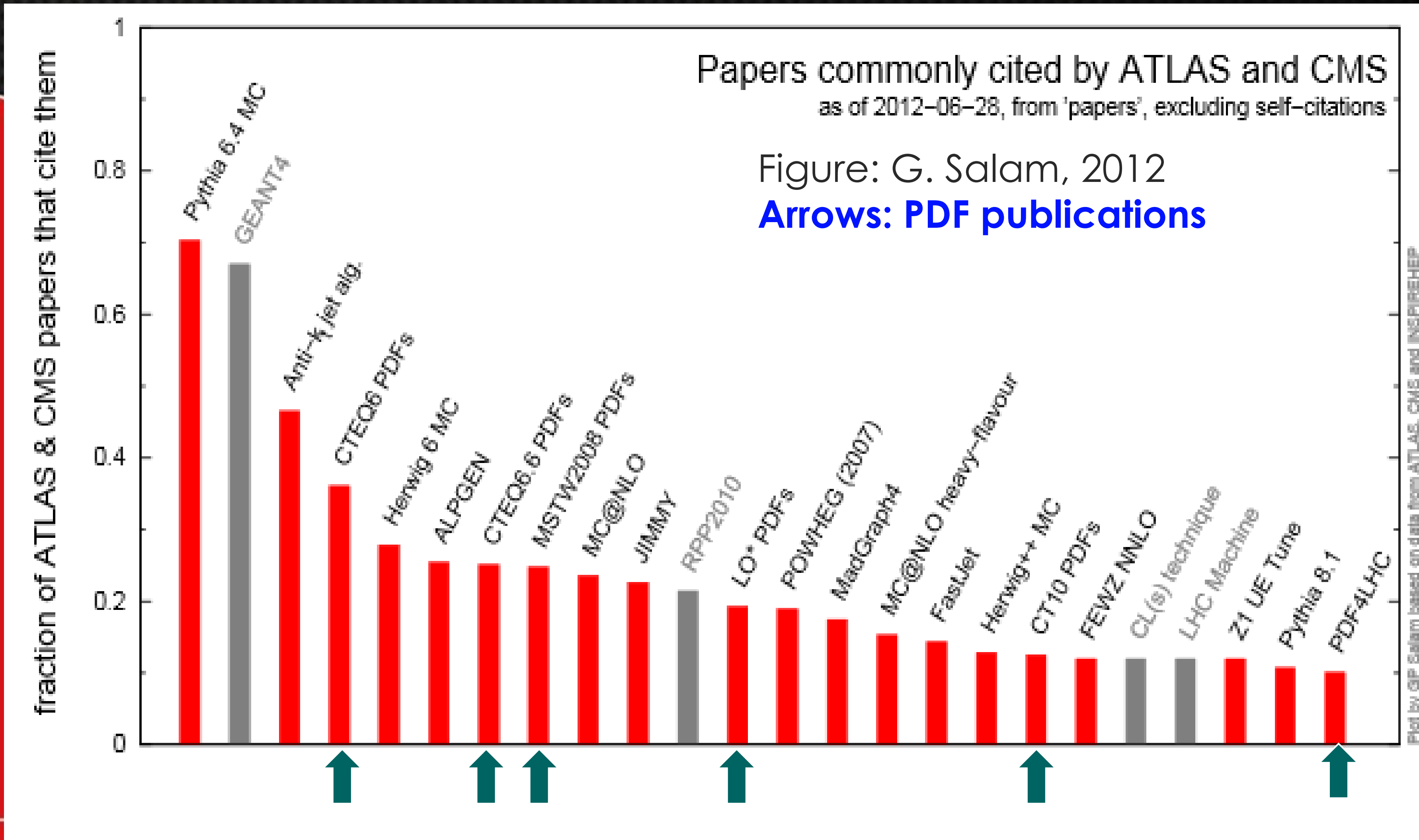
Loopfest XIII workshop, CUNY, June 18, 2014

A meta-analysis of parton distribution functions for LHC applications

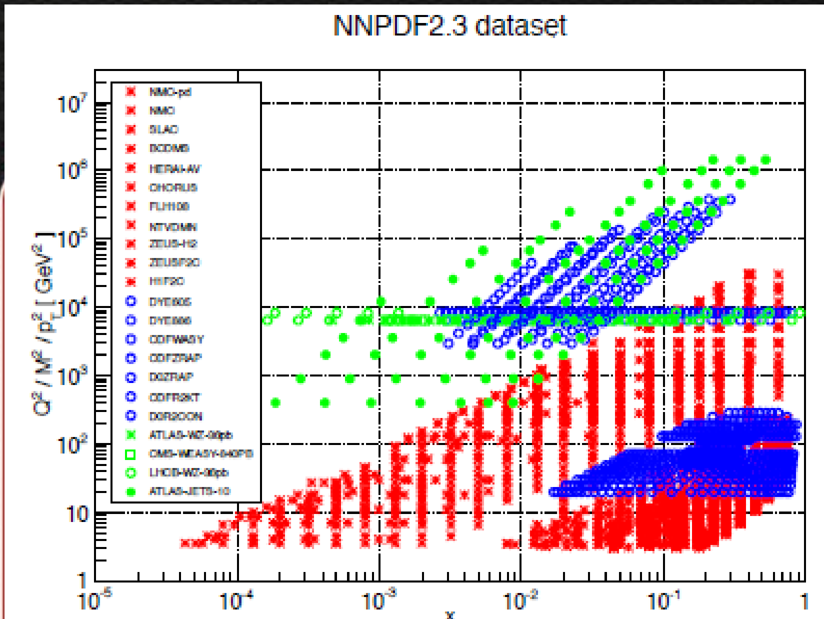
Jun Gao, Joey Huston,
Pavel Nadolsky (presenter)

arXiv:1401.0013, arXiv:1406.xxxx
http://metapdf.hepforge.org

A large fraction of theoretical simulations for hadronic experiments relies on PDFs

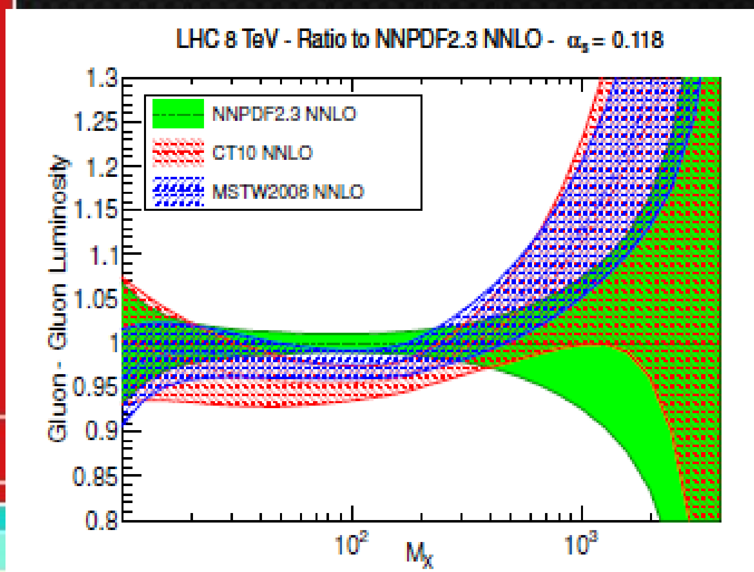


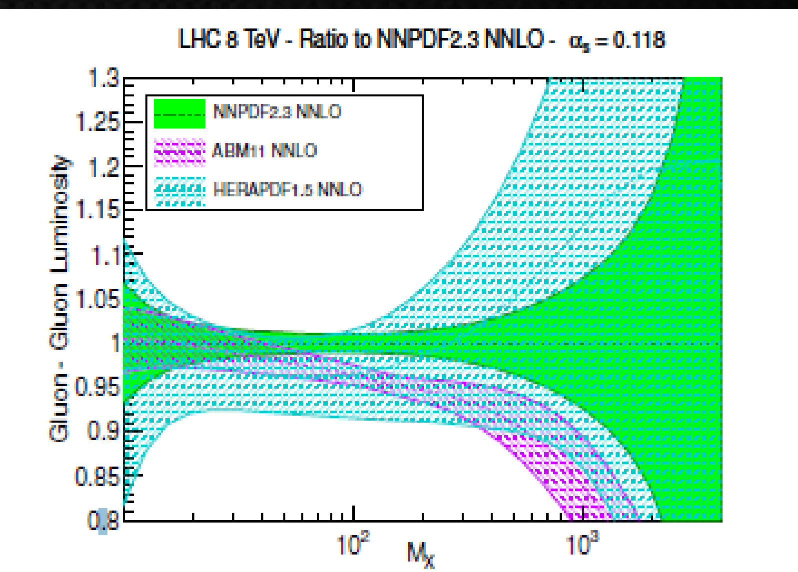
Experiments ⇒ nucleon PDFs



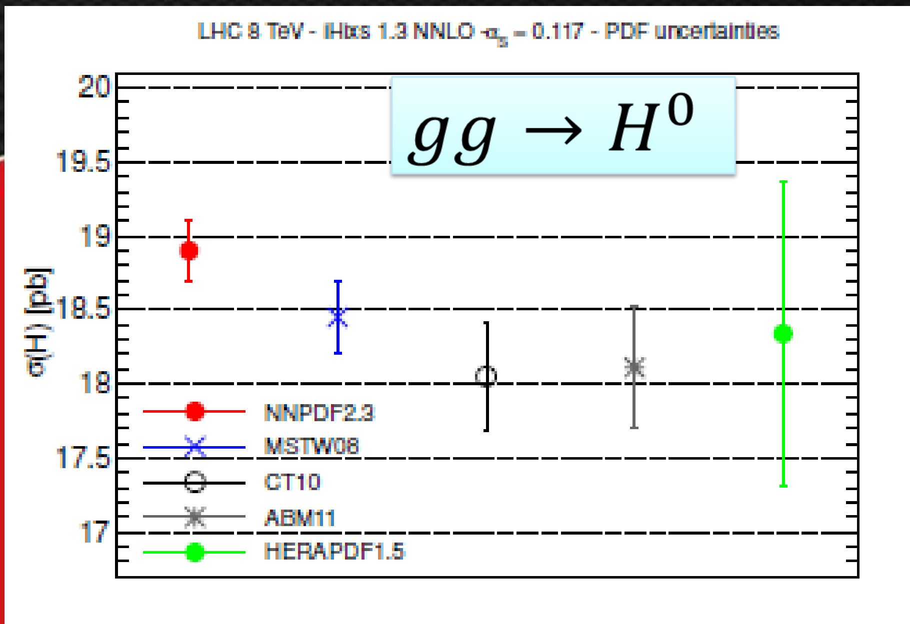
NNLO ensembles of PDF parametrizations are constructed by 5 groups using a variety of experiments and techniques

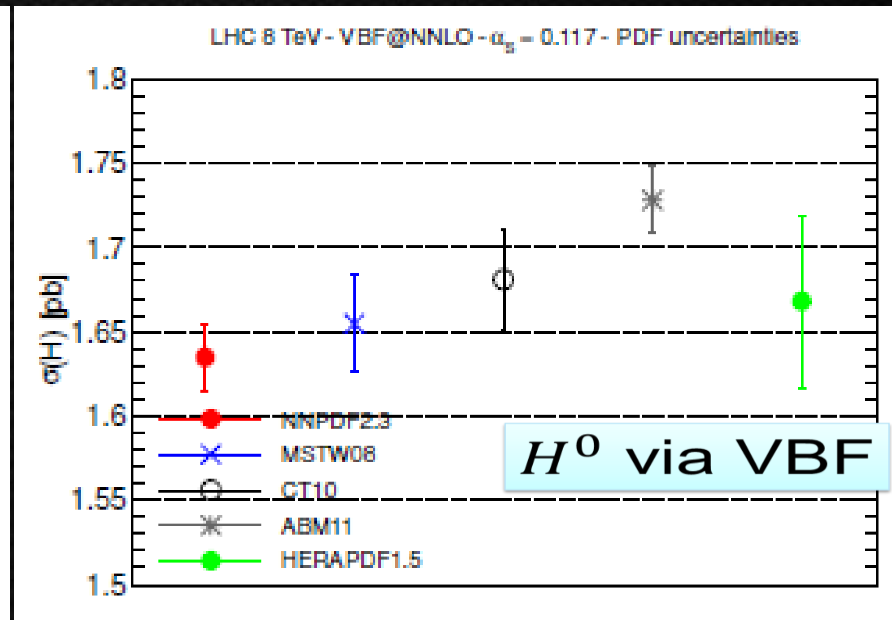
ABM: arXiv:1302.1516
CT10 NNLO: arXiv:1302.6246
NNPDF2.3: arXiv:1207.1303
MSTW'12: arXiv:1211.1215
HERAPDF1.5: arXiv:1112.2107





PDFs ⇒ LHC experiments





Each PDF ensemble (ABM, CT,..) provides 30-100 PDF member sets to compute the **central value** and **PDF uncertainty** for the QCD observable

CT10, MSTW'08, NNPDF2.3 predictions are combined according to the **PDF4LHC convention** (M. Botje et al., (2011), arxiv:1101.0538; R. Ball et al., **arXiv:1211.5142)**

Downside: The same QCD observable is computed multiple times. Many PDF sets contribute little to the total PDF uncertainty

Let's try to compute the PDF uncertainty more efficiently

The approach of PDF meta-analysis reduces the number of PDF sets needed to compute the combined PDF uncertainty

200 CT10, MSTW'08, NNPDF2.3 error sets

101 Hessian META sets for most LHC applications at GeV (the META 1.0 PDF set)

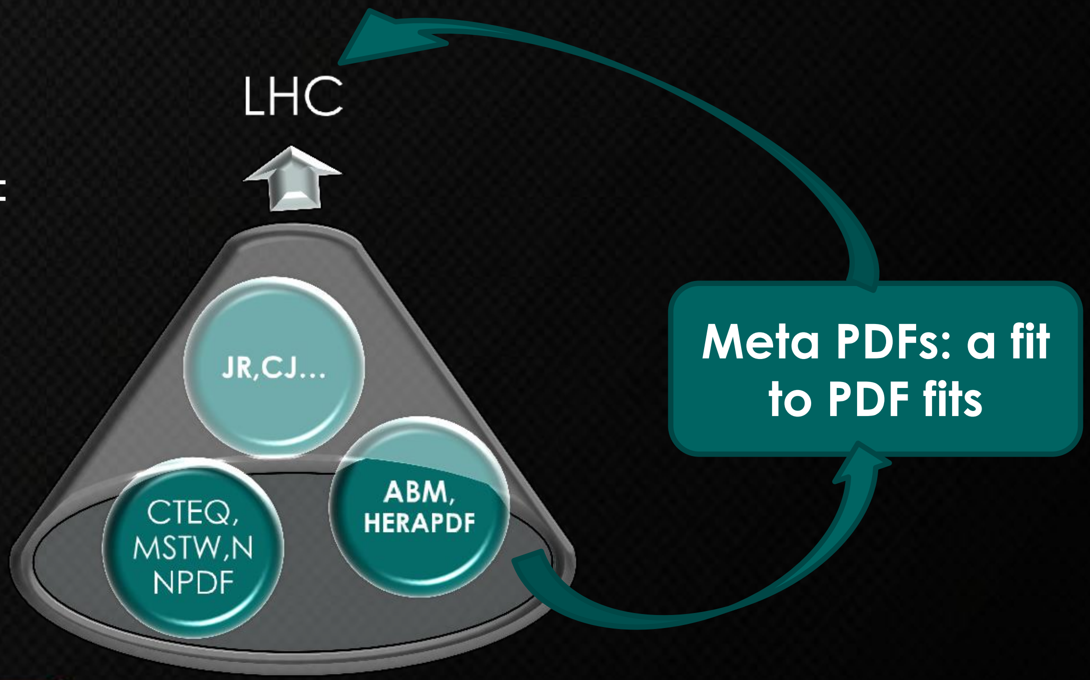
13 META sets for LHC Higgs production observables (the LHCH meta-PDF set)

The meta-PDF sets are available on metapdf.hepforge.org

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



Four steps of the meta-analysis

See arXiv:1401.0013 for details

	# of PDF sets
1. Select the input PDF ensembles (CT, MSTW, NNPDF…)	**200**
2. Fit each PDF error set in the input ensembles at GeV by a common functional form (“a meta-parametrization”). **66 parameters in total.**	**200**
3. Generate many **Monte-Carlo** replicas from meta-parametrizations of each set to investigate the probability distribution (using Thorne-Watt’s method)	**300**
4. Construct a final ensemble of 68% c.l. **Hessian eigenvector sets** to propagate the PDF uncertainty found from the combined ensemble of replicated meta-parametrizations into LHC predictions.	**101 or 13**

Combination into the final META ensemble

We constructed a META ensemble from CT10, MSTW2008, NNPDF2.3 input PDFs.

These ensembles…

… are obtained in the global PDF analysis

… have comparable numbers of experimental data points and comparable **nominal** PDF uncertainties

…assume the central $\alpha_s(M_Z) \approx 0.118$

…are realized in similar heavy-quark schemes

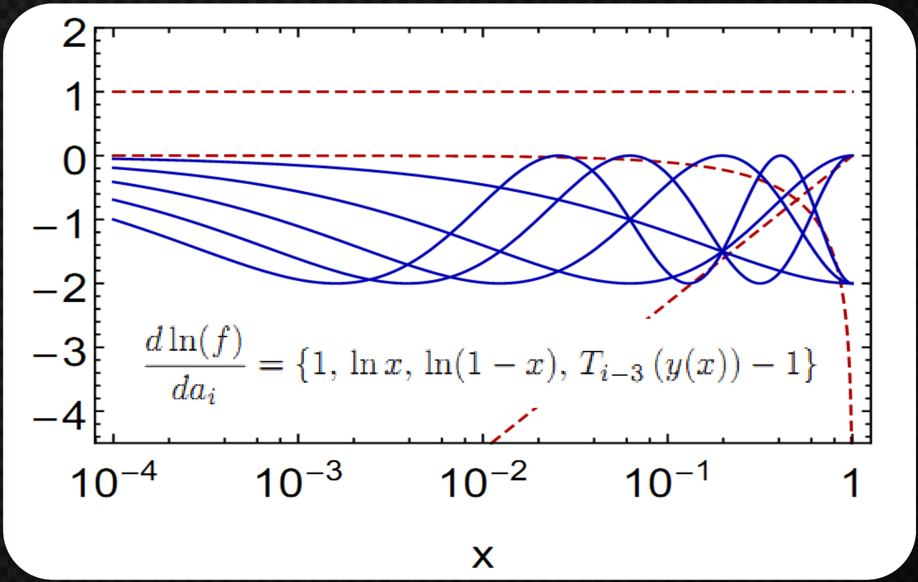
… evolve with $N_f = 5$ at $Q > Q_0 = 8$ GeV

The functional form for the meta parametrization

$$f(x, Q_0; \{a\}) = e^{a_1} x^{a_2} (1-x)^{a_3} e^{\sum_{i\geq 4} a_i \, (T_{i-3}(y(x))-1))}$$

J. Pumplin, 0909.5176, A. Glazov, et al., 1009.6170, A. Martin, et al., 1211.1215

The initial scale of DGLAP evolution is Q_0=8 GeV. $T_i(y)$ are Chebyshev polynomials with $y(x)=\cos(\pi x^{\beta})$ and β=1/4.

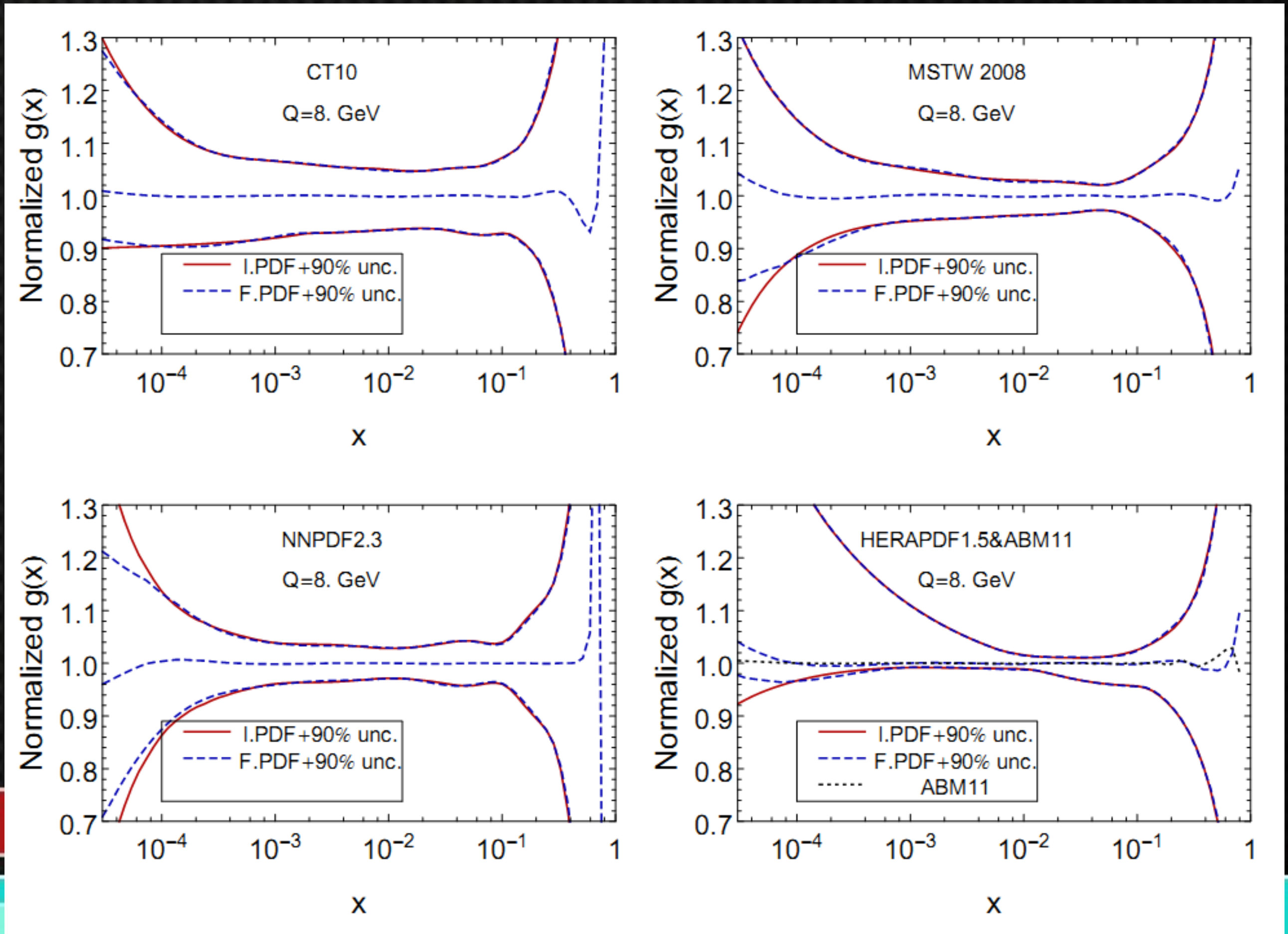


The input PDFs are fitted by this form in the x regions covered by the experimental data.

Outside these x regions, the PDFs are determined by extrapolation.

Agreement of the input (I) and fitted (F) PDFs

The PDF uncertainty of each input ensemble is reproduced by the fitted set.



This level of agreement is preserved by DGLAP evolution to $Q > 8$ GeV

Construction of the META ensemble

The ensembles are merged by averaging their meta-parameters. For CT10, MSTW, NNPDF ensembles, unweighted averaging is appropriate, 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

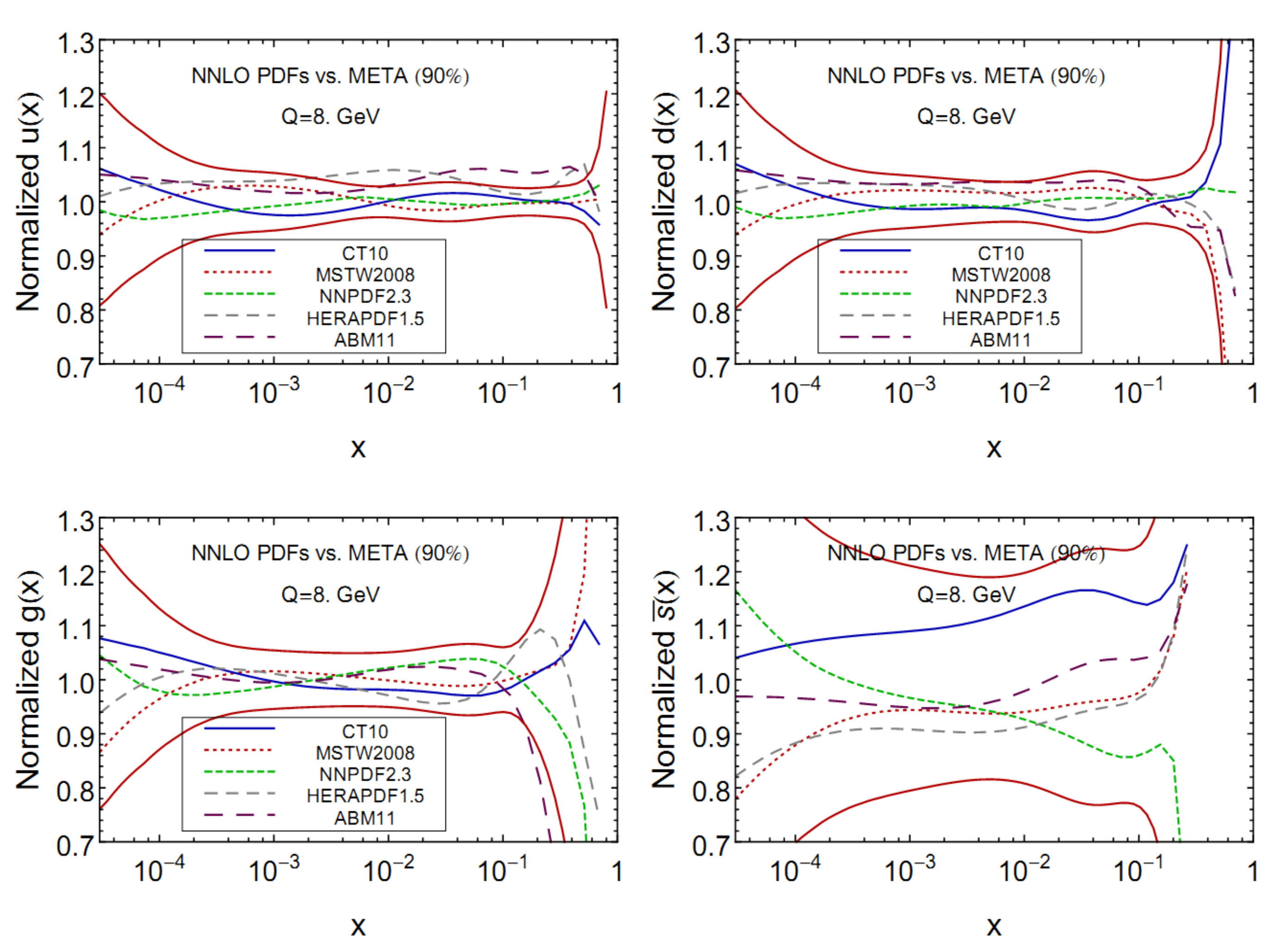
$$\mathrm{cov}(a_i, a_j)_g = \frac{N_{rep}}{N_{rep}-1} \langle (a_i - \langle a_i \rangle_g) \cdot (a_j - \langle a_j \rangle_g) \rangle_g,$$

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

Standard deviation on g

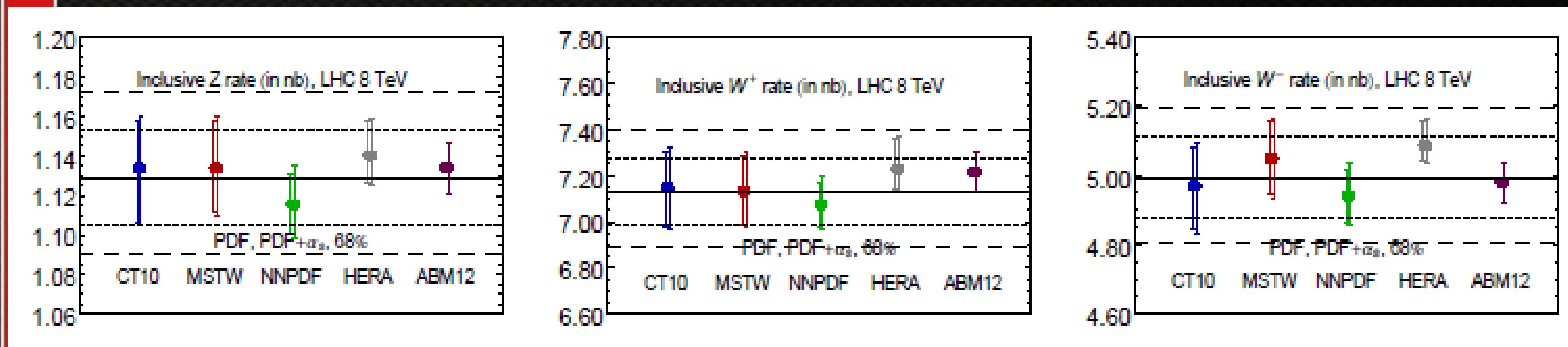
META PDFs vs. central PDFs of input ensembles

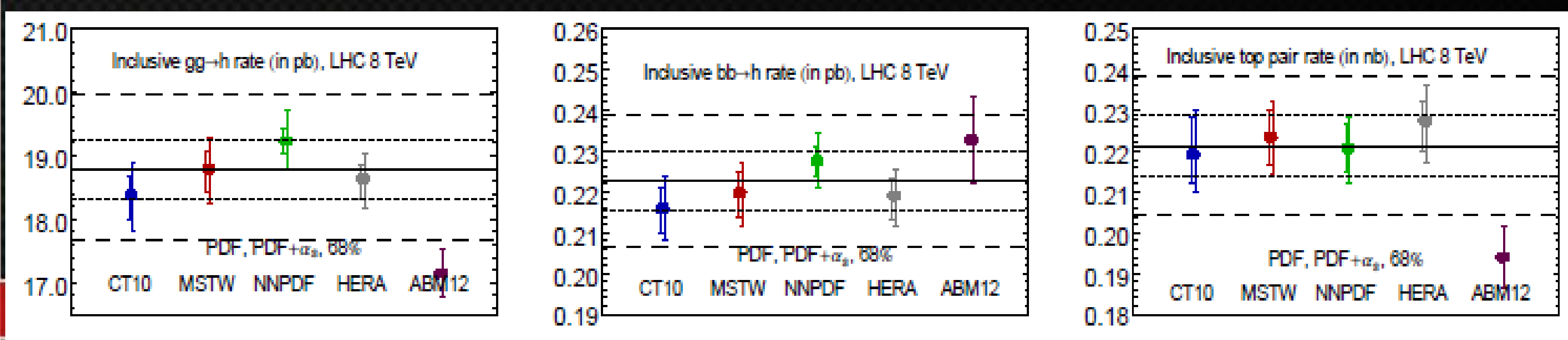
Comparisons of the META PDF with the original central PDF from each groups with $a_S (M_Z)$=0.118 at Q=8 GeV.



META predictions for the LHC

Comparisons of the LHC predictions, including central prediction, PDF uncertainties, and PDF+alphas uncertainties, at 68% C.L.. The PDF uncertainty is about the same as from the envelope prescription in the benchmark study **(R. Ball, et al., 1211.5142)**, e.g., for gg→h, 18.75±1.24 pb there, while 18.78±1.15 pb here.





PDF-induced correlations

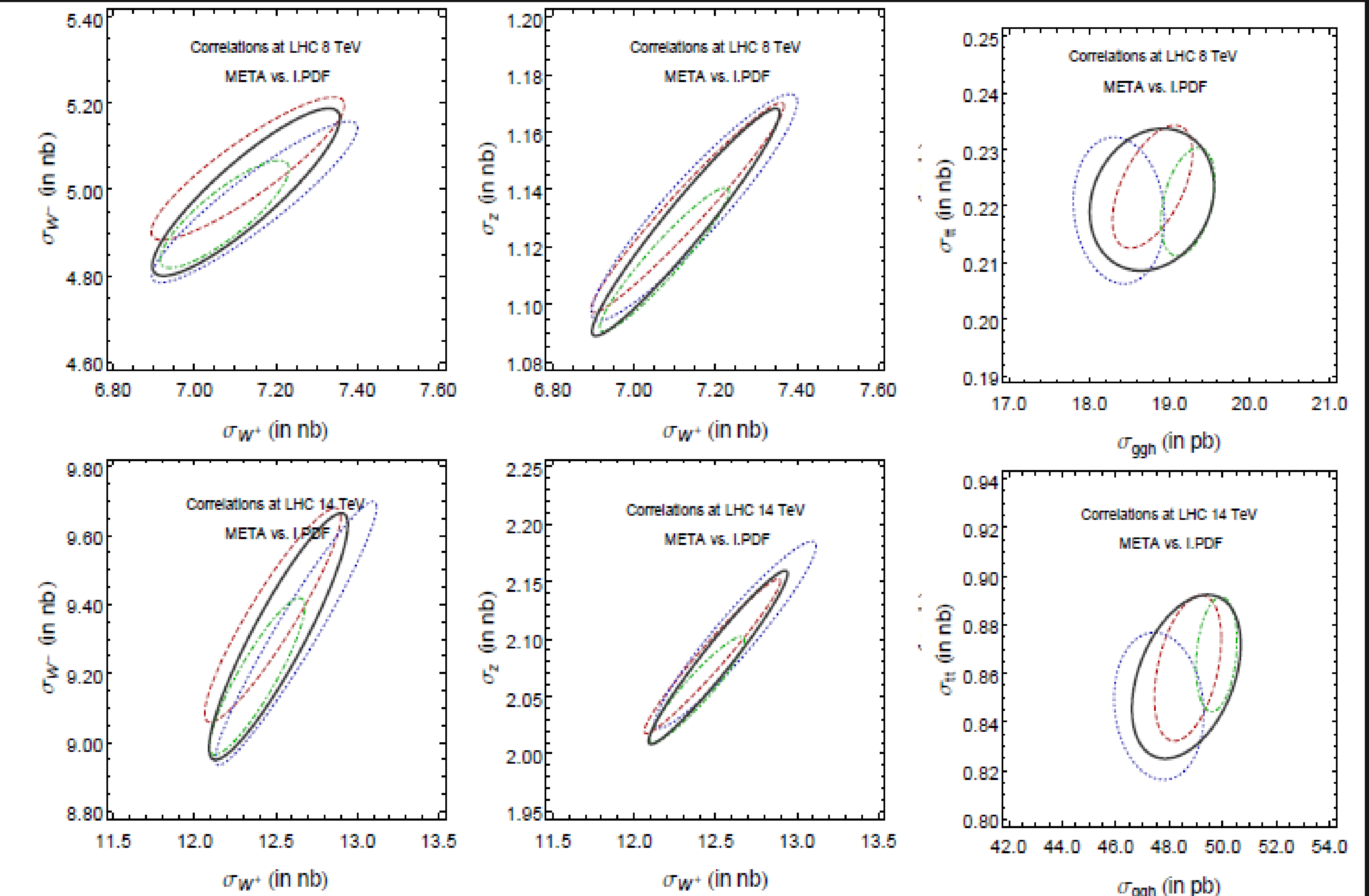


Figure 26: Comparison of 90% error ellipses of NNLO cross sections using various ensembles: META PDF(black solid), CT10(blue dotted), MSTW(red dashed), NNPDF(green dot-dashed).

To predict a class of similar observables, the number of meta-PDF replicas can be reduced from ~100 to a few

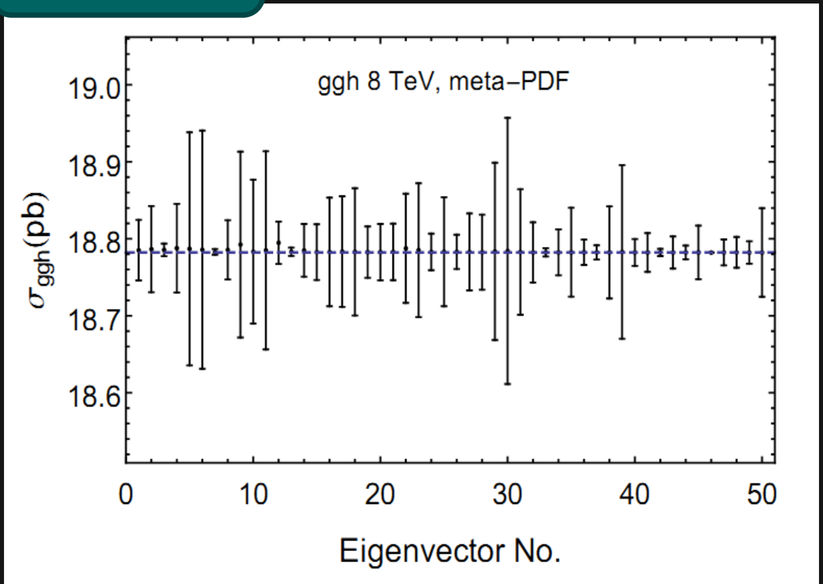
Example: data set diagonalization for Higgs boson production processes

Data set diagonalization

J. Pumplin, 2009

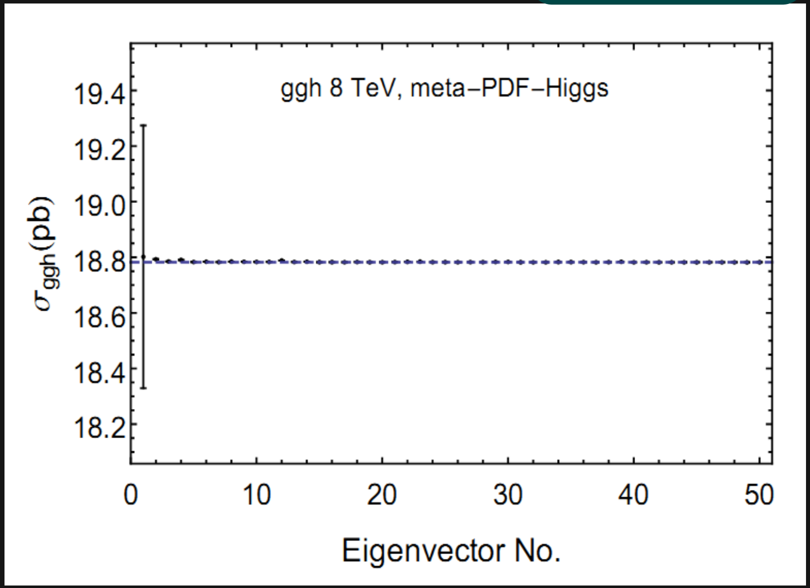
In Hessian method there is additional freedom of choosing the eigenvector basis. Basis 2 can be produced by an arbitrary orthogonal transformation of basis 1. They represent identical measurements in the PDF parameter space and predict same results for physical observables with linear dependence.

Basis 1



=

Basis 2



Benefit: fewer eigenvectors are needed to fully describe a group of physical observables, including uncertainties and correlations

Preliminary

LHCH PDF set for SM Higgs boson studies

Select 46 physical observables, including SM Higgs production cross sections at LHC 8 and 14 TeV, in gg->h, bb->h, VBF, Wh, Zh, hh, tth, h+jet channels, for inclusive rate and rate in high pt or large rapidity region of the Higgs boson.

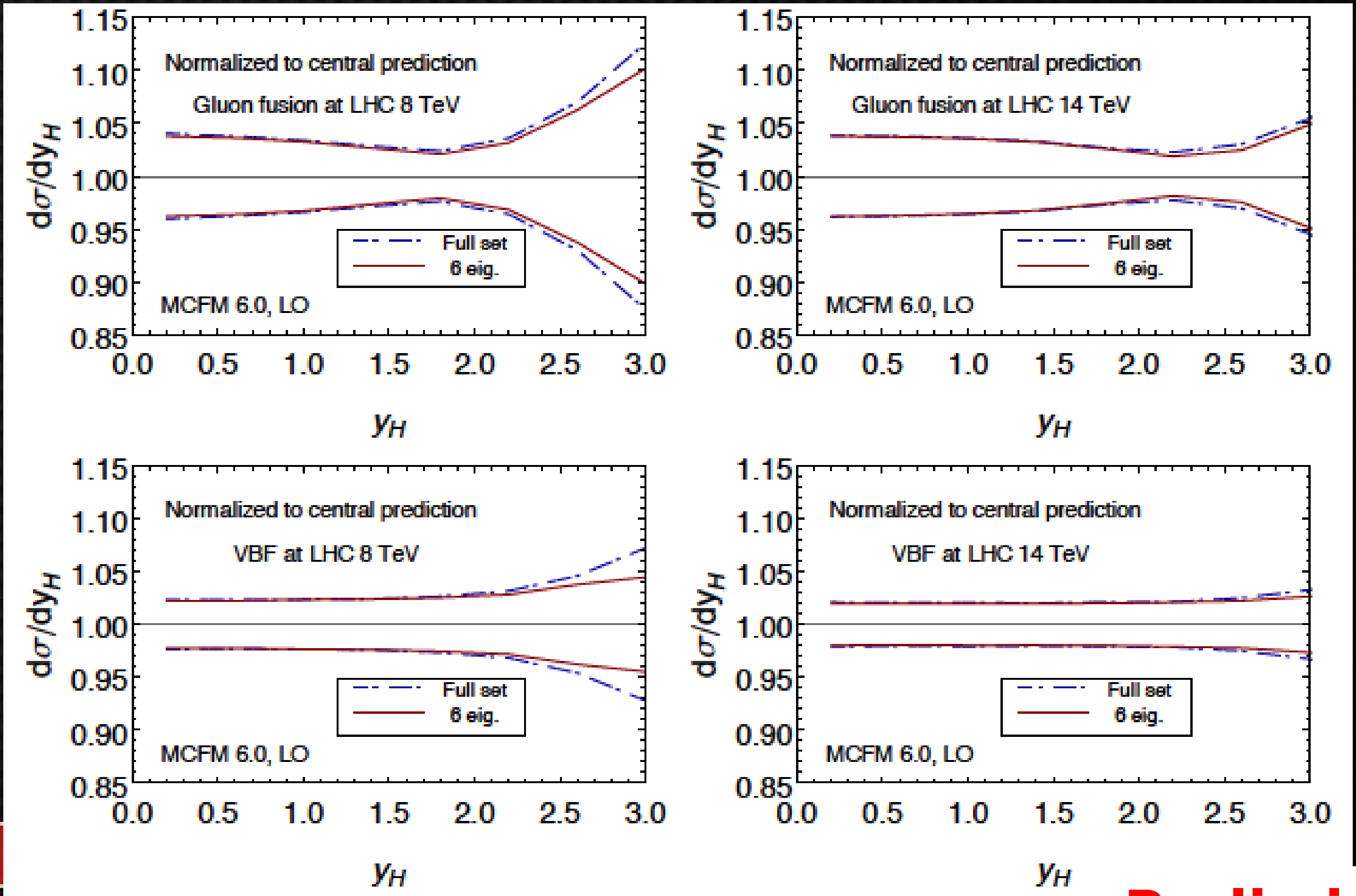
Central values, PDF uncertainties and correlations for these observables based on the full META ensembles with 50+1 eigenvectors are fully reproduced by the LHCH ensemble with 6+1 eigenvectors

Predictions of the META PDF ensemble

Predictions of the LHCH ensemble

Normalized rapidity distributions

Blue dot-dashed: META PDF 1.0 ensemble (100 error sets)
Red solid: LHCH ensemble (12 error sets)



Preliminary

PDF-induced correlations between Higgs production cross sections in various channels

	GGH inc., 8 TeV, NNLO	GGH 0j exc., 8 TeV, NLO	GGH 1j exc., 8 TeV, NLO	GGH 2j inc., 8 TeV, LO	GGH 2j exc., 8 TeV, LO	GGH 2j full mass, 8 TeV, LO	VBF inc., 8 TeV, LO	VBF exc., 8 TeV, LO	GGH inc., 14 TeV, NNLO	GGH 0j exc., 14 TeV, NLO	GGH 1j exc., 14 TeV, NLO	GGH 2j inc., 14 TeV, LO	GGH 2j exc., 14 TeV, LO	GGH 2j full mass, 14 TeV, LO	VBF inc., 14 TeV, LO	VBF exc., 14 TeV, LO
VBF exc., 14 TeV, LO	−0.43 / ***−0.44***	−0.49 / ***−0.5***	−0.3 / ***−0.33***	0.09 / ***0.09***	0.09 / ***0.09***	0.06 / ***0.09***	0.92 / ***0.93***	0.92 / ***0.93***	−0.39 / ***−0.4***	−0.42 / ***−0.44***	−0.33 / ***−0.35***	0.02 / ***0.02***	0.02 / ***0.02***	0. / ***0.***	1. / ***1.***	
VBF inc., 14 TeV, LO	−0.43 / ***−0.44***	−0.49 / ***−0.5***	−0.3 / ***−0.33***	0.09 / ***0.09***	0.09 / ***0.09***	0.06 / ***0.09***	0.92 / ***0.93***	0.92 / ***0.93***	−0.39 / ***−0.4***	−0.42 / ***−0.44***	−0.33 / ***−0.35***	0.02 / ***0.02***	0.02 / ***0.02***	0. / ***0.***		
GGH 2j full mass, 14 TeV, LO	0.45 / ***0.42***	0.25 / ***0.22***	0.72 / ***0.71***	0.96 / ***0.98***	0.96 / ***0.98***	0.96 / ***0.98***	−0.04 / ***−0.05***	−0.04 / ***−0.05***	0.31 / ***0.28***	0.08 / ***0.05***	0.47 / ***0.46***	0.99 / ***0.99***	0.99 / ***0.99***			
GGH 2j exc., 14 TeV, LO	0.43 / ***0.44***	0.22 / ***0.23***	0.71 / ***0.72***	0.97 / ***0.98***	0.97 / ***0.98***	0.97 / ***0.98***	−0.01 / ***−0.02***	−0.01 / ***−0.02***	0.29 / ***0.29***	0.07 / ***0.07***	0.46 / ***0.48***	0.99 / ***0.99***				
GGH 2j inc., 14 TeV, LO	0.43 / ***0.44***	0.22 / ***0.23***	0.71 / ***0.72***	0.97 / ***0.98***	0.97 / ***0.98***	0.97 / ***0.98***	−0.01 / ***−0.02***	−0.01 / ***−0.02***	0.29 / ***0.29***	0.07 / ***0.07***	0.46 / ***0.48***					
GGH 1j exc., 14 TeV, NLO	0.98 / ***0.98***	0.94 / ***0.94***	0.93 / ***0.94***	0.3 / ***0.33***	0.3 / ***0.33***	0.3 / ***0.33***	−0.34 / ***−0.34***	−0.34 / ***−0.34***	0.97 / ***0.97***	0.89 / ***0.9***						
GGH 0j exc., 14 TeV, NLO	0.91 / ***0.92***	0.96 / ***0.97***	0.7 / ***0.73***	−0.07 / ***−0.08***	−0.07 / ***−0.08***	−0.07 / ***−0.08***	−0.4 / ***−0.4***	−0.4 / ***−0.4***	0.97 / ***0.97***							
GGH inc., 14 TeV, NNLO	0.97 / ***0.97***	0.97 / ***0.98***	0.84 / ***0.87***	0.14 / ***0.14***	0.14 / ***0.14***	0.14 / ***0.14***	−0.38 / ***−0.39***	−0.38 / ***−0.39***								
VBF exc., 8 TeV, LO	−0.41 / ***−0.41***	−0.44 / ***−0.45***	−0.31 / ***−0.33***	0.06 / ***0.05***	0.06 / ***0.05***	0.04 / ***0.05***	1. / ***0.99***									
VBF inc., 8 TeV, LO	−0.41 / ***−0.41***	−0.44 / ***−0.45***	−0.31 / ***−0.33***	0.06 / ***0.05***	0.06 / ***0.05***	0.04 / ***0.05***										
GGH 2j full mass, 8 TeV, LO	0.27 / ***0.29***	0.06 / ***0.08***	0.57 / ***0.6***	0.99 / ***0.99***	0.99 / ***0.99***											
GGH 2j exc., 8 TeV, LO	0.27 / ***0.29***	0.06 / ***0.08***	0.57 / ***0.6***	0.99 / ***0.99***												
GGH 2j inc., 8 TeV, LO	0.27 / ***0.29***	0.06 / ***0.08***	0.57 / ***0.6***													
GGH 1j exc., 8 TeV, NLO	0.93 / ***0.93***	0.83 / ***0.83***														
GGH 0j exc., 8 TeV, NLO	0.97 / ***0.97***															
GGH inc., 8 TeV, NNLO	3.3% / ***3.3%***	3.2% / ***3.2%***	3.6% / ***3.5%***	6.9% / ***6.8%***	6.9% / ***6.8%***	7.% / ***6.8%***	2.4% / ***2.4%***	2.4% / ***2.4%***	3.3% / ***3.3%***	3.2% / ***3.2%***	3.4% / ***3.4%***	5.7% / ***5.7%***	5.7% / ***5.7%***	5.8% / ***5.8%***	2.1% / ***2.%***	2.1% / ***2.%***

Correlation table for Higgs cross sections

Red indicates $|\cos(\phi)|>0.7$

Numbers in Italic–bold (plain) for 6 eigenvecotrs (full set 50 eig.)

META PDF 1.0 (LHCH)

VBF–like cut applied for 2 or more jets final states

jet (anti–k_T, 0.4) selection with $|y|<4.5$ and $p_T>30$ GeV

including α_s uncertainty

Upper numbers: META PDF 1.0 (100 error sets)

Lower numbers: LHCH (12 error sets)

Preliminary

Conclusions

- A meta-analysis combines PDF ensembles at the level of their parametrizations, not at the level of LHC observables. META PDF ensembles are available to compute the PDF+ uncertainty for LHC observables at NNLO for the combination of CT10, MSTW'08, and NNPDF2.3 ensembles.
- The META representation reproduces the PDF uncertainties and correlations of the input ensembles. The common parametrization simplifies statistical analysis and combination. Only 13 meta-PDFs (LHCH) are required to compute PDF uncertainties and correlations in various Higgs production channels

Possible extensions

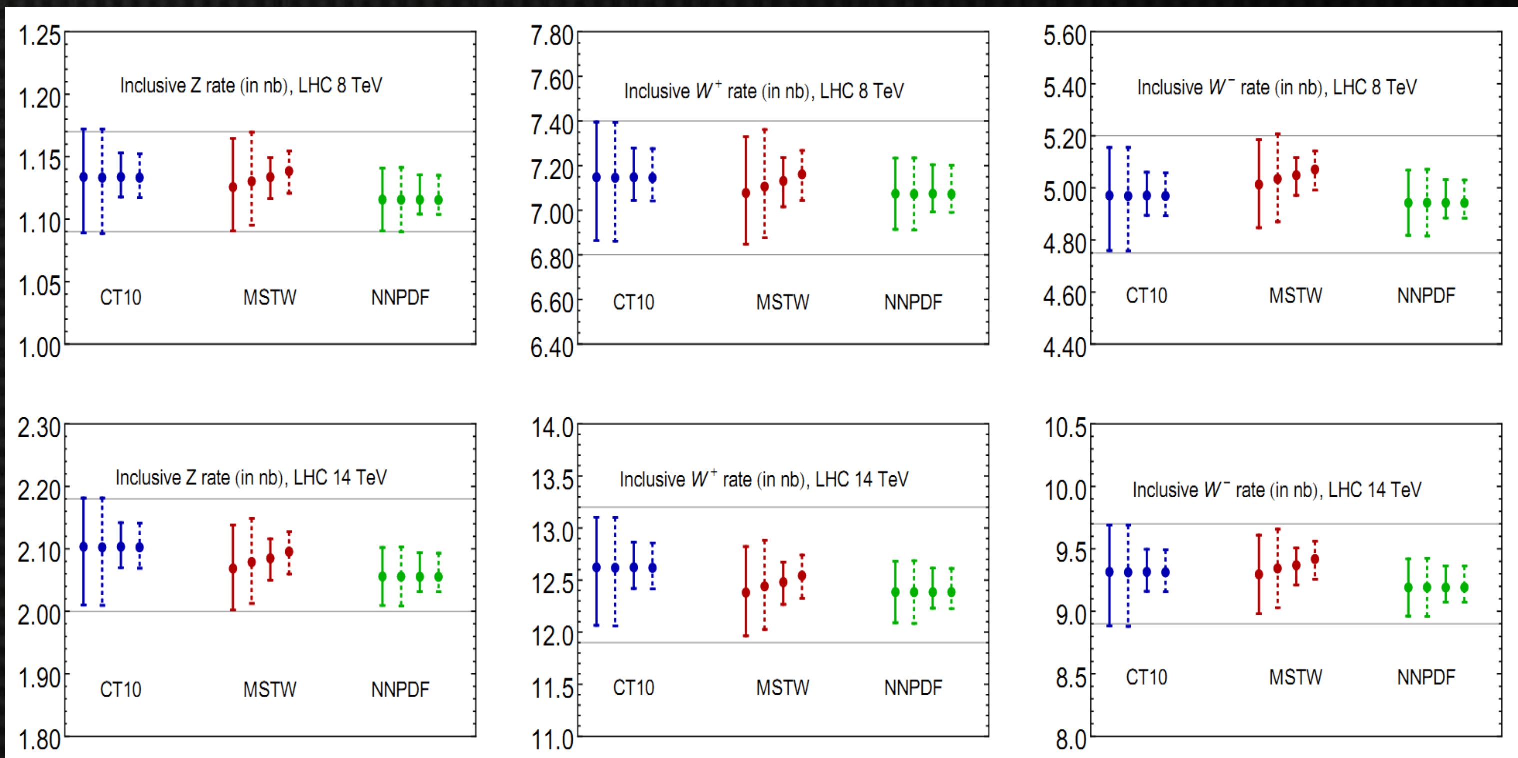
- Include more ensembles into META PDFs
  - Compatibility of data sets, $\alpha_s(M_Z)$, central values, definitions of PDF uncertainties
  - More refined combination procedure, e.g., including weights to account for data sample sizes, etc.
- Include new experimental measurements by PDF reweighting (according to Giele-Keller or NNPDF approach)

Back-up slides

25

Agreement on predictions for benchmark LHC processes

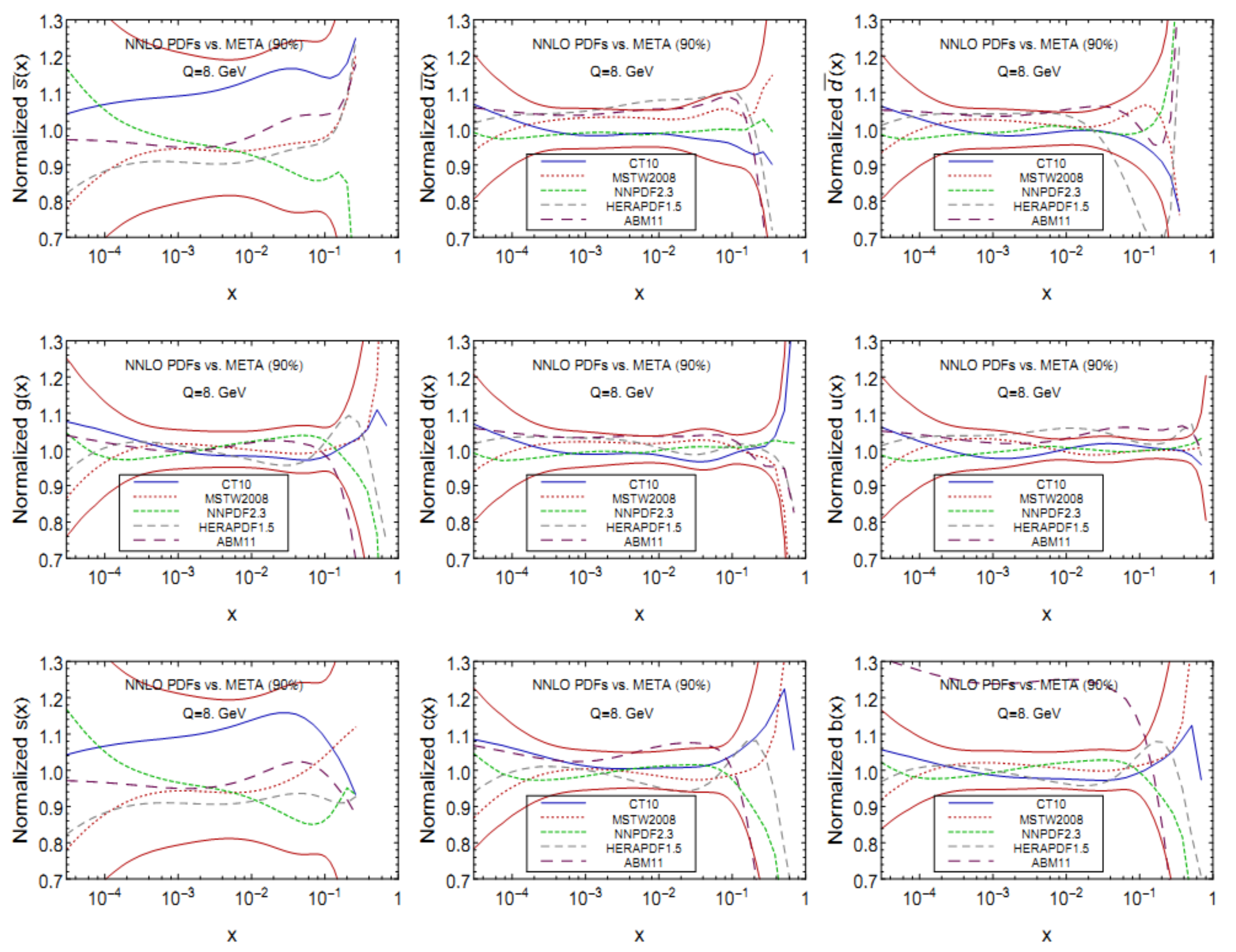
For NNLO inclusive rates of W, Z, Higgs, top quark pair production, NLO jet cross sections in different kinematic bins, at the LHC 8 and 14 TeV, the fitted PDFs can well reproduce predictions of the original PDFs including the PDF induced correlations.



blue(CT10), red(MSTW), green(NNPDF), solid(dotted) for input (fitted) PDFs

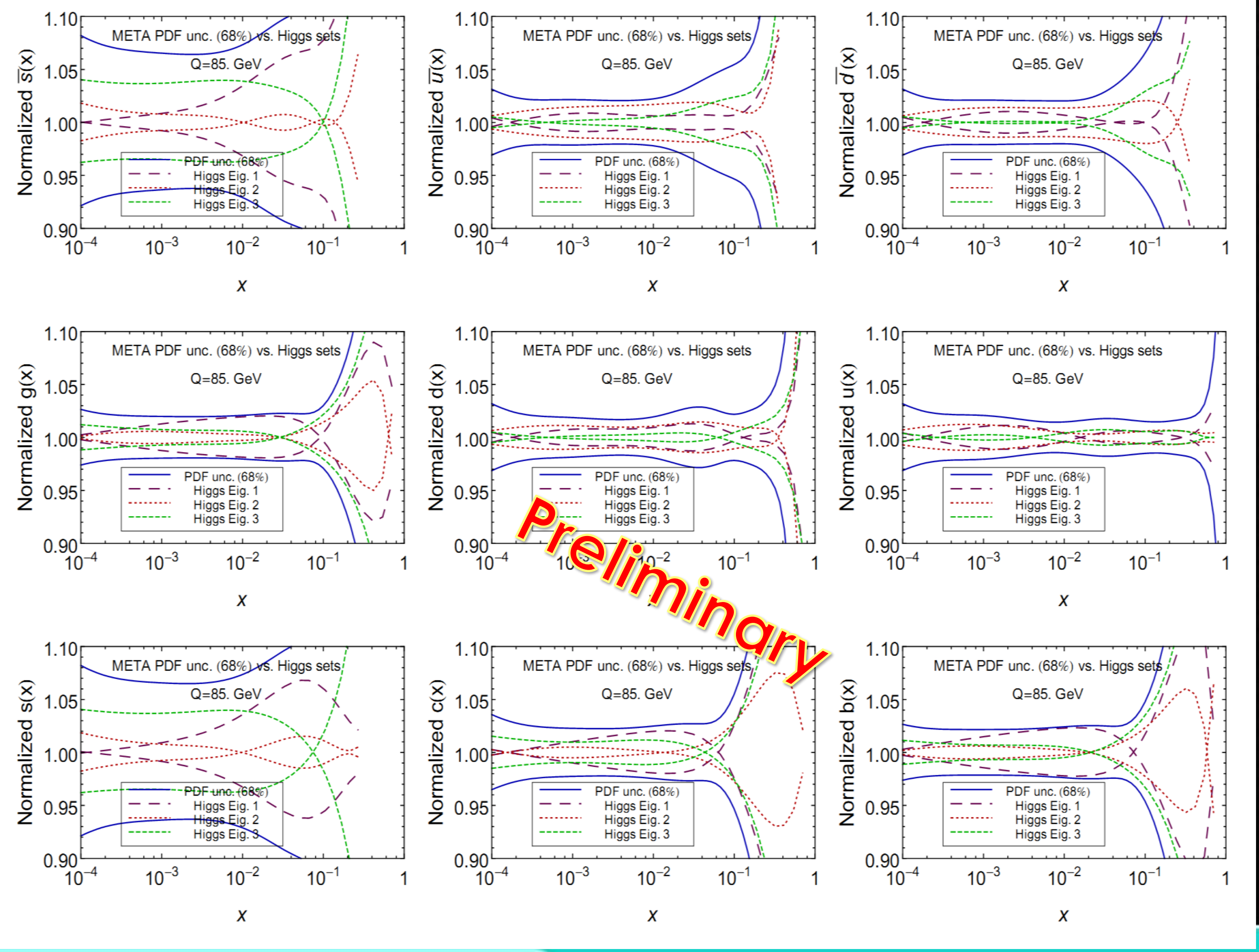
The META PDFs

Comparisons of the META PDF with the original central PDF from each groups with $\alpha_S(M_Z)=0.118$ at Q=8 GeV.



The LHCH set

PDFs from the first 6 eigenvectors



The LHCH set

PDFs from the first 6 eigenvectors