Seminar, LAL Orsay, May 22, 2015

CT14 and META2.0 parton distributions

Pavel Nadolsky
Southern Methodist University

On behalf of CTEQ-TEA group
S. Dulat (Xinjiang), J. Gao (Argonne), M. Guzzi (Manchester), T.-J. Hou, P.N. (SMU), J. Huston, J. Pumplin, C. Schmidt, D. Stump, C. -P. Yuan (Michigan State U.)

About the title

1. **CT14** are new parton distribution functions (PDFs) for general-purpose applications at the LHC and other experiments. They are developed by CTEQ-TEA (**T**ung **E**t. **A**l.) group

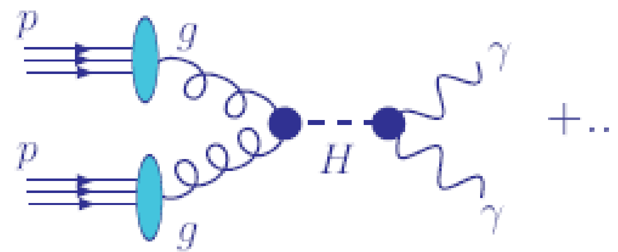
   CT14 are provided at LO, NLO, and NNLO and include LHC Run-1 experimental data; will be submitted to LHAPDF within a week

2. **Meta-parametrizations** are introduced to combine PDFs from several groups (CT, MMHT, NNPDF, …) in a variety of LHC applications. **META** parametrizations (version 2.0) offer a versatile framework for combination of PDF+α_s uncertainties from global PDF ensembles in LHC Run-2 analyses.

PDFs are basic blocks of theoretical predictions for hadronic scattering in perturbative QCD. They cannot be computed, but their accuracy must match accuracy of hard-scattering cross sections

Example: total cross section for $gg \rightarrow \text{Higgs} \rightarrow \gamma\gamma$

Cross section $\sigma_{pp\rightarrow H\rightarrow\gamma\gamma}$ for production and decay of H, e.g., via $g + g \rightarrow H$:



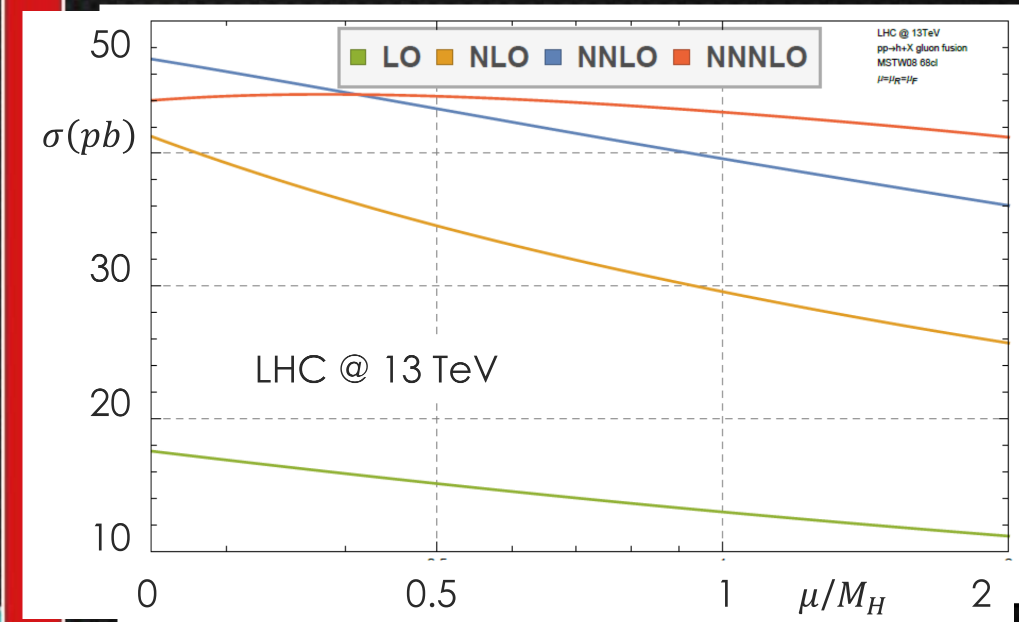
$$\sigma_{pp\rightarrow H\rightarrow\gamma\gamma X}(Q) = \sum_{a,b=g,q,\bar{q}} \int_0^1 d\xi_a \int_0^1 d\xi_b \hat{\sigma}_{ab\rightarrow H\rightarrow\gamma\gamma}\left(\frac{x_a}{\xi_a}, \frac{x_b}{\xi_b}, \frac{Q}{\mu_R}, \frac{Q}{\mu_F}; \alpha_s(\mu_R)\right) \times f_a(\xi_a, \mu_F) f_b(\xi_b, \mu_F) + O\left(\frac{\Lambda_{QCD}^2}{Q^2}\right)$$

- $\hat{\sigma}_{ab\rightarrow H\rightarrow\gamma\gamma}$ is the cross section for scattering of two partons, a and b; can be computed as a **perturbative** series in $\alpha_s(\mu_R)$, at a renormalization scale $\mu_R \gg \Lambda_{QCD}$

- $f_{a/p}\,(\xi, \mu_F)$ is the **nonperturbative** PDF for finding a parton a with the momentum fraction ξ in the proton p, at a factorization scale $\mu_F \gg \Lambda_{QCD}$

Hard-scattering cross sections for $gg \to H \to \gamma\gamma$

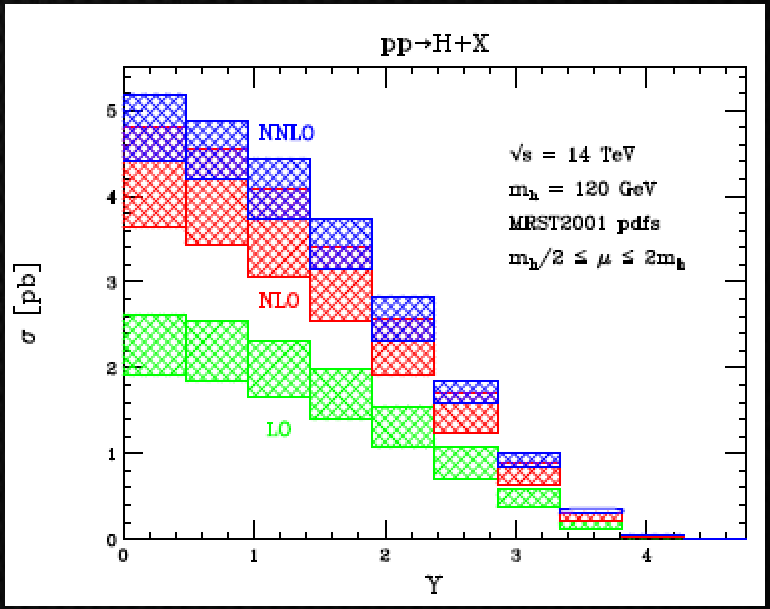
N3LO for total cross sections



Anastasiou, Duhr, Dulat, Herzog, Mistlberger, 1503.06056

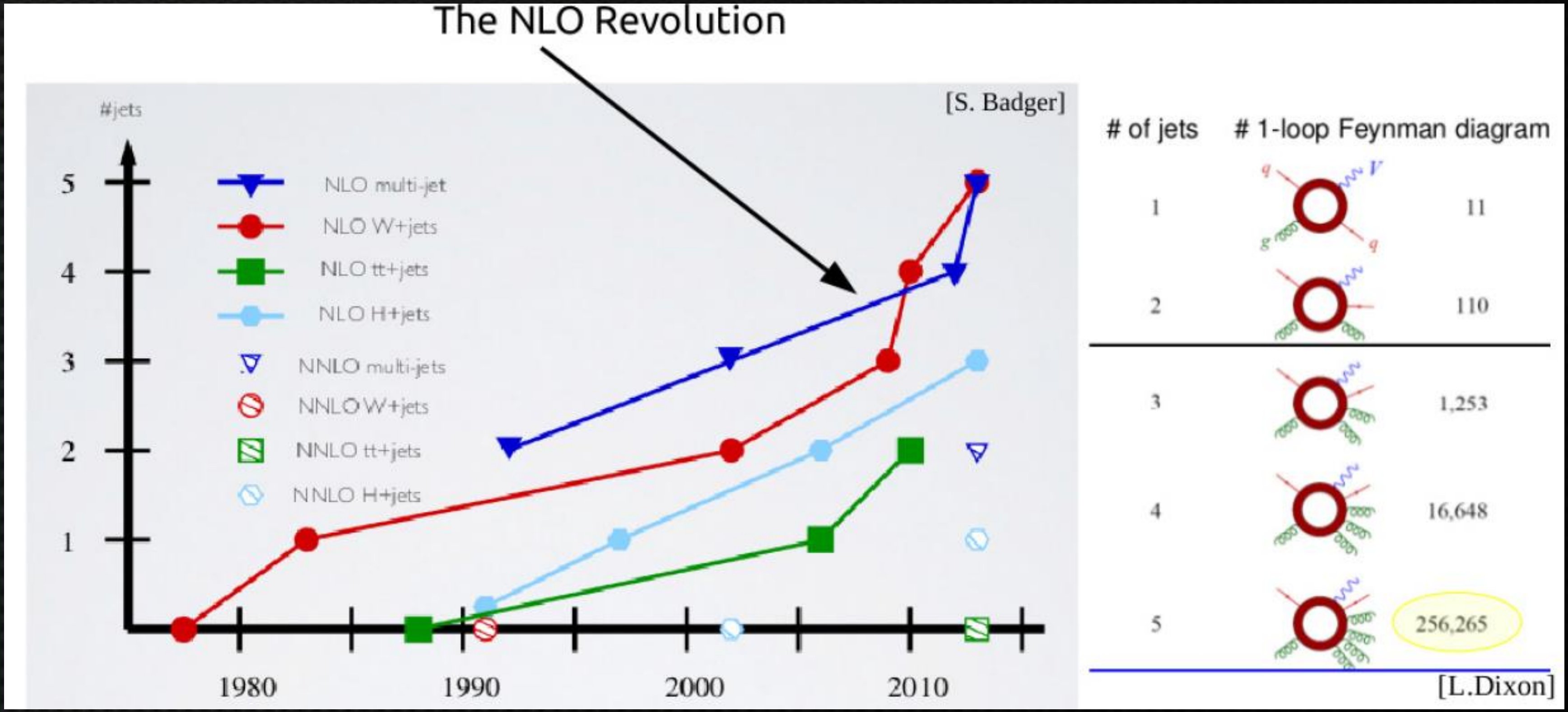
N3LO corrections are of the order of +2.2%.
The total scale variation at N3LO is 3%

NNLO for differential distributions



Anastasiou, Melnikov, Petriello, hep-ph/0409088, 0501130

Perturbative QCD loop revolution



Since 2005, generalized unitarity and related methods dramatically advanced the computations of **perturbative** NLO/NNLO/N3LO hard cross sections.

Universality of PDFs

To all orders in α_s, PDFs are **defined** as matrix elements of certain correlator functions:

$$f_{q/p}(x,\mu) = \frac{1}{4\pi}\int_{-\infty}^{\infty} dy^- e^{iy^- p^+} \langle p \left| \overline{\psi}_q(0,y^-,\vec{0}_T)\gamma^+\psi_q(0,0,\vec{0}_T) \right| p\rangle, \text{ etc.}$$

PDFs are **universal** – depend only on the type of the hadron (p) and parton (q, $\bar{q}$, g)

... can be **parametrized** as

$$f_{i/p}(x,Q_0) = a_0 x^{a_1}(1-x)^{a_2}F(a_3,a_4,...) \text{ at } Q_0 \sim 1 \text{ GeV}$$

... predicted by solving DGLAP equations at $\mu > Q_0$

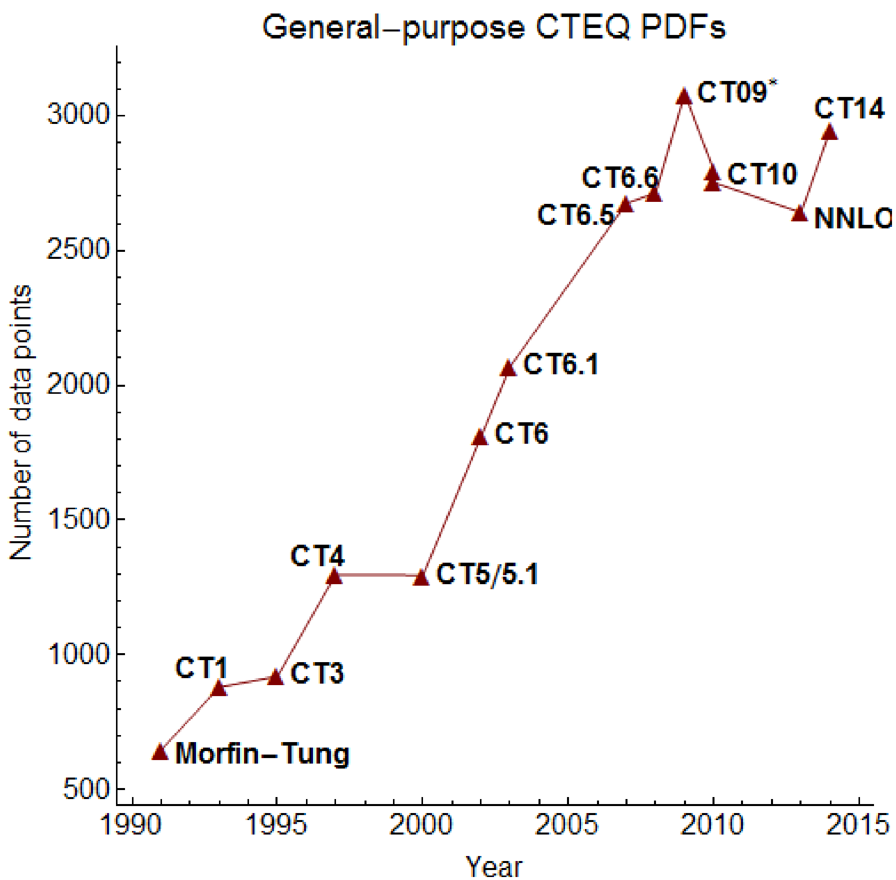
... constrained by a few "clean" QCD processes, such as inclusive DIS

Coordinated Theoretical-Experimental study of QCD

Timeline of global QCD analysis

Global analysis *(term promoted by J. Morfin & W.-K. Tung in 1990)*:

constrains PDFs or other nonperturbative functions with data from diverse hadronic experiments

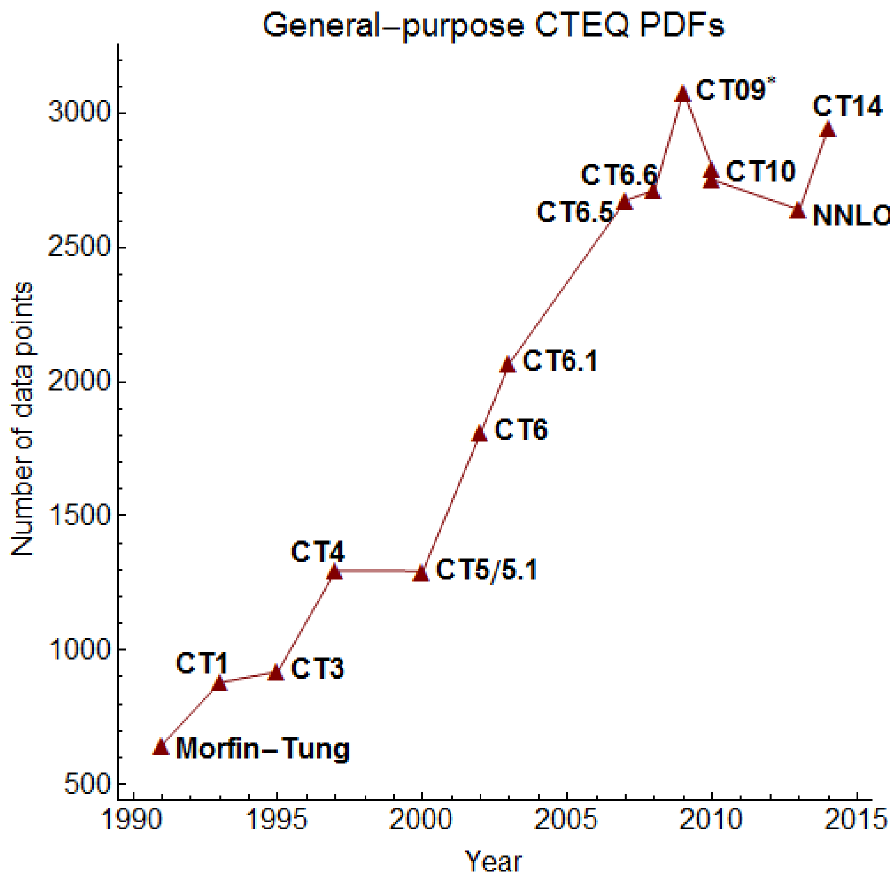


Morfin–Tung:	DIS, low–Q Drell–Yan process
CT1:	DIS, DY, direct γ
CT3:	low–x DIS, W charge asymmetry
CT4:	– direct γ; + CDF high–p_T jets
CT5/5.1:	$+\sigma_{pp}/\sigma_{pd}$ in DY process; D0 jets
CT6:	Error PDFs; correlated syst. errors
CT6.1:	Tevatron Run–1b jets
CT6.5:	GM–VFN scheme; free s(x)
CT6.6:	PDF correlations
CT09*:	+Tevatron Run–2 jets
CT10:	+NNLO, combined HERA, Run–2 W asymmetry
CT14:	+LHC Run–1 W, Z, jet production

Coordinated Theoretical-Experimental study of QCD

Timeline of global QCD analysis

Global analysis *(term promoted by J. Morfin & W.-K. Tung in 1990)*: constrains PDFs or other nonperturbative functions with data from diverse hadronic experiments



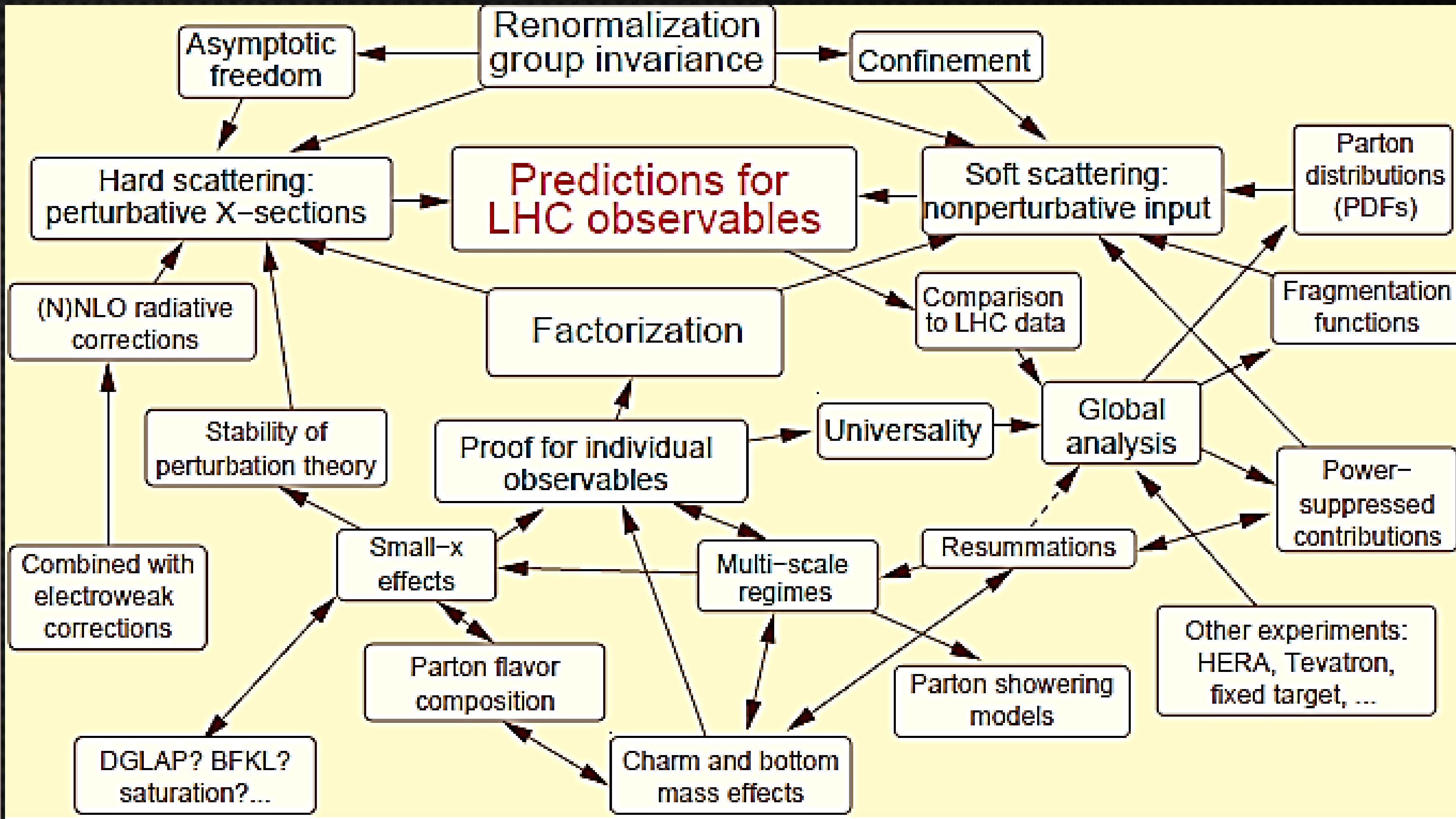
CT10/CT10W NLO [arXiv:1007.2241] and **CT10 NNLO** [arXiv:1302.6246] are in good agreement with LHC Run-1 data

The latest CT14 ensembles include several developments toward a long-term target of obtaining "**PDFs that achieve 1% accuracy**" in LHC processes

This is the most basic picture.

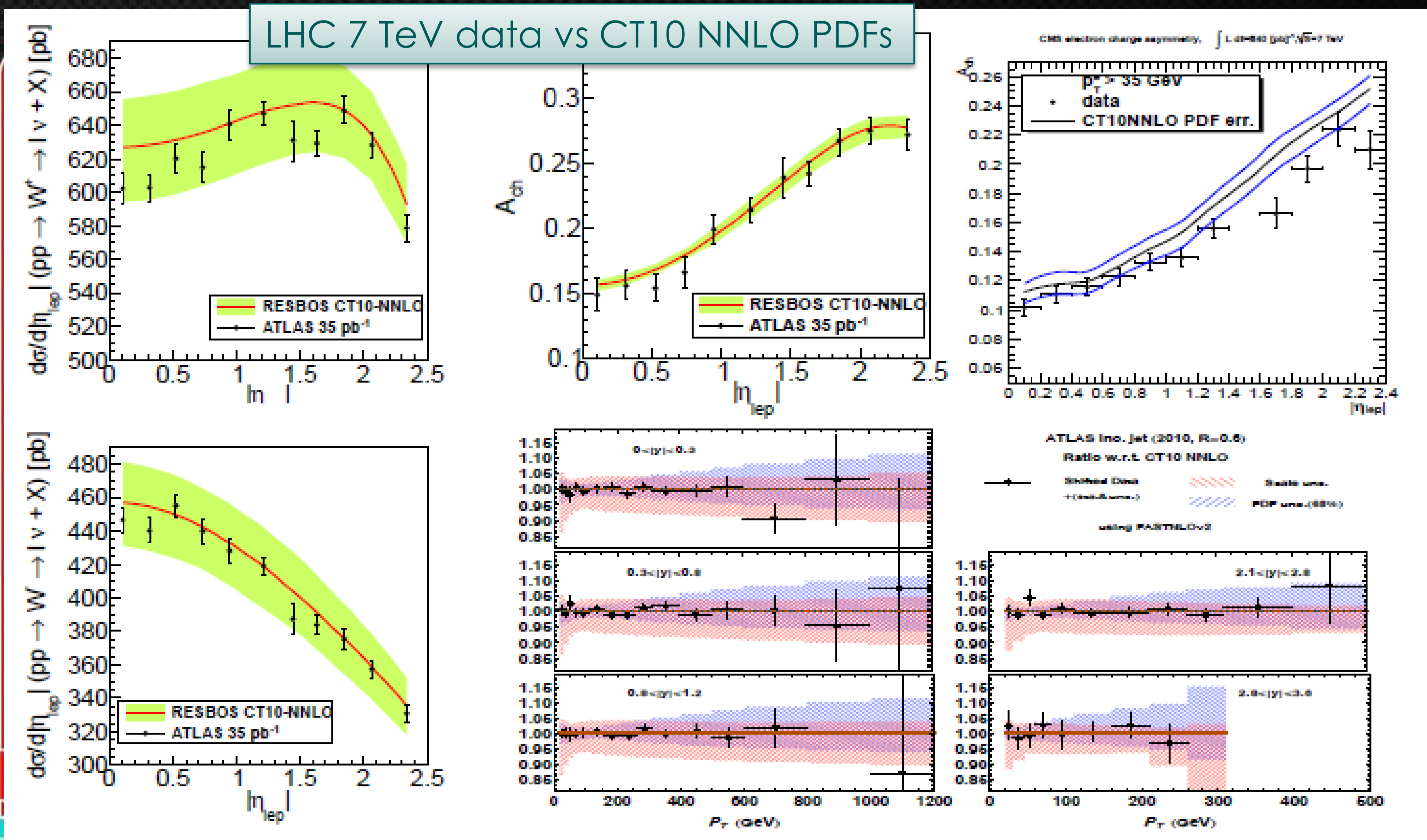
The actual story is more involved and still develops

(N)(N)NLO hard cross sections alone supply only some ingredients for the global analysis. Multiple effects contribute at comparable level.

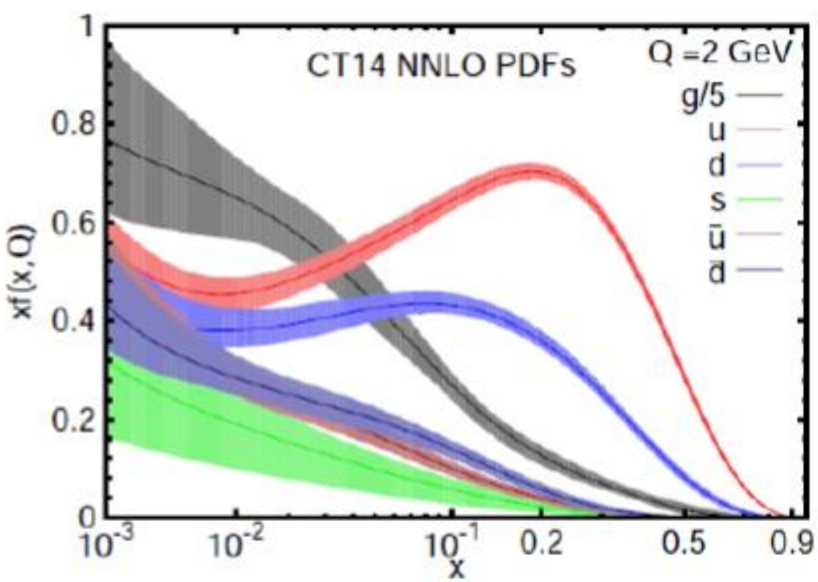


Concept map (c. 2007), even more relevant now

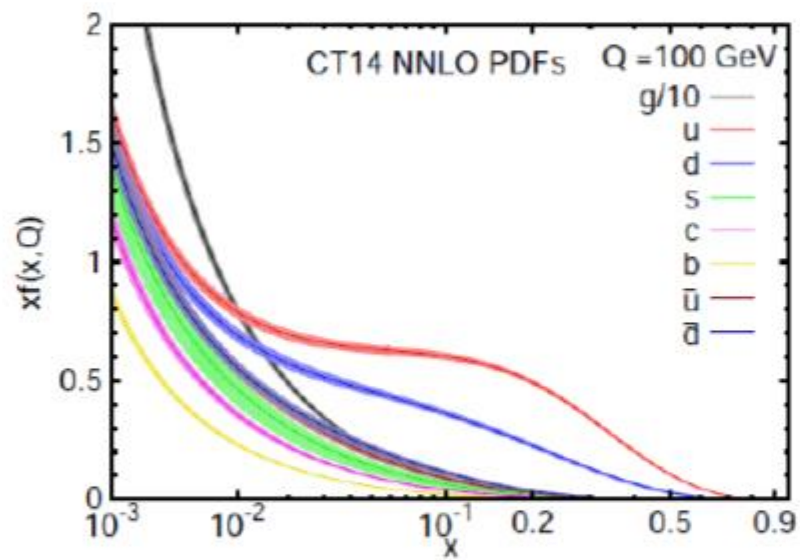
Remarkably, all components can be made to work as intended. For example, CT10 NNLO PDFs do not include LHC data, but **predict** LHC Run-1 observables very well



We will do even better with CT14 PDFs, which now include the LHC data



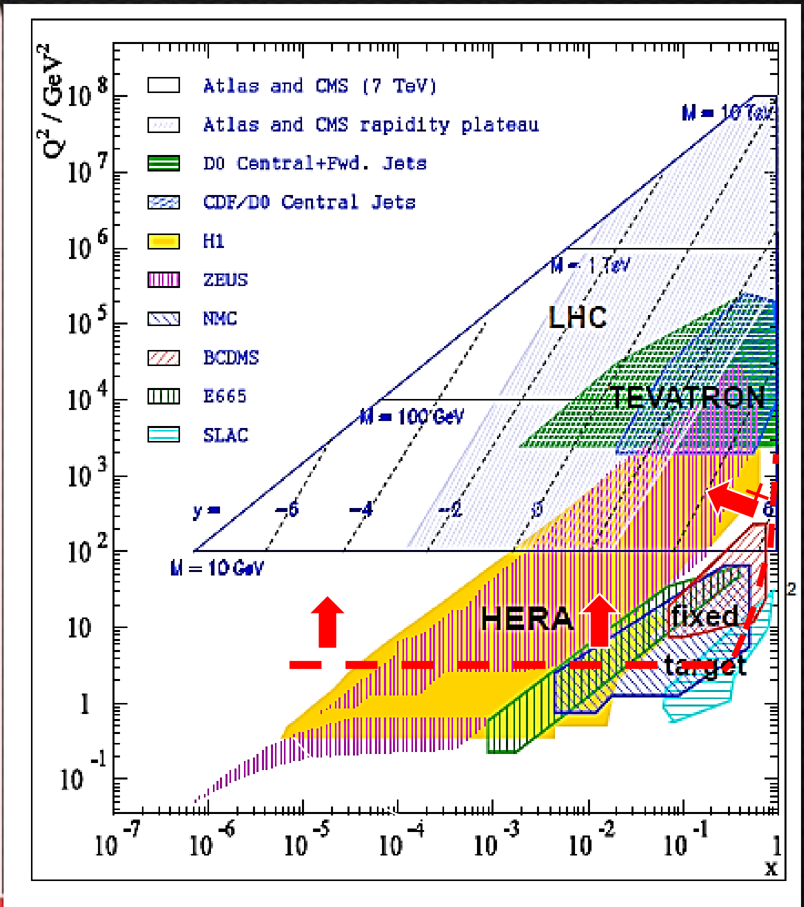
Q= 2 GeV



Q= 100 GeV

CT14: selection of experiments

Experimental measurements are selected so as to reduce dependence on any theoretical input beyond the leading power in perturbative QCD



Only DIS data with $Q^2 > 4\ GeV^2$, $W^2 > 12.25\ GeV^2$ (above the red line) are accepted to ensure stable perturbative predictions

Still using data from DIS and DY on **nuclear targets,** but are expecting to start replacing them (e.g., NMC DIS **on deuteron**) by comparable future LHC/Tevatron measurements on **the proton**

Experiments in the CT14 analysis

33 experiments; χ^2/N_{pt} = 3252/2947 =1.10

Experimental data set		N_{pt}	χ^2_e/N_{pt}
BCDMS F_2^p	[12]	337	1.14
BCDMS F_2^d	[13]	250	1.18
NMC F_2^d/F_2^p	[14]	123	1.08
NMC σ^p_{red}	[14]	201	1.85
CDHSW F_2^p	[15]	85	0.85
CDHSW F_3^p	[15]	96	0.83
CCFR F_2^p	[16]	69	1.02
CCFR xF_3^p	[17]	86	0.36
NuTeV $\nu\mu\mu$ SIDIS	[18]	38	0.62
NuTeV $\bar{\nu}\mu\mu$ SIDIS	[18]	33	1.18
CCFR $\nu\mu\mu$ SIDIS	[19]	40	0.72
CCFR $\bar{\nu}\mu\mu$ SIDIS	[19]	38	0.53
H1 σ_r^b	[20]	10	0.68
HERA charm production	[21]	47	1.26
HERA1 Combined NC and CC DIS	[22]	579	1.02
H1 F_L	[23]	9	1.92

Experimental data set		N_{pt}	χ^2_e/N_{pt}
E605 Drell-Yan process	[24]	119	0.98
E866 Drell-Yan process	[25]	15	0.87
E866 Drell-Yan process	[25]	184	1.37
CDF Run-1 electron A_{ch}	[26]	11	0.81
CDF Run-2 electron A_{ch}	[27]	11	1.24
D0 Run-2 muon A_{ch}	[29]	9	0.92
LHCb 7 TeV 35 pb^{-1} W/Z $d\sigma/dy_\ell$	[31]	14	0.7
LHCb 7 TeV 35 pb^{-1} A_{ch}, $p_{T\ell} > 20$ GeV	[31]	5	1.19
D0 Run 2 Z rapidity	[32]	28	0.59
CDF Run 2 Z rapidity	[33]	29	1.64
CMS 7 TeV 4.7 fb^{-1}, muon A_{ch}	[34]	11	0.8
CMS 7 TeV 840 pb^{-1}, electron A_{ch}	[35]	11	0.87
ATLAS 7 TeV 35 pb^{-1} W/Z cross sections and A_{ch}	[36]	41	1.11
D0 Run-2 electron A_{ch} ($9.7 fb^{-1}$)	[39]	13	1.79
CDF Run-2 inclusive jet production	[40]	72	1.45
D0 Run-2 inclusive jet production	[41]	110	1.09
ATLAS 7 TeV 35 pb^{-1} incl. jet production	[42]	90	0.55
CMS 7 TeV 5 fb^{-1} incl. jet production	[43]	133	1.33

Experiments in the CT14 analysis

33 experiments; χ^2/N_{pt} = 3252/2947 =1.10

Red arrows indicate new data sets

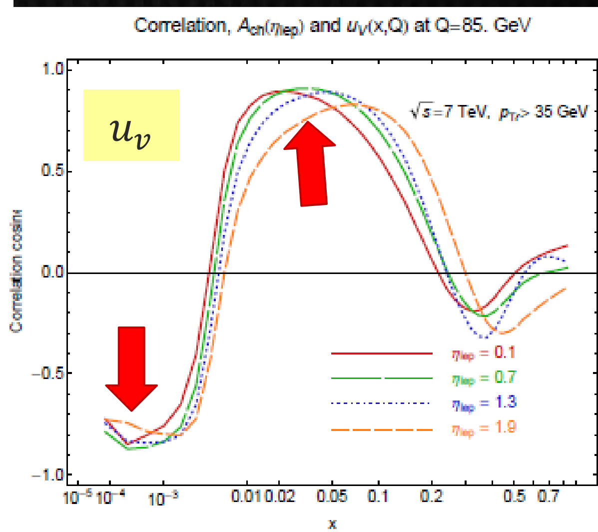
Experimental data set		N_{pt}	χ^2_e/N_{pt}
BCDMS F_2^p	[illegible]	[illegible]	[illegible]
BCDMS F_2^d	[illegible]	[illegible]	[illegible]
NMC F_2^d/F_2^p	[14]	123	1.08
NMC σ^p_{red}	[14]	201	1.85
CDHSW F_2^p	[15]	85	0.85
CDHSW F_3^p	[15]	96	0.83
CCFR F_2^p	[16]	69	1.02
CCFR xF_3^p	[17]	86	0.36
NuTeV $\nu\mu\mu$ SIDIS	[18]	38	0.62
NuTeV $\bar{\nu}\mu\mu$ SIDIS	[18]	33	1.18
CCFR $\nu\mu\mu$ SIDIS	[19]	40	0.72
CCFR $\bar{\nu}\mu\mu$ SIDIS	[19]	38	0.53
H1 σ_r^b	[20]	10	0.68
HERA charm production	[illegible]	47	1.26
HERA1 Combined NC and CC DIS	[22]	579	1.02
H1 F_L	[23]	9	1.92

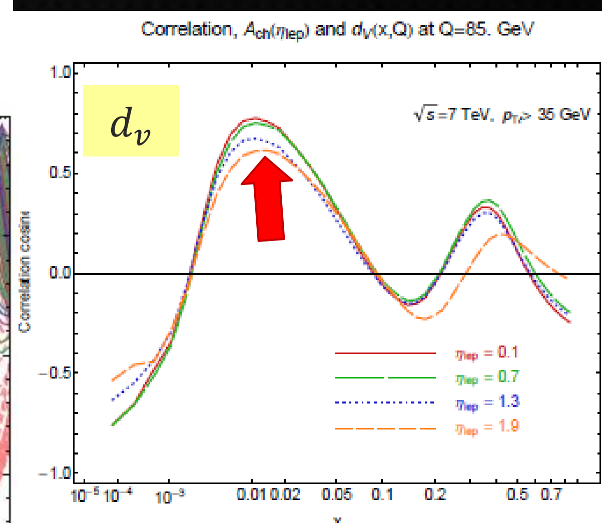
Experimental data set		N_{pt}	χ^2_e/N_{pt}
E605 Drell-Yan process	[24]	119	0.98
[illegible]	[25]	15	0.87
E866 Drell-Yan process	[25]	184	1.37
CDF Run-1 electron A_{ch}	[26]	11	0.81
CDF Run-2 electron A_{ch}	[27]	11	1.24
D0 Run-2 muon A_{ch}	[29]	9	0.92
LHCb 7 TeV 35 pb^{-1} W/Z $d\sigma/dy_\ell$	[illegible]	14	0.7
LHCb 7 TeV 35 pb^{-1} A_{ch}, $p_{T\ell} > 20$ GeV	[illegible]	5	1.19
D0 Run 2 Z rapidity	[32]	28	0.59
CDF Run 2 Z rapidity	[33]	29	1.64
CMS 7 TeV 4.7 fb^{-1}, muon A_{ch}	[illegible]	11	0.8
CMS 7 TeV 840 pb^{-1}, electron A_{ch}	[illegible]	11	0.87
ATLAS 7 TeV 35 pb^{-1} W/Z cross sections and A[illegible]	[illegible]	41	1.11
D0 Run-2 electron A_{ch} $(9.7 fb^{-1})$	[illegible]	13	1.79
CDF Run-2 inclusive jet production	[40]	72	1.45
D0 Run-2 inclusive jet production	[41]	110	1.09
ATLAS 7 TeV 35 pb^{-1} incl. jet production	[illegible]	90	0.55
CMS 7 TeV 5 fb^{-1} incl. jet production	[illegible]	133	1.33

Role of the LHC Run-1 data

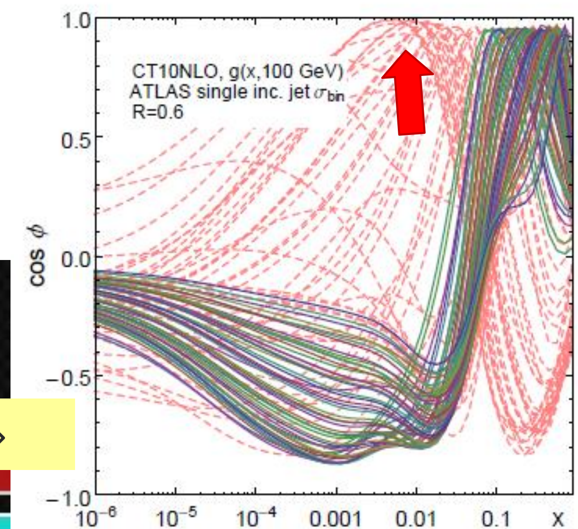
- LHC Run-1 measurements start to impose some unique constraints on parton flavor composition (on g, u_v and d_v, $\frac{s+\bar{s}}{\bar{u}+\bar{d}}$,…). This can be demonstrated by studying correlations of LHC observables with PDFs (*arXiv:0802.0007*).
- CT10 included ad hoc restrictions on the relevant combinations of PDFs.
- In CT14, these constraints are relaxed,

the combinations are constrained by the LHC data.





ATLAS inc. jets and gluon ⇒



↑↑
CMS charge asymmetry

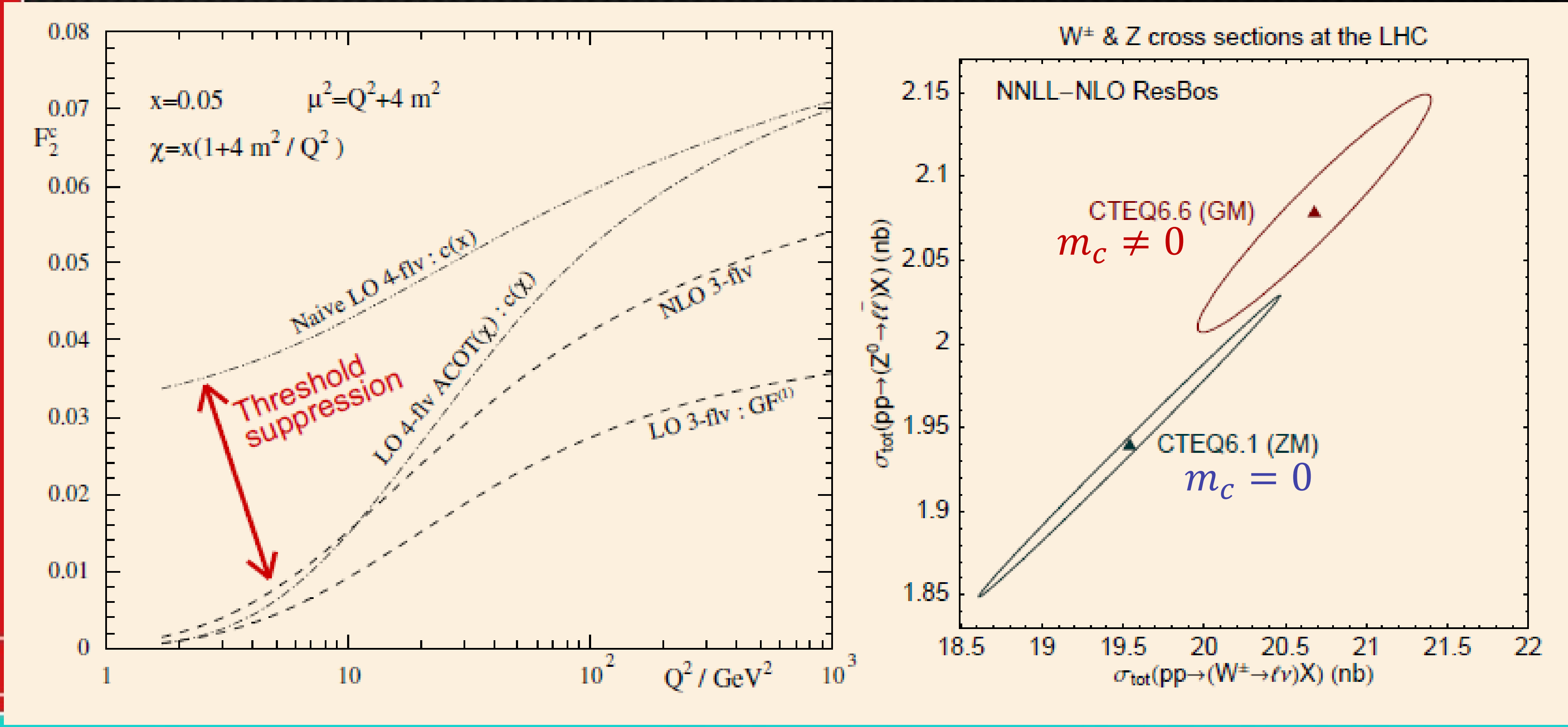
CT14: theoretical treatment

- **NNLO theory with massive heavy quarks** for neutral-current DIS, DY, W, Z production; **benchmarked NLO** for charged- current DIS and jet production
- PDFs are parametrized by **new functional forms** at **Q_0=1.3 GeV.**
- Assume central $\boldsymbol{\alpha_S(Mz) = 0.118}$, but also provide PDFs for other α_S.
- Use pole mass $m_c = 1.3$ GeV and $m_b = 4.75$ GeV
- Correlated systematic errors are included in most experiments.
- PDF uncertainties are estimated with two methods, based on Hessian matrix and Lagrange multipliers

NNLO cross sections in a general-mass scheme

NC DIS and DY cross sections are evaluated at NNLO in a general-mass scheme *(Guzzi, Lai, P.N., Yuan, **arXiv:1108.5112**)*.

Dependence on m_c in DIS propagates into predicted W/Z cross sections at the LHC via $u(x,Q)$, $d(x,Q)$ *(CTEQ6.5, **hep-ph/0611254**)*



Benchmark comparisons of theoretical cross sections

Key cross sections in CT14, MMHT, NNPDF3.0 analyses were benchmarked against cross sections from other groups.

This is important. Some changes in $g(x, Q)$ and $s(x, Q)$ in the CT14 ensemble are caused by the improved numerical calculation of CC DIS cross sections and NLO jet cross sections

Benchmark comparisons of PDF analyses

1. J. Gao et al., MEKS: a program for computation of inclusive jet cross sections at hadron colliders , arXiv:1207.0513

Codes for NLO jet production

2. R. Ball et al., Parton Distribution benchmarking with LHC data, arXiv:1211.5142

(N)NLO LHC cross sections

3. S. Alekhin et al., ABM11 PDFs and the cross section benchmarks in NNLO, arXiv:1302.1516; The ABM parton distributions tuned to LHC data; arXiv:1310.3059

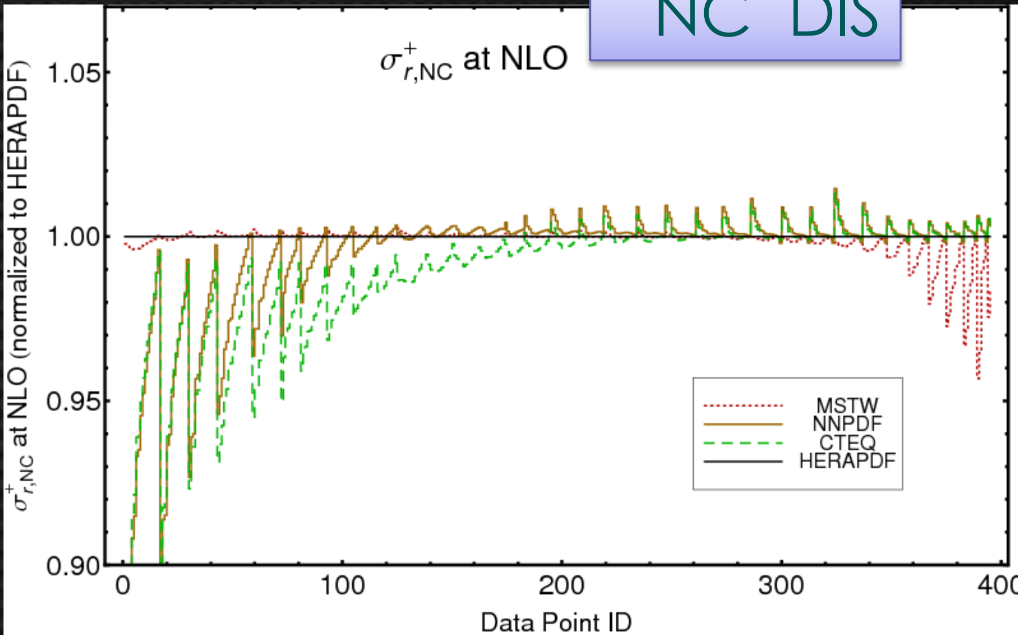
W/Z, $t\bar{t}$,...

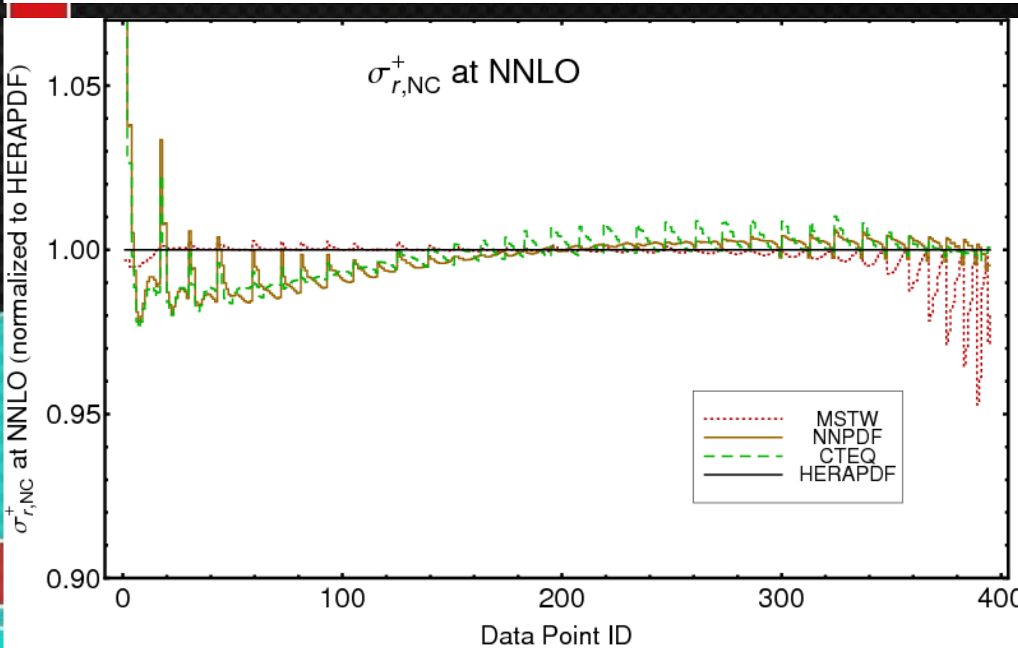
4. A.Cooper-Sarkar et al., PDF dependence of the Higgs production cross section in gluon fusion from HERA data, 2013 Les Houches Proceedings, arXiv:1405.1067, p. 37

NC DIS; CC DIS (in progress)

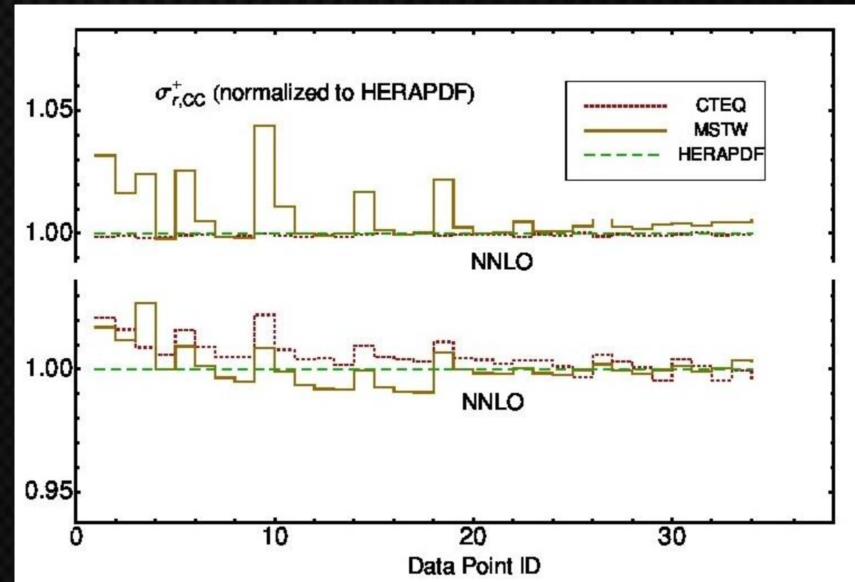
5. S. Forte and J. Rojo, Dataset sensitivity of the gg->H cross-section in the NNPDF analysis, arXiv:1405.1067, p. 56

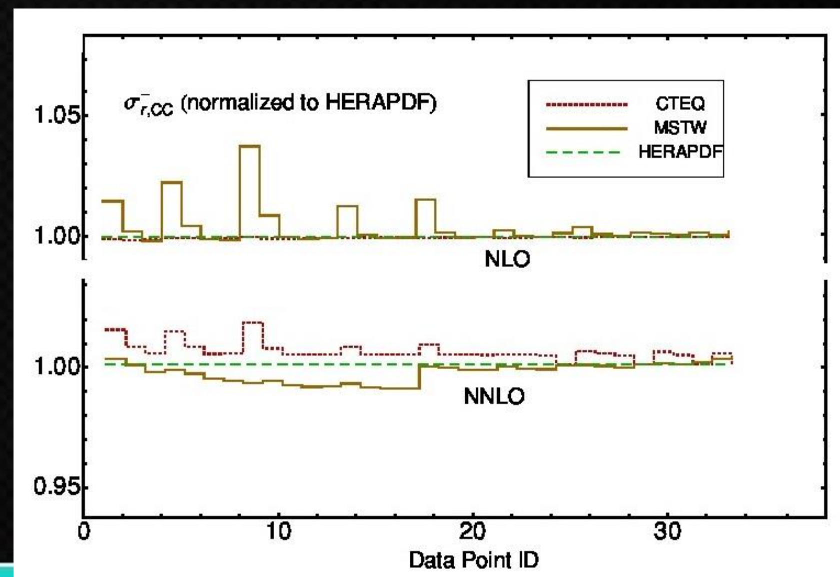
NC DIS



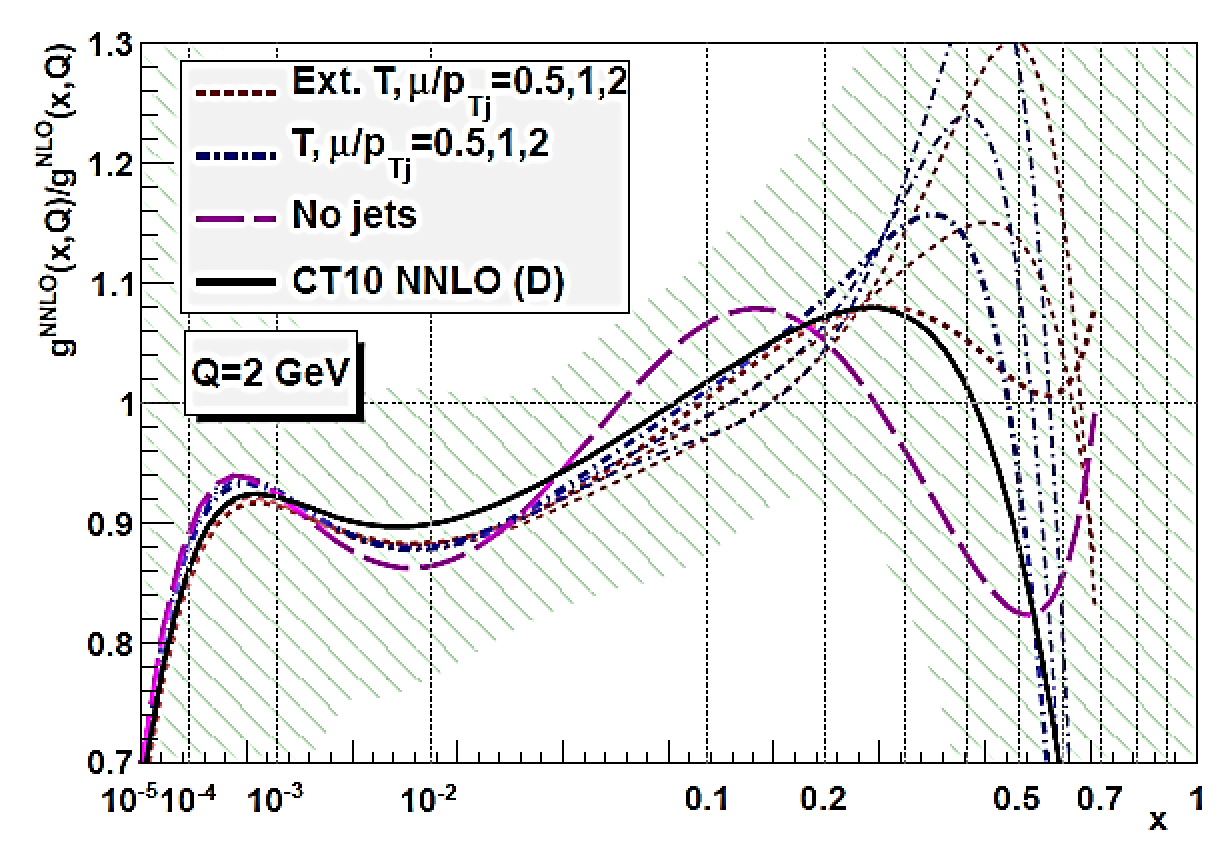


CC DIS (preliminary)





Role of correlated systematic errors



One of the objectives of the CT10 NNLO study was to investigate the role of correlated systematic errors and theoretical uncertainties

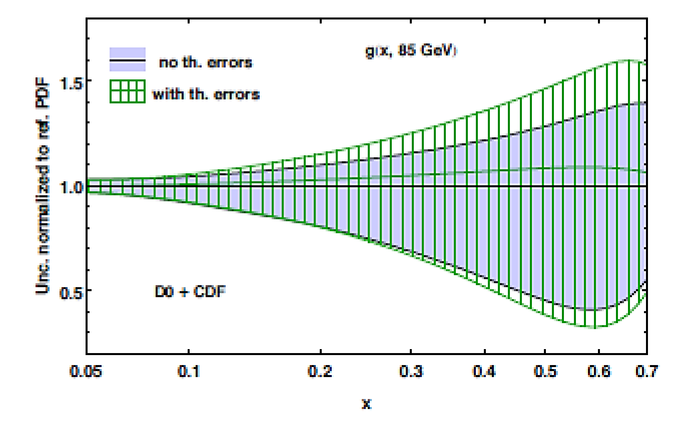
For example, the large-x g(x,Q) depends on the implementation of corr. syst. errors in Tevatron jet experiments, as well as on the assumptions about QCD scales. The CT10 NNLO gluon error sets are constructed so as to span the full range of uncertainty due to experimental errors, corr. syst. errors, and various scale choices. Similarly in CT14.

Residual uncertainty in **NLO** cross sections

CC DIS and jet production hard cross sections are still computed at NLO

In the CT14 study, we estimate the theoretical uncertainty in the PDFs from the QCD scale dependence and normalization variations in the jet cross sections due to the missing NNLO contributions.

The NLO scale uncertainty in these cases is small compared to the experimental uncertainty.



- About 20% increase of the gluon PDF uncertainty in large-x region and 10% in the Higgs mass region, for a fit with only Tevatron jet data included (+DIS+…)
- Similar results are observed when also including the LHC jet data or using different criteria for the determination of PDF uncertainties

Jun Gao, 2014

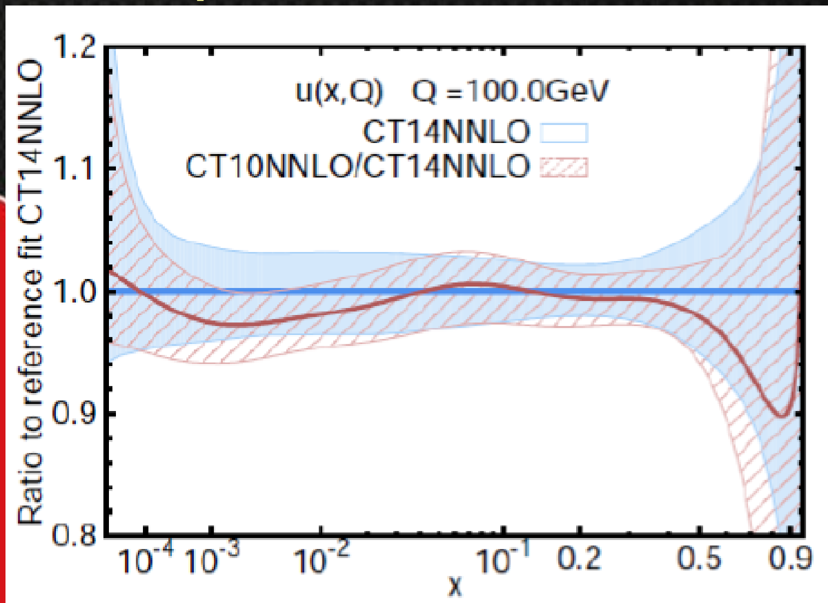
CT14: new parametrization forms

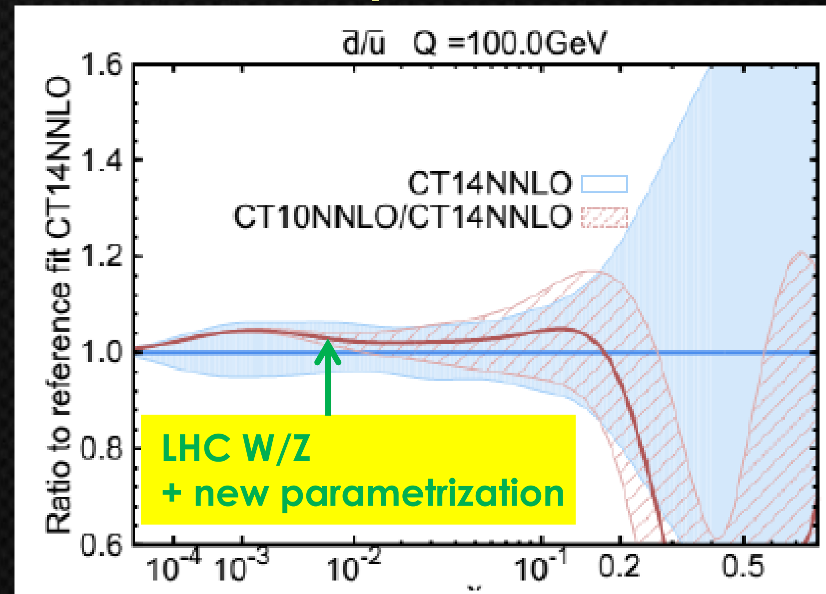
- CT14 relaxes restrictions on several PDF combinations that were enforced in CT10. [These combinations were not constrained by the pre-LHC data.]
  - The assumptions $\frac{\bar{d}(x,Q_0)}{\bar{u}(x,Q_0)} \to 1$, $u_v(x,Q_0) \sim d_v(x,Q_0) \propto x^{A_{1v}}$ with $A_{1v} \approx -\frac{1}{2}$ at $x < 10^{-3}$ are relaxed once LHC W/Z data are included
  - CT14 parametrization for $s(x,Q)$ includes extra parameters
- CT14 fits have 28 free parameters
- In general, $\mathrm{f_a}(\mathrm{x}, \mathrm{Q_0}) = \mathrm{Ax^{a_1}}(1-\mathrm{x})^{\mathrm{a_2}} \mathrm{P_a}(\mathrm{x})$
- CT10 assumed $P_a(x) = \exp(a_0 + a_3\sqrt{x} + a_4 x + a_5\, x^2)$
  - exponential form conveniently enforces positive definite behavior
  - but power law behaviors from a_1 and a_2 may not dominate
- In CT14, $\mathrm{P_a}(\mathrm{x}) = \mathrm{G_a}(\mathrm{x})\mathrm{F_a}(\mathrm{z})$, where $\mathrm{G_a}(\mathrm{x})$ is a smooth factor
  - $\mathrm{z} = 1 - 1(1-\sqrt{\mathrm{x}})^{\mathrm{a_3}}$ preserves desired Regge-like behavior at low x and high x (with a_3>0)
- Express $F_a(z)$ as a linear combination of Bernstein polynomials:

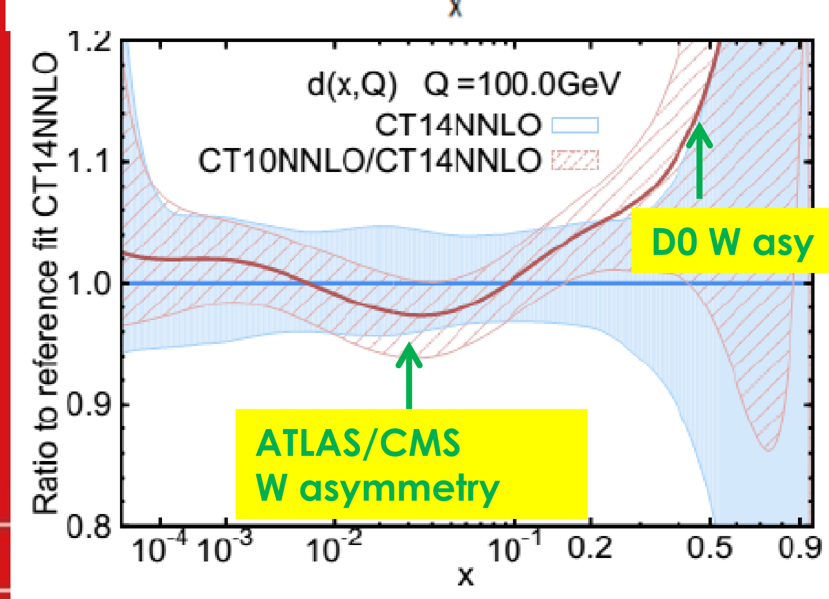
$$z^4, 4z^3(1-z), 6z^2(1-z)^2, 4z(1-z)^3, (1-z)^4$$

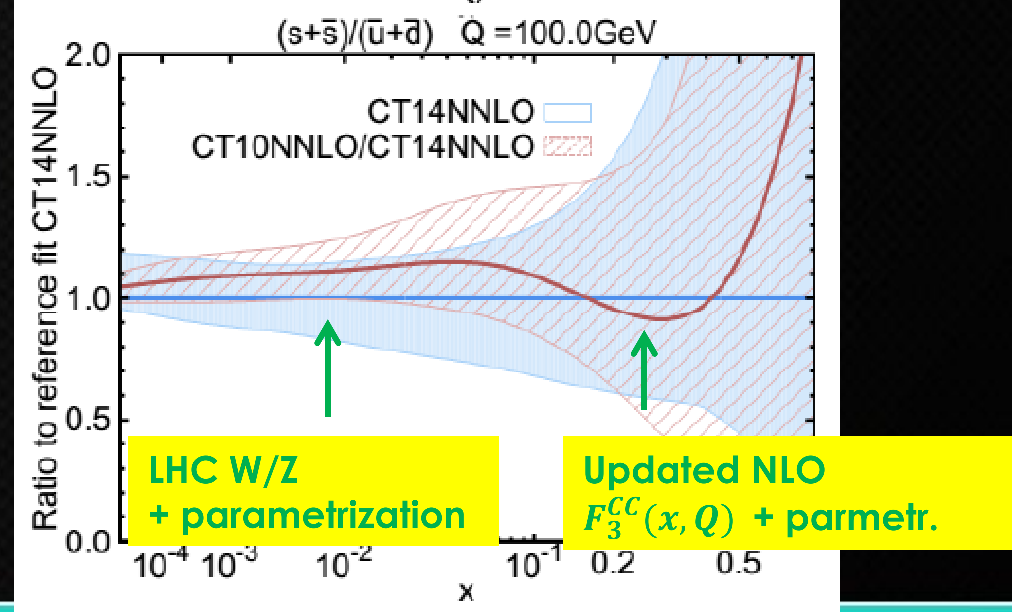
  - each basis polynomial has a single peak, with peaks at different values of z; reduces correlations among parameters

Compare CT14 and CT10 quark PDFs

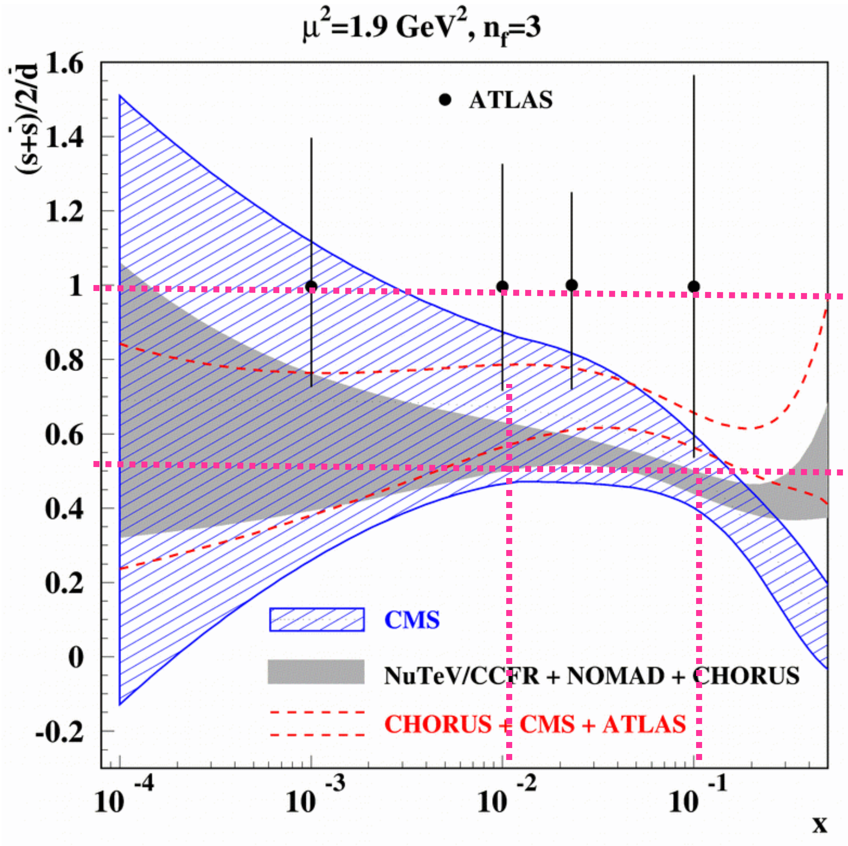


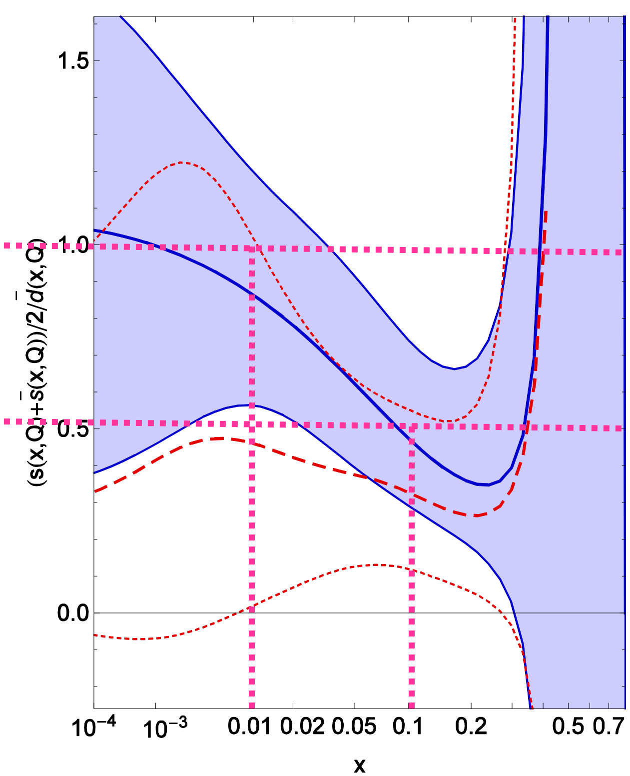






Strangeness PDF from CT14 and ABM

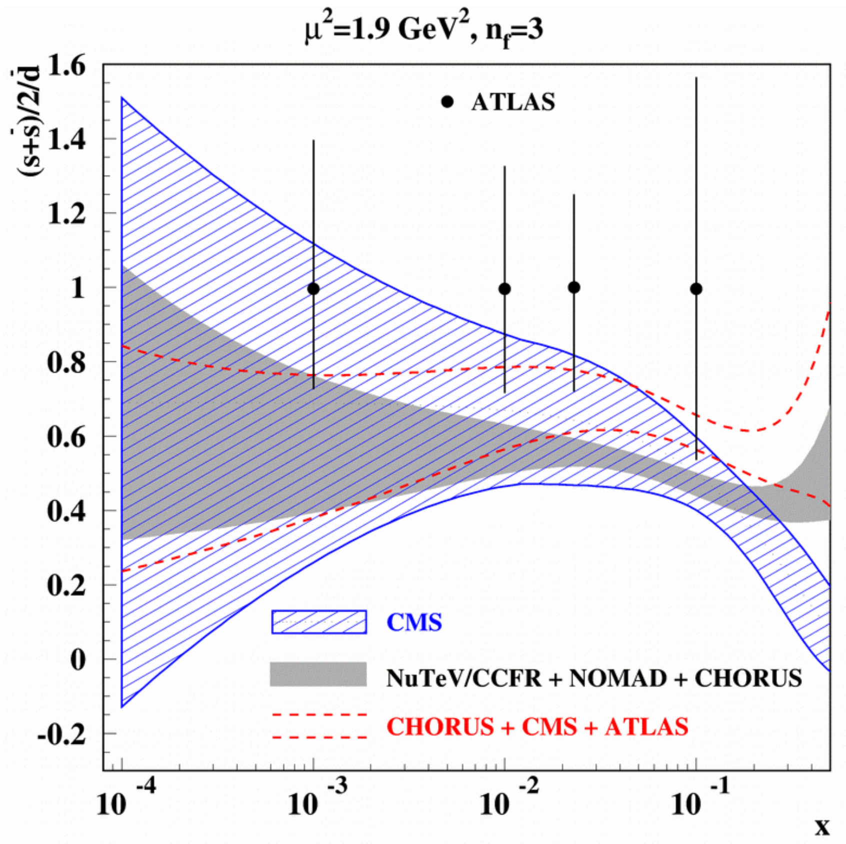




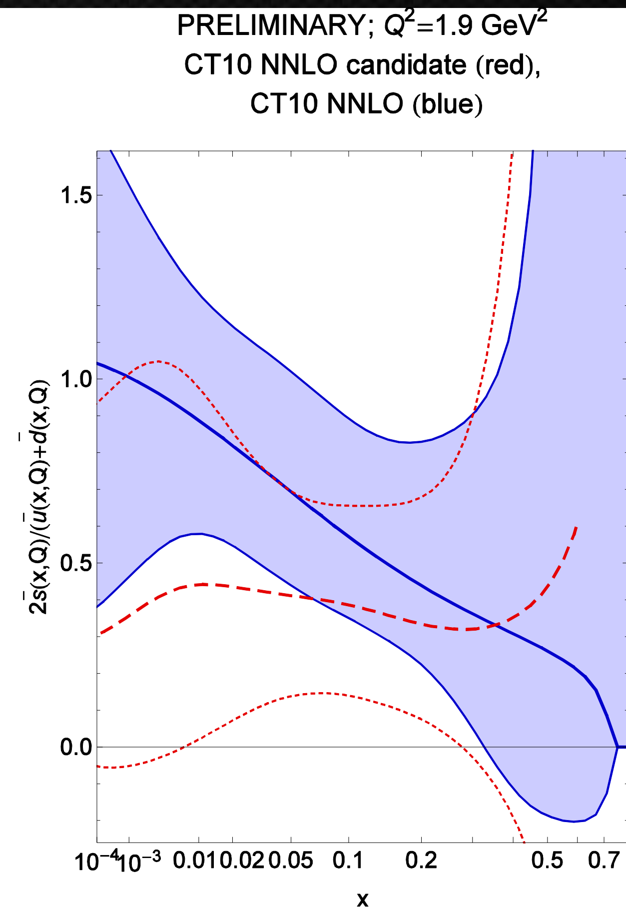
Alekhin et al., hep-ph/1404.6469
68%c.l. errors, $\Delta\chi^2 = 1$

$(s+\bar{s})/(2\bar{d})$.
90% c.l. errors

Strangeness PDF from ABM and CT14



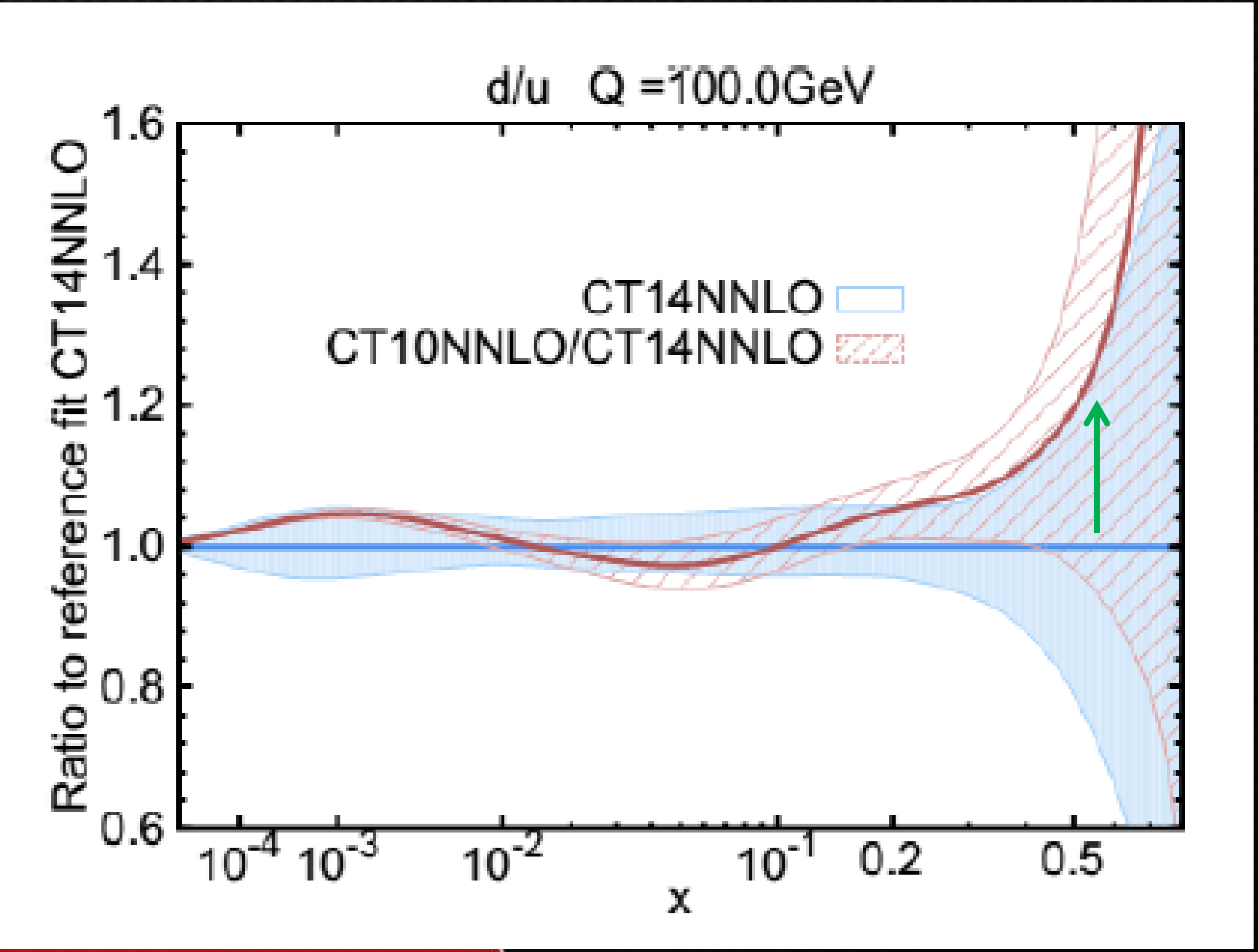
Alekhin et al., hep-ph/1404.6469



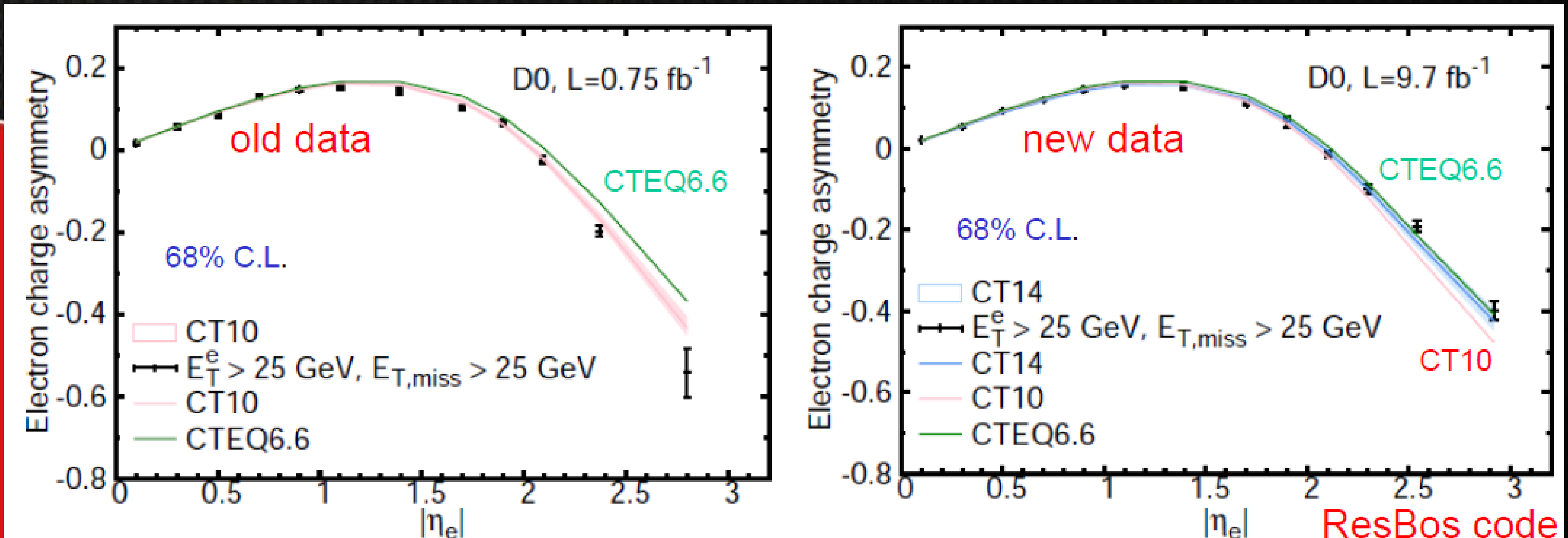
$(s+\bar{s})/(\bar{u}+\bar{d})$

CT14 vs. CT10: $d(x,Q)/u(x,Q)$

d/u is reduced in CT14 at $x \approx 0.2$, compared to CT10, as a result of using updated D0 Run-2 charge asymmetry data in the electron channel

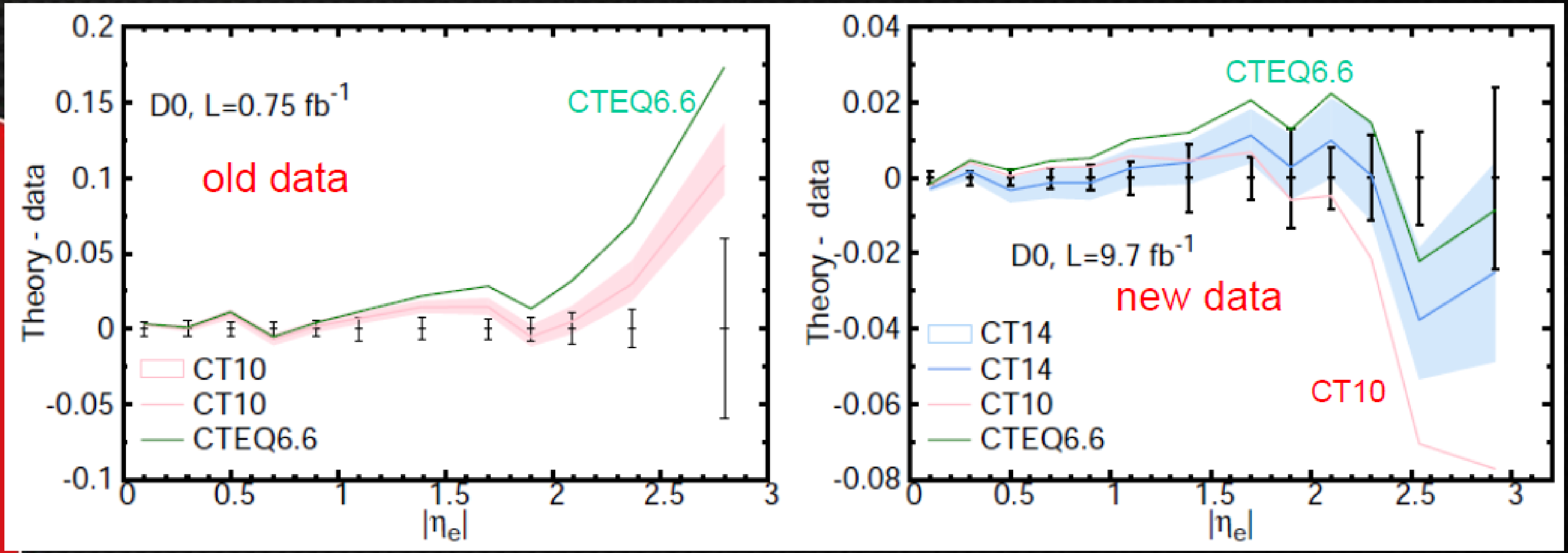


D0 Run-2 electron charge asymmetry in CT10 and CT14



- CT10 was fitted to the old $(0.75\ fb^{-1})$ D0 A_{ch}^{ele} data ⇒ harder d/u than in CTEQ6.6
- CT14 is fitted to the $(9.7\ fb^{-1})$ D0 A_{ch}^{ele} data. It prefers predictions that are closer to CTEQ6.6

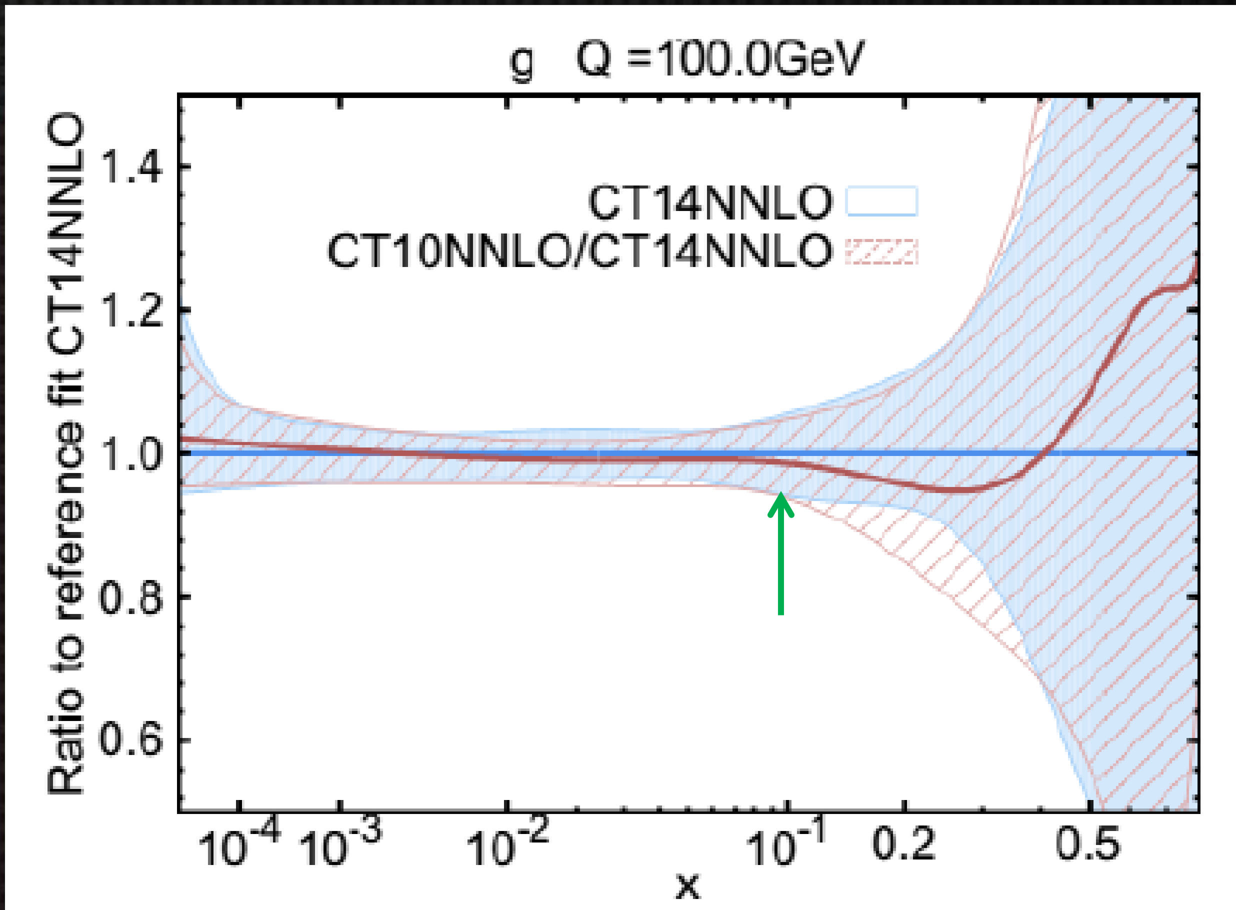
D0 Run-2 electron charge asymmetry in CT10 and CT14



- CT10 was fitted to the old $(0.75\ fb^{-1})$ D0 A_{ch}^{ele} data ⇒ harder d/u than in CTEQ6.6
- CT14 is fitted to the $(9.7\ fb^{-1})$ D0 A_{ch}^{ele} data. It prefers predictions that are closer to CTEQ6.6

CT14 vs. CT10: the gluon PDF

$g(x, Q)$ is slightly higher in CT14 at $x \sim 0.05$ because of several effects.



CT14 Higgs cross sections increase compared to CT10 by about 1-2%

CT14 among other world PDFs

- Comparison
- Combination

Recent NNLO PDF ensembles
(other than CT)

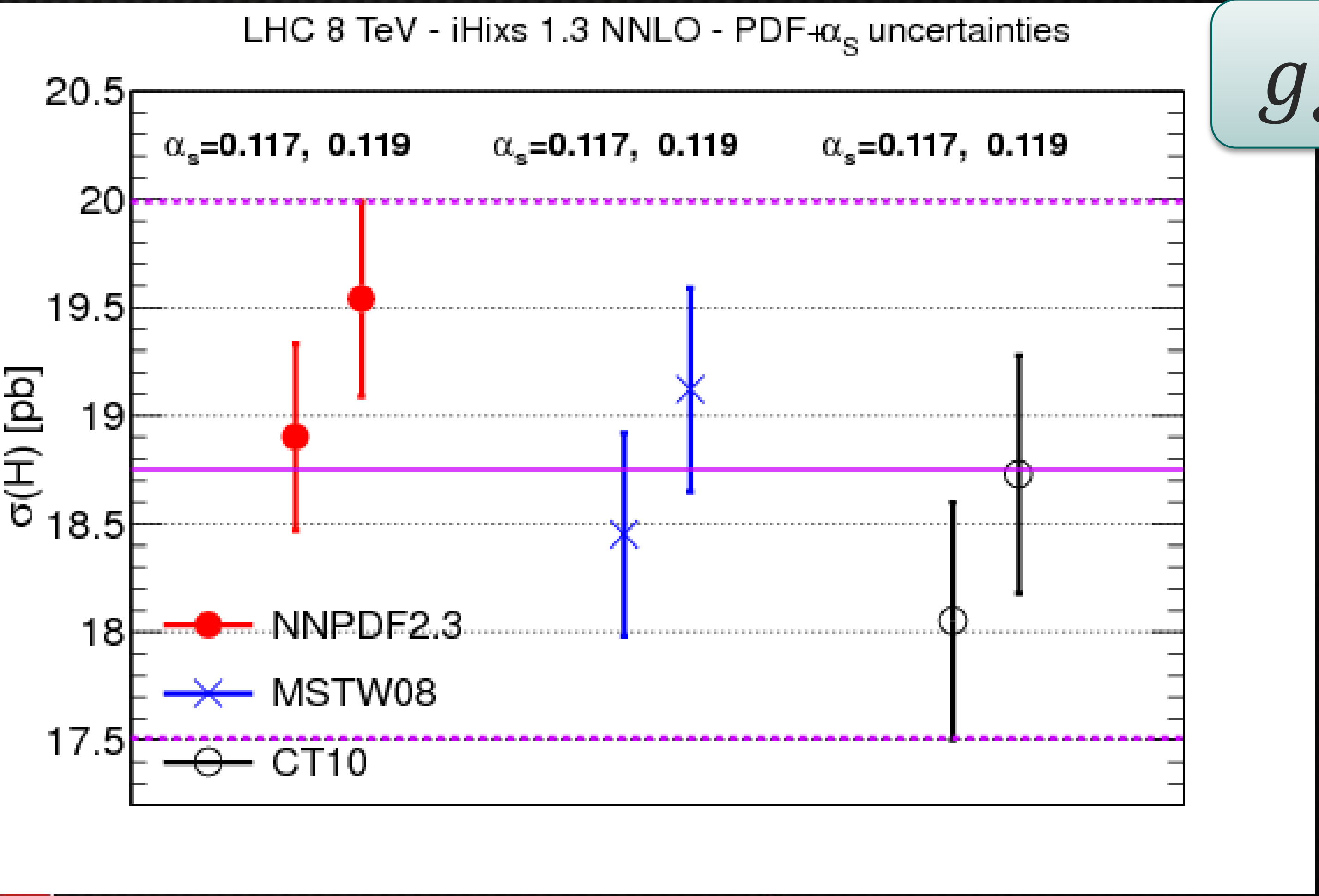
A. Global PDF analyses

- **MMHT'14:** Harland-Lang, Martin, Motylinski, Thorne, *arXiv:1412.3989*
- **Neural-network PDF 3.0:** R. Ball et al. *arXiv:1410.8849*

B. Non-global PDF analyses

- **ABM:** Alekhin, Blümlein, Moch, *arXiv:1202.2281,….*
  Based on compatible DIS, DY, W+c measurements
- **HERAPDF 1.5 /2.0:** *ZEUS-prel-11-002, …*
  Based on DIS and jet production at HERA

2014: the typical NNLO PDF+α_s uncertainty is larger than 1%

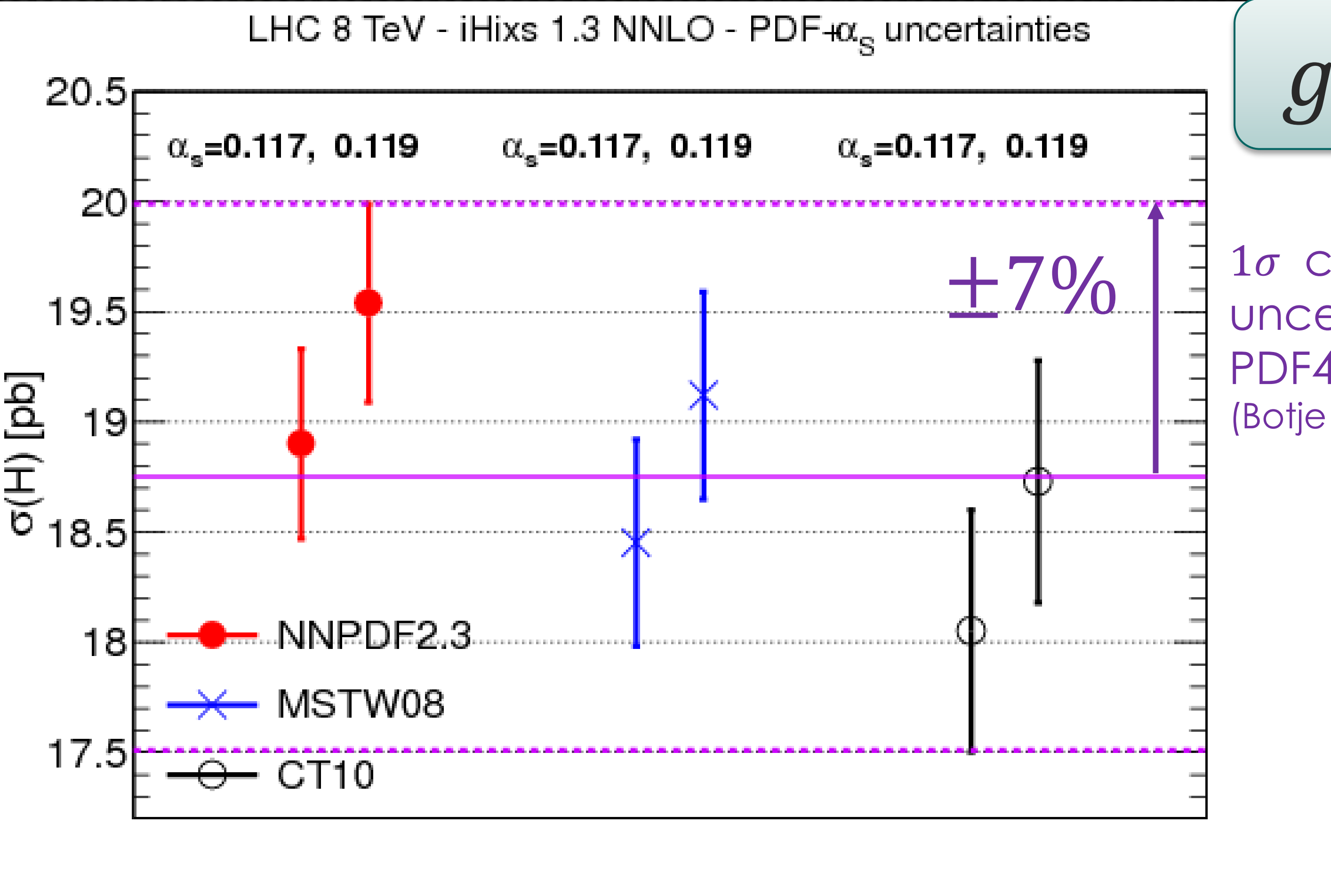


$gg \rightarrow H^0_{SM}$

R. Ball et al., Parton Distribution benchmarking with LHC data **arXiv:1211.5142**

2014: the typical NNLO PDF+α_s uncertainty is larger than 1%

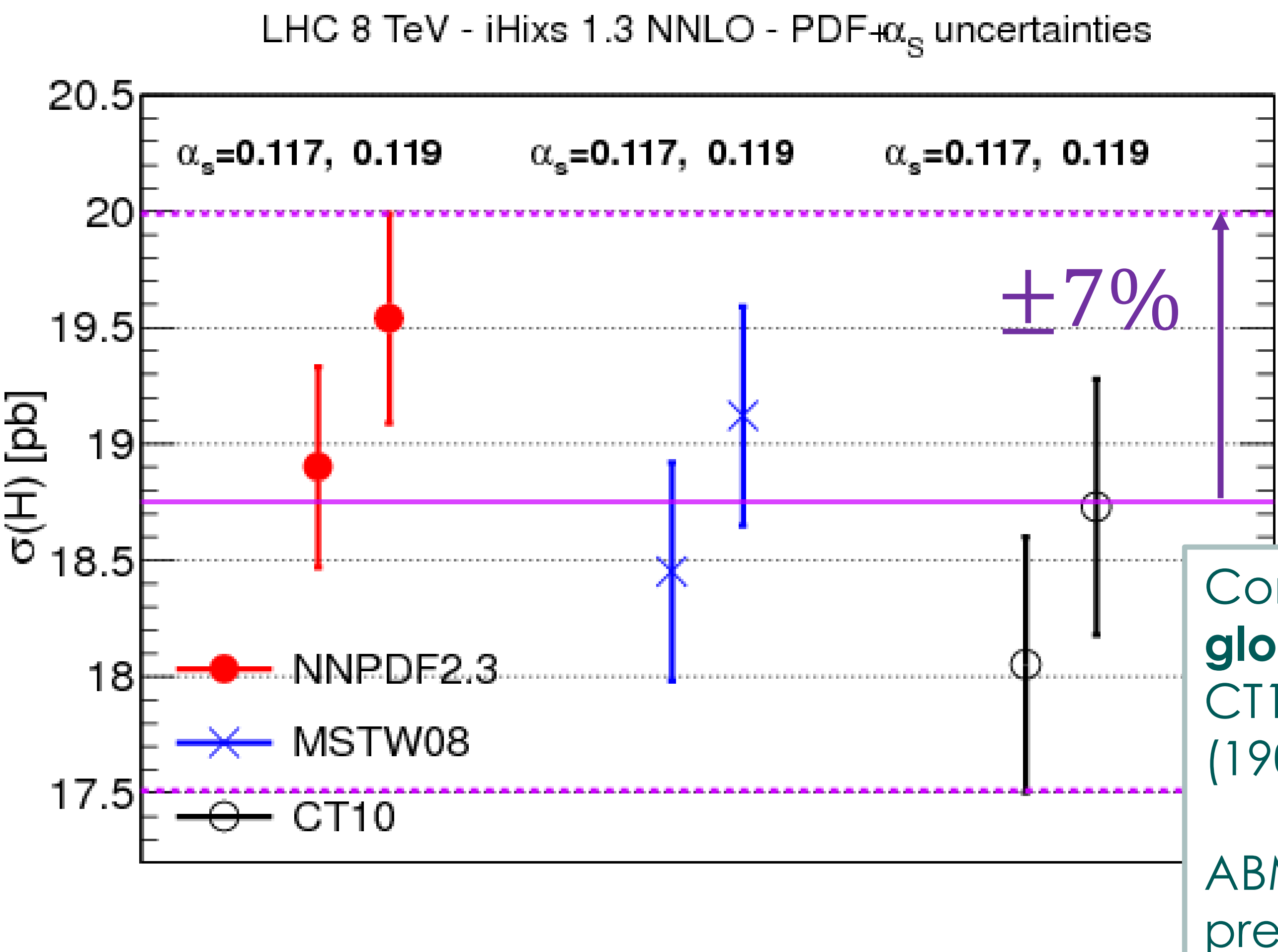
$$gg \to H^0_{SM}$$



1σ combined PDF+α_s uncertainty, using PDF4LHC convention (Botje et al., arxiv:1101.0538)

R. Ball et al., Parton Distribution benchmarking with LHC data **arXiv:1211.5142**

2014: the typical NNLO PDF+α_s uncertainty is larger than 1%



$gg \to H^0_{SM}$

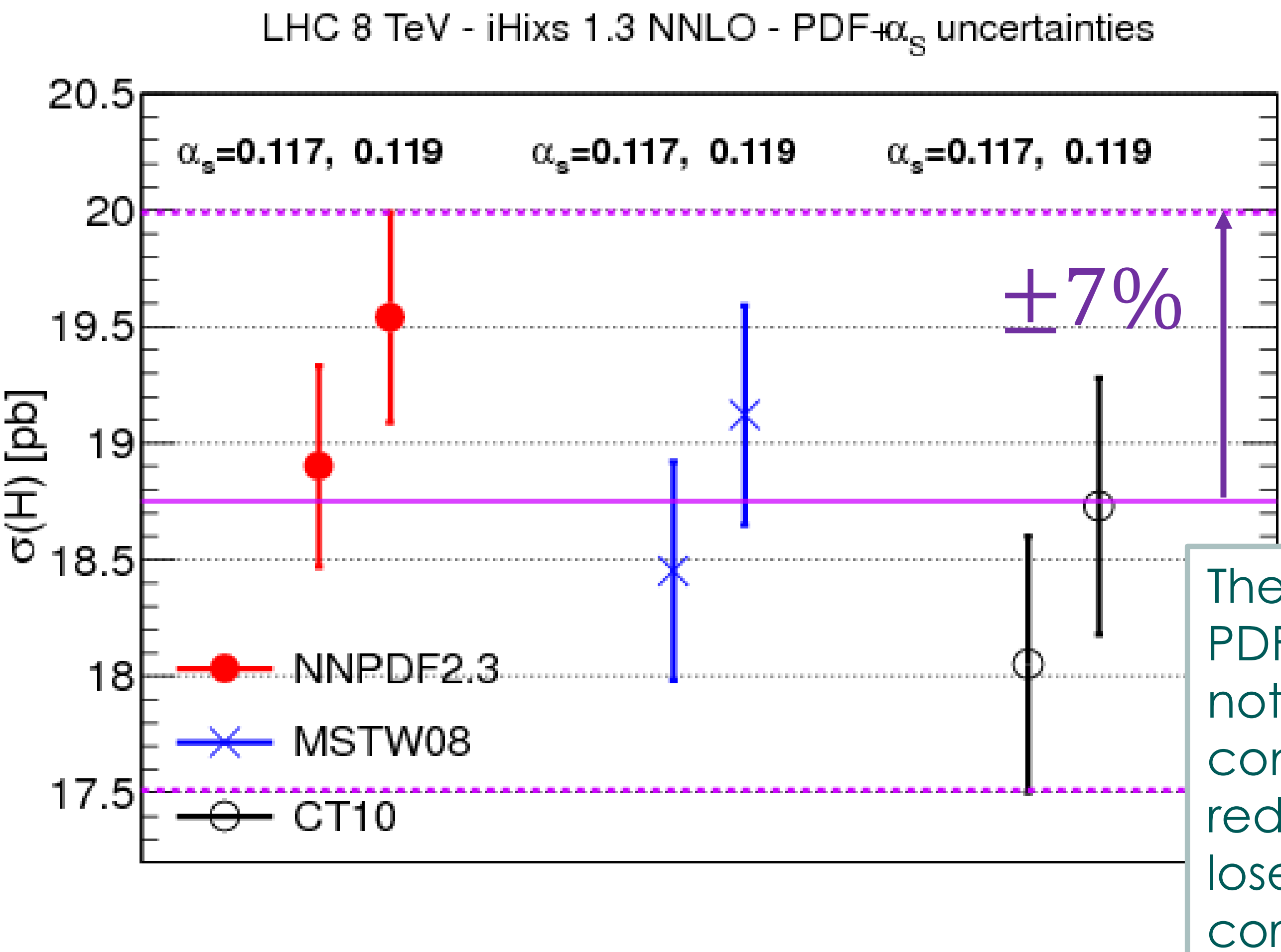
1σ combined PDF+α_s uncertainty, using PDF4LHC convention (Botje et al., arxiv:1101.0538)

Combination of three **global** PDF ensembles CT10, MSTW08, NNPDF2.3 (190 error sets)

ABM, CJ, GJR, HERA PDF predictions not included

R. Ball et al., Parton Distribution benchmarking with LHC data **arXiv:1211.5142**

2014: the typical NNLO PDF+α_s uncertainty is larger than 1%



$gg \to H^0_{SM}$

1σ combined PDF+α_s uncertainty, using PDF4LHC convention (Botje et al., arxiv:1101.0538)

The 2012 combination of PDF+α_s uncertainties is not efficient: requires to compute $\sigma(H)$ for many redundant PDF error sets, loses PDF-driven correlations

R. Ball et al., Parton Distribution benchmarking with LHC data **arXiv:1211.5142**

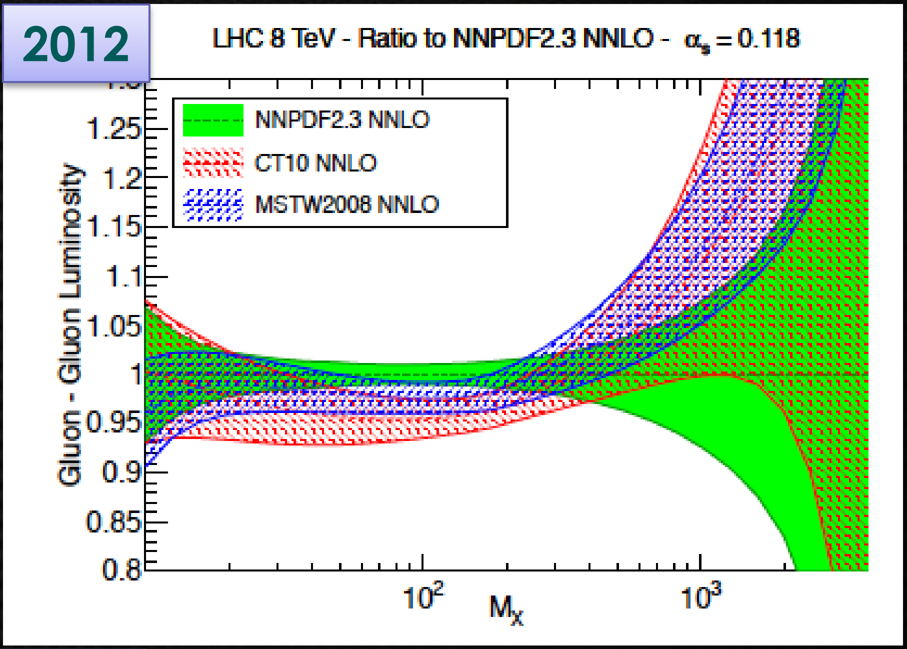
Fast forward to 2015

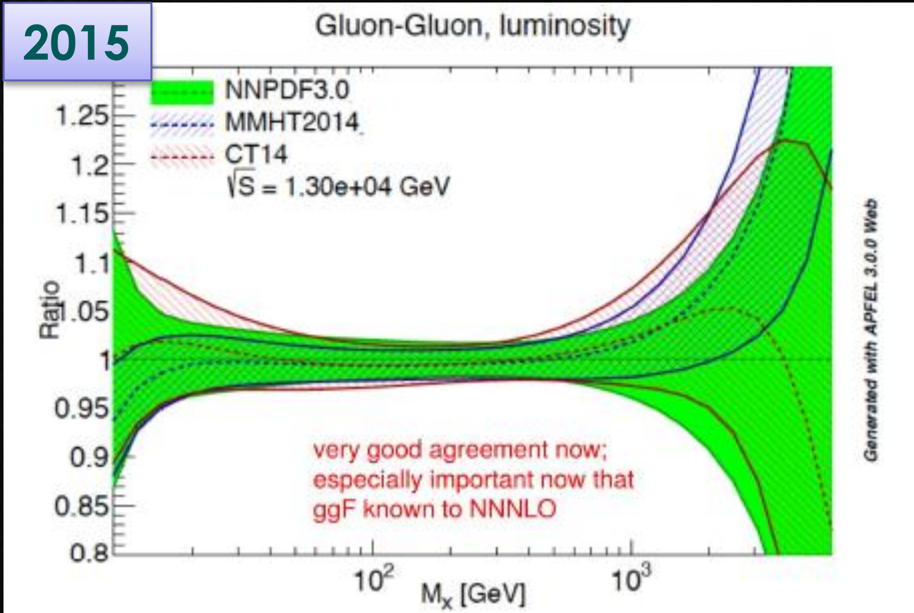
- Agreement between CT14, MMHT14, NNPDF3.0 improved for most flavors. Now very good agreement between $gg \to H$ cross sections, VBF, for many other observables

$\sigma(gg \to H^0)$ at NNLO

	CT14	MMHT2014	NNPDF3.0
8 TeV	18.66 pb -2.2% +2.0%	18.65 pb -1.9% +1.4%	18.77 pb -1.8% +1.8%
13 TeV	42.68 pb -2.4% +2.0%	42.70 pb -1.8% +1.3%	42.97 pb -1.9% +1.9%

J.Huston, PDF4LHC, April 2015





Why NNLO PDFs of the new generation are in better agreement than ever

Since 2012, PDF analysis groups carried out a series of benchmarking exercises for key processes of DIS and jet production in PDF fits

Methodologies of all groups were cross-validated and improved.

Now that PDFs are in good agreement, we can combine them by more efficient methods than the 2010 PDF4LHC prescription

In *arXiv:1401.0013*, Jun Gao and I proposed a method for such combination, based on meta-parametrizations of parton distribution functions

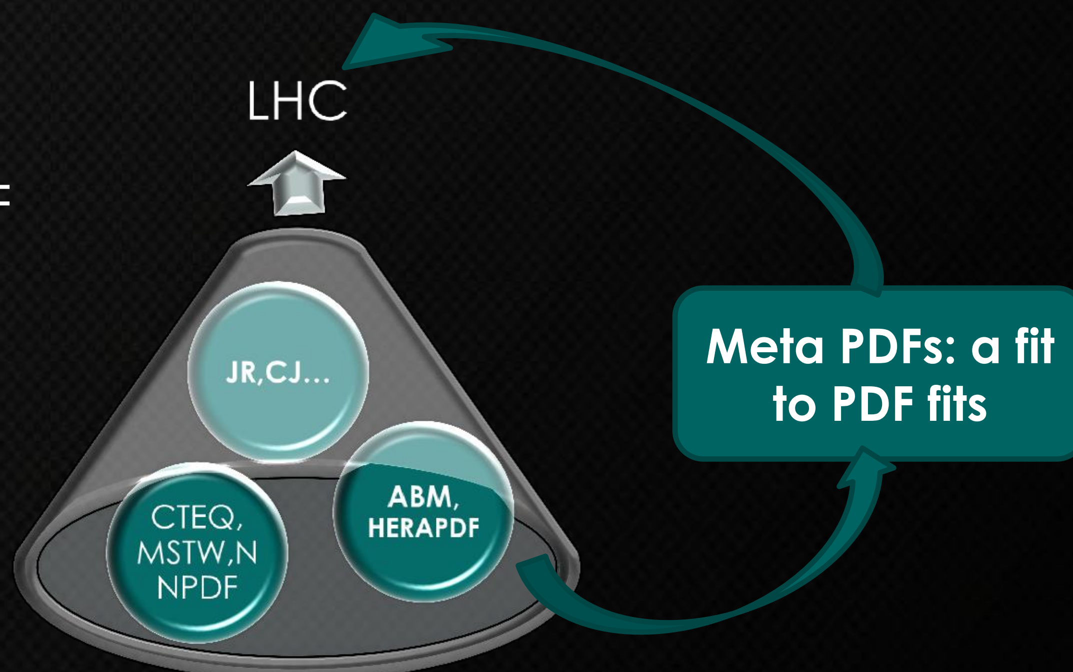
A META1.0 PDF ensemble including CT10, MSTW'08, and NNPDF2.3 was released.

We *(Gao, Huston, P.N.)* just finished the META 2.0 ensemble with advanced properties, including CT14, MMHT'14, and NNPDF3.0.

What is the PDF meta-analysis?

A meta-analysis compares and combines LHC predictions based on several PDF ensembles. It serves the same purpose as the PDF4LHC prescription. It combines the PDFs directly in space of PDF parameters. It can significantly reduce the number of error PDF sets needed for computing PDF uncertainties and PDF-induced correlations.

The number of input PDF ensembles that can be combined is almost unlimited



META1.0 PDFs: A working example of a meta-analysis

See arXiv:1401.0013 for details

1. Select the input PDF ensembles (CT, MSTW, NNPDF…)
2. Fit each PDF error set in the input ensembles by a common functional form (“**a meta-parametrization**”)
3. Generate many Monte-Carlo replicas from meta-parametrizations of each set to investigate the probability distribution on the ensemble of all meta-parametrizations (as in Thorne, Watt, 1205.4024)
4. Construct a final ensemble of 68% c.l. **Hessian eigenvector sets** to propagate the PDF uncertainty from the combined ensemble of replicated meta-parametrizations into LHC predictions.

Only in the META set (item 2)

Only in the META set (item 4)

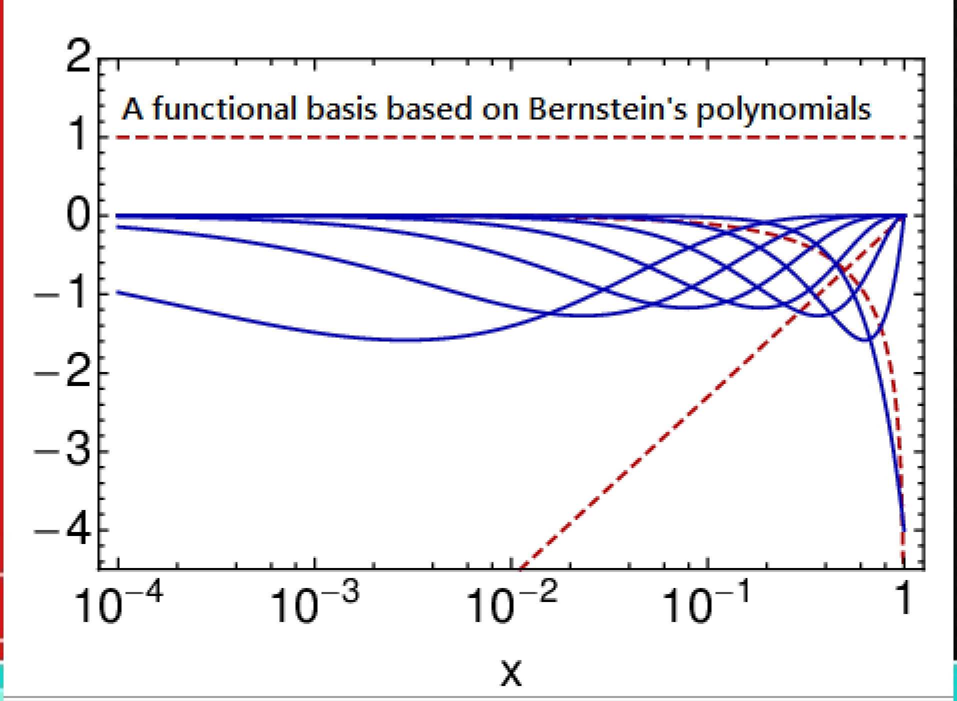
META PDFs: functional forms

New

v. 1.0: Chebyshev polynomials **(Pumplin, 0909.5176, Glazov, et al., 1009.6170, Martin, et al., 1211.1215)**

v 2.0: Bernstein polynomials ⇒ more faithful reproduction of the full ensemble of MC replicas

The initial scale of DGLAP evolution is Q_0=8 GeV.



The meta-parametrizations are fitted to the input PDFs at $x > 3 \cdot 10^{-5}$ for all flavors ; $x < 0.4$ for $\bar{u}$, $\bar{d}$; $x < 0.3$ for s, $\bar{s}$; and $x < 0.8$ for other flavors. PDFs outside these x regions are determined entirely by extrapolation.

The logic behind the META approach

Emphasize simplicity and intuition

When expressed as the meta –parametrizations, PDF functions can be combined by averaging their meta-parameter values

Standard error propagation is more feasible, e.g., to treat the meta-parameters as discrete data in the linear (Gaussian) approximation for small variations

The Hessian analysis can be applied to the combination of all input ensembles in order to optimize uncertainties and eliminate "noise"

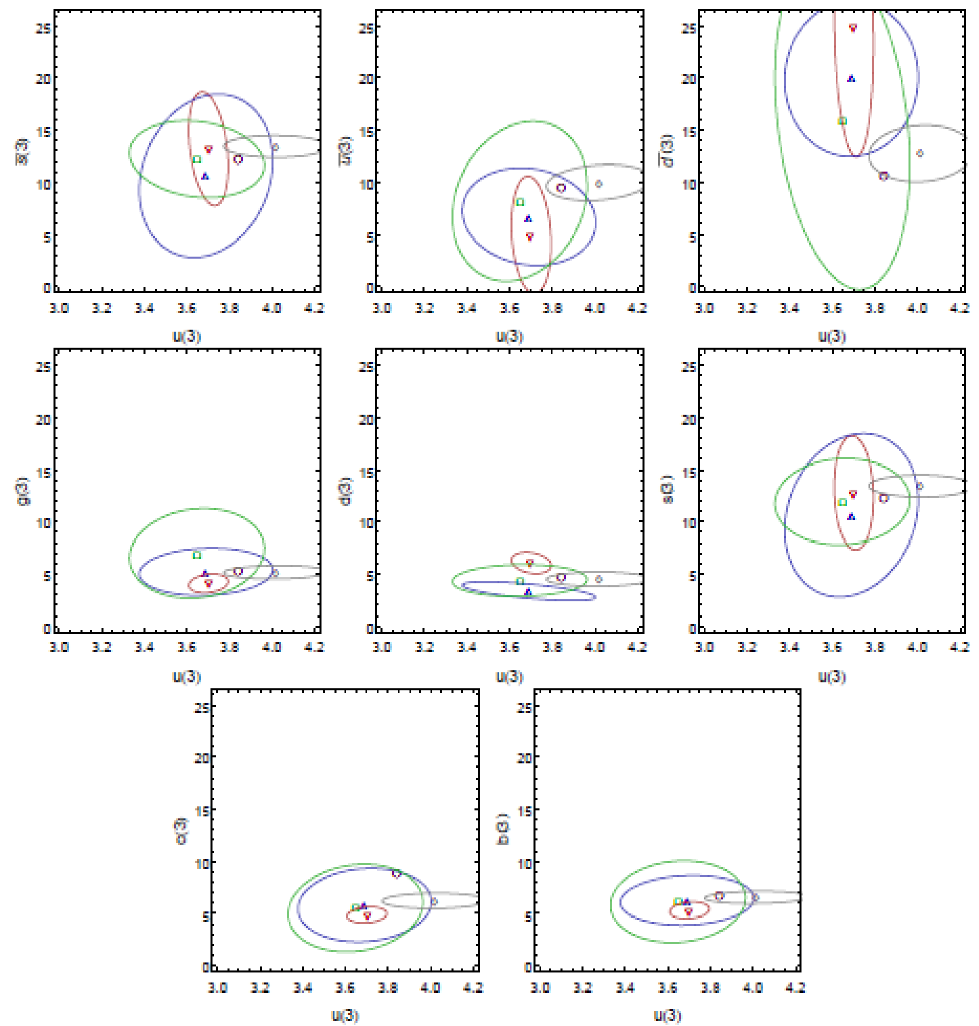


Figure 10: Fitted PDF parameters and 90% c.l. ellipses for CT10 (blue up triangle), MSTW08 (red down triangle), NNPDF2.3 (green square), HERAPDF1.5 (gray diamond) and ABM11 (magenta circle).

Meta-parameters of 5 sets and META PDFs

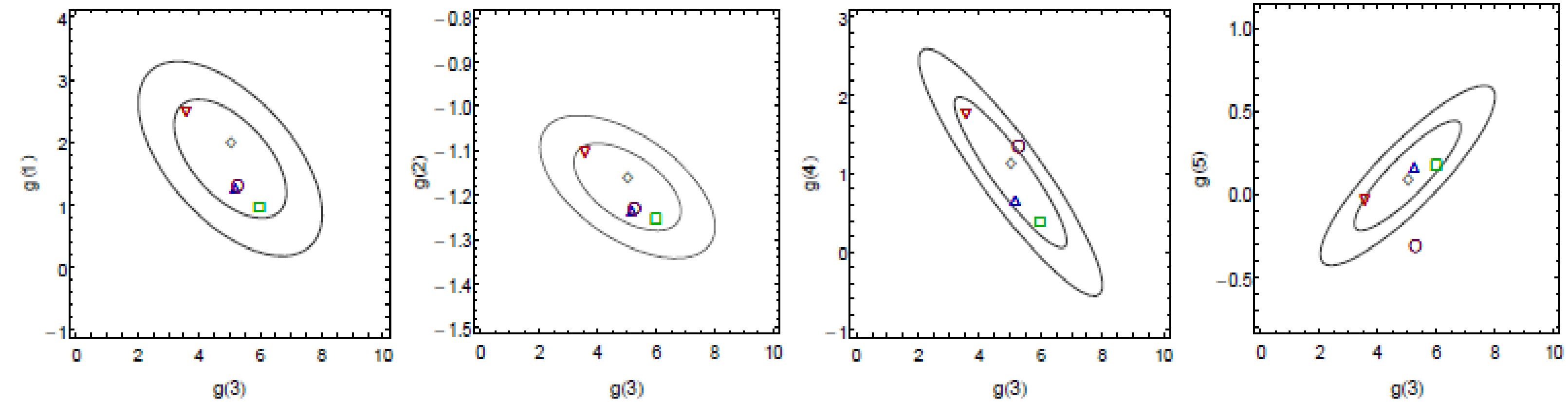


Figure 16: Comparison of META PDF confidence intervals with central NNLO PDFs of the input PDF ensembles in space of meta-parameters a_{1-5} for the gluon PDF. Up triangle, down triangle, square, diamond, and circle correspond to the best-fit PDFs from CT10, MSTW, NNPDF, HERAPDF, and ABM respectively. The ellipses correspond to 68 and 90% c.l. ellipses of META PDFs.

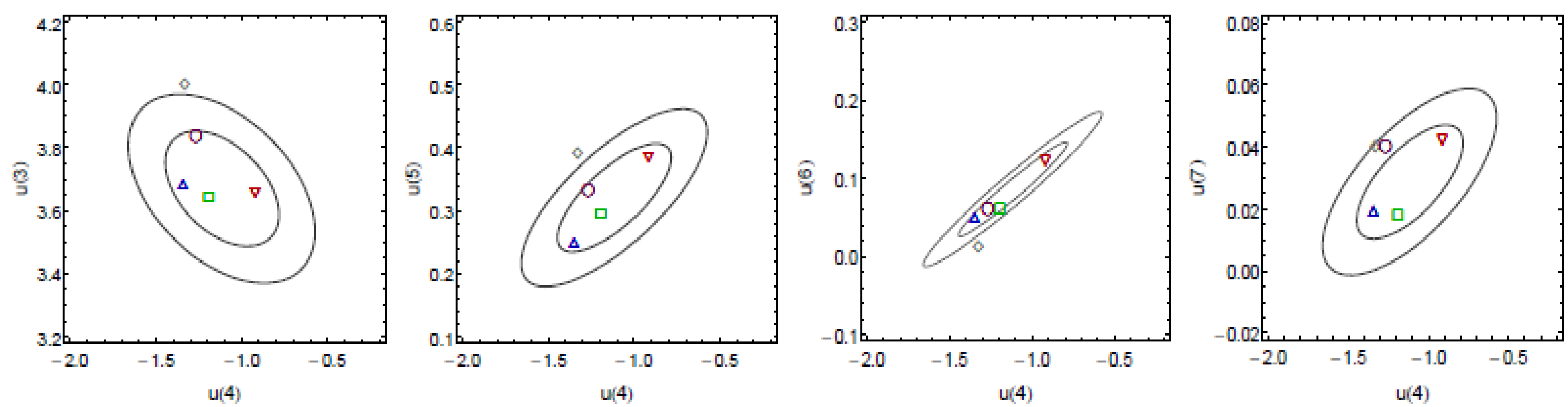


Figure 17: Same as Fig. 16, for a_{3-7} of the u quark PDF.

Merging PDF ensembles

The ensembles can be merged by averaging their meta-parameters. For CT10, MSTW, NNPDF ensembles, unweighted averaging is reasonable, given their similarities.

For any parameter a_i, ensemble g with N_{rep} initial replicas:

$$\langle a_i \rangle_g = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} a_i(k),$$

Central value on g

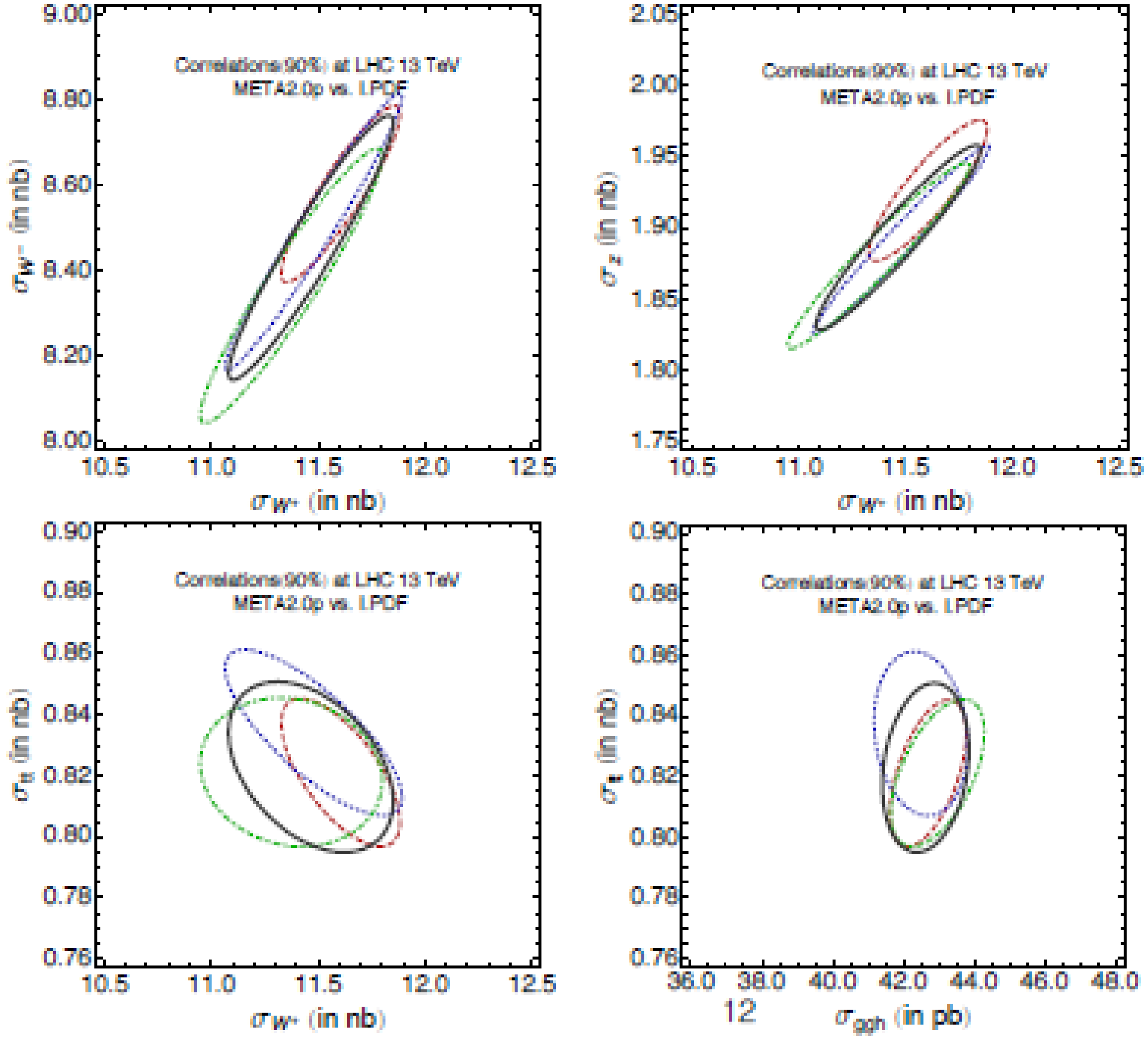
$$\mathrm{cov}(a_i, a_j)_g = \frac{N_{rep}}{N_{rep}-1} \langle (a_i - \langle a_i \rangle_g) \cdot (a_j - \langle a_j \rangle_g) \rangle_g,$$

$$(\delta a_i)_g = \sqrt{\mathrm{cov}(a_i, a_i)_g}.$$

Standard deviation on g

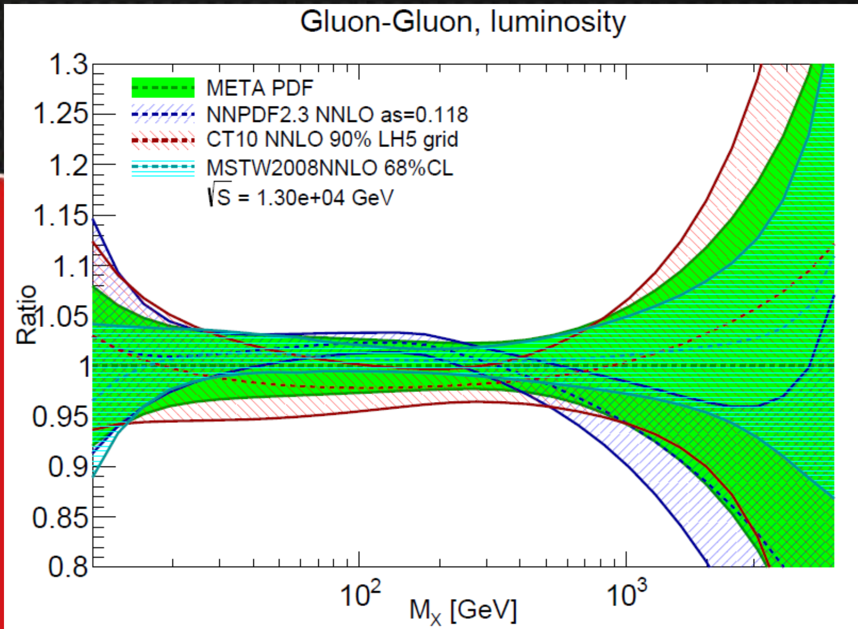
META2.0 predictions for LHC observables

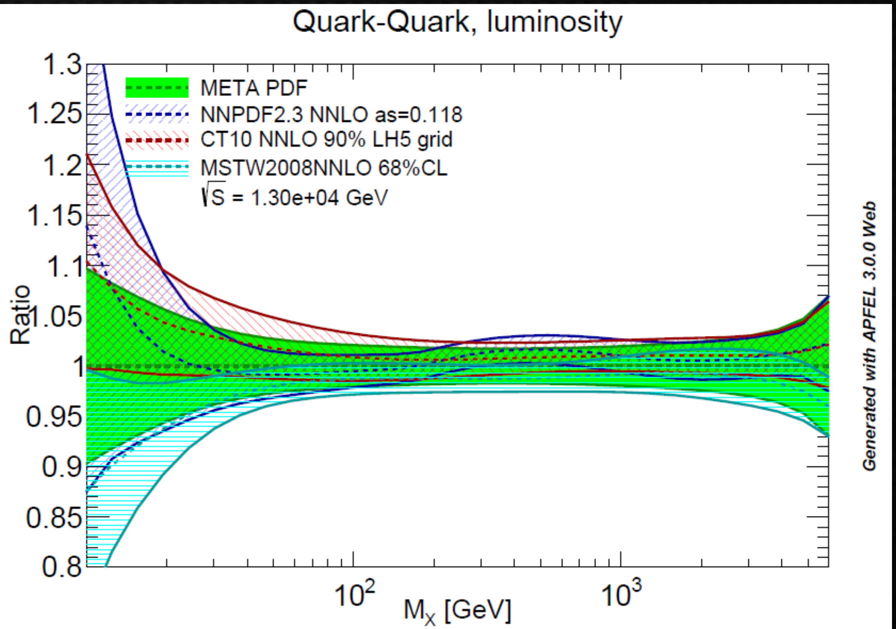
- Currently only have results for META NNLO v2.0p, will add later for v2.1, inclusive observables at 13 TeV



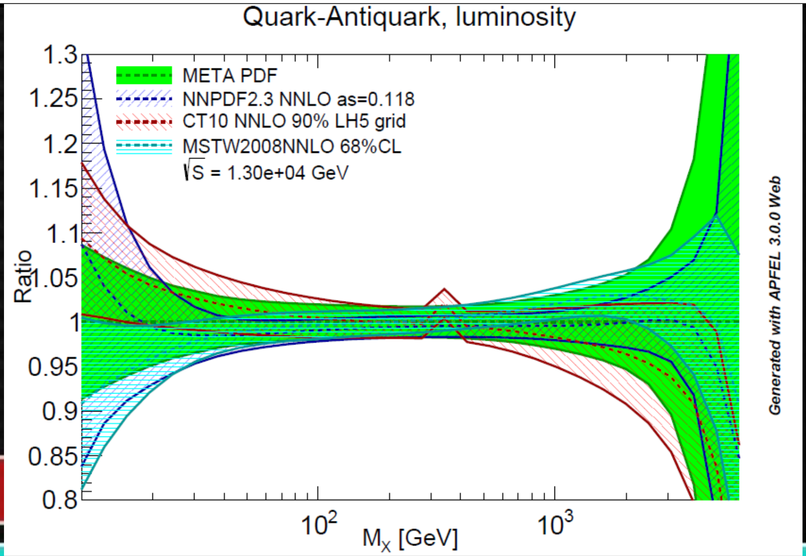
- Blue, CT14p, red, MMHT14, green, NNPDF3.0, black, METAv2.0p, error ellipse at 90% cl; using Vrap0.9, iHixs1.3, and top++2.0

Some parton luminosities





PRELIMINARY



Plots are made with APFEL WEB (apfel.mi.infn.it; Carrazza et al., 1410.5456)

Reduction of the error PDFs

The number of final error PDFs can be much smaller than in the input ensembles

In the META2.0 study:
200 CT, MSTW, NNPDF error sets

⇒ 600 MC replicas for reconstructing the combined probability distribution

⇒ 40-100 Hessian META sets for most LHC applications (**general-purpose** ensemble META2.0)

⇒ 13 META sets for LHC Higgs production observables (**reduced ensemble** META LHCH)

New

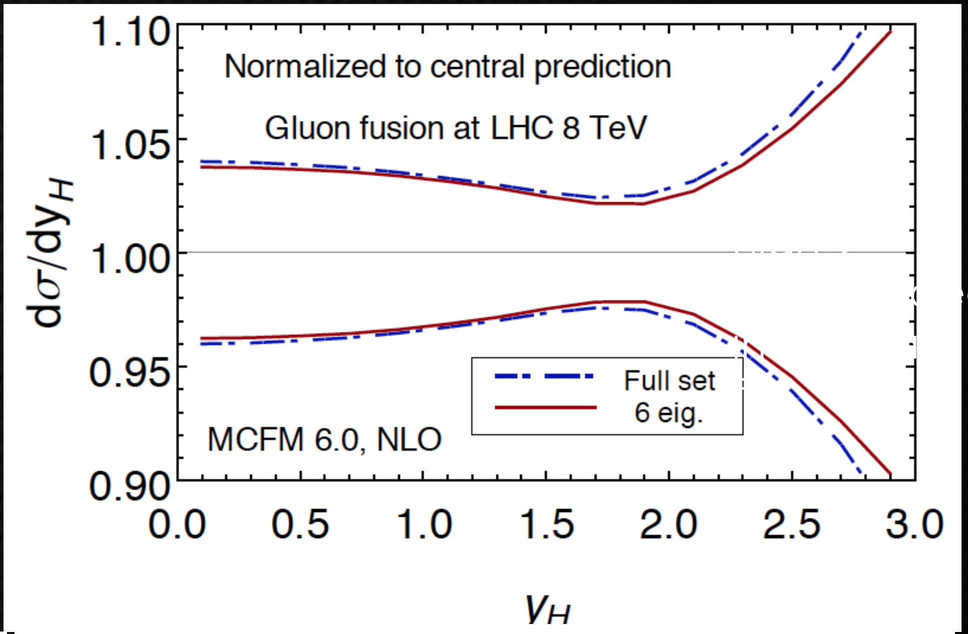
Reduced META ensemble

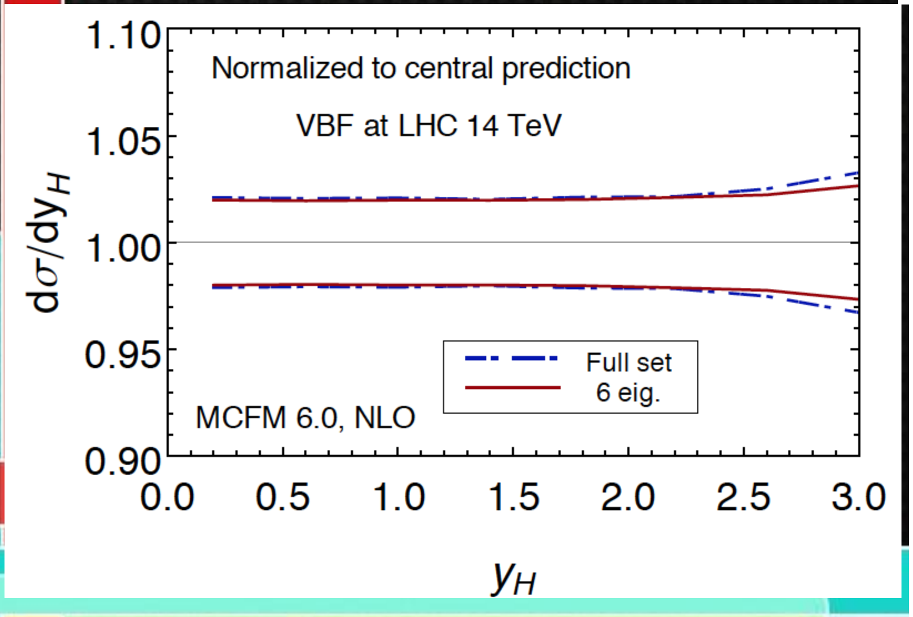
- Already the general-purpose ensemble reduced the number of error PDFs needed to describe the LHC physics; but we can further perform a data set diagonalization to pick out eigenvector directions important for Higgs physics or another class of LHC processes
- Select global set of Higgs cross sections at 8 and 14 TeV (46 observables in total; more can be easily added if there is motivation)

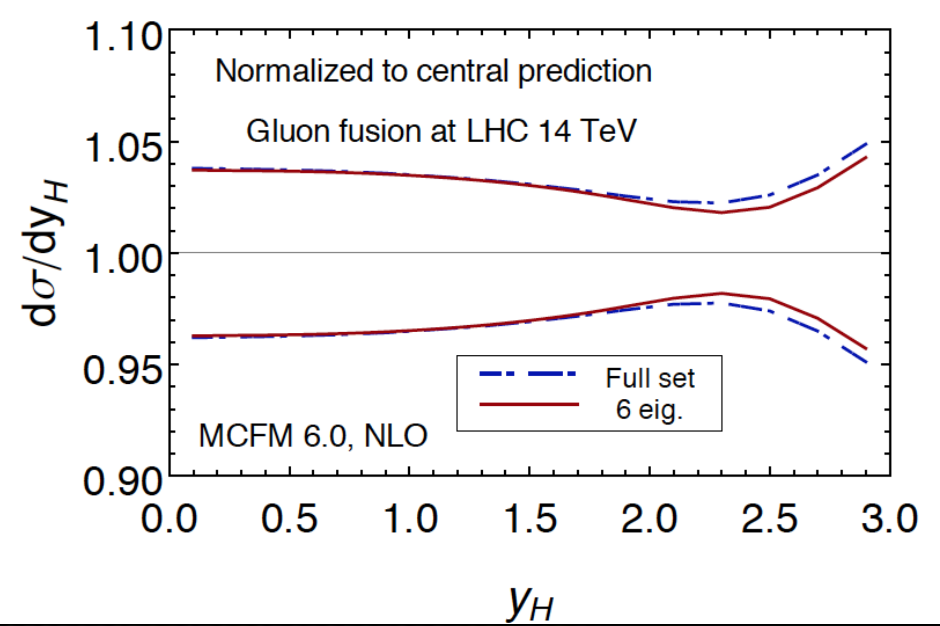
production channel	$\sigma(inc.)$	$\sigma(\lvert y_H\rvert > 1)$	$\sigma(p_{T,H} > m_H)$	scales
$gg \to H$	iHixs1.3 [32] at NNLO	MCFM6.3 [33] at LO	—	m_H
$b\bar{b} \to H$	iHixs at NNLO	—	—	m_H
VBF	VBFNLO2.6 [34] at NLO	same	same	m_W
HZ	VHNNLO1.2 [35] at NNLO	CompHEP4.5 [36] at LO	CompHEP at LO	$m_Z + m_H$
$HW^{\pm}$	VHNNLO at NNLO	—	—	$m_W + m_H$
HW^{+}	CompHEP at LO	same	same	$m_W + m_H$
HW^{-}	CompHEP at LO	same	same	$m_W + m_H$
$H + 1jet$	MCFM at LO	same	same	m_H
$Ht\bar{t}$	MCFM at LO	CompHEP at LO	CompHEP at LO	$2m_t + m_H$
HH	Hpair [37] at NLO	—	—	$2m_H$

Higgs eigenvector set

- The reduced META eigenvector set does a good job of describing the uncertainties of the full set for *typical* processes such as ggF or VBF
- But actually does a good job in reproducing PDF-induced correlations and describing those LHC physics processes in which $g,\ \bar{u},\ \bar{d}$ drive the PDF uncertainty (see next slide)







process	$\sigma_{cen.}$	δ_{Full}	$\delta_{Diag.}$	$\sigma^{\alpha_s}_{0.116}$	$\sigma^{\alpha_s}_{0.12}$
$gg \to H$ [pb]	18.77	$^{+0.48}_{-0.46}$	$^{+0.48}_{-0.44}$	18.11	19.4
	43.12	$^{+1.13}_{-1.07}$	$^{+1.13}_{-1.04}$	41.68	44.6
VBF [fb]	302.5	$^{+7.8}_{-6.7}$	$^{+7.6}_{-6.7}$	303.1	301.
	878.2	$^{+19.7}_{-17.9}$	$^{+19.2}_{-17.3}$	877.3	878.
HZ [fb]	396.3	$^{+8.4}_{-7.3}$	$^{+8.1}_{-7.4}$	393.0	399.
	814.3	$^{+14.8}_{-13.2}$	$^{+13.8}_{-13.0}$	806.5	823.
$HW^\pm$ [fb]	703.0	$^{+14.4}_{-14.4}$	$^{+14.3}_{-14.1}$	697.4	708.
	1381	$^{+28}_{-22}$	$^{+26}_{-22}$	1368	139
HH [fb]	7.81	$^{+0.33}_{-0.30}$	$^{+0.33}_{-0.30}$	7.50	8.1
	27.35	$^{+0.78}_{-0.72}$	$^{+0.78}_{-0.68}$	26.48	28.2
$t\bar{t}$ [pb]	248.4	$^{+9.1}_{-8.2}$	$^{+9.2}_{-8.1}$	237.1	259.
	816.9	$^{+21.4}_{-19.6}$	$^{+21.4}_{-18.4}$	785.5	848.
$Z/\gamma^*(l^+l^-)$ [nb]	1.129	$^{+0.025}_{-0.023}$	$^{+0.024}_{-0.023}$	1.113	1.14
	1.925	$^{+0.043}_{-0.041}$	$^{+0.040}_{-0.037}$	1.897	1.95
$W^+(l^+\nu)$ [nb]	7.13	$^{+0.14}_{-0.14}$	$^{+0.14}_{-0.13}$	7.03	7.2
	11.64	$^{+0.24}_{-0.23}$	$^{+0.22}_{-0.21}$	11.46	11.8
$W^-(l^-\bar{\nu})$ [nb]	4.99	$^{+0.12}_{-0.12}$	$^{+0.12}_{-0.11}$	4.92	5.0
	8.59	$^{+0.21}_{-0.20}$	$^{+0.19}_{-0.18}$	8.46	8.74
W^+W^- [pb]	4.14	$^{+0.08}_{-0.08}$	$^{+0.08}_{-0.07}$	4.04	4.2
	7.54	$^{+0.15}_{-0.14}$	$^{+0.14}_{-0.12}$	7.39	7.57
ZZ [pb]	0.703	$^{+0.016}_{-0.014}$	$^{+0.015}_{-0.014}$	0.695	0.71
	1.261	$^{+0.026}_{-0.024}$	$^{+0.024}_{-0.022}$	1.256	1.27
W^+Z [pb]	1.045	$^{+0.019}_{-0.018}$	$^{+0.019}_{-0.017}$	1.039	1.06
	1.871	$^{+0.033}_{-0.031}$	$^{+0.029}_{-0.027}$	1.850	1.89
W^-Z [pb]	0.788	$^{+0.020}_{-0.019}$	$^{+0.019}_{-0.018}$	0.780	0.79
	1.522	$^{+0.034}_{-0.032}$	$^{+0.033}_{-0.031}$	1.509	1.54

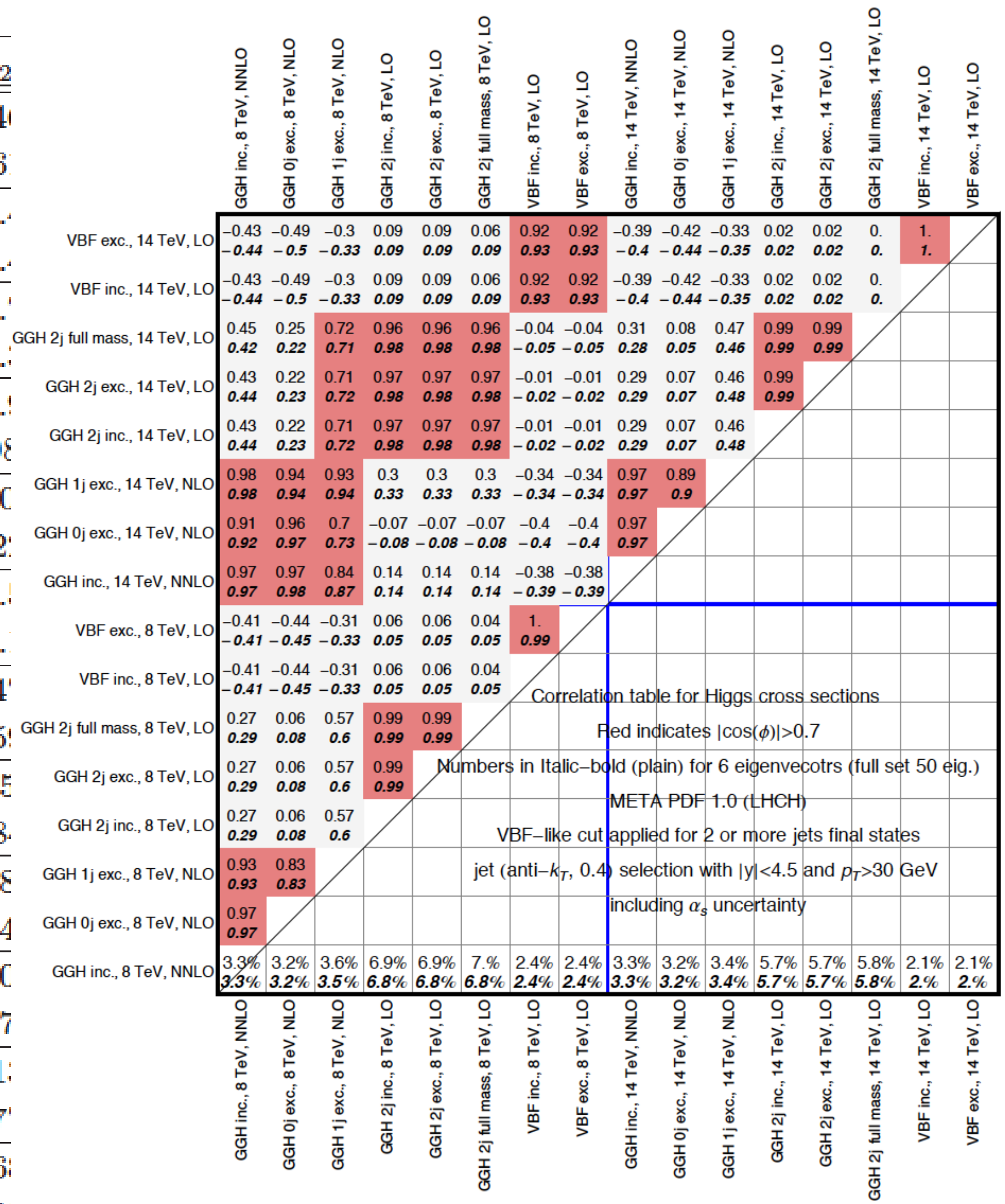
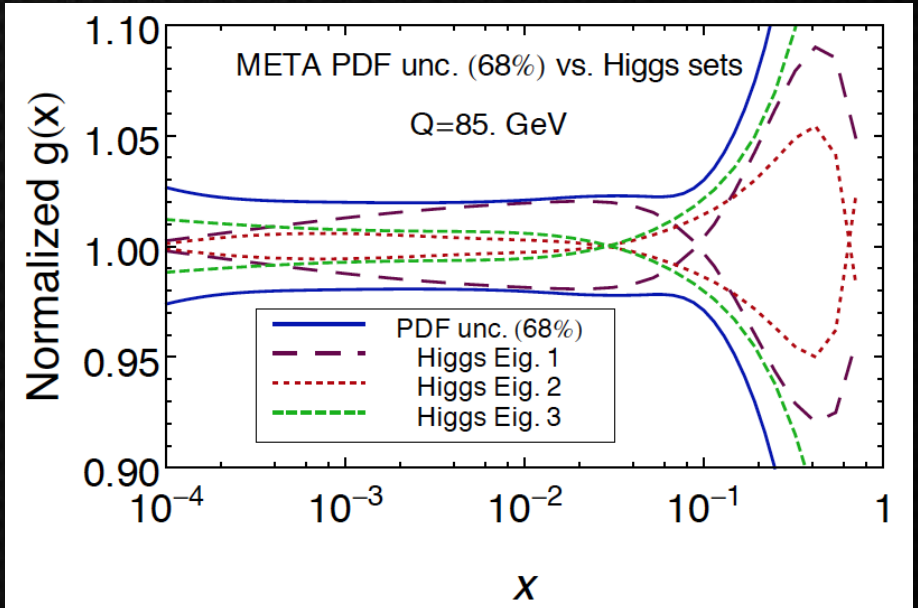


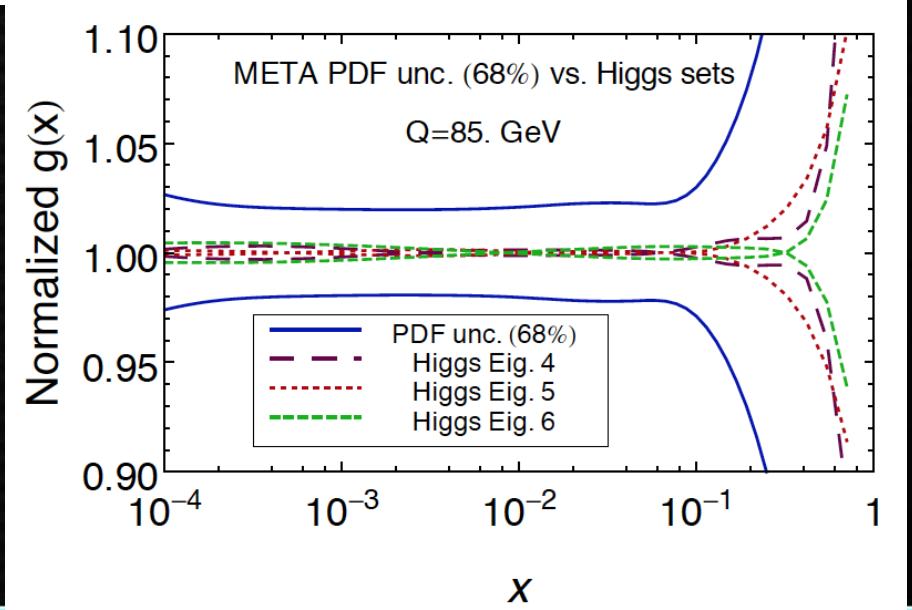
FIG. 7: Same as Fig. 5, with α_s uncertainties included by adding in quadrature.

Re-diagonalized eigenvectors…

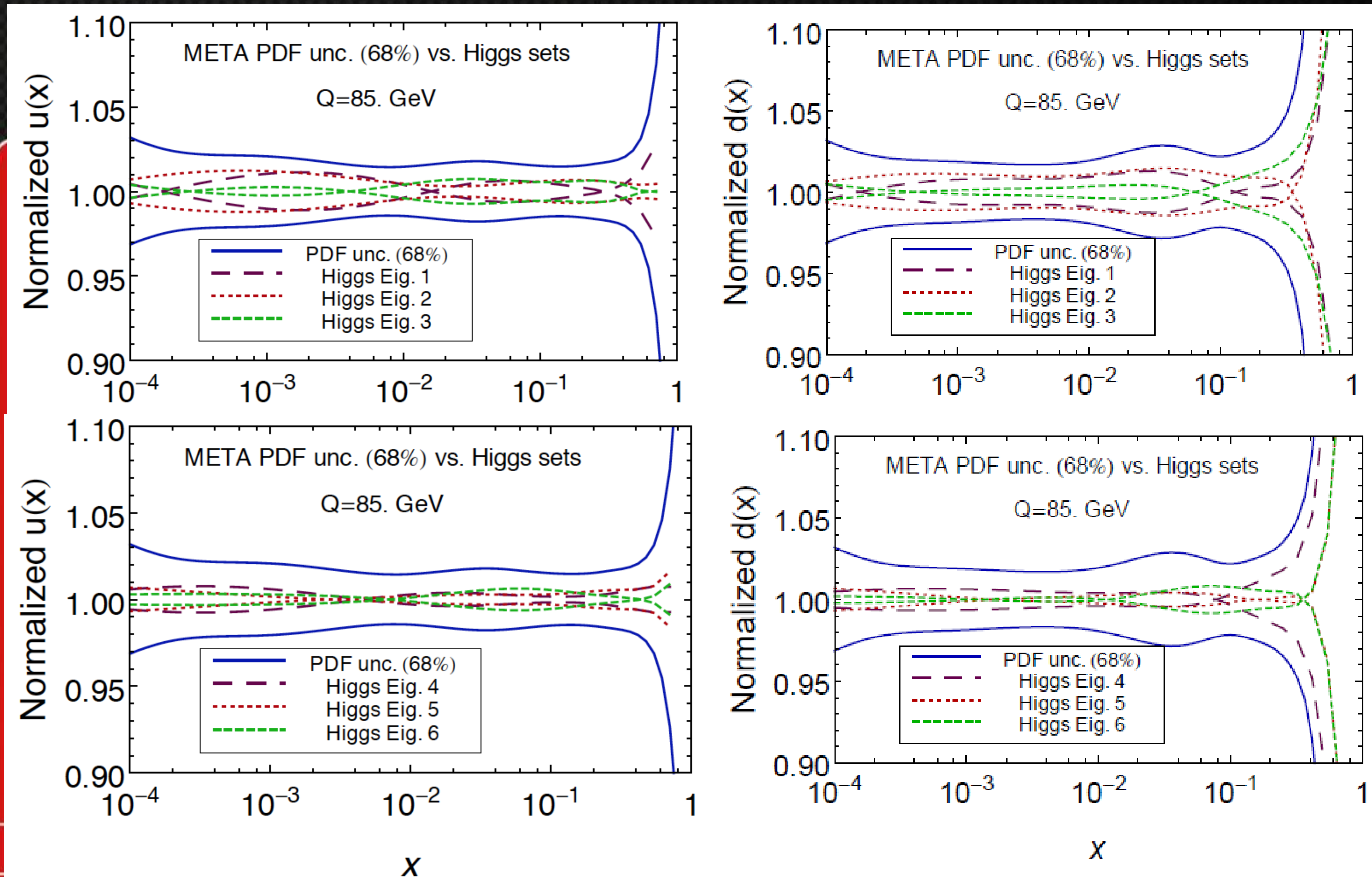
…are associated with the parameter combinations that drive the PDF uncertainty in Higgs, W/Z production at the LHC

- Eigenvectors 1-3 cover the gluon uncertainty. They also contribute to $\bar{u}, \bar{d}$ uncertainty.
- Eigenvector 1 saturates the uncertainty for most of the $gg \to H$ range.





u, d quark uncertainties are more distributed



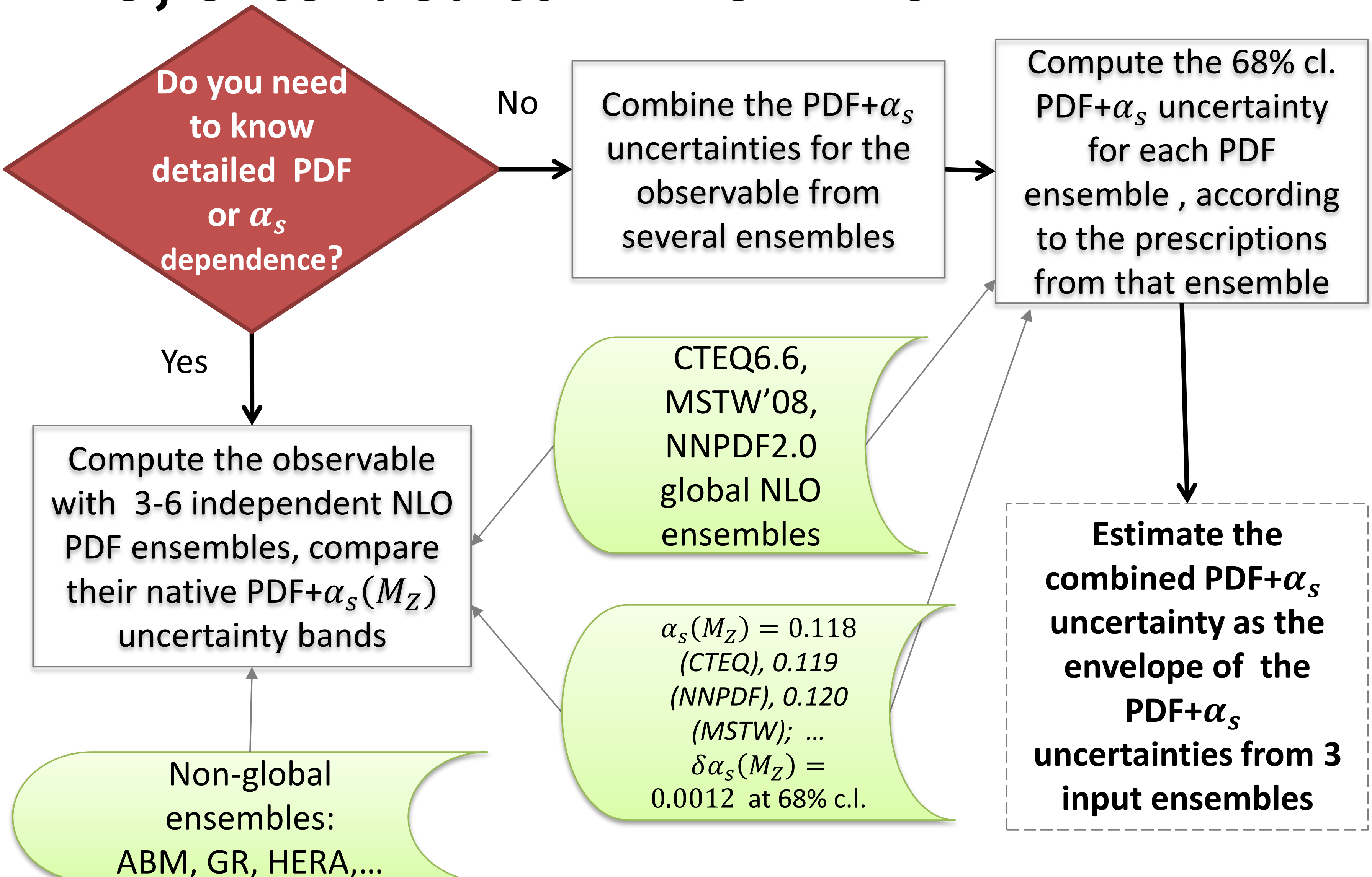
To summarize, the meta-parametrization and Hessian method facilitate the combination of PDF ensembles even when the MC replicas are introduced at the intermediate stage

- A general and intuitive method. Implemented in a public Mathematica module MP4LHC
- The PDF parameter space of all input ensembles is visualized explicitly.
- Data combination procedures familiar from PDG can be applied to each meta-PDF parameter
- Asymmetric Hessian errors can be computed, similar to CT14 approach
- Effective in data reduction; makes use of diagonalization of the Hessian matrix in the Gaussian approximation. Reproduces correlations between Higgs signals and backgrounds with just 13 META –LHCH PDFs.
- Is considered as a candidate combination method of the 2015 PDF4LHC prescription

Back-up slides

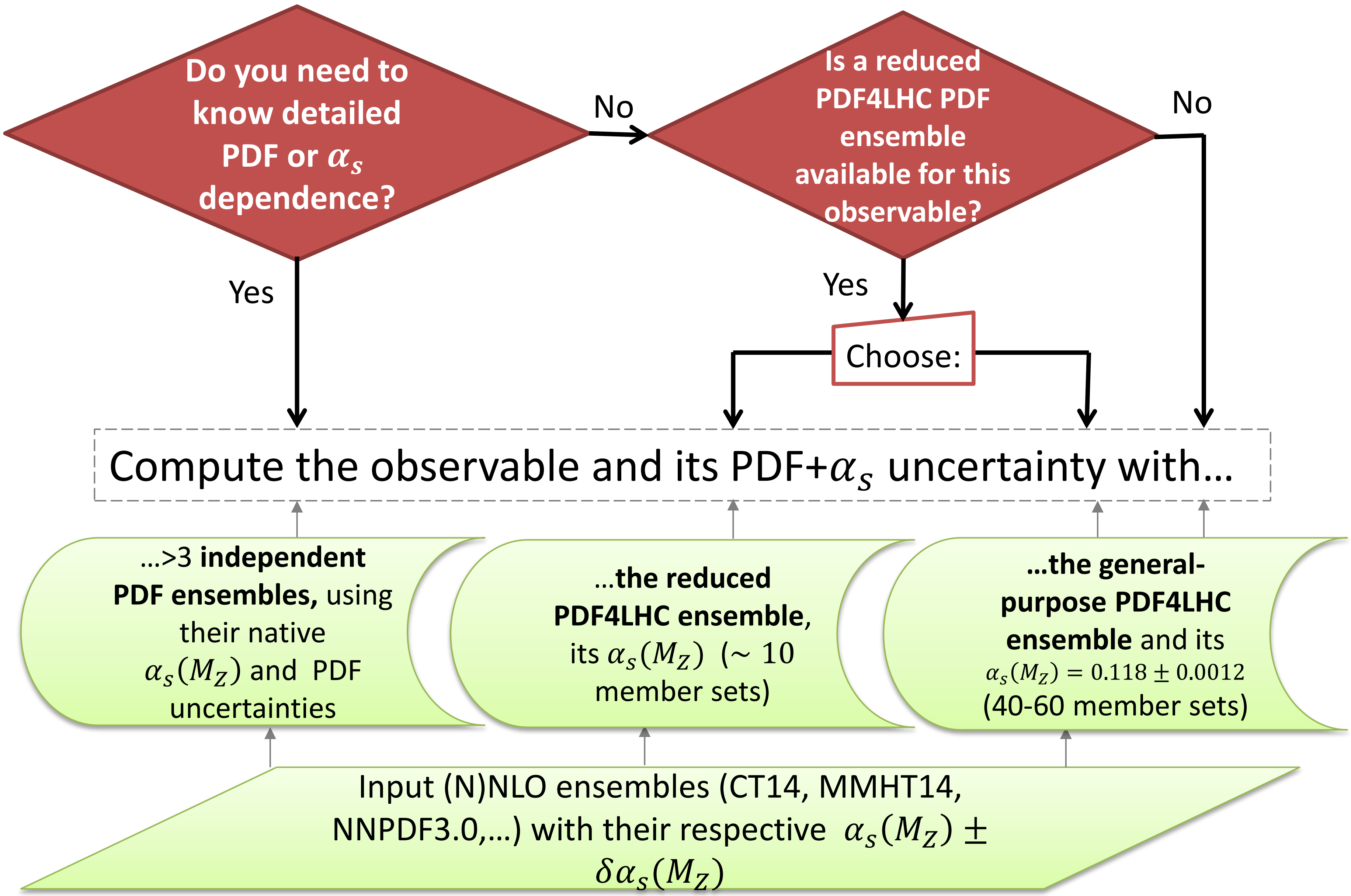
57

2010 PDF4LHC recommendation for an LHC observable:
NLO; extended to NNLO in 2012



M. Botje et al., arXiv:1101.0538

2015: A concept for a new PDF4LHC recommendation



This procedure applies both at NLO and NNLO

Combination of the PDFs into the future PDF4LHC ensemble

PDFs from several groups are combined into a PDF4LHC ensemble of error PDFs **before** the LHC observable is computed. This simplifies the computation of the PDF+α_s uncertainty and will likely cut down the number of the PDF member sets and the CPU time needed for simulations.

The same procedure is followed at NLO and NNLO. The combination was demonstrated to work for global ensembles (CT, MSTW, NNPDF). It still needs to be generalized to allow inclusion of non-global ensembles.

The PDF uncertainty at 68% c.l is computed from error PDFs at central $\alpha_s(M_Z)$.

Two additional error PDFs are provided with either PDF4LHC ensemble to compute the α_s uncertainty using $\alpha_s(M_Z) = 0.118 \pm 0.0012$ at the 68% c.l.

Progress in developing the combination procedure

Two methods for combination of PDFs were extensively compared, with promising results:

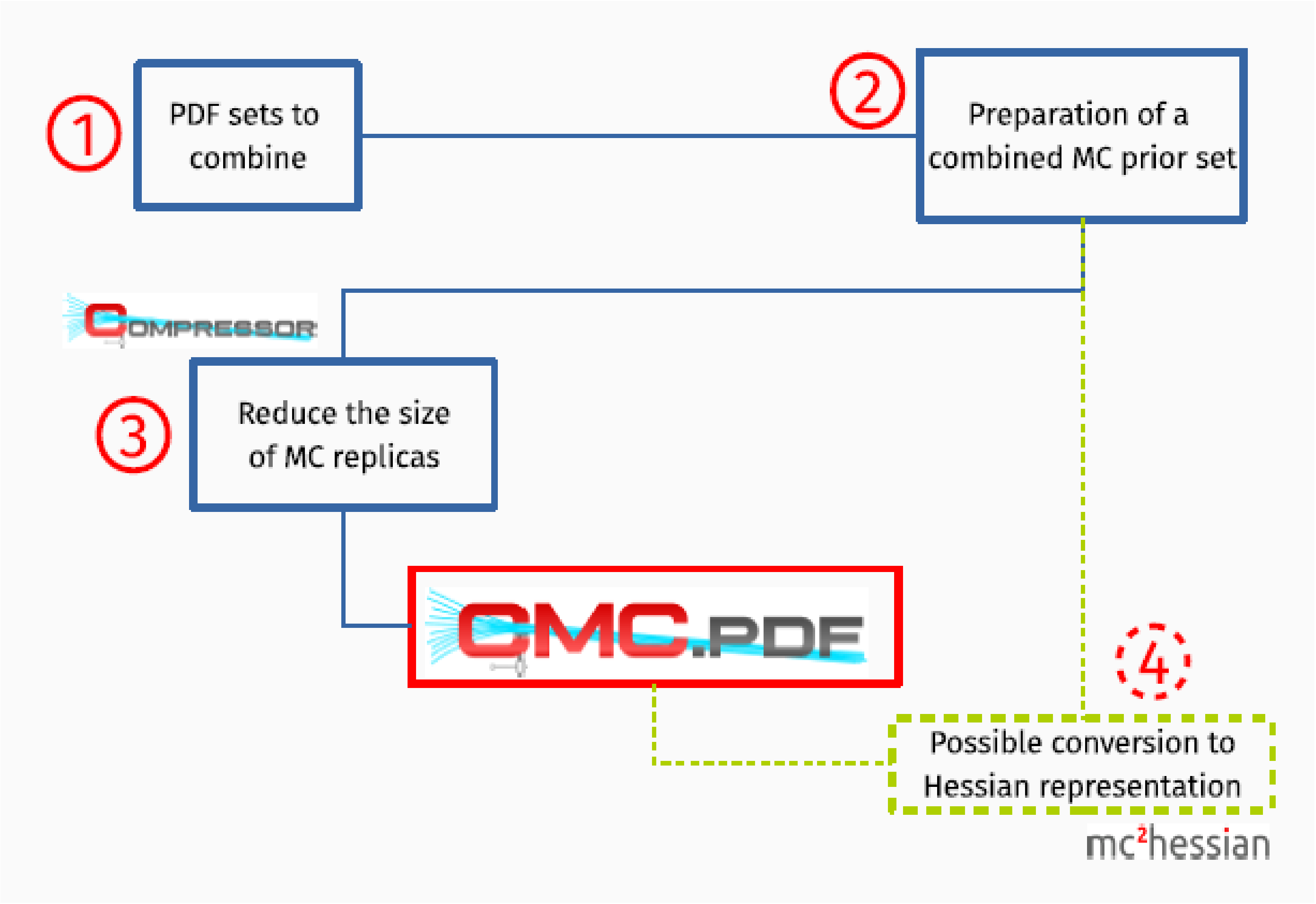
1. **Meta-parametrizations + MC replicas + Hessian data set diagonalization**
(J. Gao, J. Huston, P. Nadolsky, 1401.0013)

2. Compression of Monte-Carlo replicas
(Carazza, Latorre, Rojo, Watt, 1504:06469)

Both procedures start by creating a combined ensemble of MC replicas from all input ensembles (G. Watt, R. Thorne,1205.4024; S. Forte, G. Watt, 1301.6754). They differ at the second step of reducing a large number of input MC replicas (~ 300) to a smaller number for practical applications (13-100 in the META approach; 40 in the CMC approach). The core question is how much input information to retain in the reduced replicas in each Bjorken-x region.

CMC PDFs

S. Carrazza, Feb. 2015



THE COMPRESSION STRATEGY

We define **statistical estimators** for the MC prior set:

1. **moments:** central value, variance, skewness and kurtosis
2. **statistical distances:** the Kolmogorov distance
3. **correlations:** between flavors at multiple x points

These estimators are them **compared** to subsets of replicas **interactively** driven by an *error function*, i.e.

$$\mathrm{ERF}_{\mathrm{tot}} = \sum_n \frac{1}{N_n} \sum_i \left(\frac{C_i^{(n)} - O_i^{(n)}}{O_i^{(n)}} \right)^2$$

where n runs over the number of statistical estimators and

- N_i is a normalization factor extracted from random realizations
- $O_i^{(n)}$ is the value of the estimator for the prior
- $C_i^{(n)}$ is the corresponding value for the compressed set

Benchmark comparisons of CMC and META PDFs

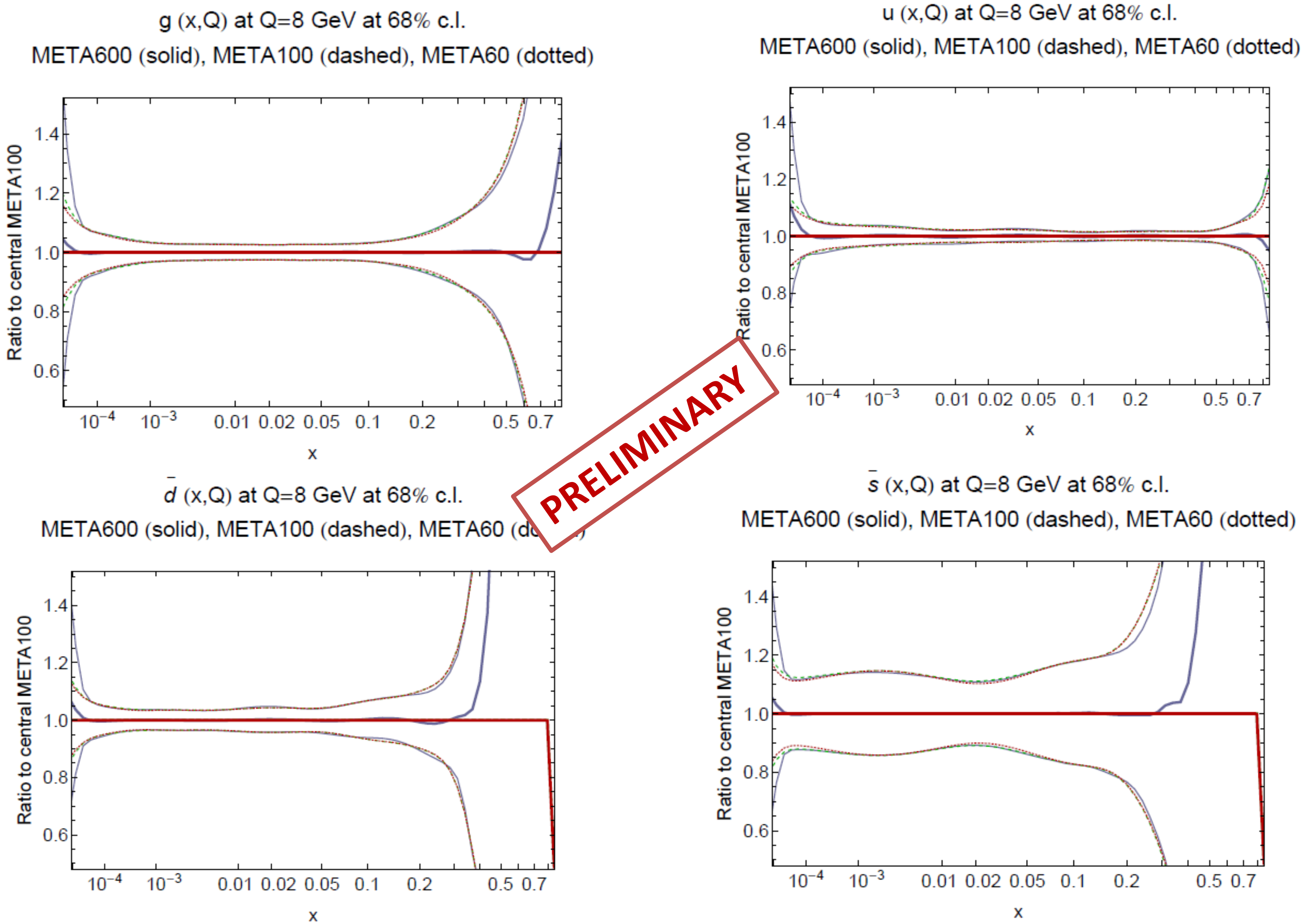
CMC ensembles with 40 replicas and META ensembles with 40-100 replicas are compared with the full ensembles of 300-600 MC replicas.

Accuracy of both combination procedures is already competitive with the 2010 PDF4LHC procedure, can be further fine-tuned by adjusting the final number of replicas.

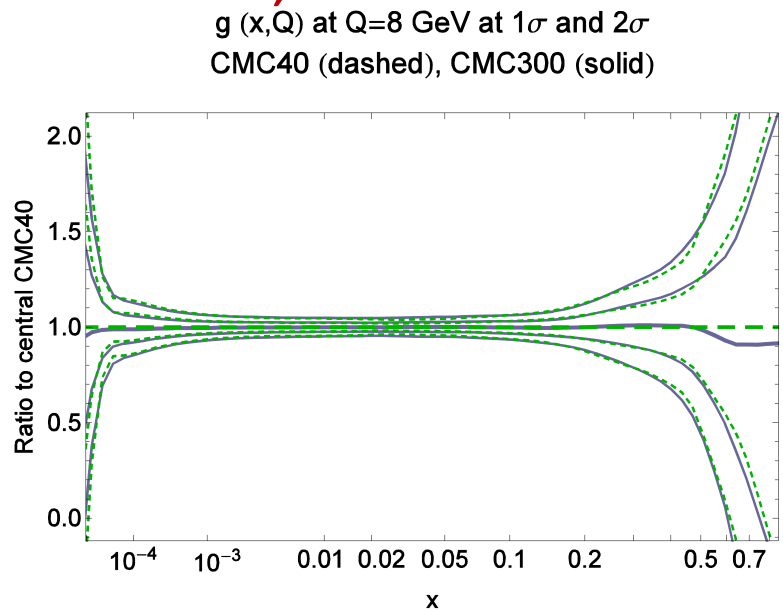
Error bands:

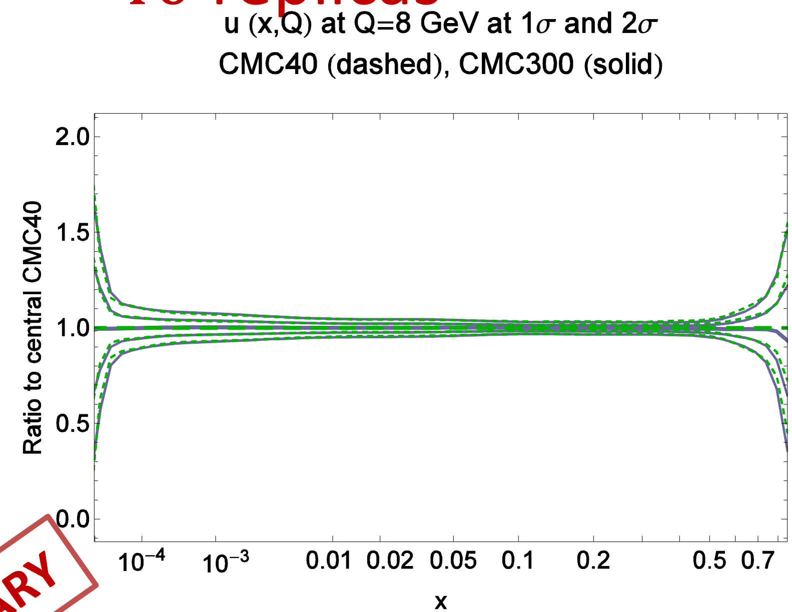
In the (x, Q) regions covered by the data, the agreement of 68%, 95% c.l. intervals is excellent. The definition of the central PDFs and c.l. intervals is ambiguous in extrapolation regions, can differ even within one approach. E.g., differences between mean, median, mode “central values”.

Reduction, META ensemble: 600 → 100 → 60 error sets

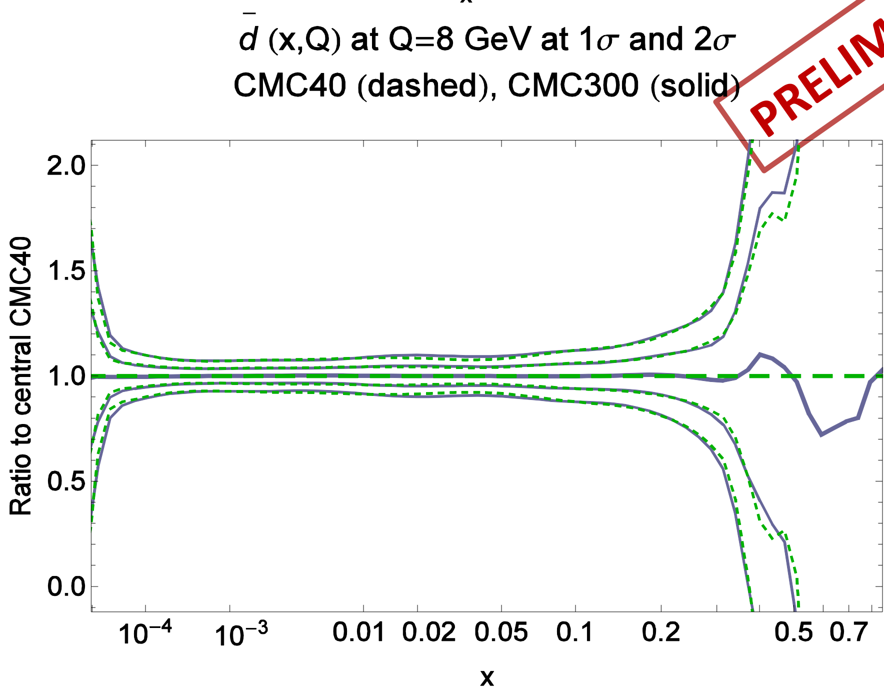


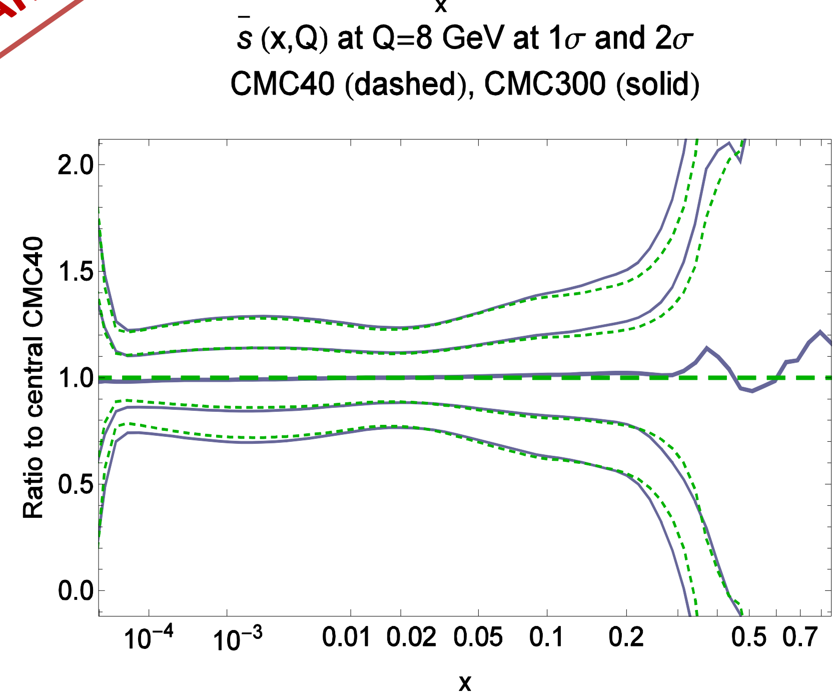
Reduction, CMC ensemble: 300 → 40 replicas





PRELIMINARY





Benchmark comparisons, general observations II

PDF-PDF correlations:

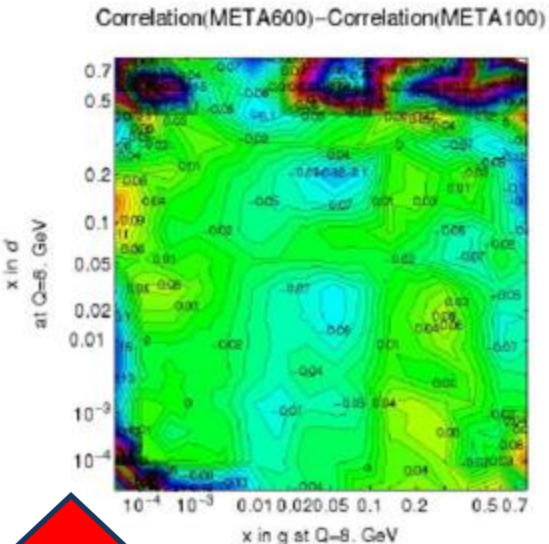
Correlations of META300 and CMC300 ensembles differ by up to ± 0.2 as a result of fluctuations in replica generation

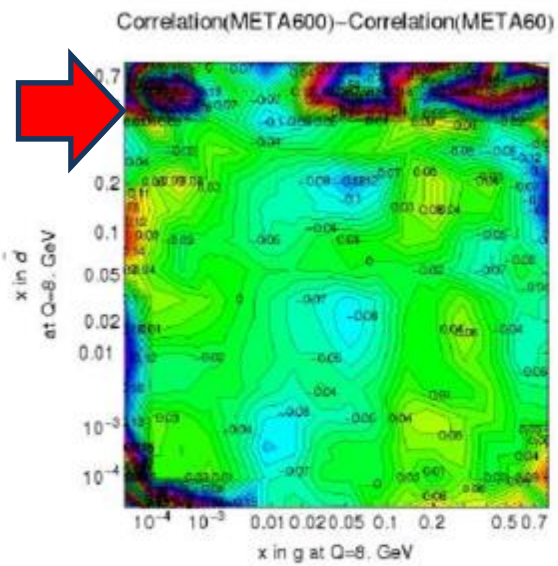
META40 PDFs faithfully reproduce PDF-PDF correlations of the META600 PDFs in the regions with data; fail to reproduce correlations in extrapolation regions ⇒ *next slide, upper row*

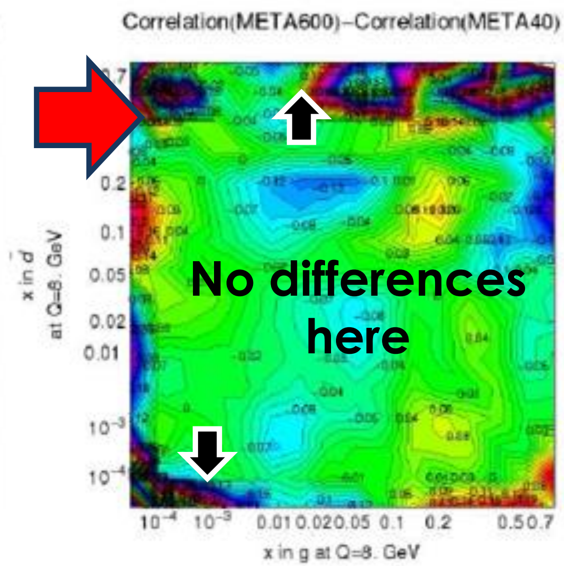
CMC40 PDFs better reproduce correlations of CMC300 in extrapolation regions; lose more accuracy in (x, Q) regions with data, but still within acceptable limits ⇒ *next slide, lower row*

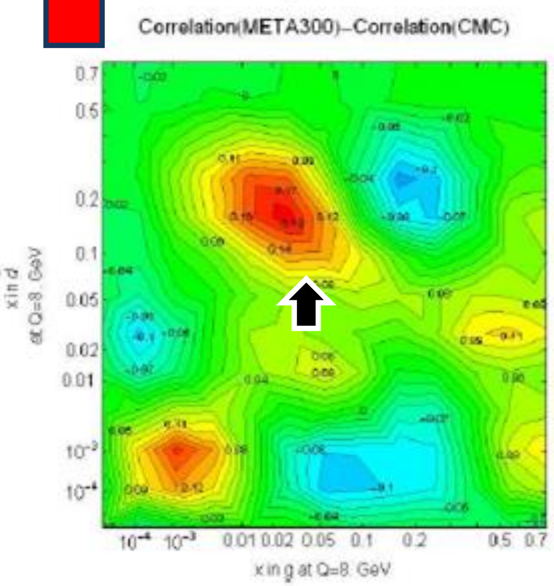
These patterns of correlations persist at the initial scale $Q_0 = 8$ GeV as well as at EW scales

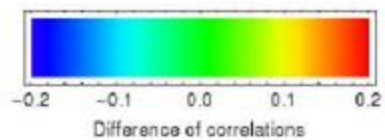
PDF-PDF correlation, example: $\bar{d}(x, Q)$ vs $g(x, Q)$ at $Q = 8\ GeV$





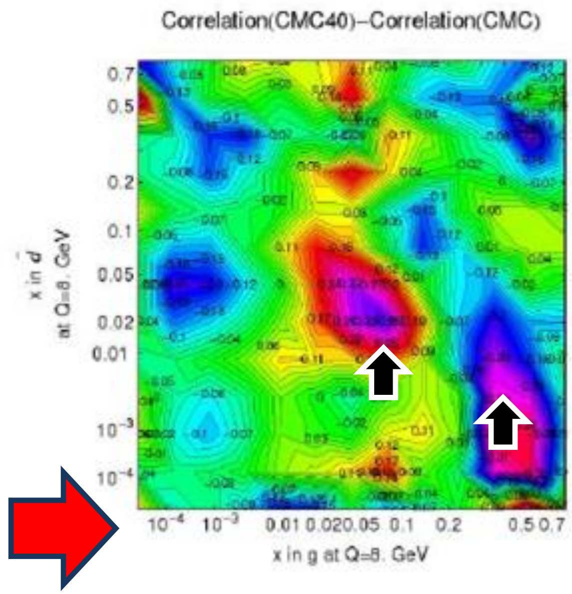






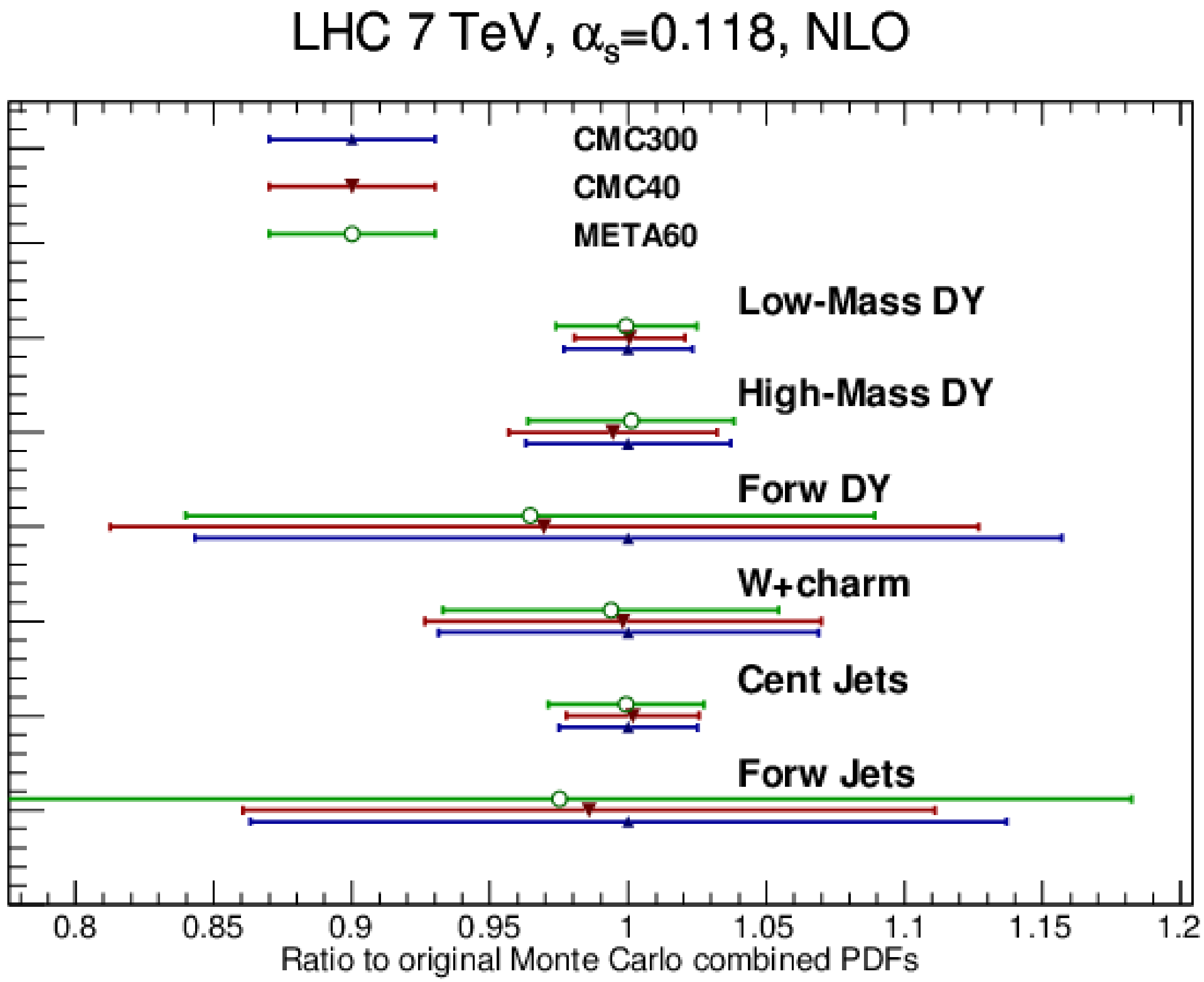
Upper row: META600->60->40

Lower row: CMC300->CMC40



PRELIMINARY

Agreement at the level of benchmark cross sections



CMC-META benchmark cross sections are consistent in the x regions constrained by data

There are moderate differences in extrapolation regions. Either reduced ensemble only partly captures non-Gaussianity of the full MC ensemble at such x

J. Rojo