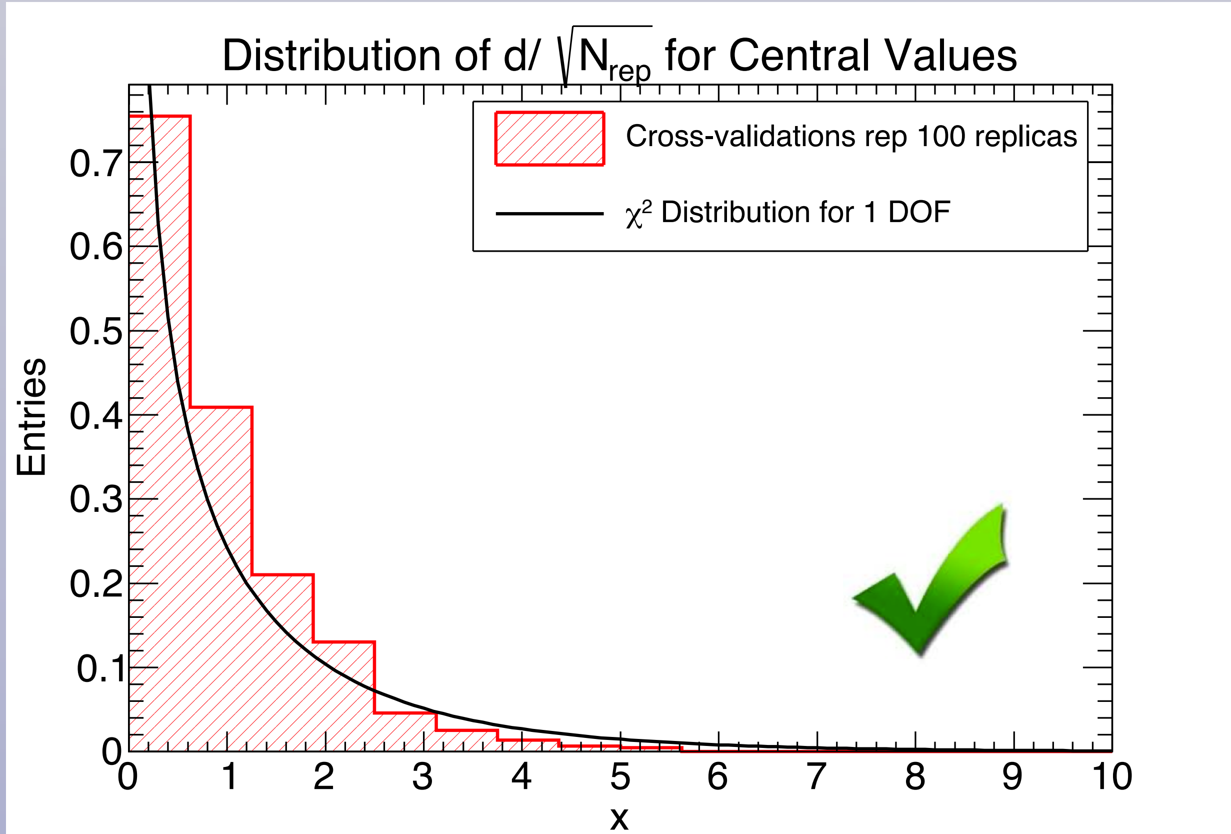


Note large uncertainties on gluon from missing jet data

PDF reweighting

- **Statistically identical results** should be obtained when **refitting** or when **reweighting,** even more so in a **closure test**
- Thus **Bayesian inference** can be used as the **ultimate closure test**, sensitive to all the higher moments of the refitted / reweighting PDFs
- As an illustration, compare effects of **CDF, D0 and ATLAS jet data** included by refitting and by reweighting. Compare also **NNPDF** and **Giele-Keller** prescriptions.

Distribution of distances between refit and RW



PDF reweighting

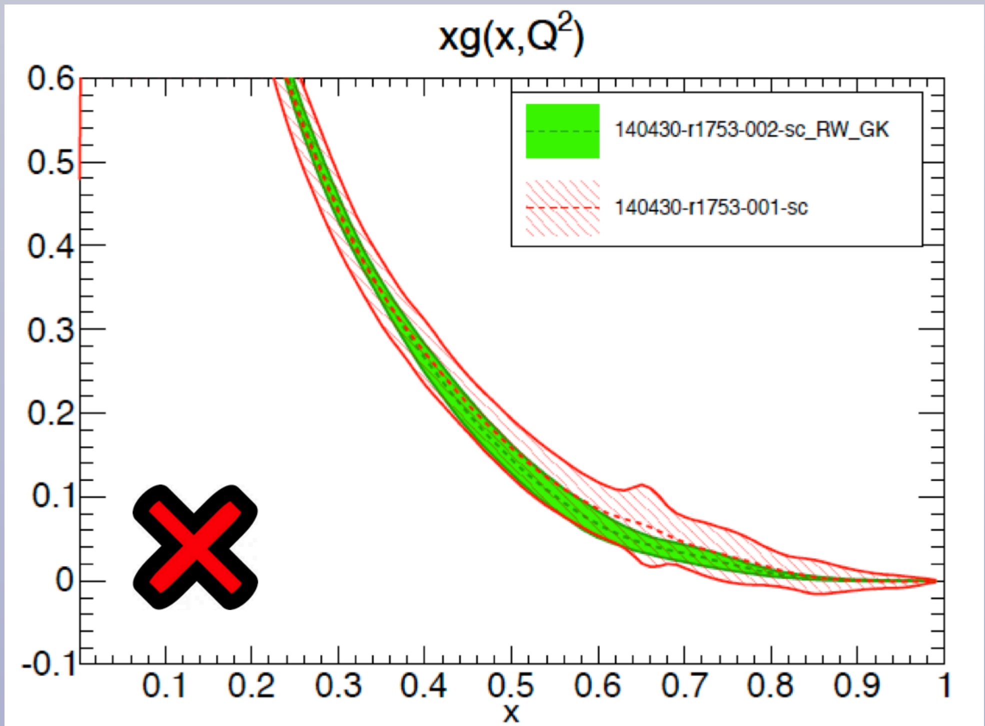
- **Statistically identical results** should be obtained when **refitting** or when **reweighting,** even more so in a **closure test**
- Thus **Bayesian inference** can be used as the **ultimate closure test**, sensitive to all the higher moments of the refitted / reweighting PDFs
- As an illustration, compare effects of **CDF, D0 and ATLAS jet data** included by refitting and by reweighting. Compare also **NNPDF** and **Giele-Keller** prescriptions.

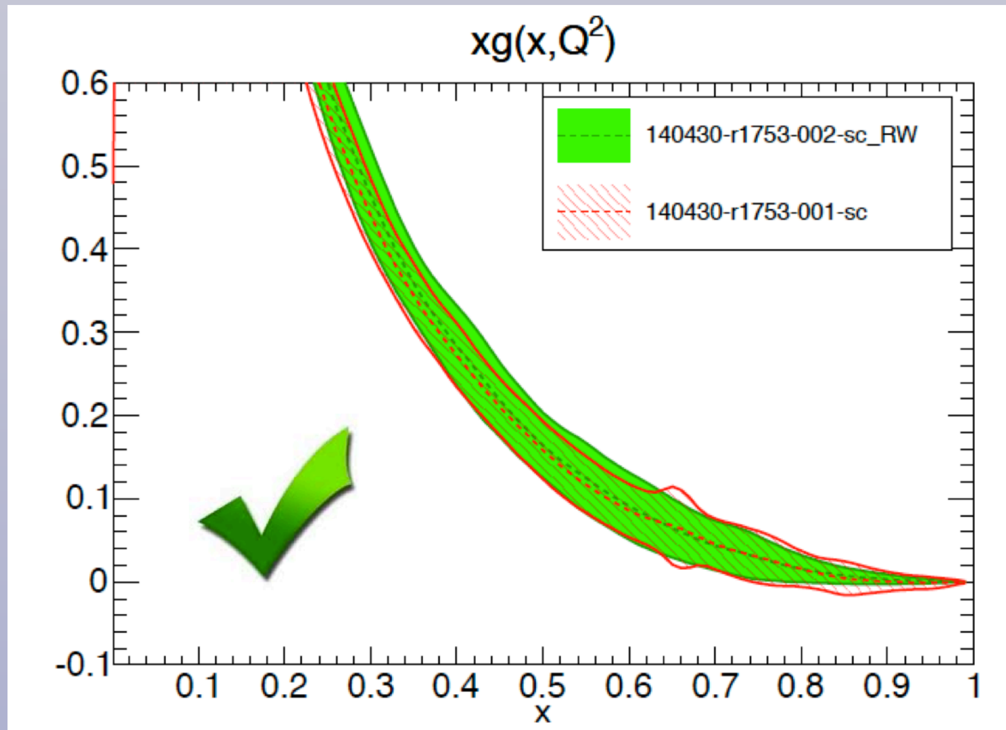
Giele-Keller

$$w_k^G = \frac{e^{-\frac{1}{2}\chi_k^2}}{\frac{1}{N}\sum_{k=1}^N e^{-\frac{1}{2}\chi_k^2}}.$$

NNPDF

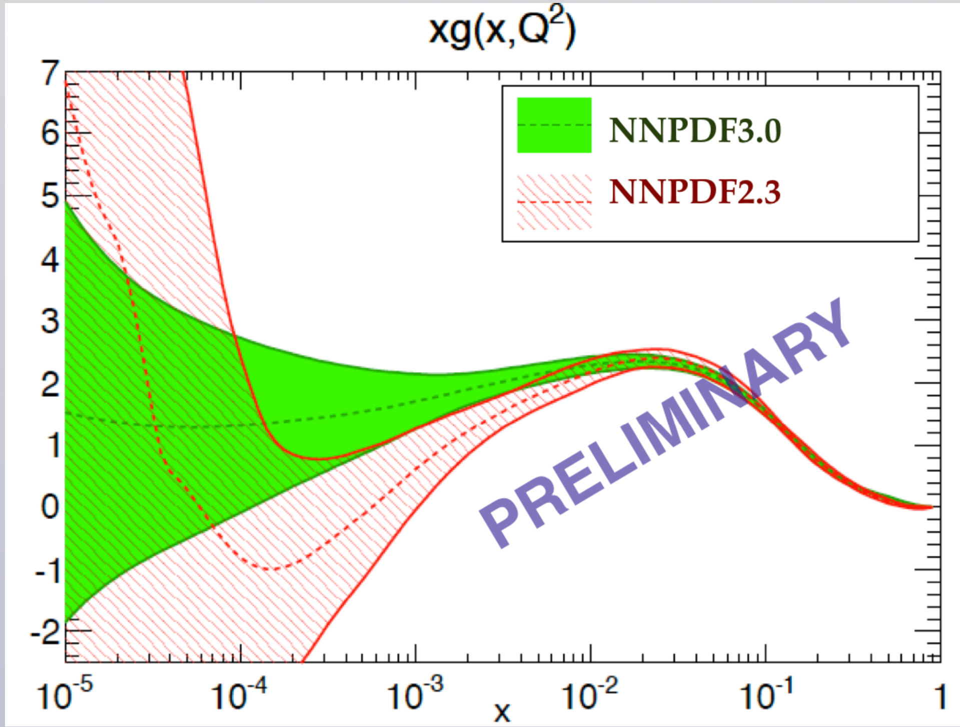
$$w_k = \frac{(\chi_k^2)^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}\chi_k^2}}{\frac{1}{N}\sum_{k=1}^N (\chi_k^2)^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}\chi_k^2}}$$



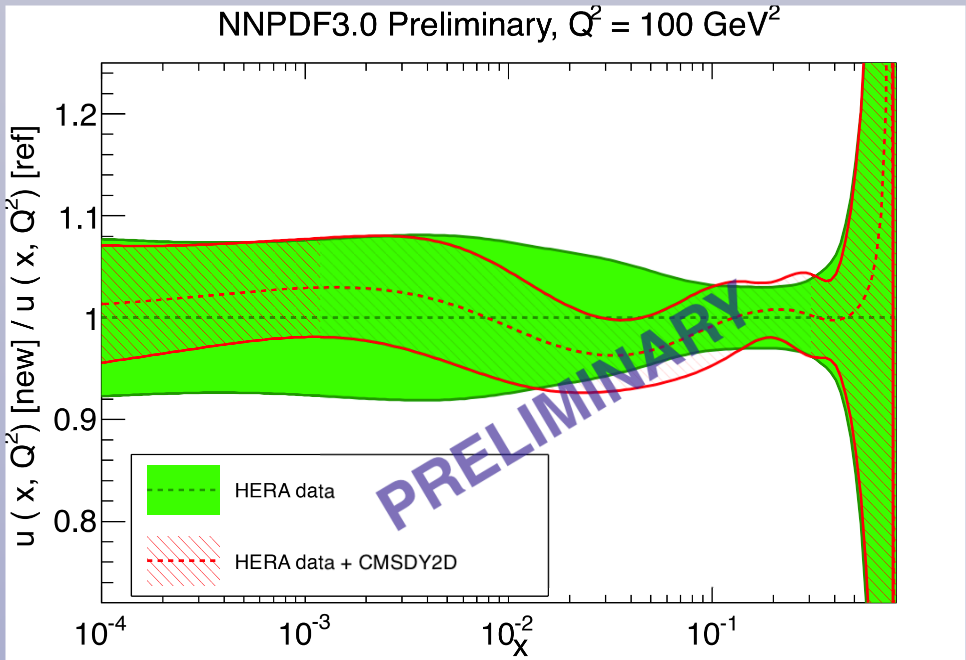


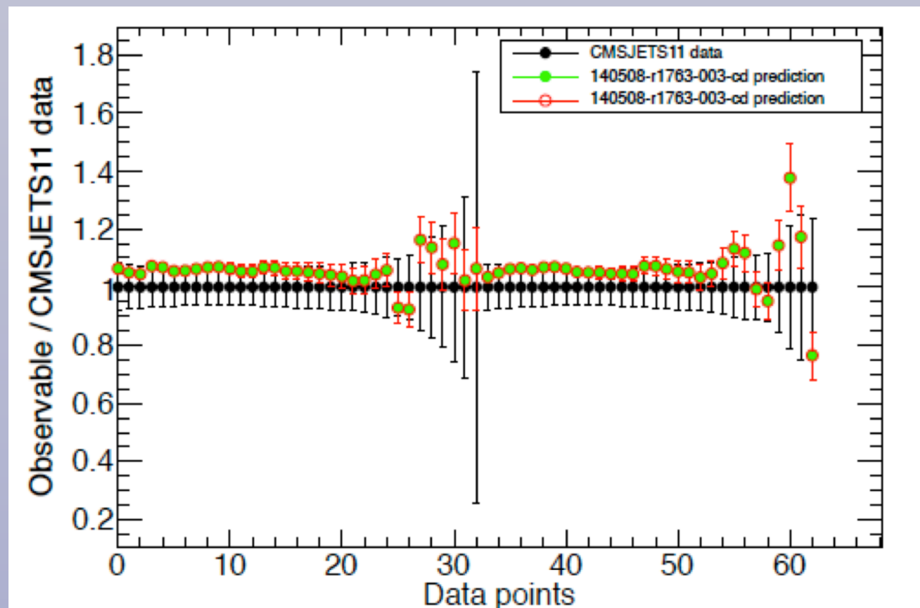
NNPDF3.0
Preliminary results

NNPDF3.0 preliminary



- Reasonable agreement with all **the new datasets** included in the NNPDF3.0 fit
- Differences between 2.3 and 3.0prel between the 1-sigma and 2-sigma levels: impact of **new data** and of **updated theory and methodology**
- Currently finalizing the fits and exploring **implications for LHC phenomenology**
- Public release by summer





Summary and outlook

The **NNPDF3.0** is the new upcoming release from the NNPDF collaboration.

It represents a **substantial improvement over NNPDF2.3** both in terms of data, theory and methodology

- **Data:** all available **H1 and ZEUS HERA-II data included**, and many new **LHC** measurements from **ATLAS, CMS, LHCb** including W asymmetry, W+charm, inclusive jets, high and low mass Drell-Yan and top quark production, and other datasets in the pipeline
- **Theory:** Improved **approximate NNLO K-factors for jet data** based on the partial exact NNLO results, **electroweak corrections** included for all relevant data, FONLL-B for NLO sets
- **Methodology:** fitting strategy validated using **closure tests,** optimized Genetic Algorithms, extended positivity, fast Bayesian regularization, PDF fitting basis independence ....

NNPDF3.0 is now being finalized and the corresponding **LHC phenomenology** studied

LO, NLO and NNLO sets for a range of α_S **values** will become available in **LHAPDF6**

In the medium and long term, **NNPDF development plans** include

- Include all relevant **LHC Run I data**: Complete set of 8 TeV measurements, high pT Z+jets, direct photon production. Then from next years also add **LHC Run II data**
- **Upgrade theory calculations** as they become available: NNLO for top quark differential distributions, exact NNLO for jets, for Z+jets
- Produce NNPDF3.0 sets with **QED corrections**, **intrinsic charm**, threshold and high-energy **resummation,** as well as PDF sets specific for Monte Carlo event generators