META parton distributions for LHC applications

Pavel Nadolsky
Southern Methodist University

With Jun Gao (Argonne),
Joey Huston (Michigan State)

arXiv:1401.0013, **1507.XXXX**

Easy computations of PDF uncertainties

Meta-parametrizations combine PDFs from several groups (CT, MMHT, NNPDF, …) in a variety of LHC applications. They simplify computations of PDF uncertainties, while preserving key information provided by input PDFs.

META parametrizations (version 2.0) will be released with a new update of the LHAPDF6 library. They offer a versatile framework for combination of PDF+α_s uncertainties from global PDF ensembles in LHC Run-2 analyses.

Table by A.Accardi, DIS'2015 workshop

	JLab	HERA I+II *Wichmann*	Tevatron new W,Z	LHC	di-μ	Nucl.	HT TMC	Flex *d*	closure
HERAPDF2.0 → *Myronenko, Brandt*		✓	¤						
CT14 → *Nadolski*			✓ ¤¤	✓	✓			✓	
MMHT14 → *Thorne*			✓ ¤¤	✓	✓	✓			
NNPDF3.0 → *Deans*				✓	✓				✓
[GJR14]	✓			✓	✓	✓	✓		
CJ12 * (→ CJ15) → *Melnitchouk*	✓	(✓)	(✓)		✗	✓	✓	✓	
ABM12 **					✓	✓	✓		

* NLO only ** No jet data ¤ but see 1503.05221 ¤¤ no reconstructed W

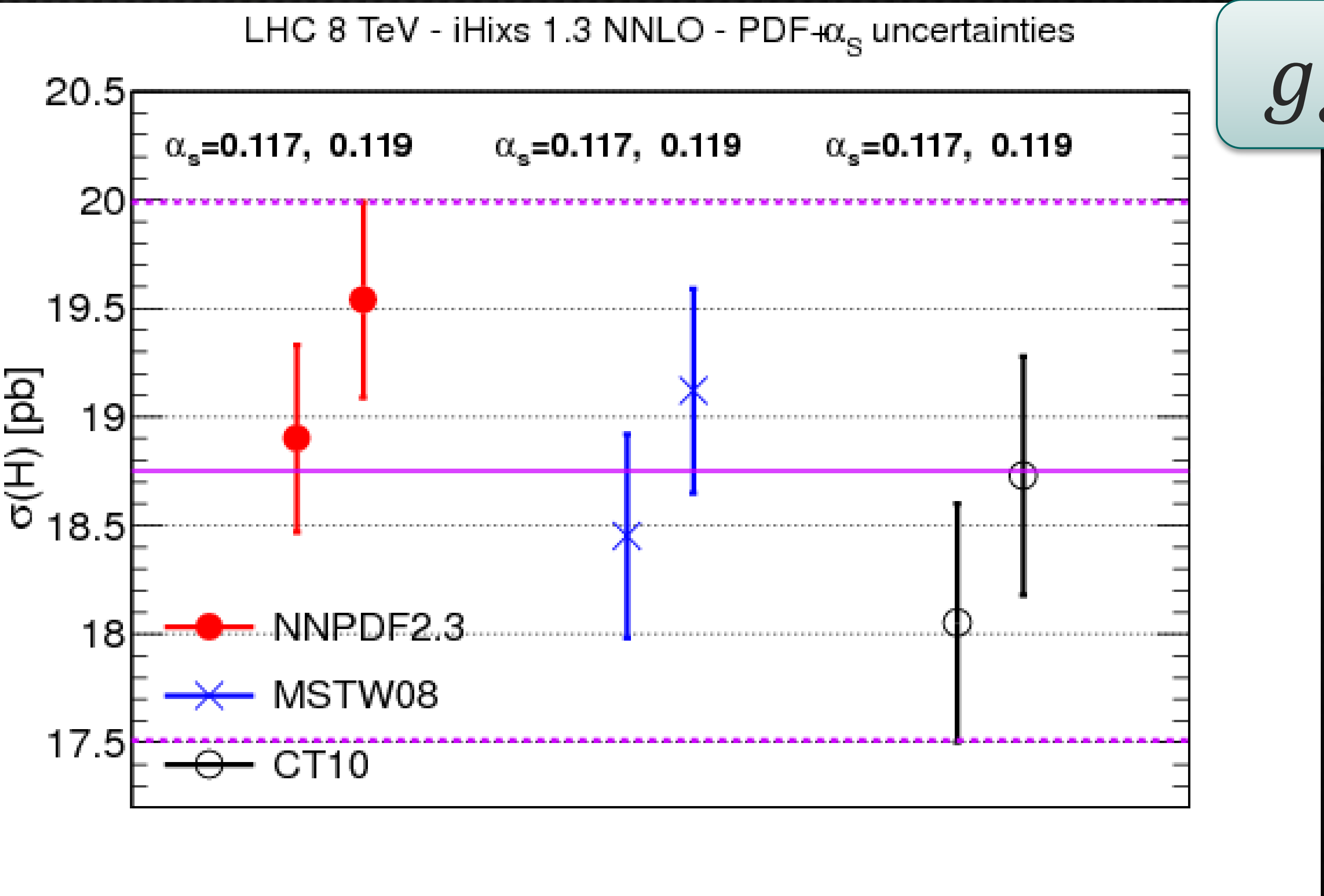
NNLO PDFs are now available from 5 groups. Their accuracy steadily advances to keep up with (N)(N)NLO hard cross sections.

Typical questions asked by PDF users

- Which PDFs should be used in a given experimental study?
- Are all predictions compatible?
- Can/should one combine PDF uncertainties from various groups?
- How to compute PDF uncertainties efficiently?
  1. Interfaces for fast NLO computations *(Applgrid, FastNLO, aMCFast)*
  2. Combination at the PDF level *(META, CMC)*

Progress in finding the answers

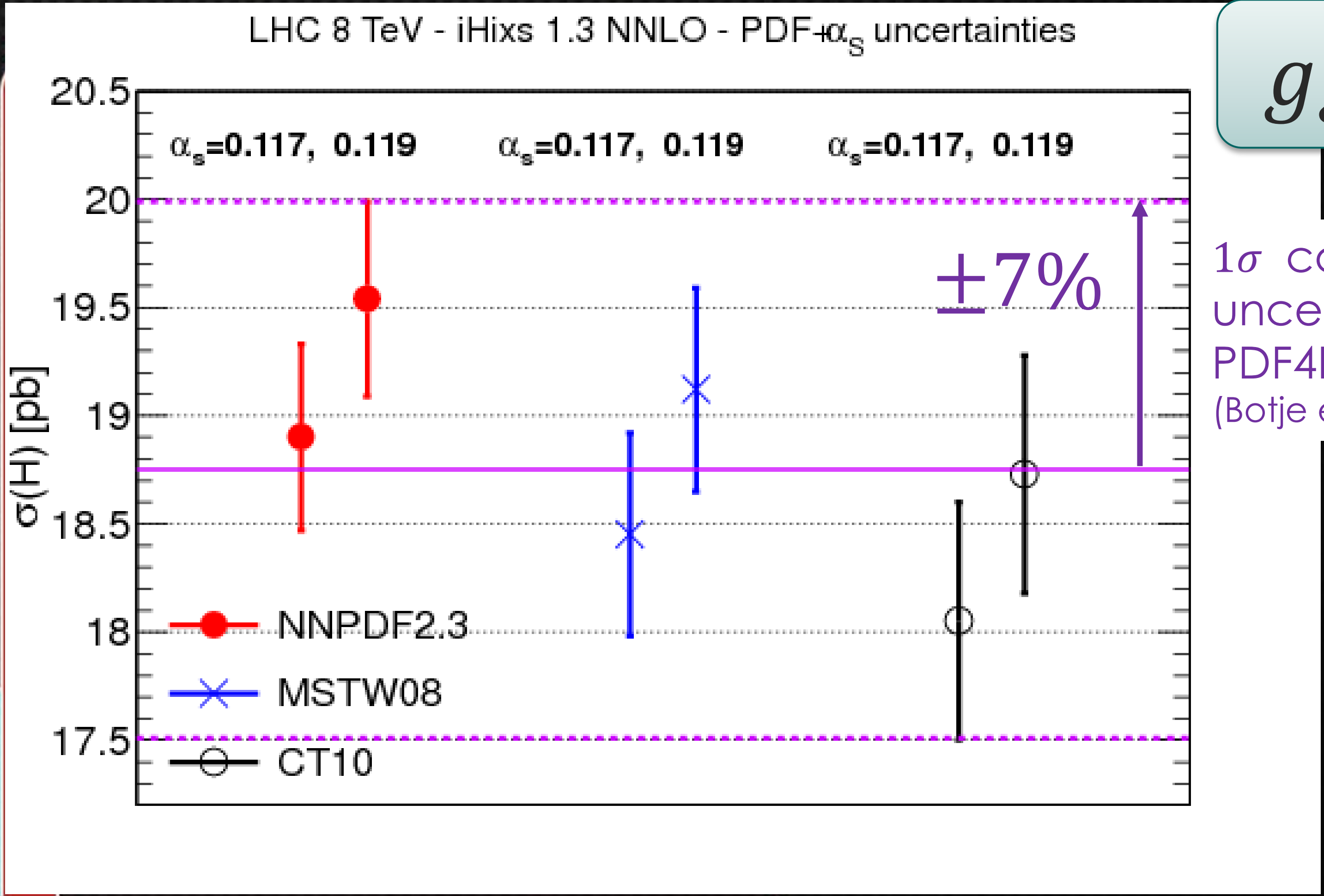
2012: the typical NNLO PDF+α_s uncertainty is larger than 2-3%



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

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

2012: the typical NNLO PDF+α_s uncertainty is larger than 1%

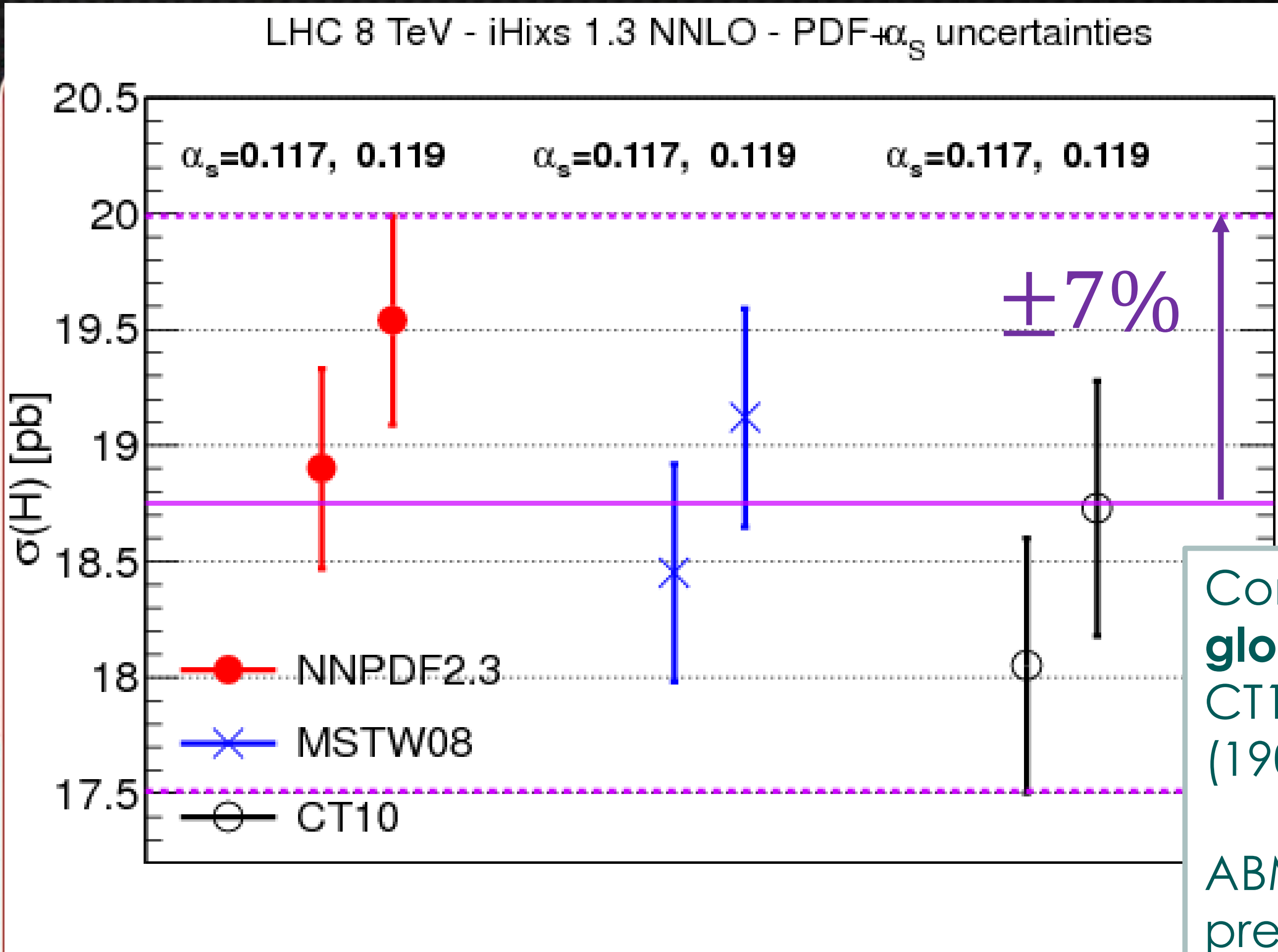


$$gg \to H^0_{SM}$$

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**

2012: the typical NNLO PDF+α_s uncertainty is larger than 1%



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

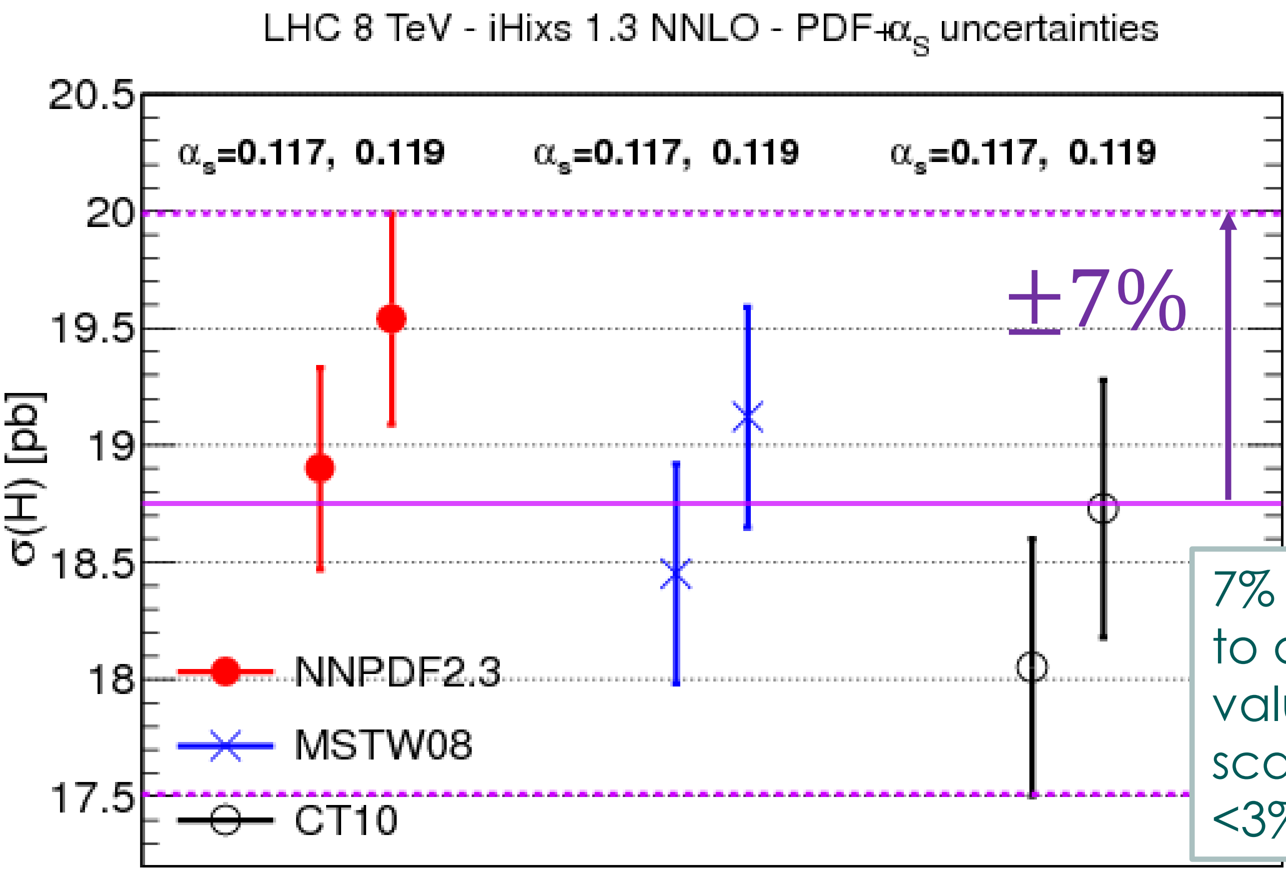
1σ combined PDF+α_s uncertainty, using PDF4LHC convention (Botje et al., arxiv:1101.0538)

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**

2012: the typical NNLO PDF+α_s uncertainty is larger than 1%



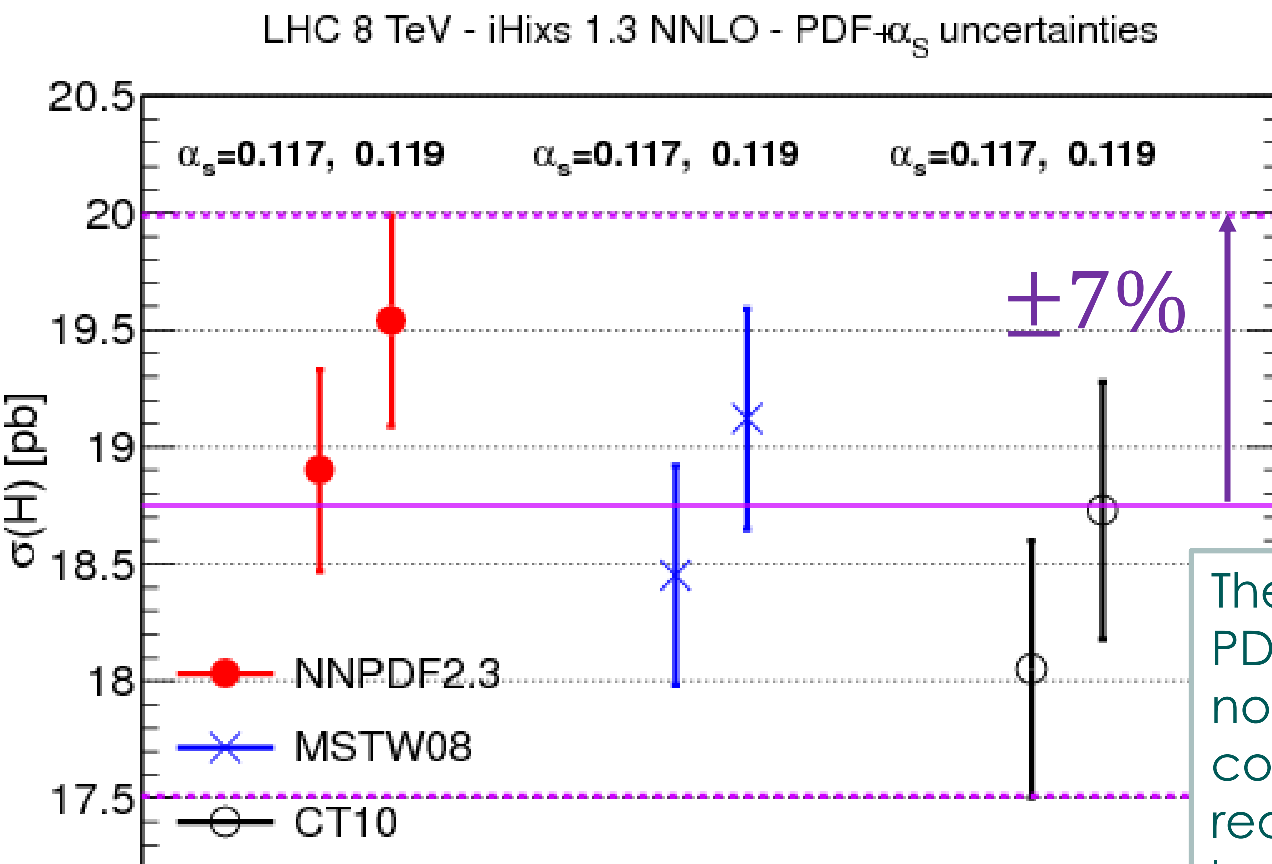
$gg \to H^0_{SM}$

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

7% PDF uncertainty is due to discrepancy in central values. Larger than N3LO scale dependence of <3%

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

2012: 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 100+ 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…

Four groups released new generations of NNLO PDFs

Global PDF analyses with LHC data

CT'14: in LHAPDF6, the paper undergoes final revisions

MMHT'14: Harland-Lang, Martin, Motylinski, Thorne, *arXiv:1412.3989*

Neural-network PDF 3.0: R. Ball et al. *arXiv:1410.8849*

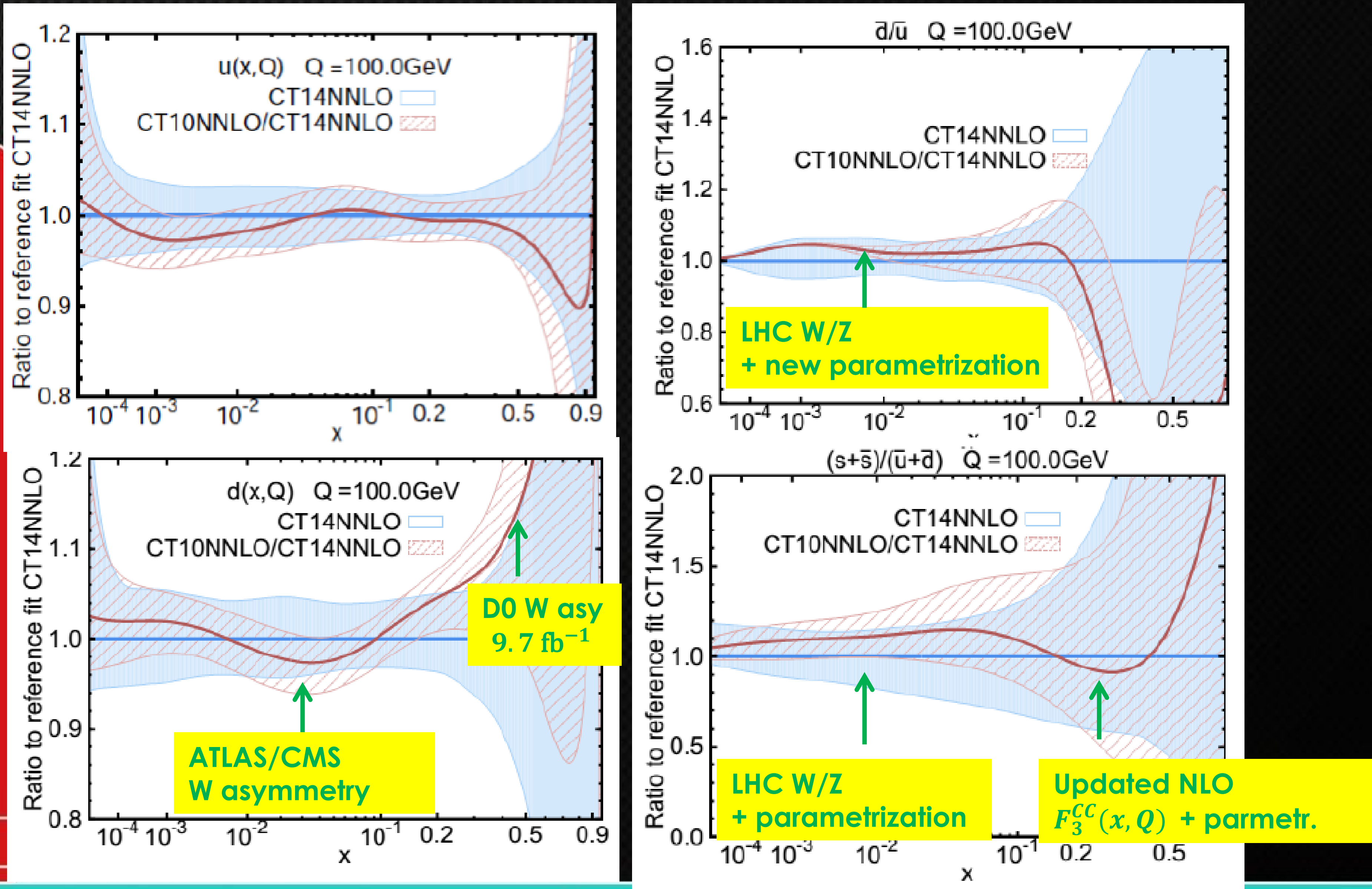
HERA2.0: in preparation

Based on DIS and jet production at HERA

CT14 PDFs: an extensive update of CT10

- Submitted to LHAPDF6; the paper is under final revisions.
- Includes LHC data from W, Z, jet production; D0 electron charge asymmetry ($9.7\ \mathrm{fb}^{-1}$); combined HERA data
- **NNLO theory with massive heavy quarks** for neutral-current DIS, DY, W, Z production; **benchmarked NLO** for charged- current DIS and jet production
- **New functional forms for PDFs** at $\boldsymbol{Q_0}$**=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

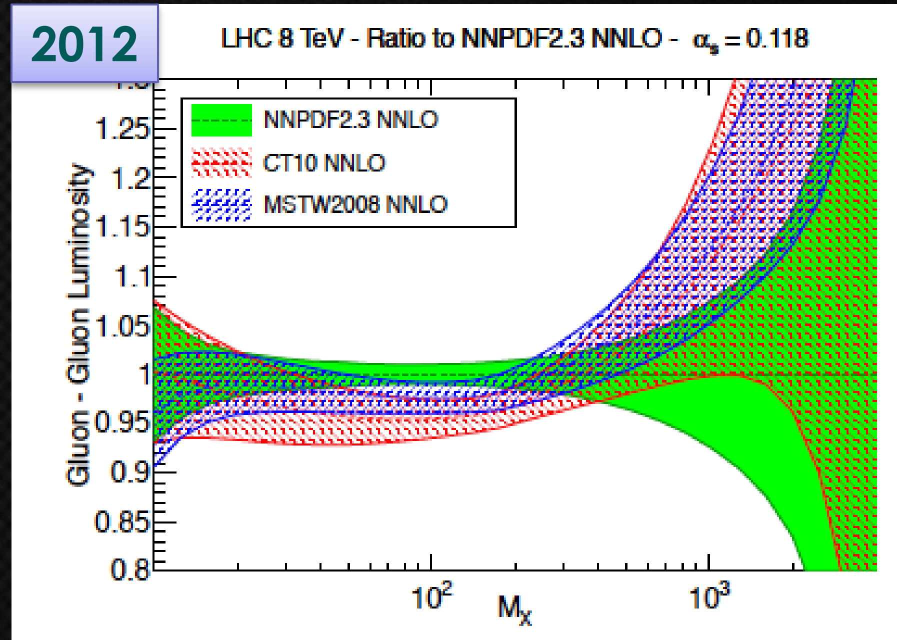
Compare CT14 and CT10 quark PDFs



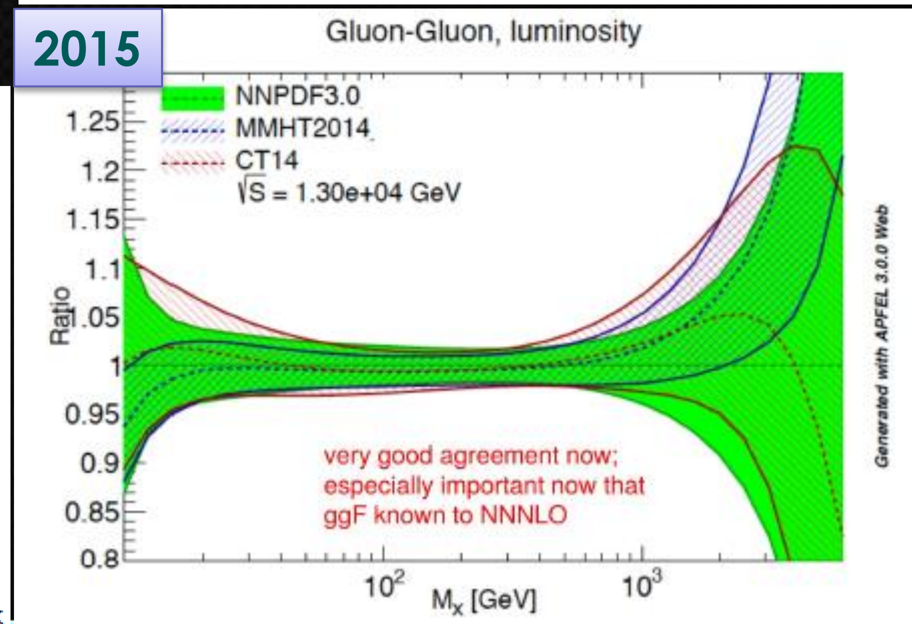
Status in 2015

- 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

2012



2015



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

NNLO PDFs of the latest generation are in better agreement because of methodological advances

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

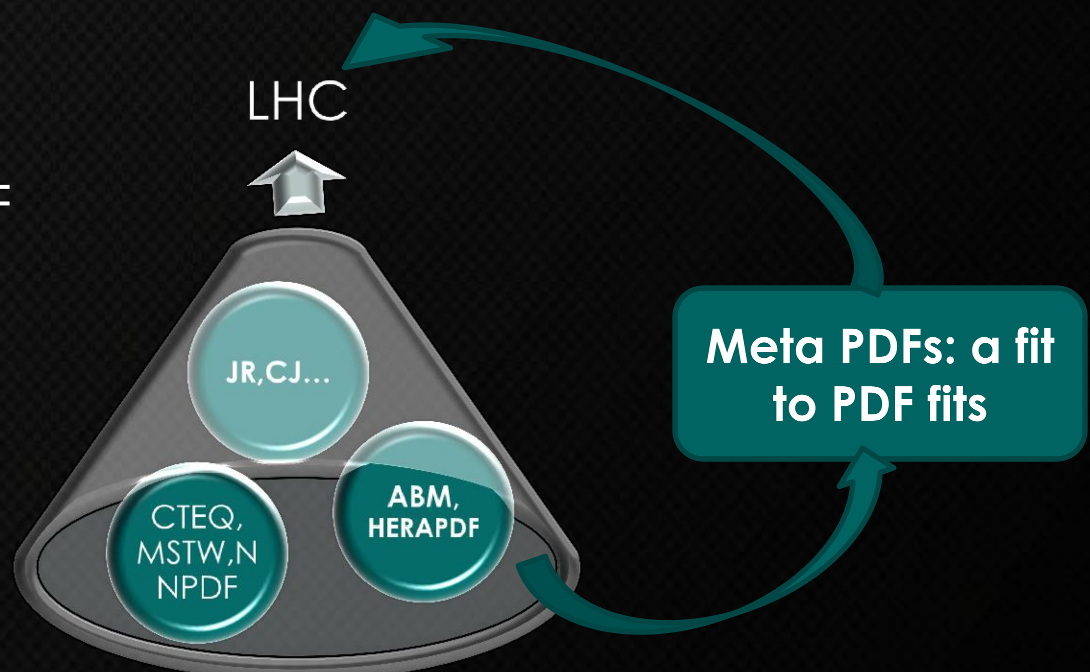
Methodologies of all groups were cross-validated and improved.

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

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



Reduction of the error PDFs

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

In the META2.0 study:
208 CT'14, MMHT'14, NNPDF3.0 error sets

⇒ 600 MC replicas for reconstructing the combined probability distribution

⇒40, 60, or 100 Hessian META sets for most LHC applications (**general-purpose** ensembles META2.0)

⇒ 13 META sets for LHC Higgs production observables (**reduced ensemble** META LHCH, obtained using the method of data set diagonalization)

META PDFs: A working example of a meta-analysis

See arXiv:1401.0013 for details

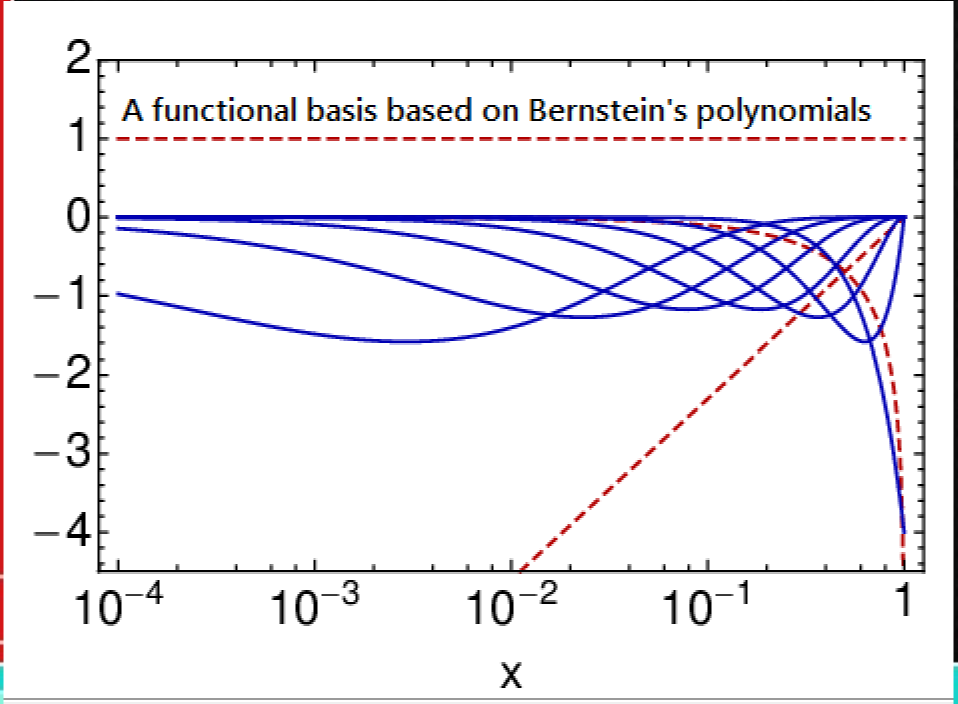
1. Select the input PDF ensembles (CT, MSTW, NNPDF…)
2. Fit each PDF error set in the input ensembles by a common functional form ("**a meta-parametrization**")
3. Generate many Monte-Carlo replicas from meta-parametrizations of each set to investigate the probability distribution on the ensemble of 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.

META PDFs: functional forms

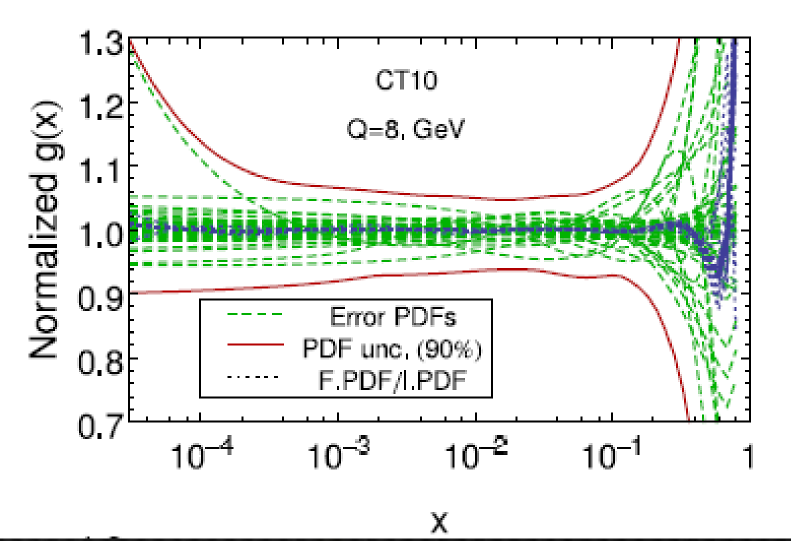
v. 1.0: Chebyshev polynomials (**Pumplin, 0909.5176, Glazov, et al., 1009.6170, Martin, et al., 1211.1215)**

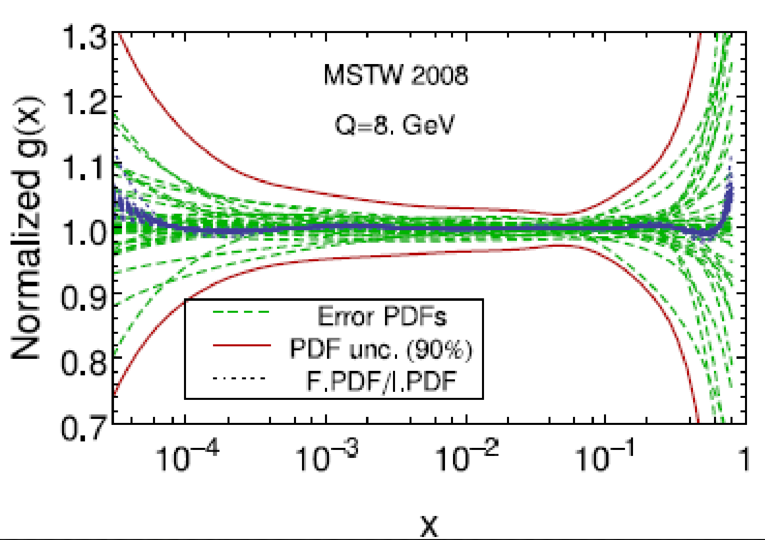
v 2.0: Bernstein polynomials ⇒ more faithful reproduction of the full ensemble of MC replicas. (**Pumplin**) Peaks occur at different x, reducing correlations between PDF parameters.

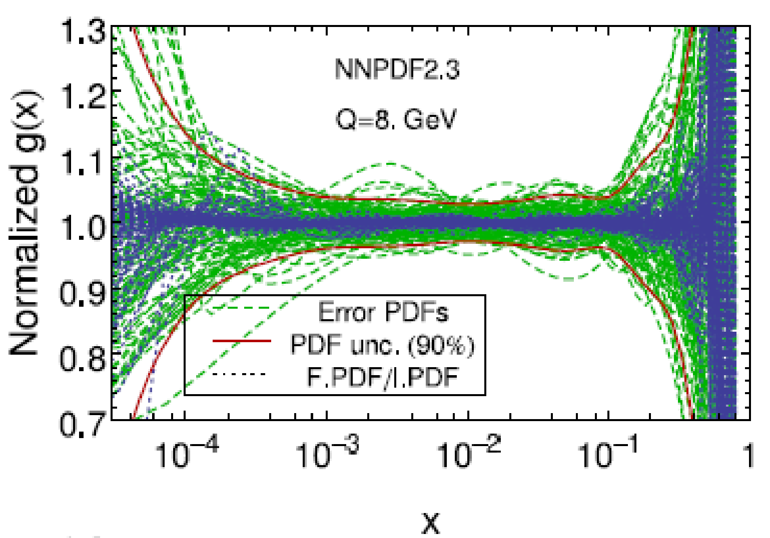
The initial scale of DGLAP evolution is Q_0=8 GeV. N_f=5.



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 final META1.0 parameterizations (F.PDF) reproduce well individual input error PDFs (I.PDF) from all groups

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"

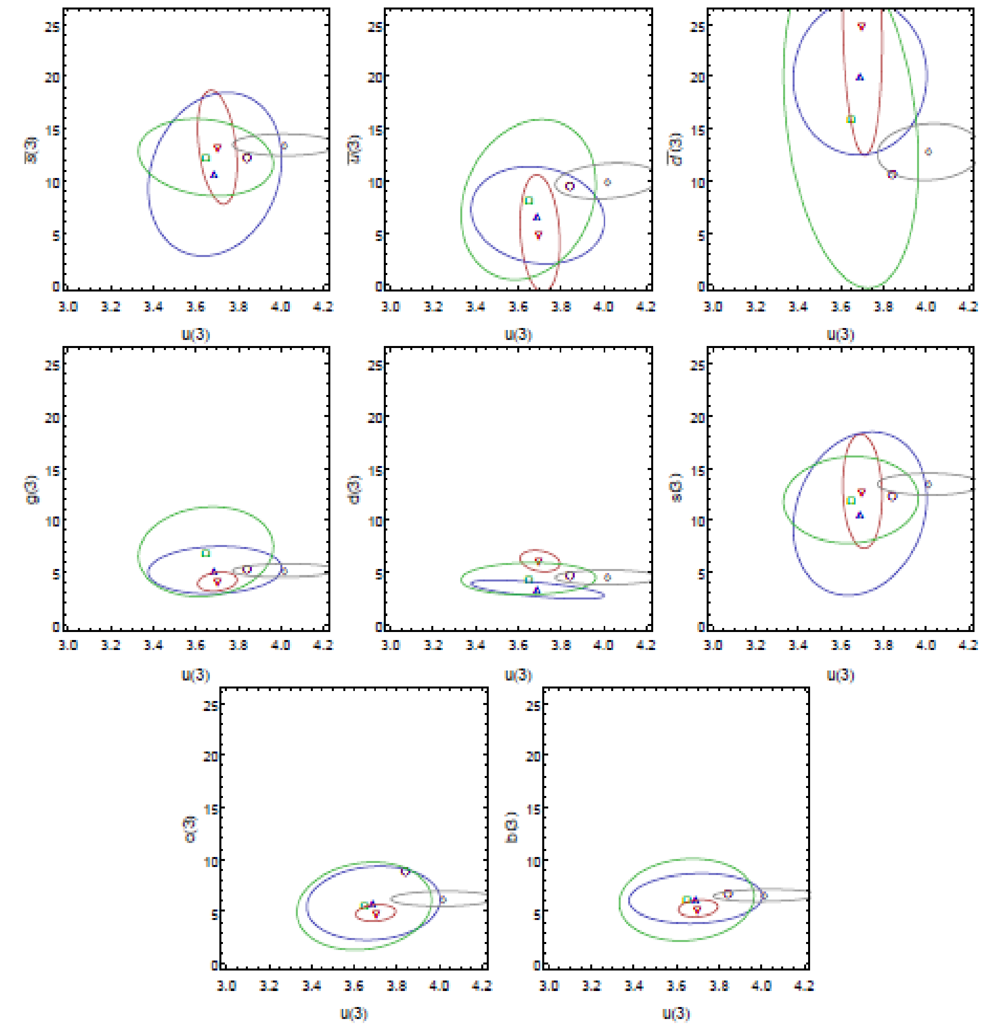


Figure 10: Fitted PDF parameters and 90% c.l. ellipses for CT10 (blue up triangle), MSTW08 (red down triangle), NNPDF2.3 (green square), HERAPDF1.5 (gray diamond) and ABM11 (magenta circle).

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:

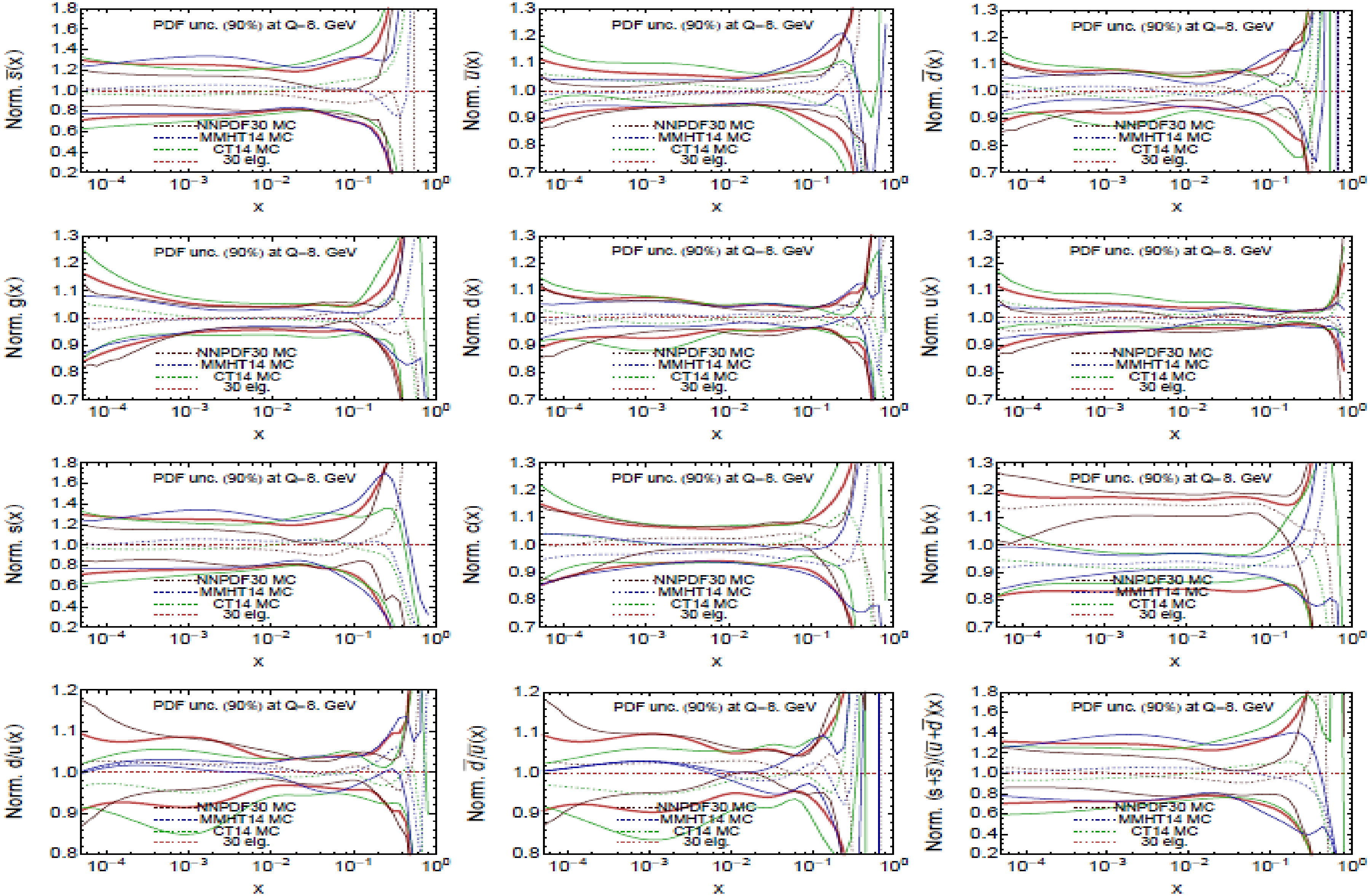
$$\langle a_i \rangle_g = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} a_i(k),$$

Central value on g

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

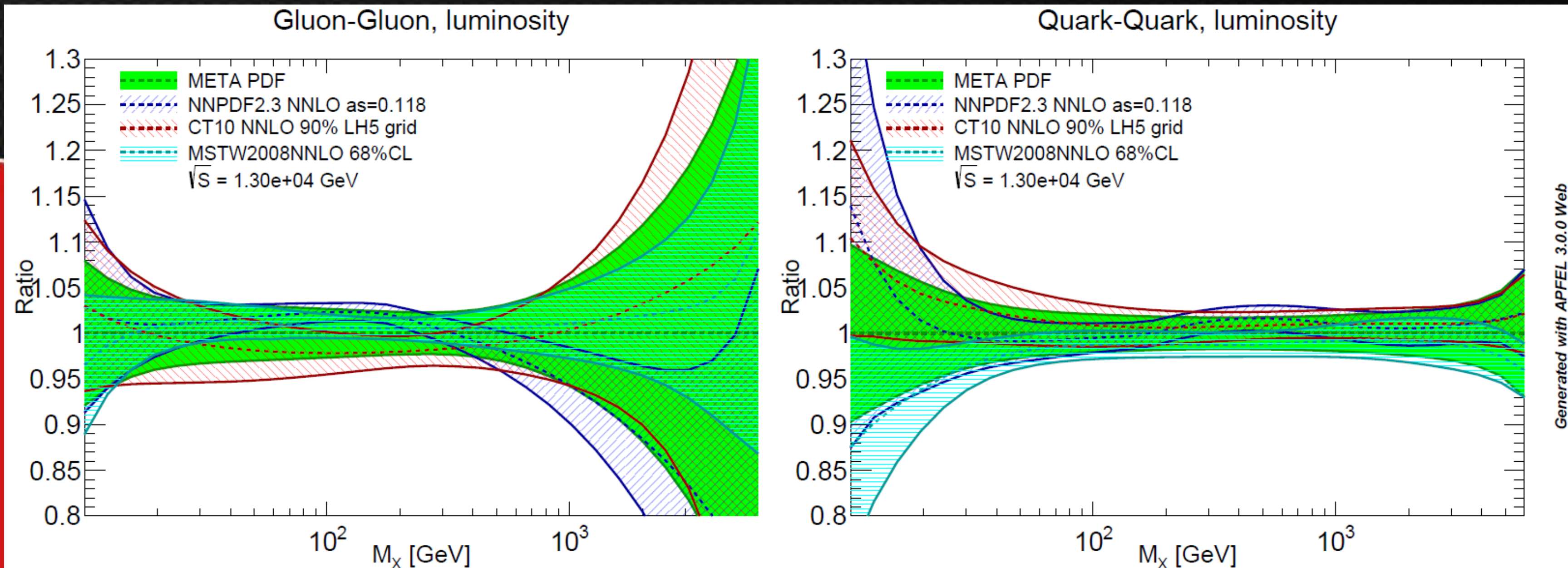
$$(\delta a_i)_g = \sqrt{\mathrm{cov}(a_i, a_i)_g}.$$

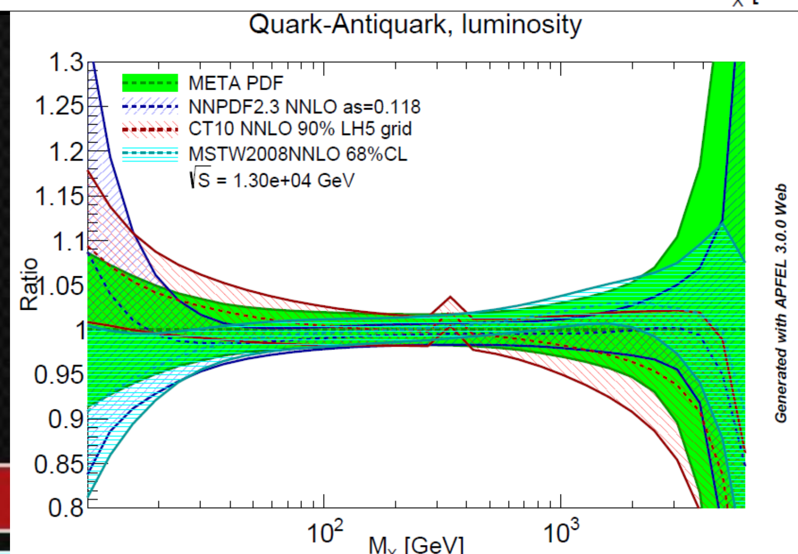
Standard deviation on g



The META60 ensemble “averages out” non-Gaussian features of input PDFs and their ratios from CT, MMHT, NNPDF MC sets

Some parton luminosities





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

- More illustrations of the META approach are in backup slides.
- The META methodology is very flexible. Special META ensembles can be constructed that reproduce the PDF classes for large classes of LHC observables (such as in Higgs production) with a small number of error PDFs

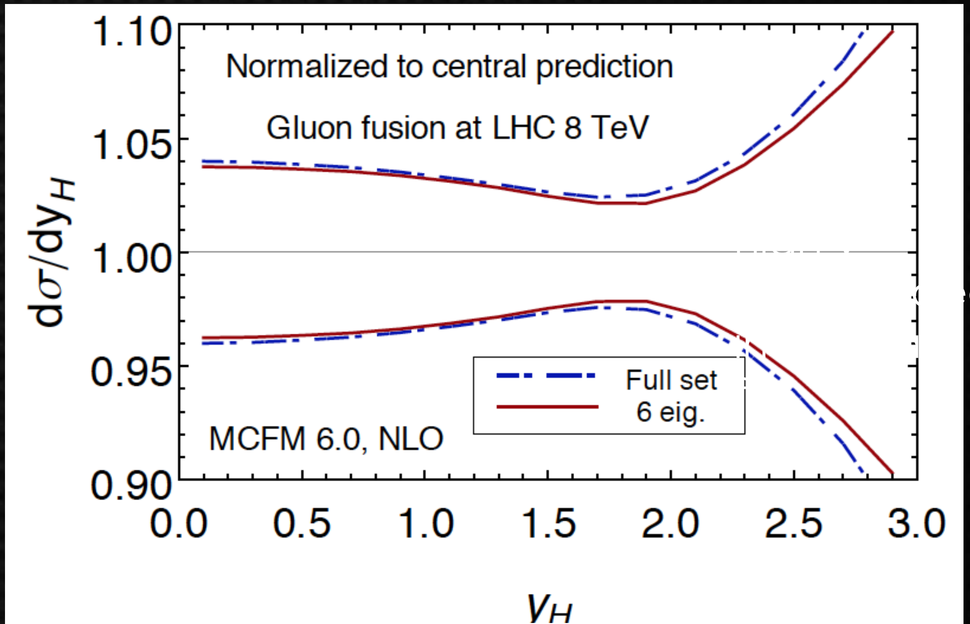
Reduced META ensemble

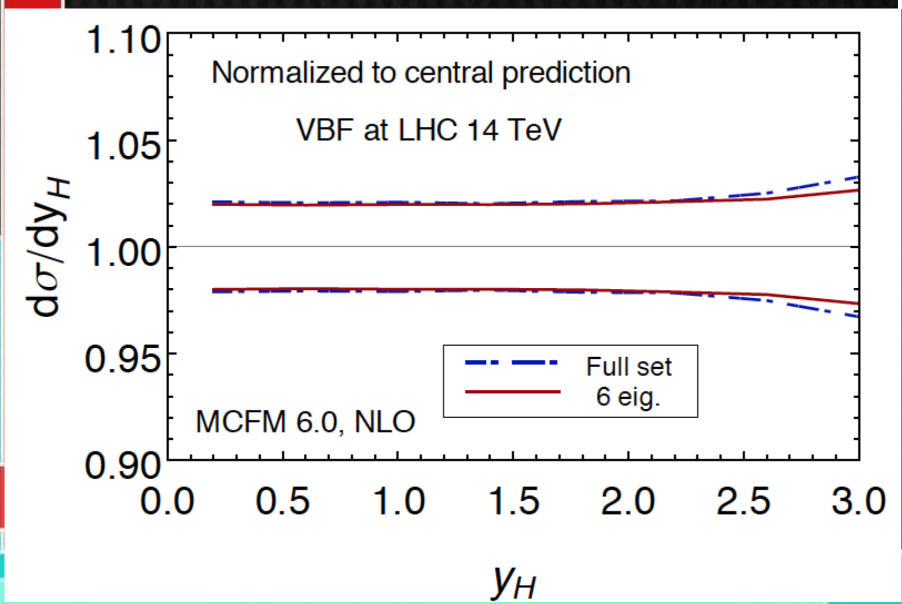
- Already the general-purpose ensemble reduced the number of error PDFs needed to describe the LHC physics; but we can further perform a data set diagonalization to pick out eigenvector directions important for Higgs physics or another class of LHC processes
- Select global set of Higgs cross sections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

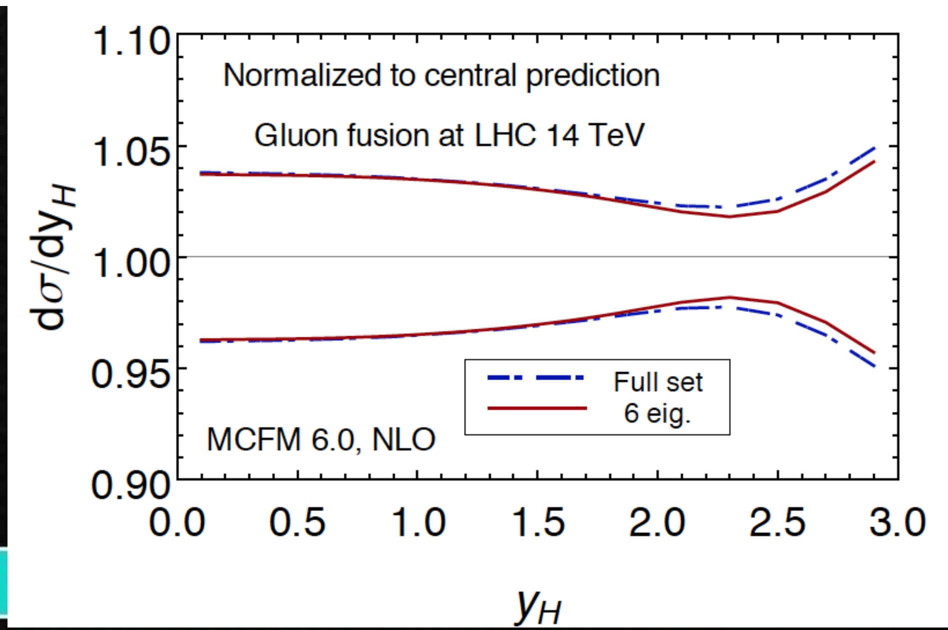
production channel	$\sigma(inc.)$	$\sigma(\lvert y_H\rvert > 1)$	$\sigma(p_{T,H} > m_H)$	scales
$gg \to H$	iHixs1.3 [32] at NNLO	MCFM6.3 [33] at LO	—	m_H
$b\bar{b} \to H$	iHixs at NNLO	—	—	m_H
VBF	VBFNLO2.6 [34] at NLO	same	same	m_W
HZ	VHNNLO1.2 [35] at NNLO	CompHEP4.5 [36] at LO	CompHEP at LO	$m_Z + m_H$
$HW^\pm$	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	8.10
	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	7.25
	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	5.08
	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	4.20
	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

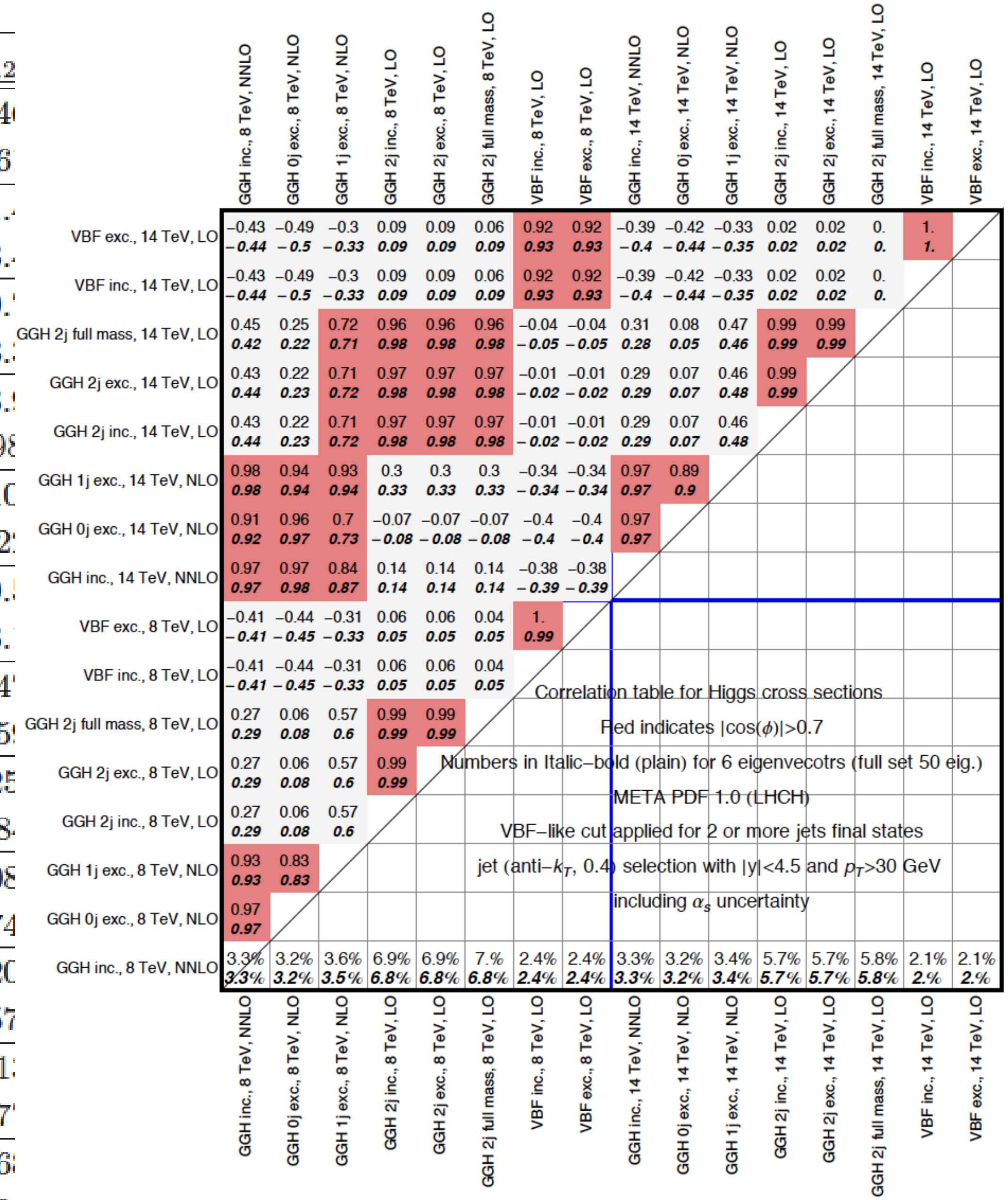


FIG. 7: Same as Fig. 5, with α_s uncertainties included by adding in quadrature.

Looking forward: Combination of the PDFs into the future PDF4LHC ensemble

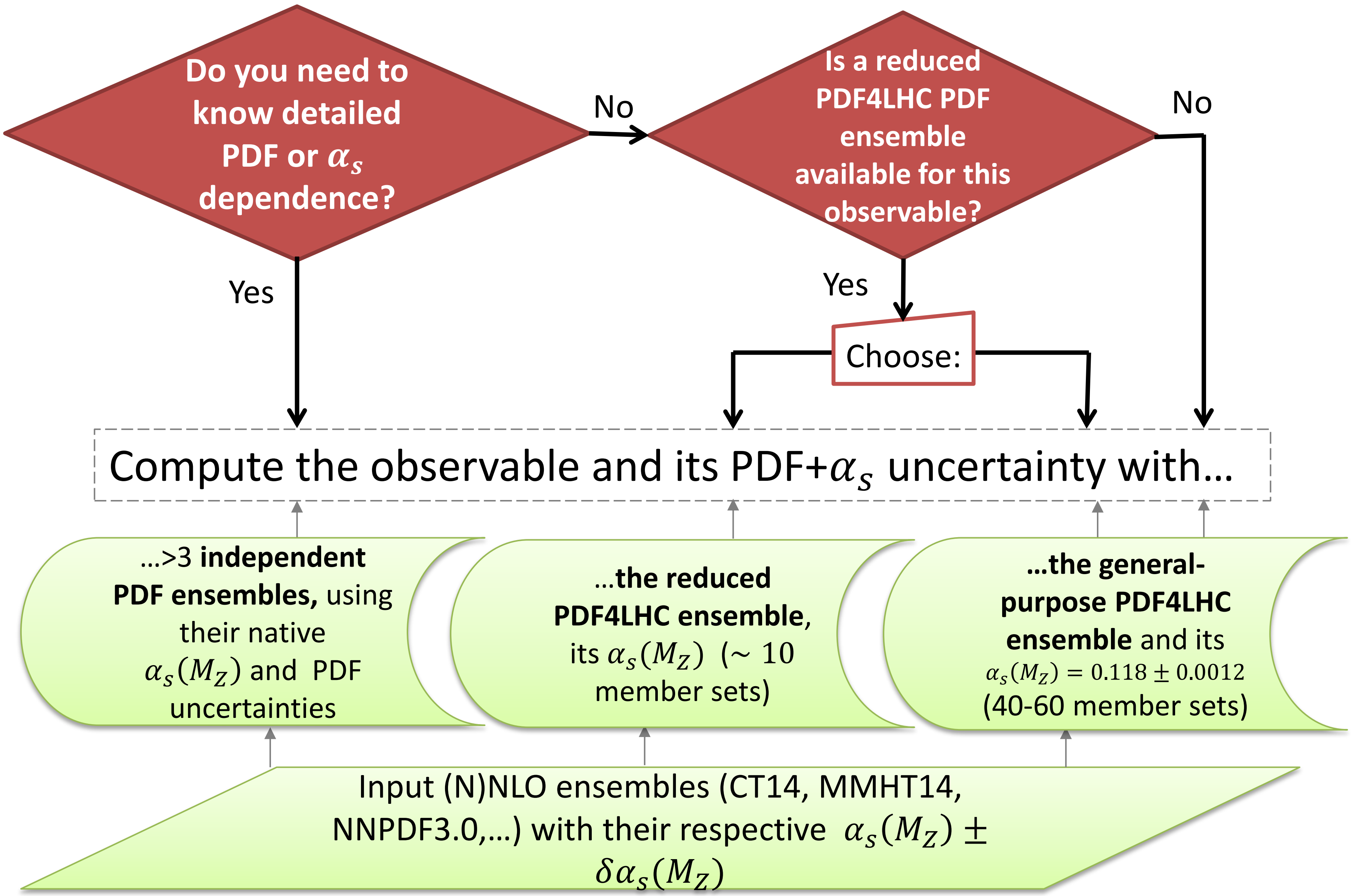
PDFs from several groups are combined into a PDF4LHC ensemble of error PDFs **before** the LHC observable is computed. This simplifies the computation of the PDF+α_s uncertainty, cuts down the number of the PDF member sets needed for simulations.

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.

2015: A concept for a new PDF4LHC recommendation



This procedure applies both at NLO and NNLO

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)

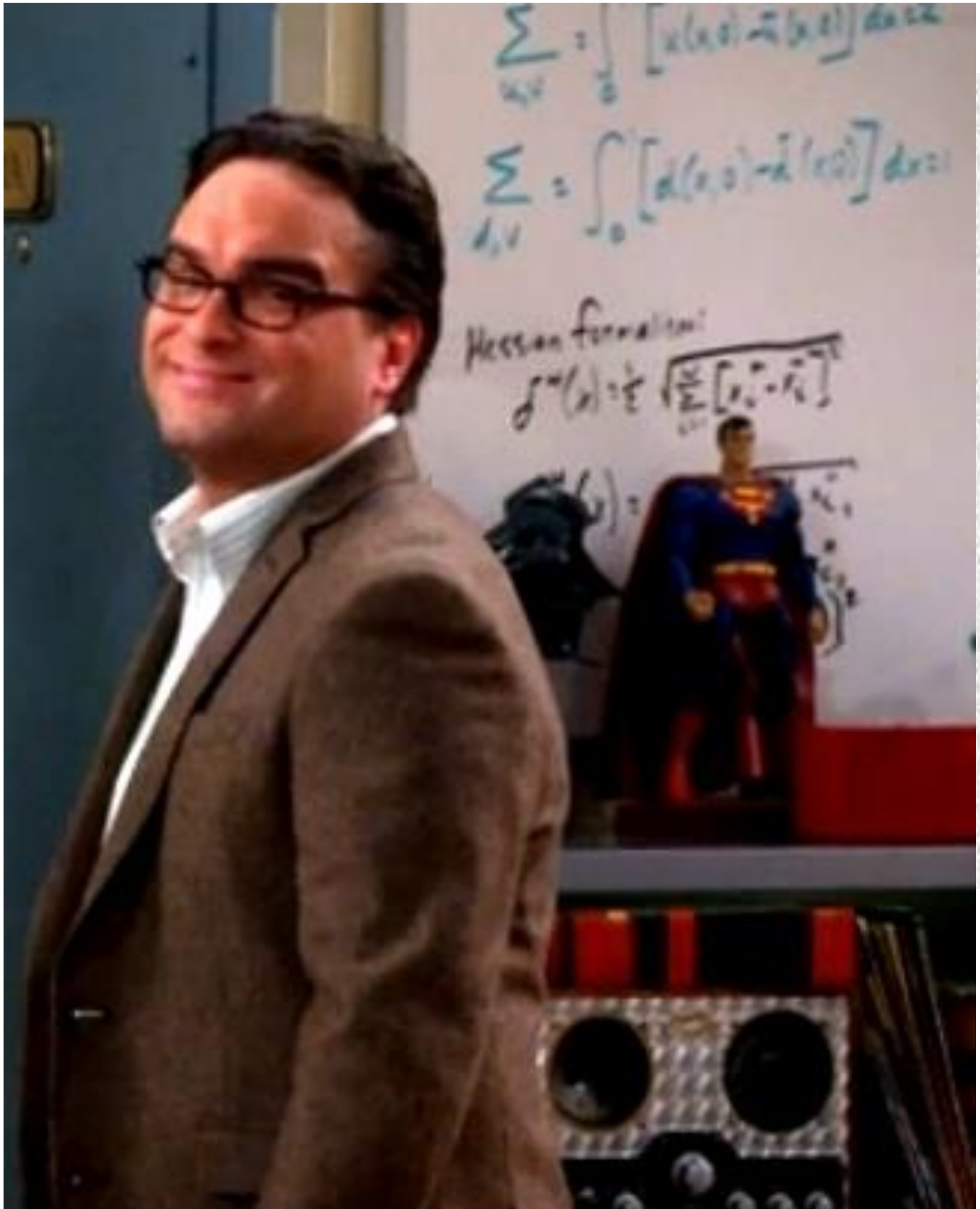
Both procedures start by creating a combined ensemble of MC replicas from all input ensembles (G. Watt, R. Thorne,1205.4024; S. Forte, G. Watt, 1301.6754). They differ at the second step of reducing a large number of input MC replicas (~ 300) to a smaller number for practical applications (13-100 in 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.

Strong sides of both methods were examined in great detail (cf. backup slides) and discussed at several PDF4LHC meetings and within collaborations

The Big Bang Theory. The Anxiety Optimization. S8, Ep. 13

/

Sheldon accepts the META PDF's, but only at the 68% c.l. He claims a low $\alpha_s(M_Z) = 0.1118$ at 6 resummed loops. We'll



Both META and CMC ensembles need of order 40 final members

Hessian error PDFs are preferable for estimates of systematic uncertainties in many situations. Reduced Hessian sets will enable studies of PDF-induced correlations between signal and background processes

A Mathematica package MP4LHC

- Includes all tools for the META analysis
- Will be public

```
In[1]:= (*Load  the Module*)
SetDirectory[NotebookDirectory[]];
<< MP4LHC.m

      MP4LHC v2.0 (META PDFs for LHC st

   Jun Gao, Pavel Nadolsky, Joey Huston (11/20/2014)

          arXiv: 1401.xxxx, 1410.xxxx

============================================================

|*   A notebook code for operations on META PDFs

|*

|*   Including the following functions:

|*  (1) Perform META fit of the MC samples

|*  (2) Generate META Hessian set based on META fit

|*  (3) Rediagonalization of the META Hessian set

============================================================

(*Generation of the META PDFs from MC samples*)

In[3]:= (*Load MC Sample1, arguments [Dir,name,Nmc,0/1 to initiate/load weight file]*)
(*can add "nostrangenessLM" in last arg. to disable LM penalty on strangeness*)
mctag = {"mcCT14j365", "mcMMHT1416", "mcNN3016", "mcHE2016"};
astag = Table[mctag[[i]] <> "A", {i, 1, Length[mctag]}];
```

To summarize, the meta-parametrizations and Hessian method have been thoroughly validated

- A general and intuitive framework. 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.

Back-up slides

META 2.0 PDFs

META parameters of PDFs

- The core idea of the meta-analysis is to cast all input PDFs into a shared parametric representation, then keep only relevant parameters upon diagonalization of the Hessian matrix
- Two methods to obtain Hessian meta-parameterizations have been developed
  - **META2.0:** By fitting $f_i(x, Q)$ by flexible functions $F_i(\{a\}; x, Q)$, such as those based on Bernstein polynomials *(our approach)*
  - **MC2Hessian:** By treating the PDF values themselves as parameters, $f_i(x_j, Q_l) \equiv f_{ijl}$ *(Carrazza et al., 1505.06736).*

This is equivalent to selecting $\mathrm{F_i(\{a\}; x, Q) = f_{ijl}\delta(x - x_j)\delta(Q - Q_l)}$.

Meta-parameters of 5 sets and META PDFs

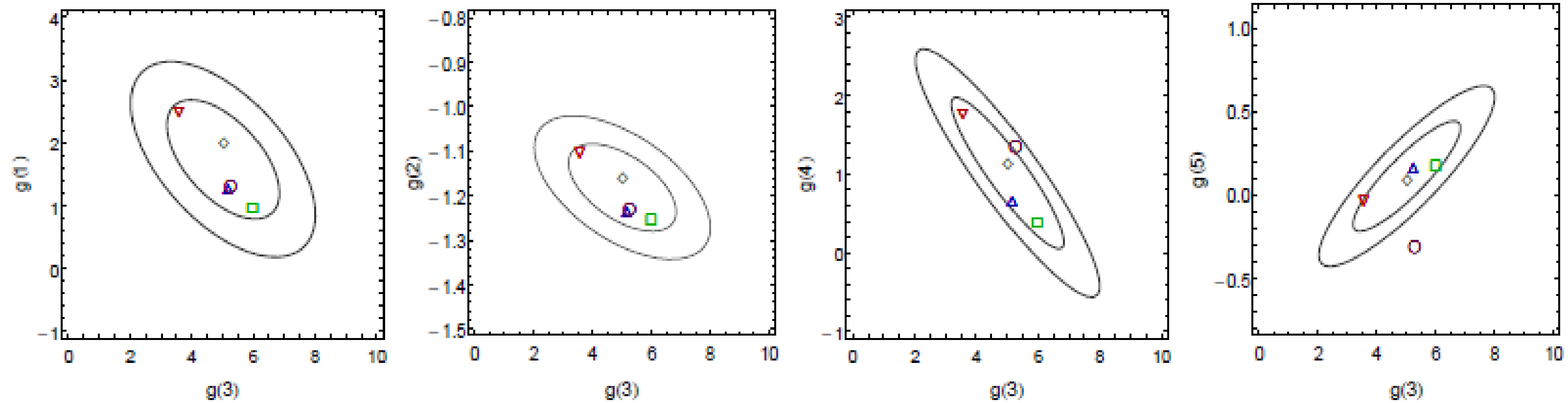


Figure 16: Comparison of META PDF confidence intervals with central NNLO PDFs of the input PDF ensembles in space of meta-parameters a_{1-5} for the gluon PDF. Up triangle, down triangle, square, diamond, and circle correspond to the best-fit PDFs from CT10, MSTW, NNPDF, HERAPDF, and ABM respectively.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

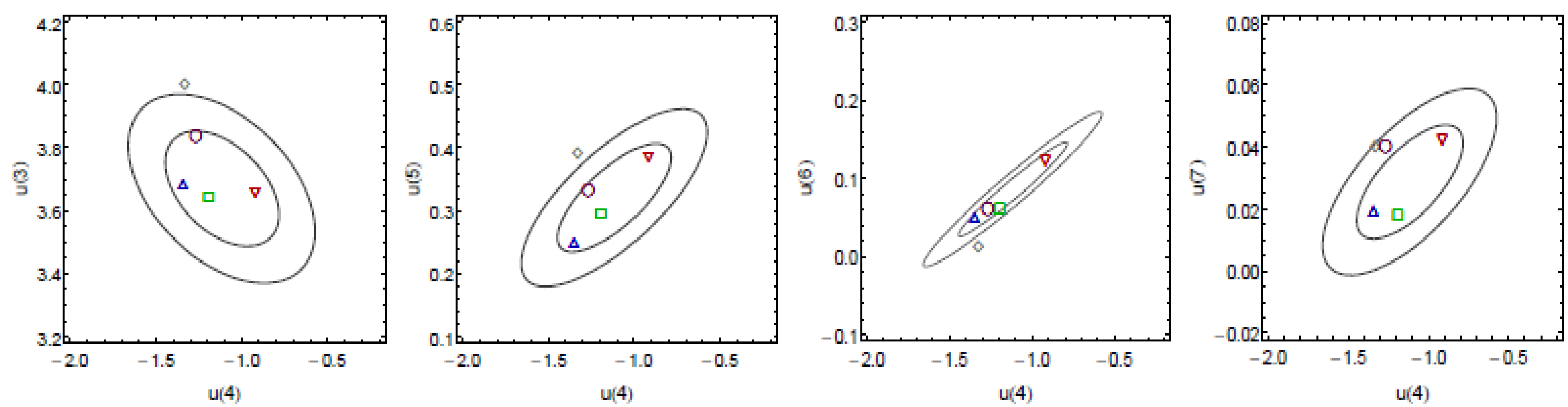
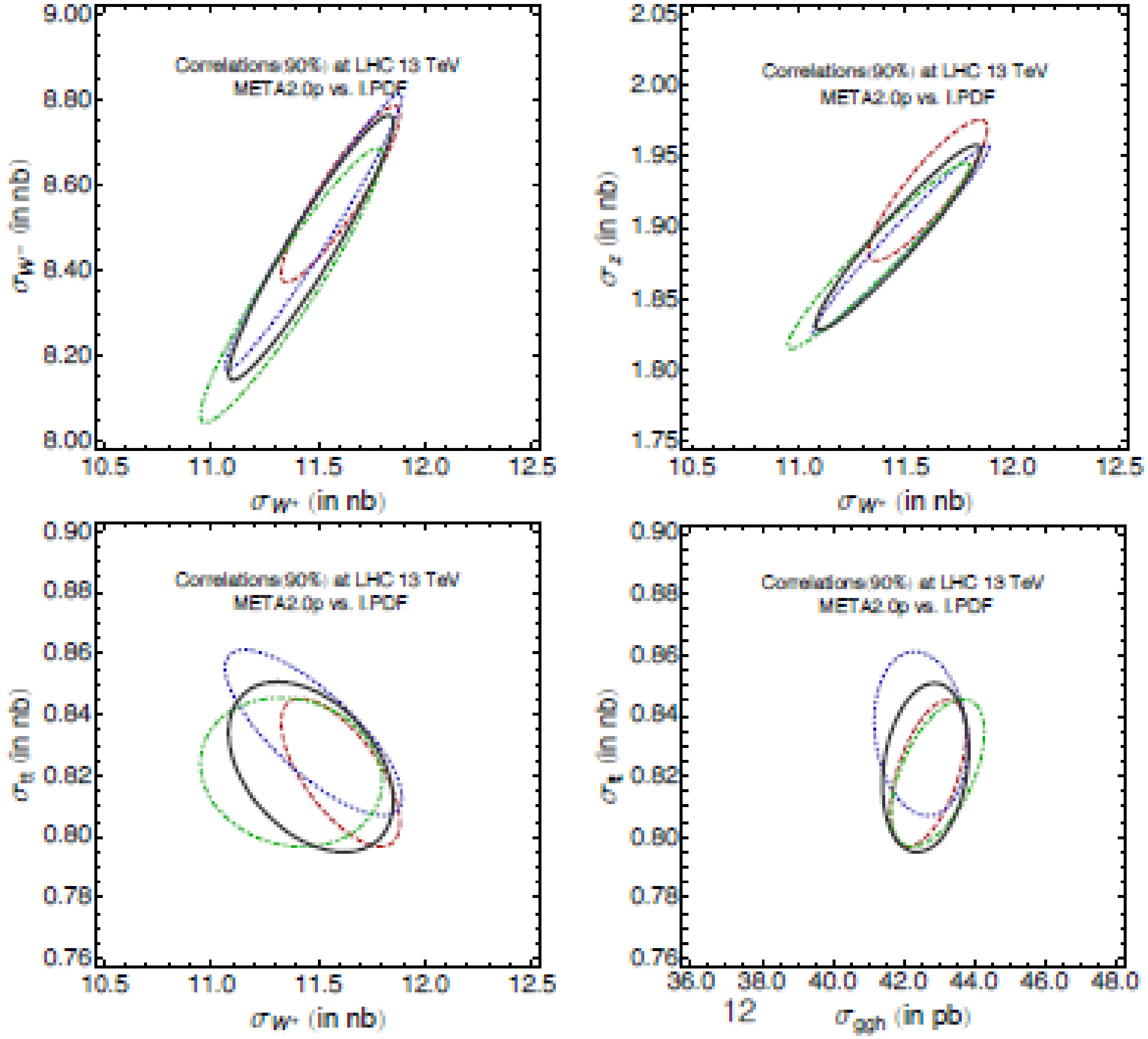


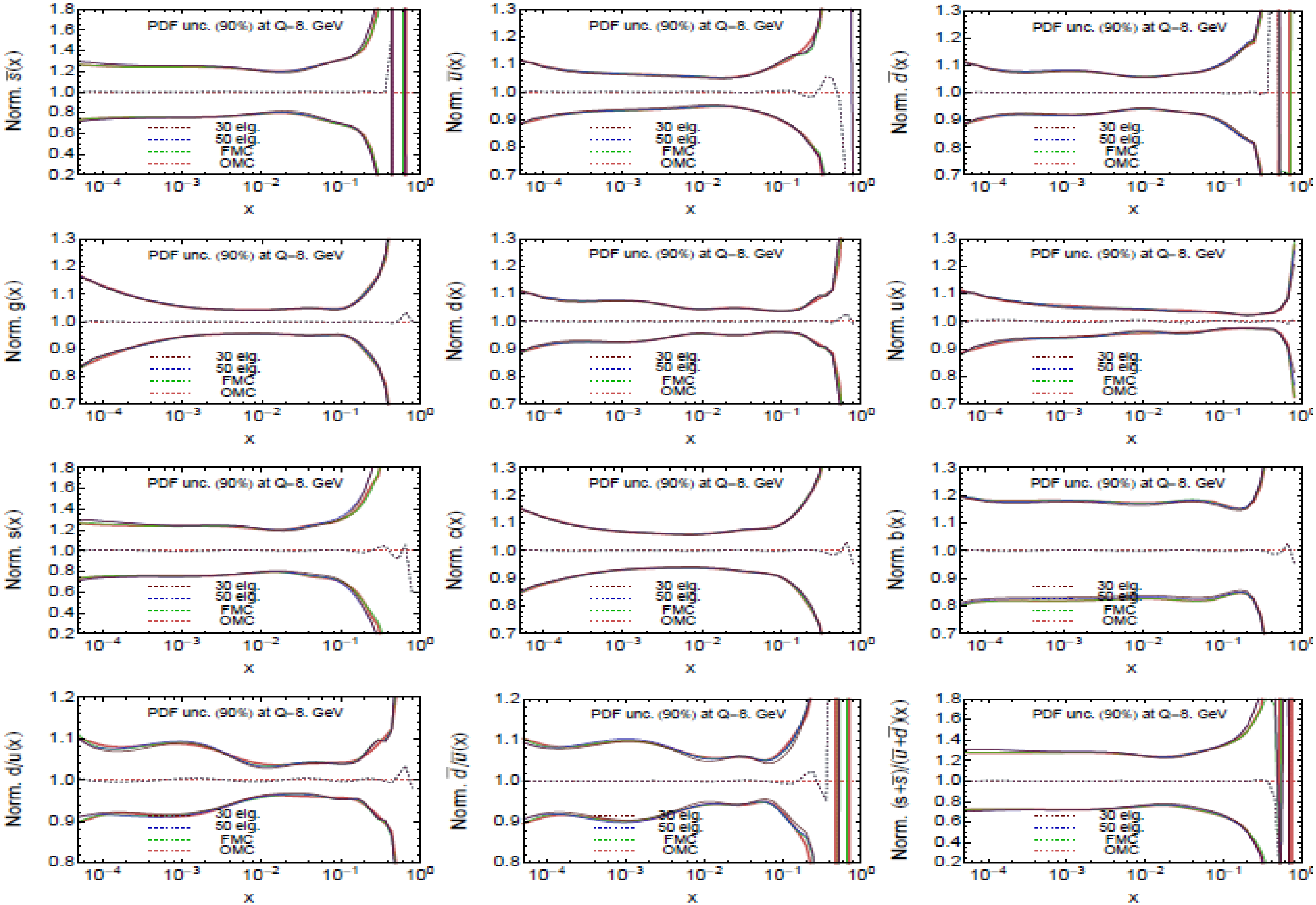
Figure 17: Same as Fig. 16, for a_{3-7} of the u quark PDF.

META2.0 predictions for LHC observables

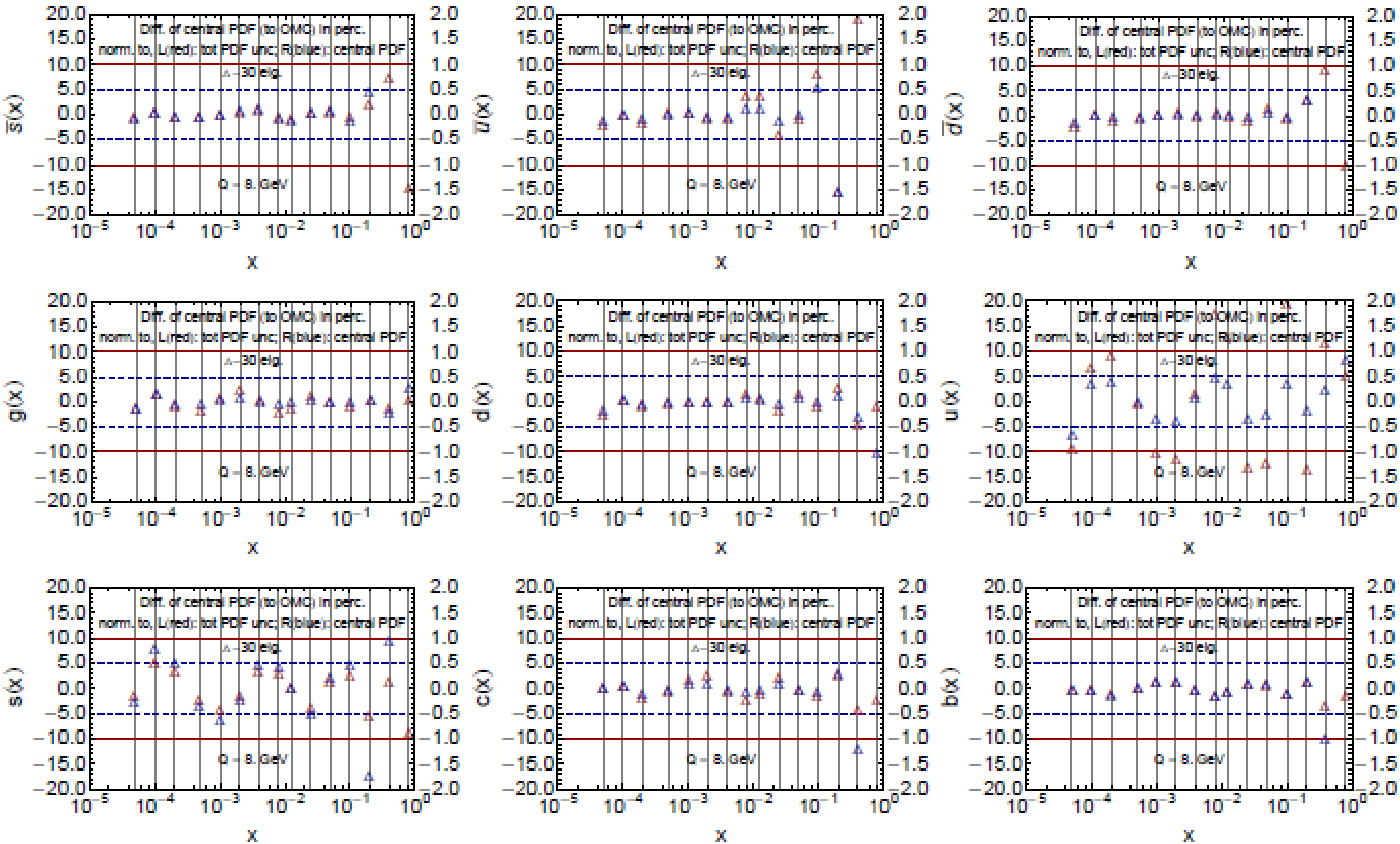
- Currently only have results for META NNLO v2.0p, will add later for v2.1, inclusive observables at 13 TeV



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



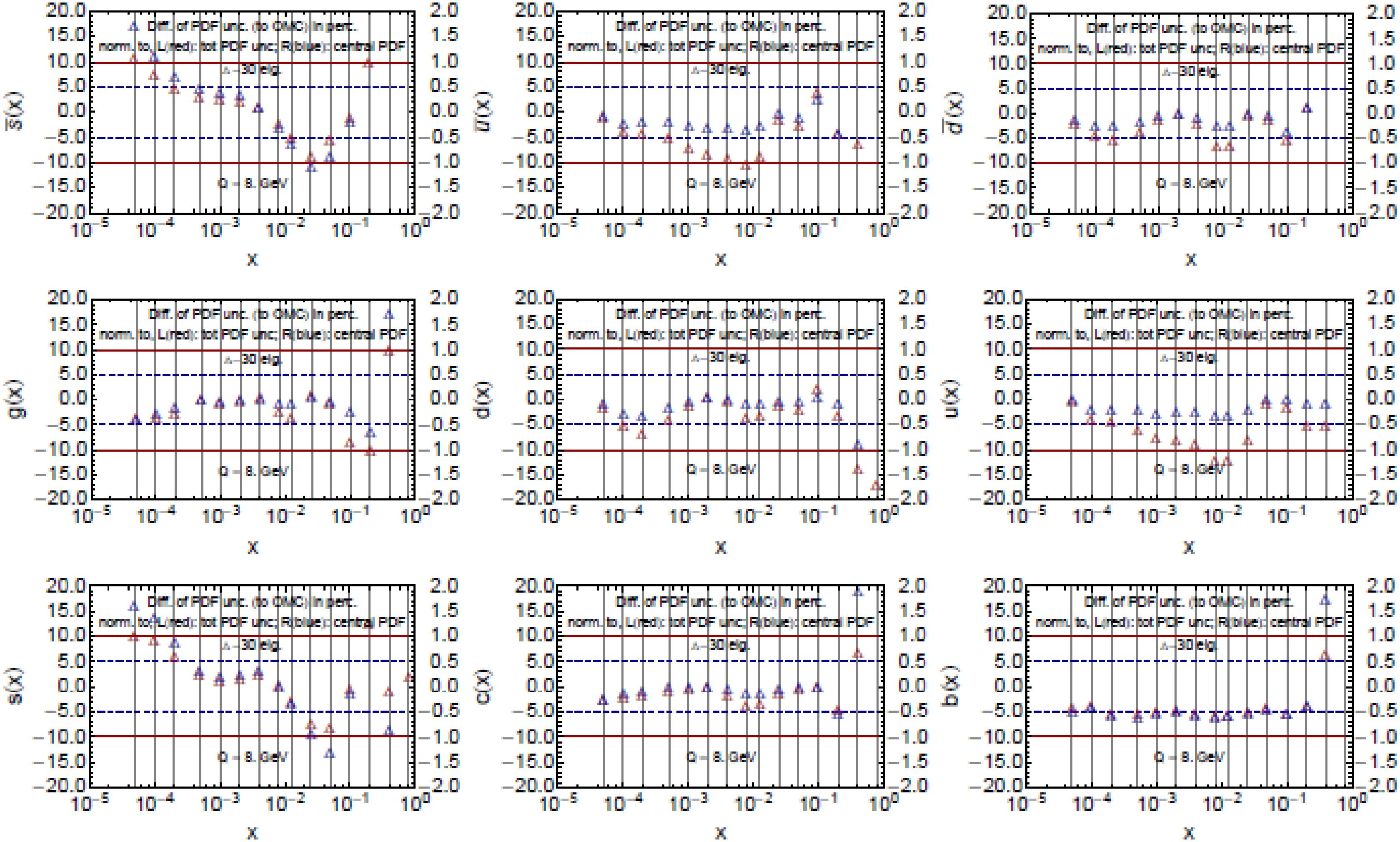
PDF uncertainty bands from original 600 MC replicas (OMC), fitted MC replicas (FMC), META60 and META100. **NO FITTING BIAS OBSERVED!**



Differences in central PDFs between META60 and MC600.

Left axis, red triangles: as percentages of PDF uncertainty

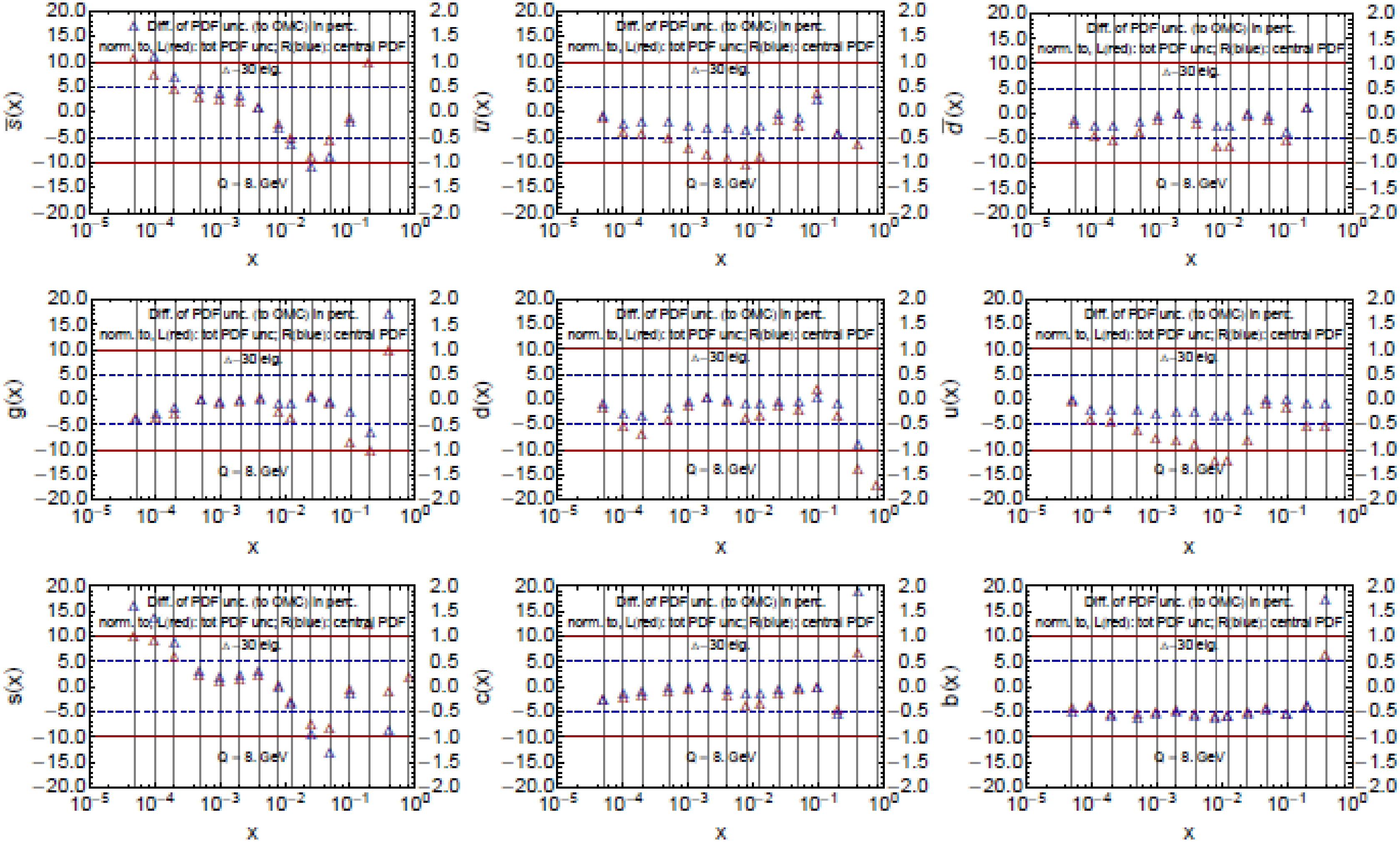
Right axis, blue triangles: as percentages of the central PDF



Differences of PDF uncertainties between META60 and MC600.

Left axis, red triangles: as percentages of the PDF uncertainty

Right axis, blue triangles: as percentages of the central PDF



Differences of PDF uncertainties between META60 and MC600.

Left axis, red triangles: as percentages of the PDF uncertainty

Right axis, blue triangles: as percentages of the central PDF

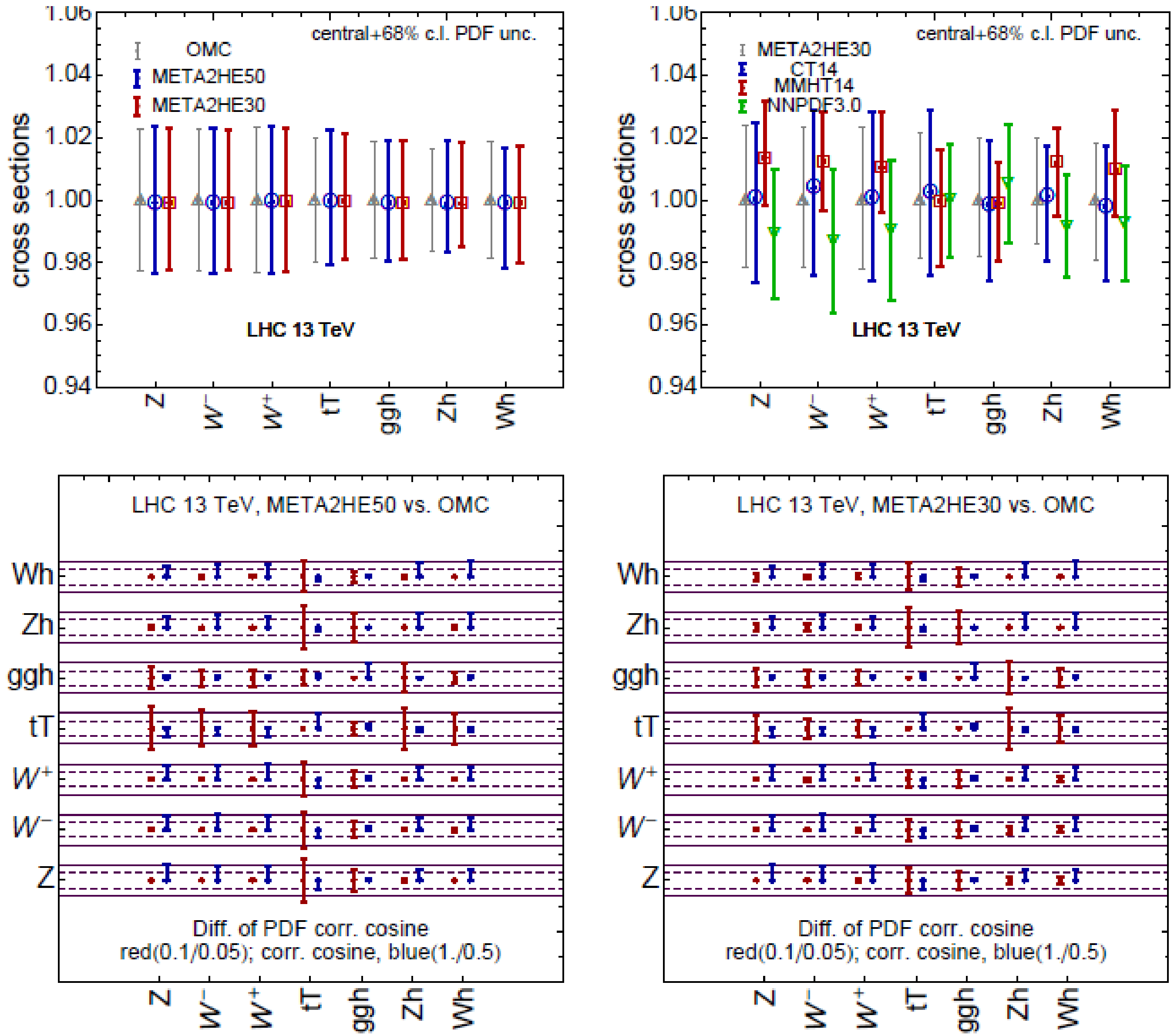


FIG. 9: Comparison on inclusive cross sections for META2HE vs. OMC, and individual PDFs vs. META2HE30; their correlations for META2HE vs. OMC.

Data set diagolization for Higgs observables

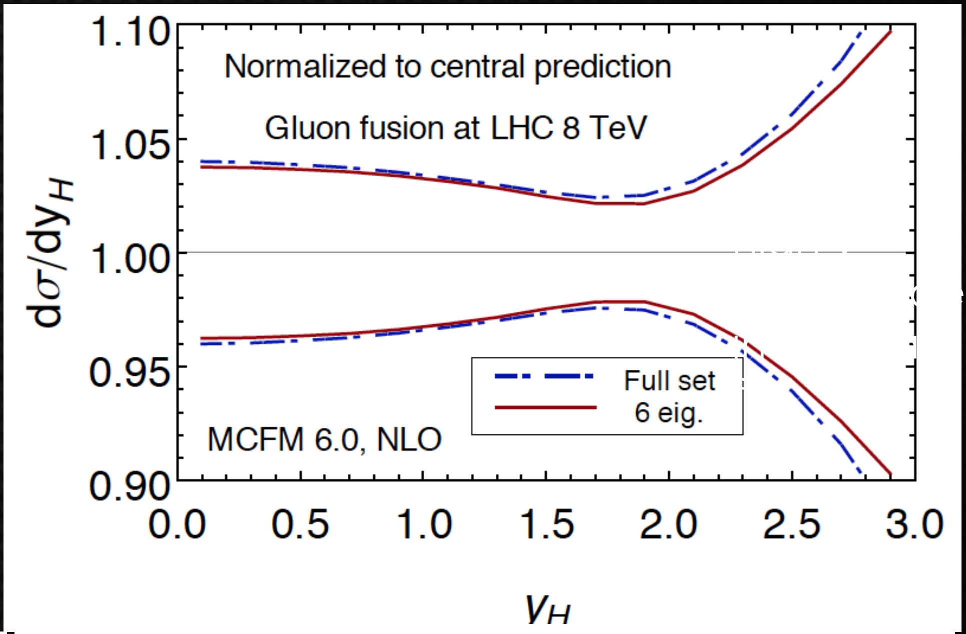
Reduced META ensemble

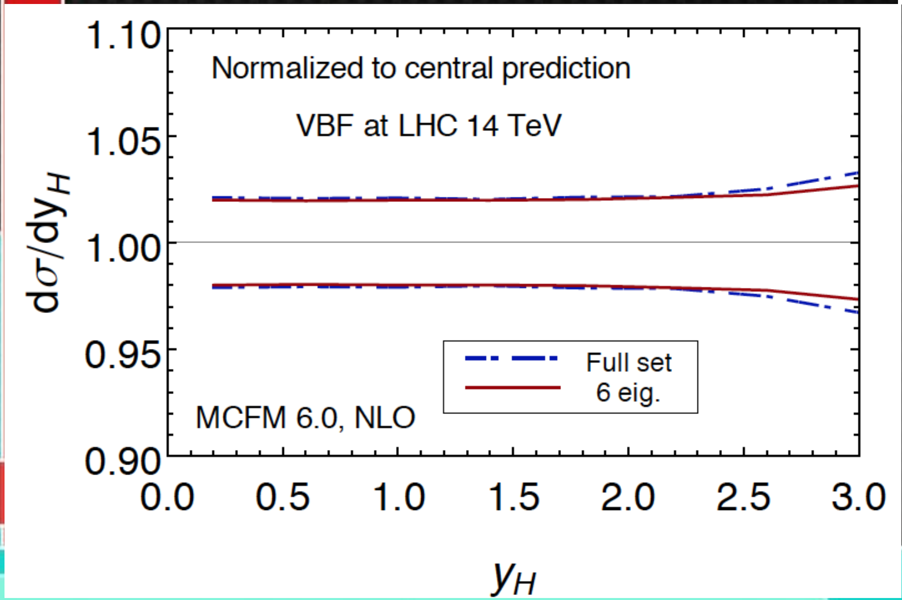
- Already the general-purpose ensemble reduced the number of error PDFs needed to describe the LHC physics; but we can further perform a data set diagonalization to pick out eigenvector directions important for Higgs physics or another class of LHC processes
- Select global set of Higgs cross sections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

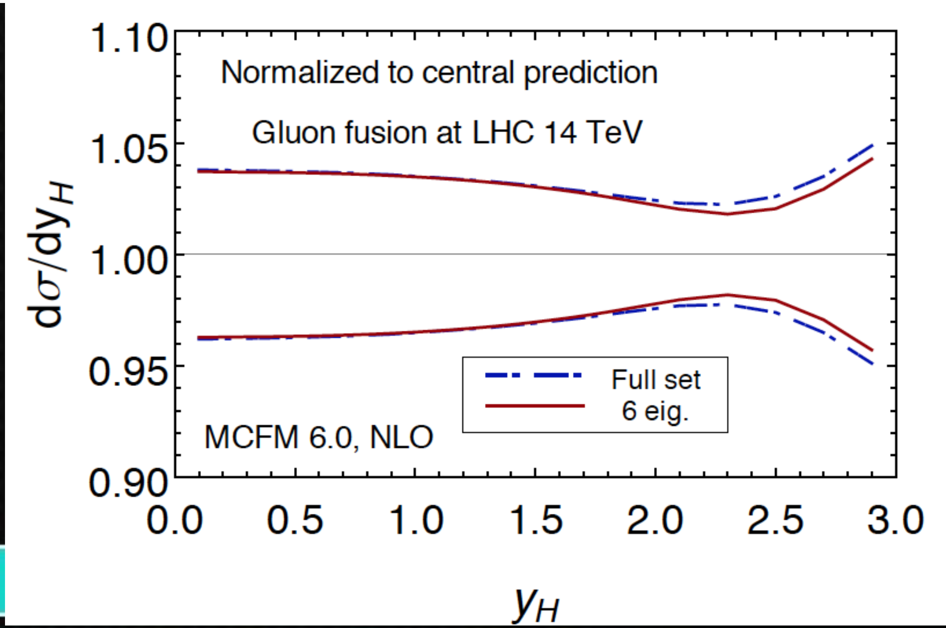
production channel	$\sigma(inc.)$	$\sigma(\lvert y_H \rvert > 1)$	$\sigma(p_{T,H} > m_H)$	scales
$gg \to H$	iHixs1.3 [32] at NNLO	MCFM6.3 [33] at LO	—	m_H
$b\bar{b} \to H$	iHixs at NNLO	—	—	m_H
VBF	VBFNLO2.6 [34] at NLO	same	same	m_W
HZ	VHNNLO1.2 [35] at NNLO	CompHEP4.5 [36] at LO	CompHEP at LO	$m_Z + m_H$
$HW^{\pm}$	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	8.7
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	7.5
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

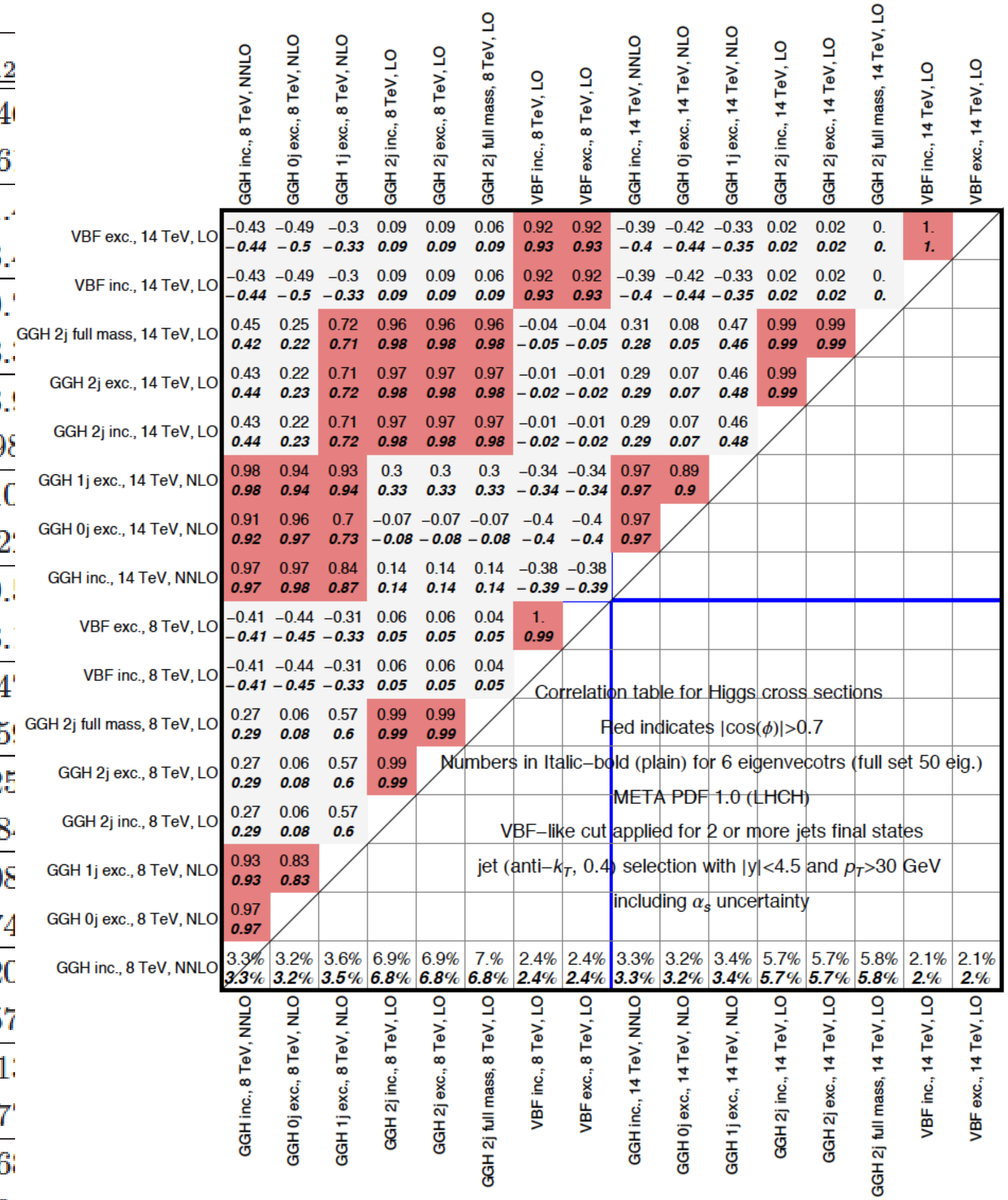
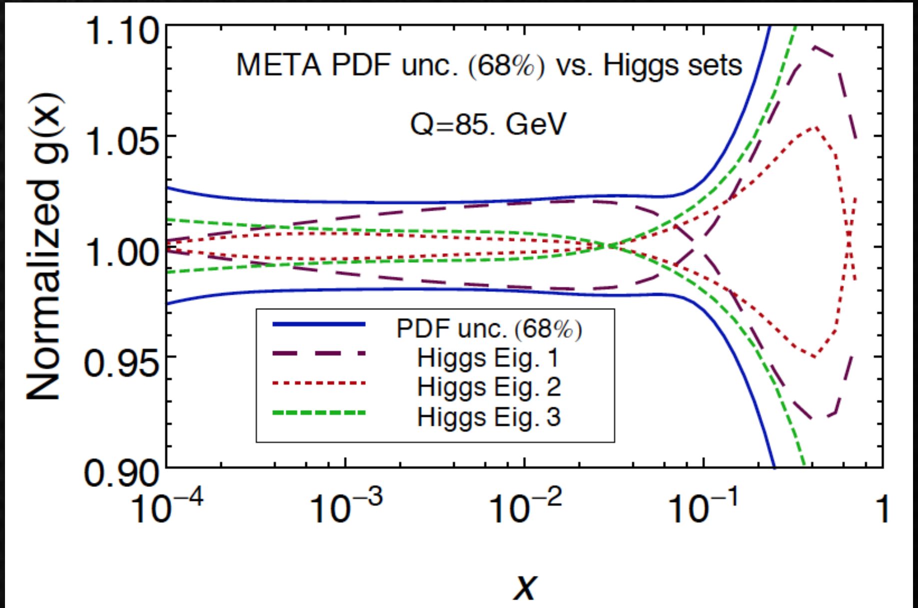


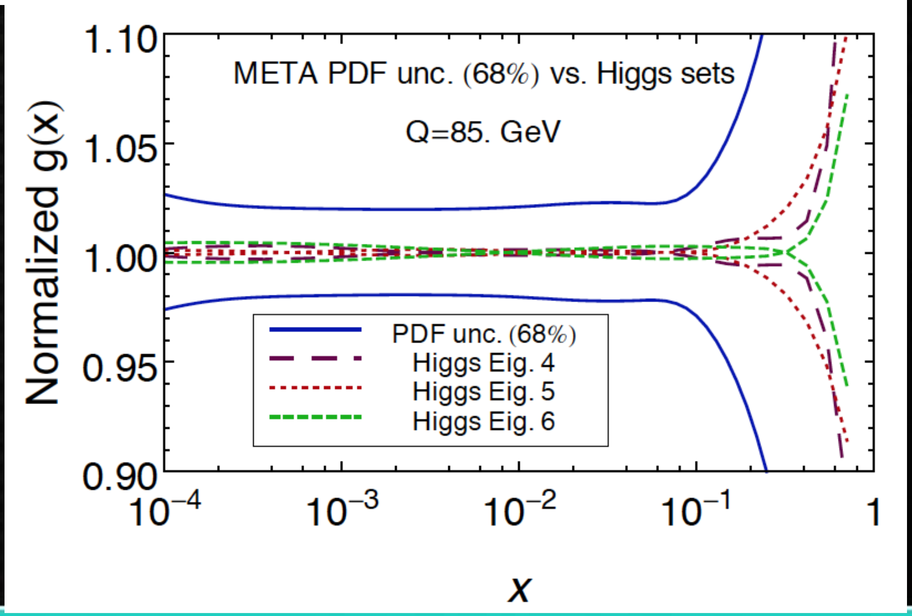
FIG. 7: Same as Fig. 5, with α_s uncertainties included by adding in quadrature.

Re-diagonalized eigenvectors…

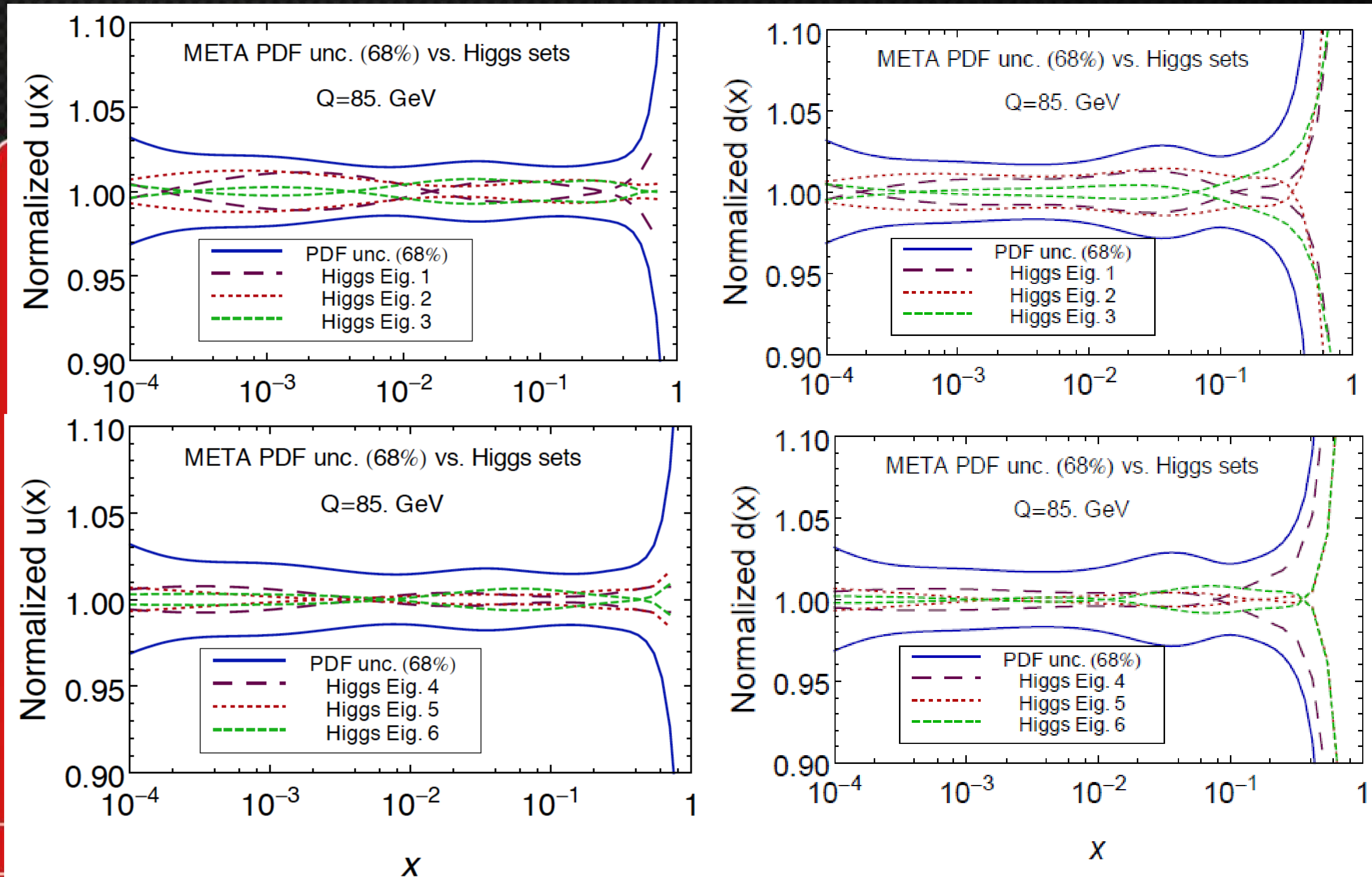
…are associated with the parameter combinations that drive the PDF uncertainty in Higgs, W/Z production at the LHC

- 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



Comparisons of CMC and META approaches

Progress in developing the combination procedure

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

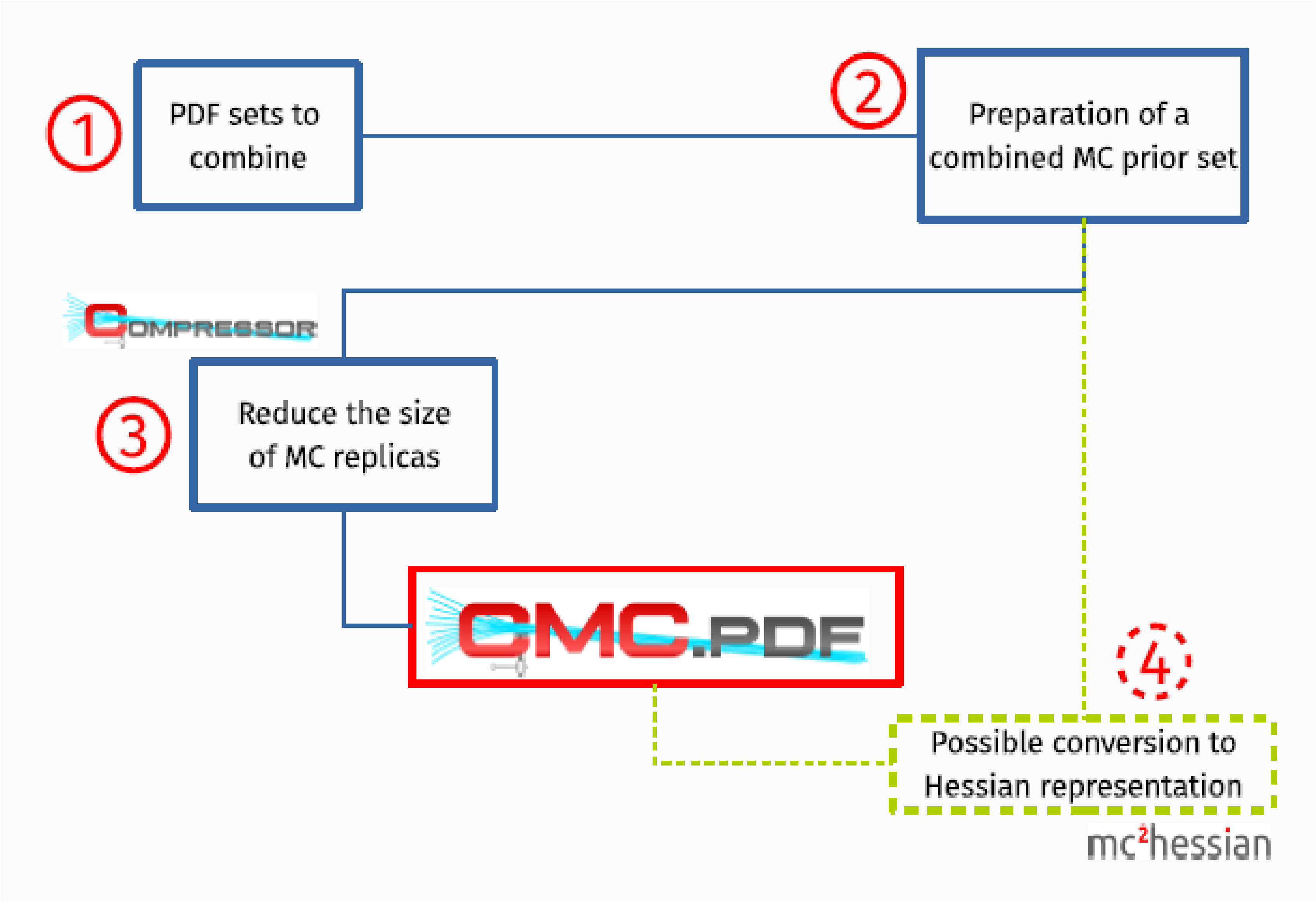
1. **Meta-parametrizations + MC replicas + Hessian data set diagonalization**
(J. Gao, J. Huston, P. Nadolsky, 1401.0013)

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

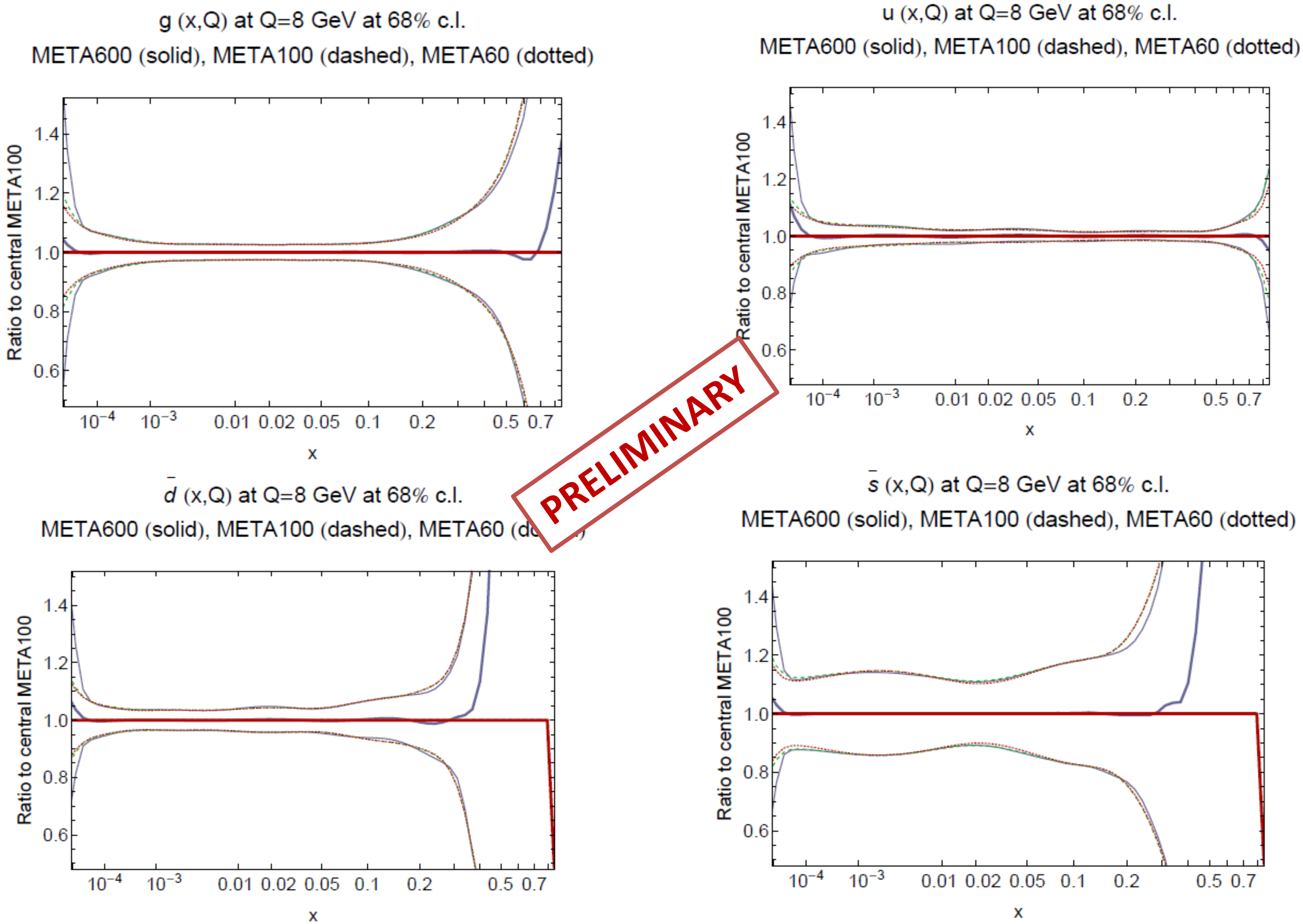
CMC ensembles with 40 replicas and META ensembles with 40-100 replicas are compared with the full ensembles of 300-600 MC replicas.

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.

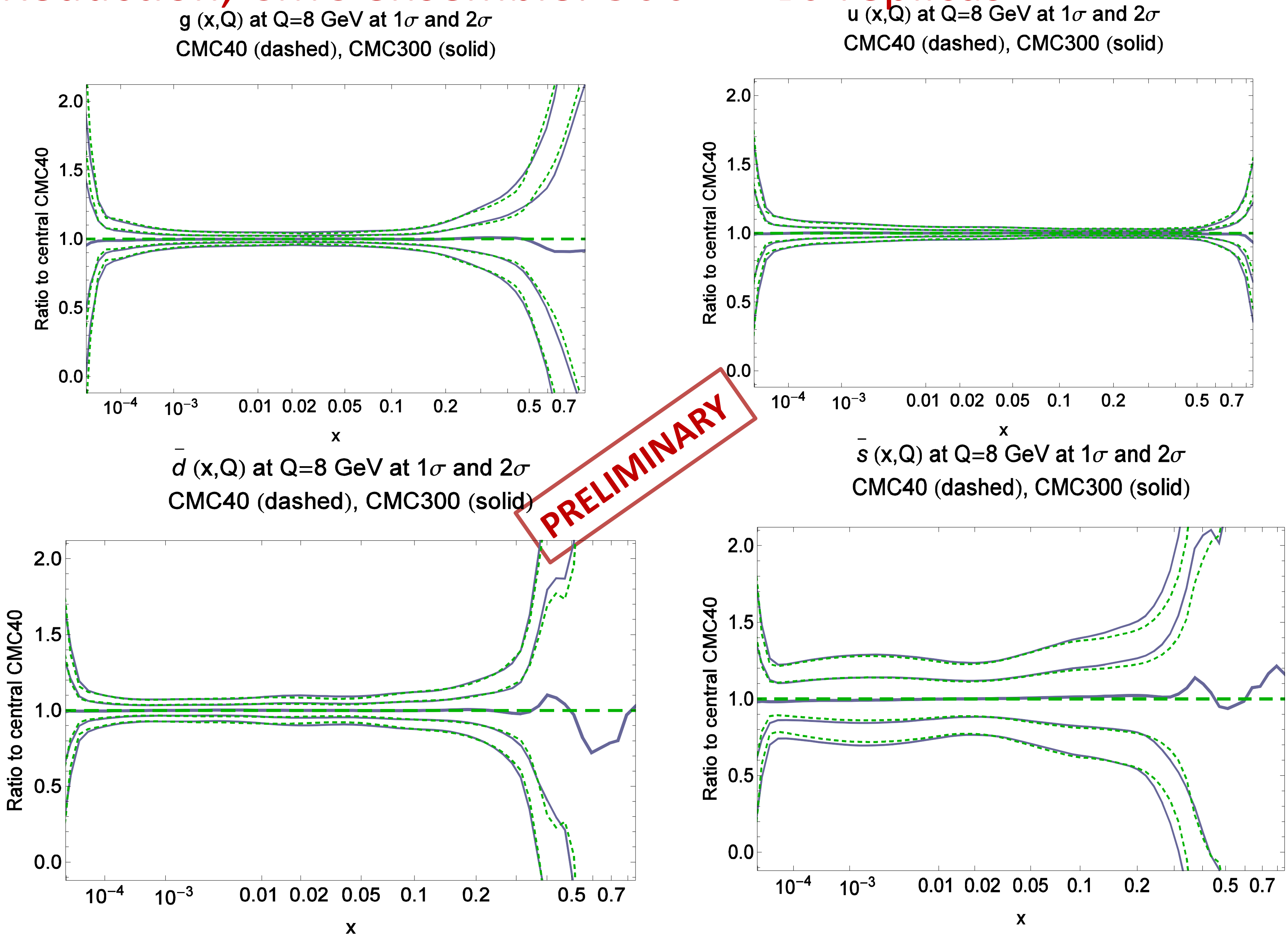
Error bands:

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



Benchmark comparisons, general observations II

PDF-PDF correlations:

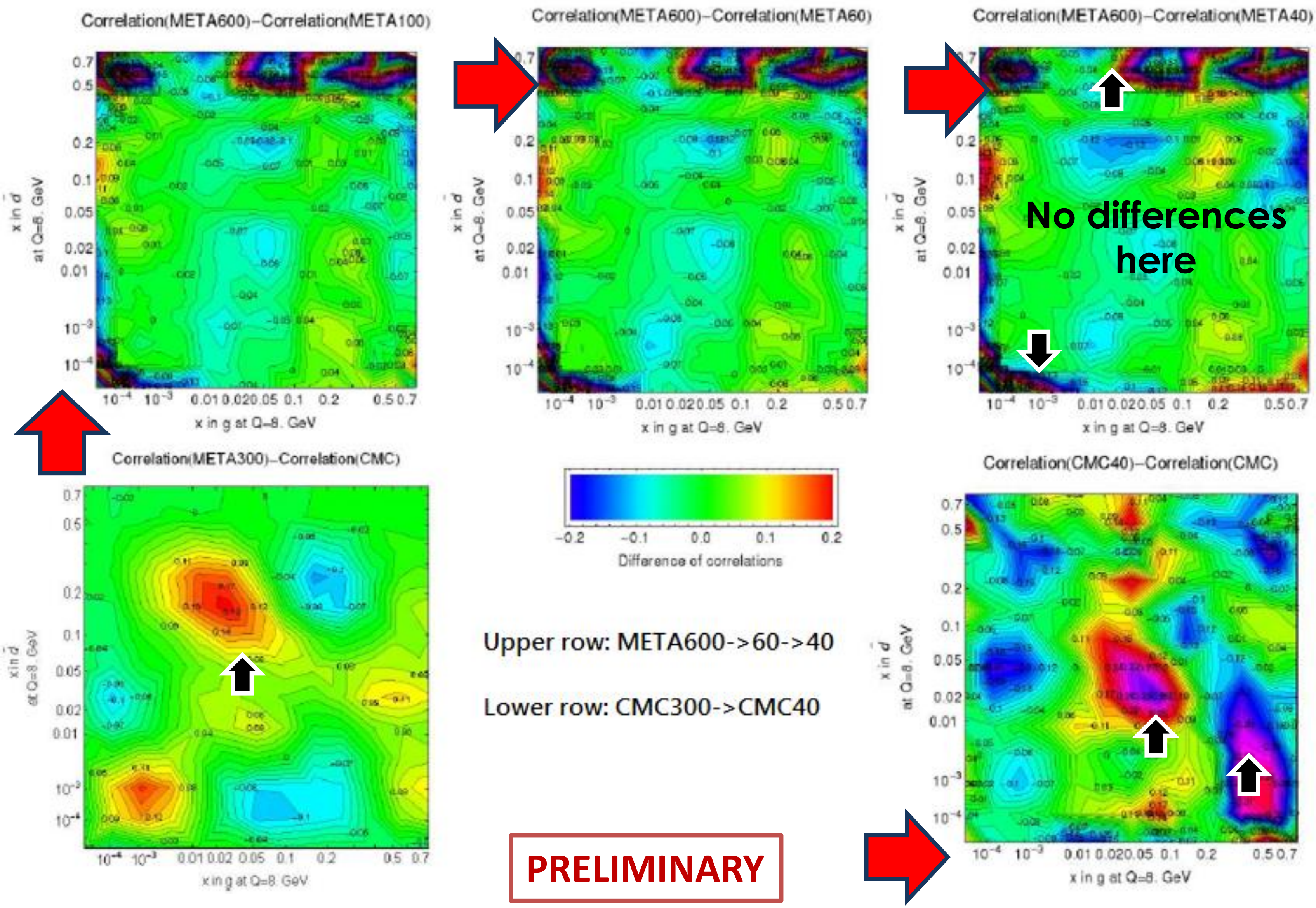
Correlations of META300 and CMC300 ensembles differ by up to ± 0.2 as a result of fluctuations in replica generation

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 $Q = 8\ GeV$

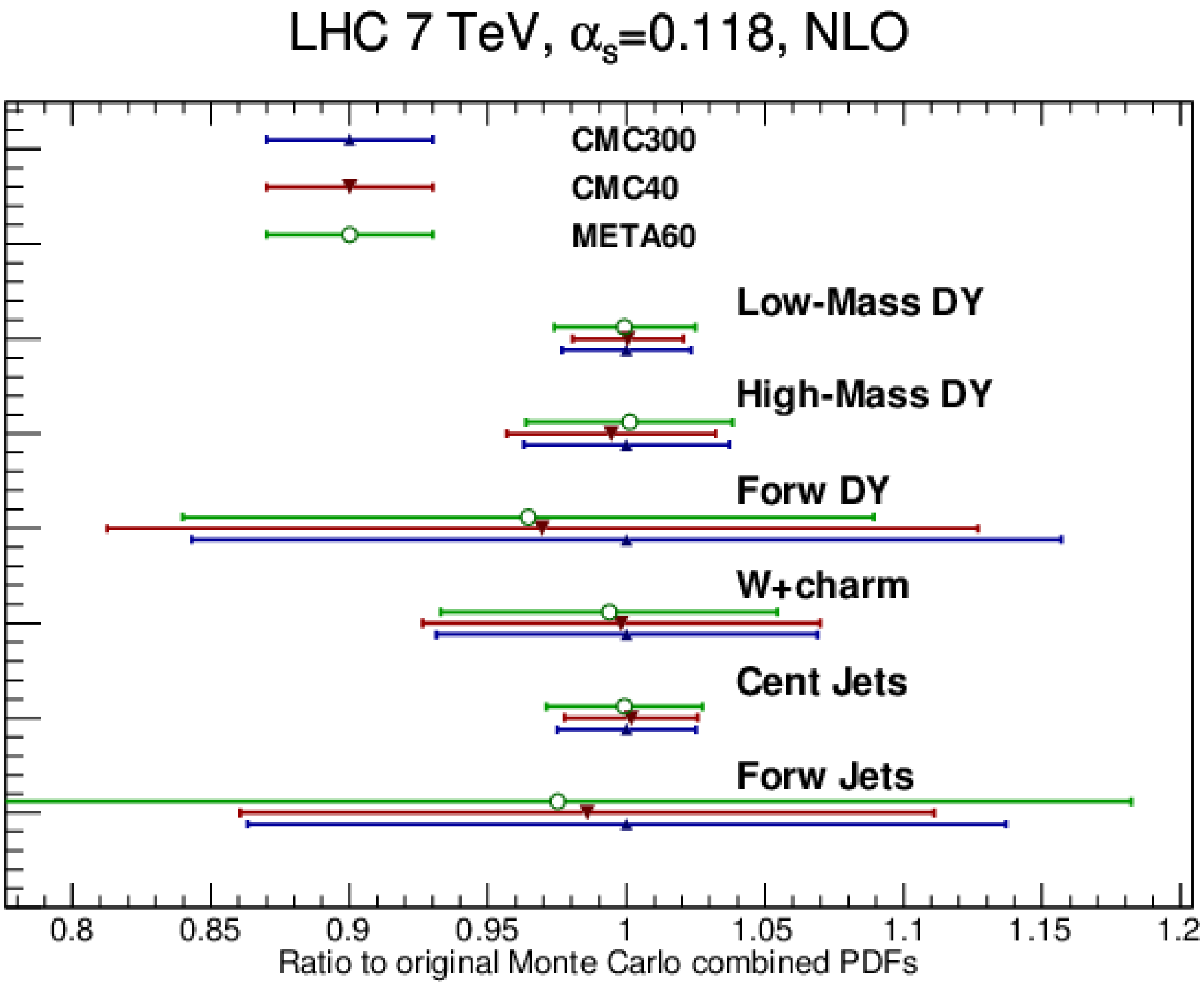


Upper row: META600->60->40

Lower row: CMC300->CMC40

PRELIMINARY

Agreement at the level of benchmark cross sections



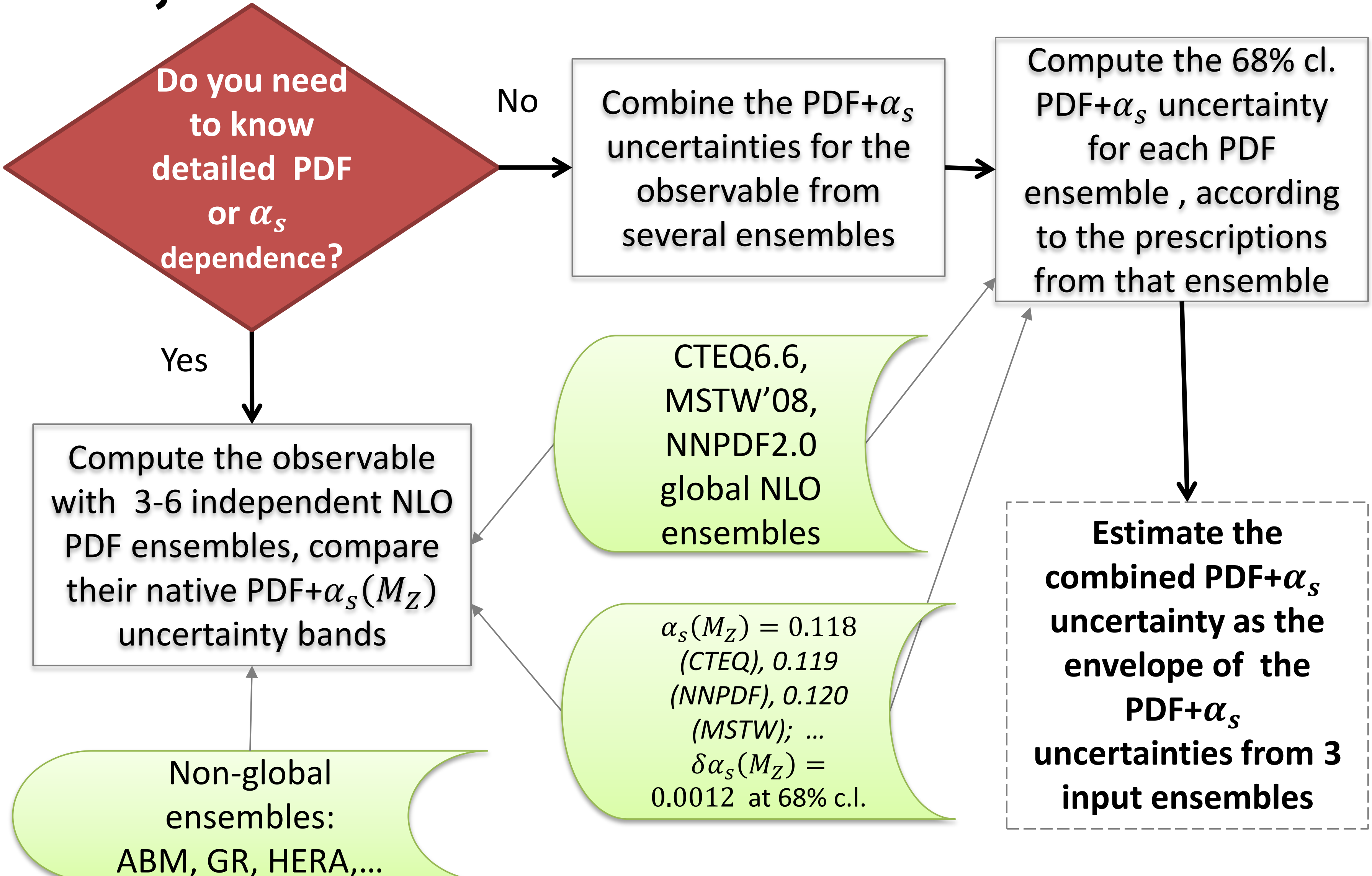
J. Rojo

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

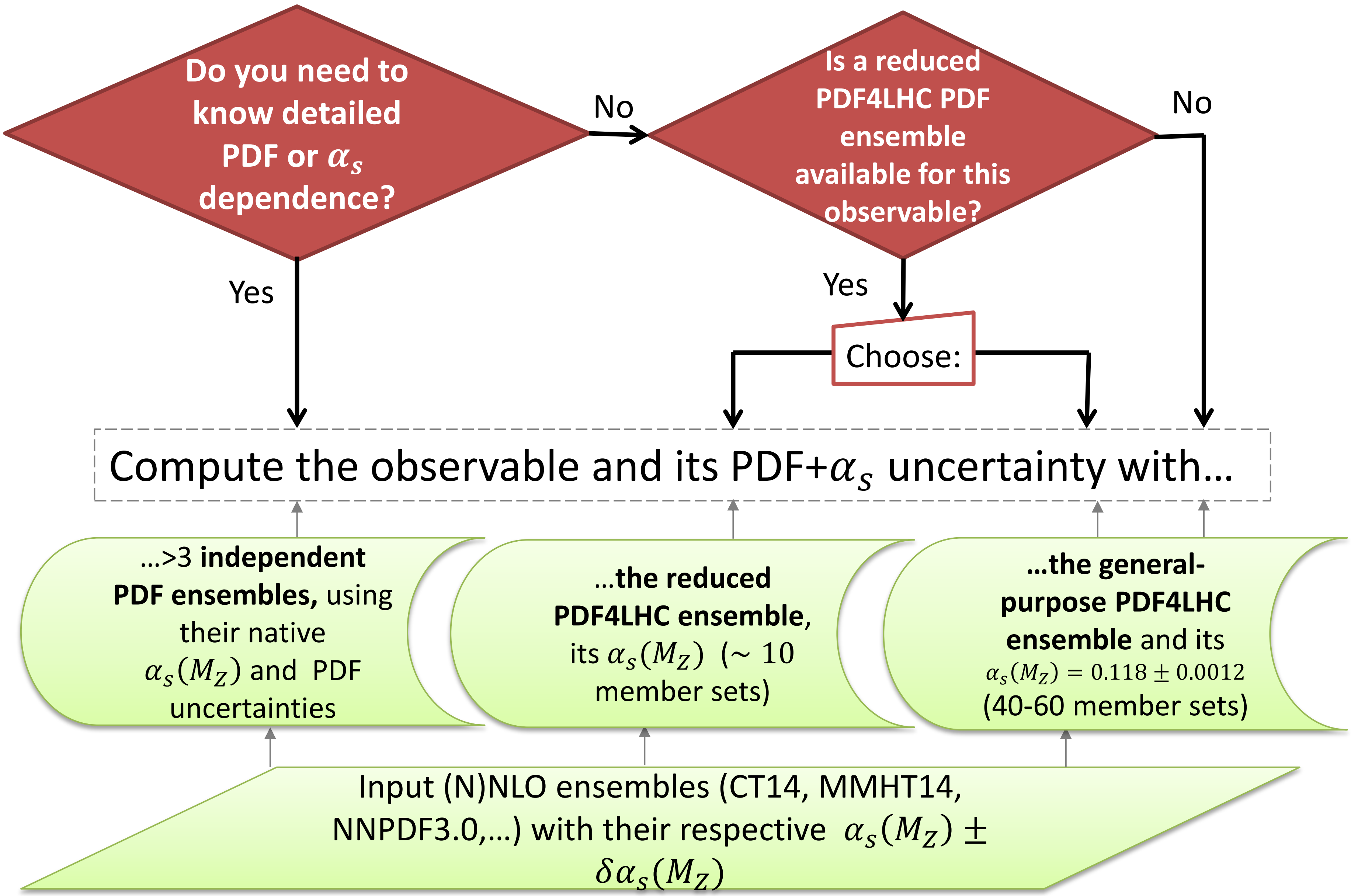
Blueprint for the 2015 PDF4LHC prescription

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.

Benchmark comparisons of two combination methods.
Work plan (from Benasque workshop)

Input MC ensemble: **NNPDF3.0+CT14+MMHT14 NNLO**, with **alphas(M_Z)=0.118**

Convert to 300 replicas in LHAPDF6 format at $Q_0 = 8$ GeV (above the bottom mass), using two independent codes (JR and JG). Cross-check that results are identical.

Done. The results from two groups agree. Mild differences are due to random variations in the generation of MC replicas.

In each approach, reduce the number of replicas to the minimal number that retains 1% or 5% accuracy in reproducing the following properties of the input ensemble:

- Means, 68%c.l. PDF uncertainties, higher moments and asymmetry (skewness), PDF-PDF correlations.

  Done. Ensembles with 40-100 META PDFs and 40 CMC replicas broadly agree.

- Predictions for the standard candle LHC observables used in the META paper: ggHiggs, ttbar, W,Z [Jun]

  Done. Broad agreement.

- Differential LHC distributions using NNPDF3.0 applgrids, supplemented with some new aMCfast grids [Juan]

A variety of comparisons collected at http://bit.ly/1KFoSTq

Benchmark comparisons of CMC and META PDFs

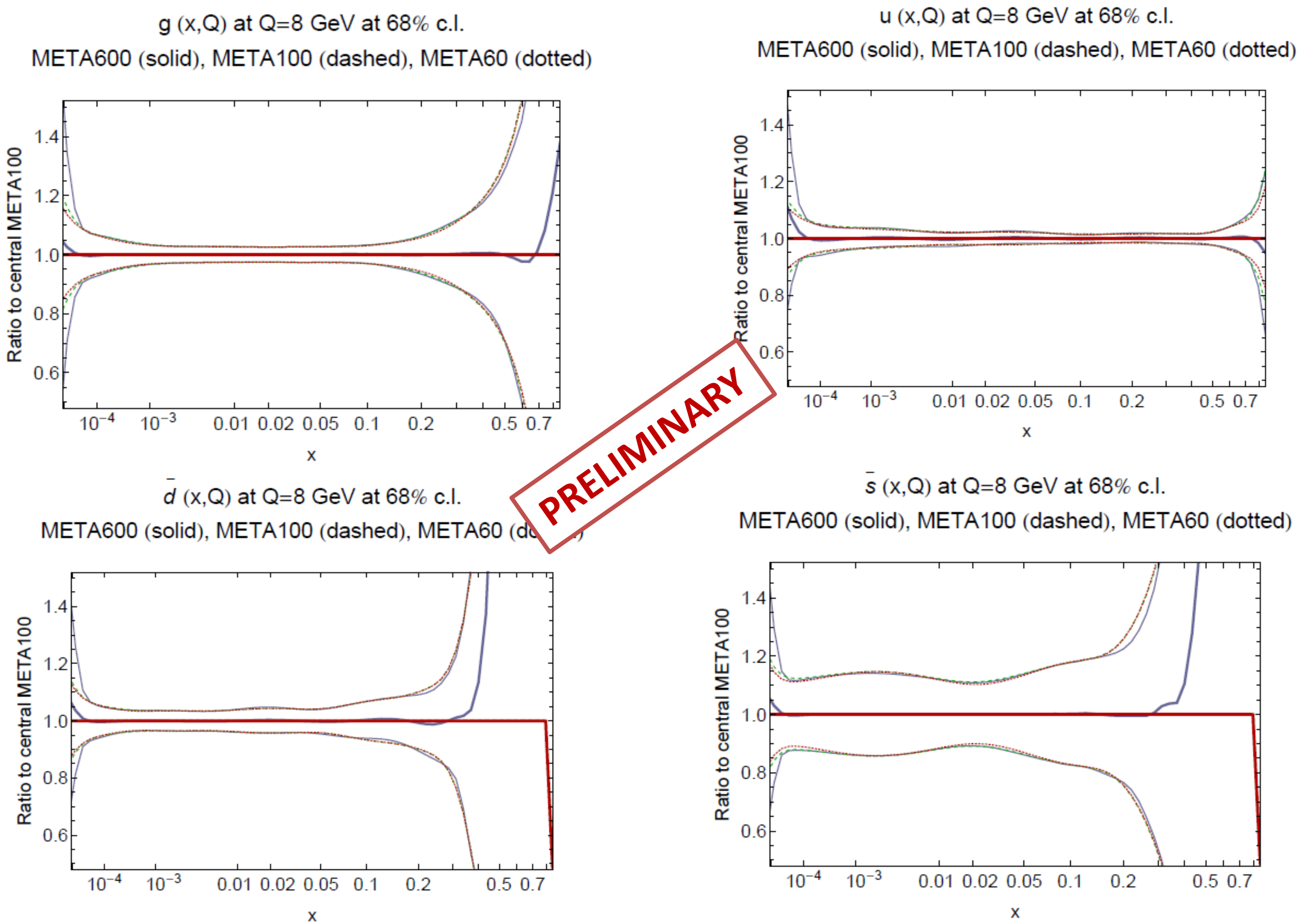
CMC ensembles with 40 replicas and META ensembles with 40-100 replicas are compared with the full ensembles of 300-600 MC replicas.

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.

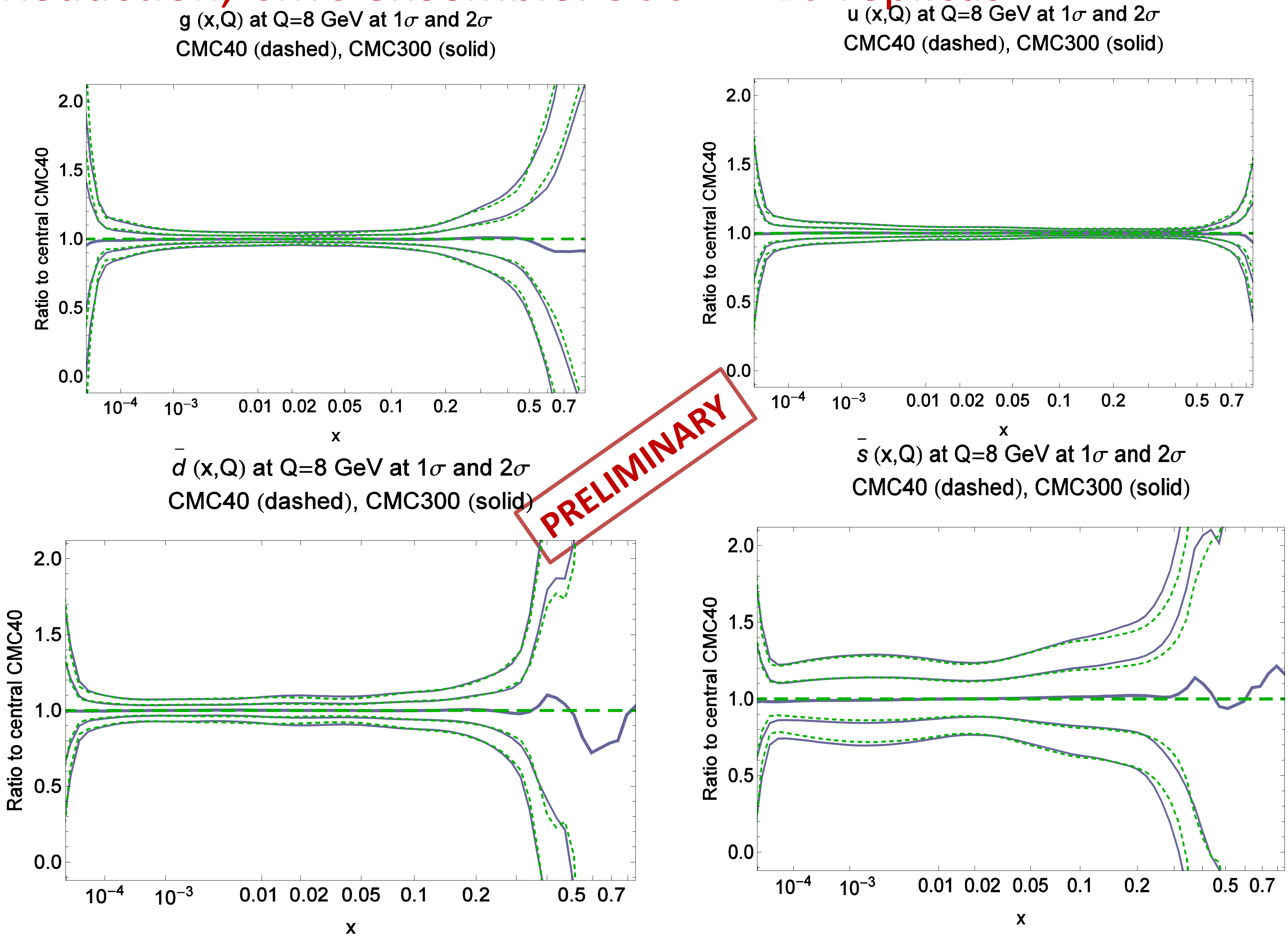
Error bands:

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



Benchmark comparisons, general observations II

PDF-PDF correlations:

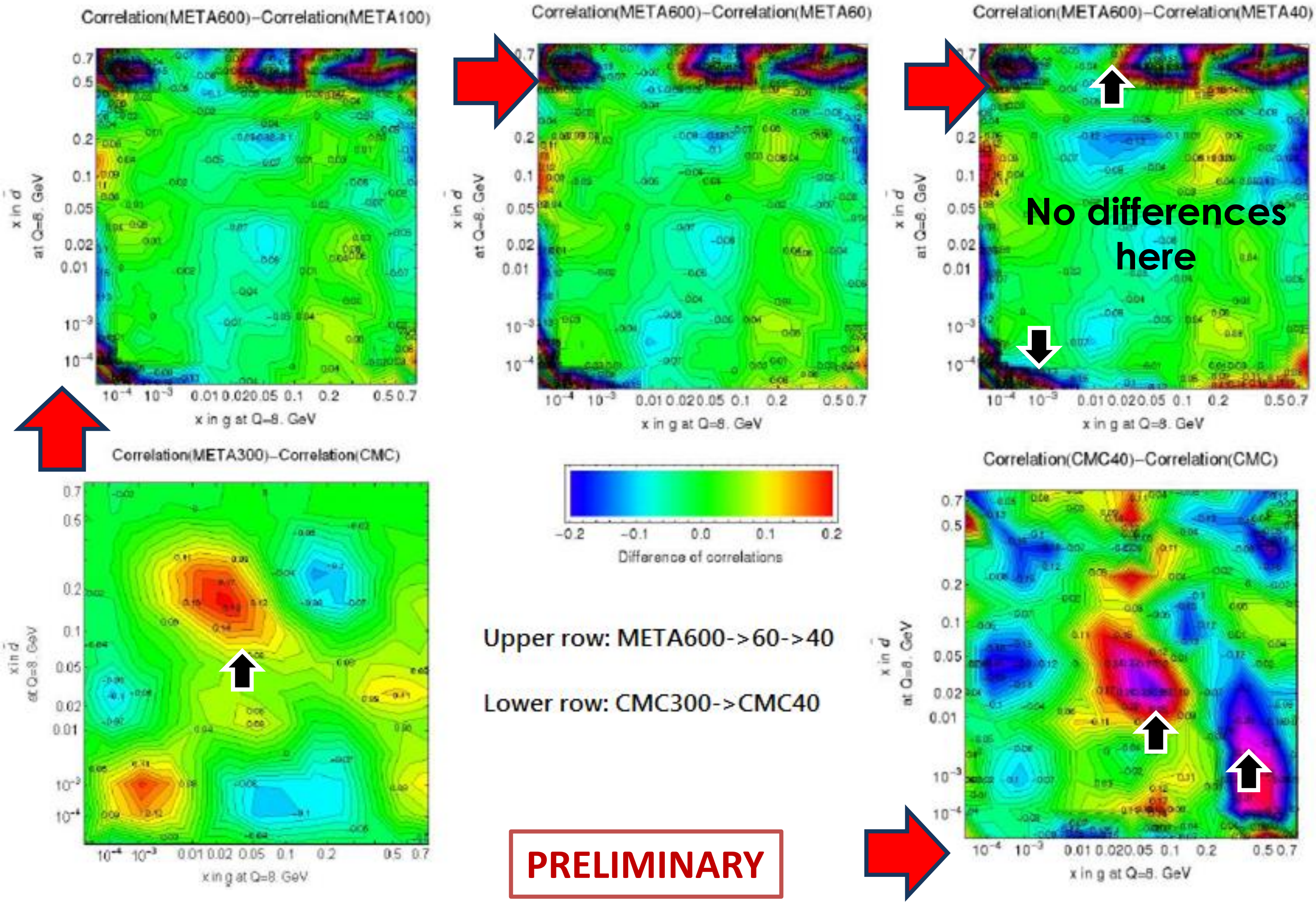
Correlations of META300 and CMC300 ensembles differ by up to $\pm\, 0.2$ as a result of fluctuations in replica generation

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