PDF4LHC meeting, CERN, November 3, 2014

META PDFs as a framework for PDF4LHC recommendations

Jun Gao, Joey Huston,
Pavel Nadolsky (presenter)

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

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

Only in the META set

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"

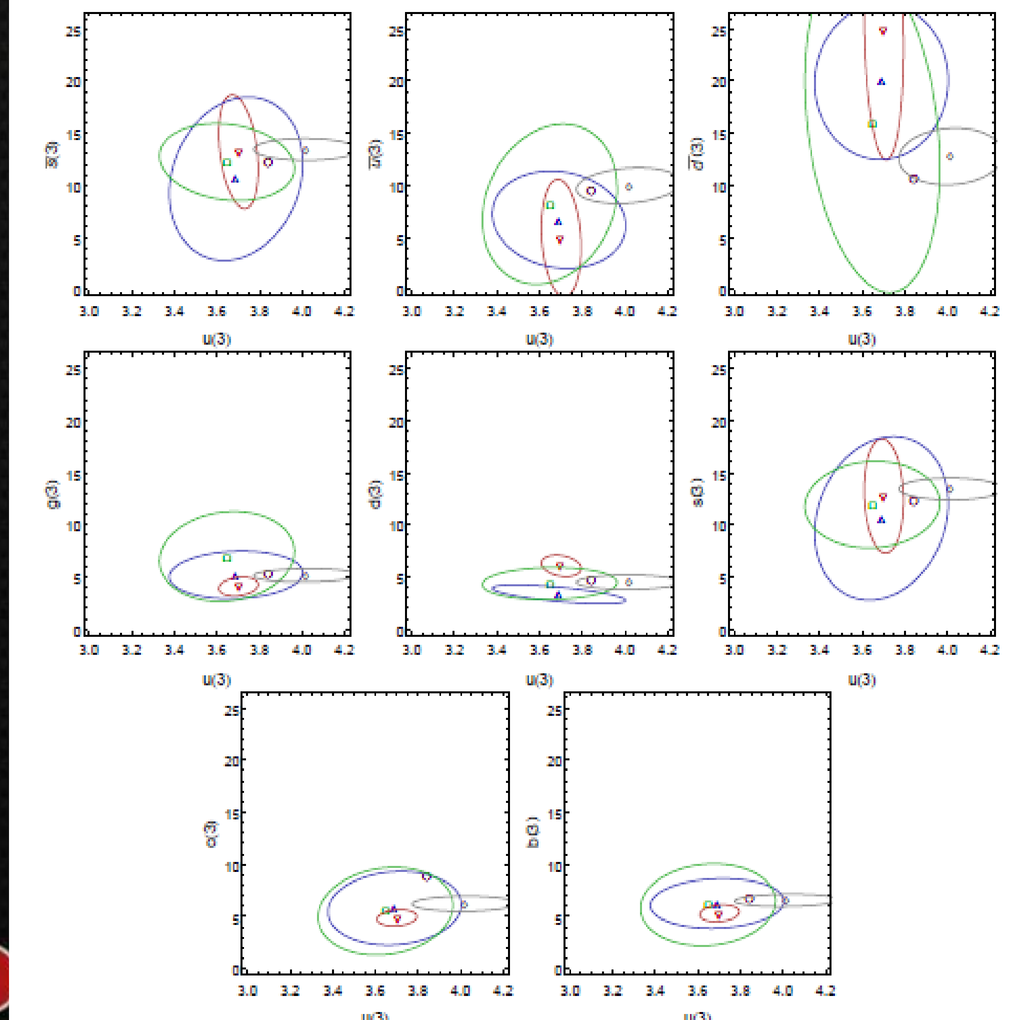


Figure 10: Fitted PDF parameters and 90% c.l. ellipses for CT10 (blue up triangle), MSTW08 (red down triangle), NNPDF2.3 (green square), HERAPDF1.5 (gray diamond) and ABM11 (magenta circle).

Specific realization (META1.0)

- Take NNLO (or NLO) PDFs

NNLO	*Initial scale*	a_s	*Error type*	*Error sets*
CT10	*1.3*	*0.118*	*Hessian*	*50*
MSTW'08	*1.0*	*0.1171*	*Hessian*	*40*
NNPDF2.3	*1.414*	*0.118*	*MC*	*100*

- **Choose a meta-parametrization of PDFs at initial scale of 8 GeV (away from thresholds) for 9 PDF flavors (66 parameters in total)**

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

- Generate MC replicas for all 3 groups and merge with equal weights, finding meta parameters for each of the replicas by fitting PDFs in x ranges probed at LHC
  - PDF uncertainties for above three groups are all reasonably consistent with each other (see next slide)
  - can add other sets, perhaps with non-equal weights to take into account more limited data sets and thus larger uncertainties

Compare relative luminosity uncertainties

good agreement in size of uncertainties between the 3 global PDFs

larger uncertainties of HERAPDF1.5 apparent

ABM11 uncertainties smaller at high mass

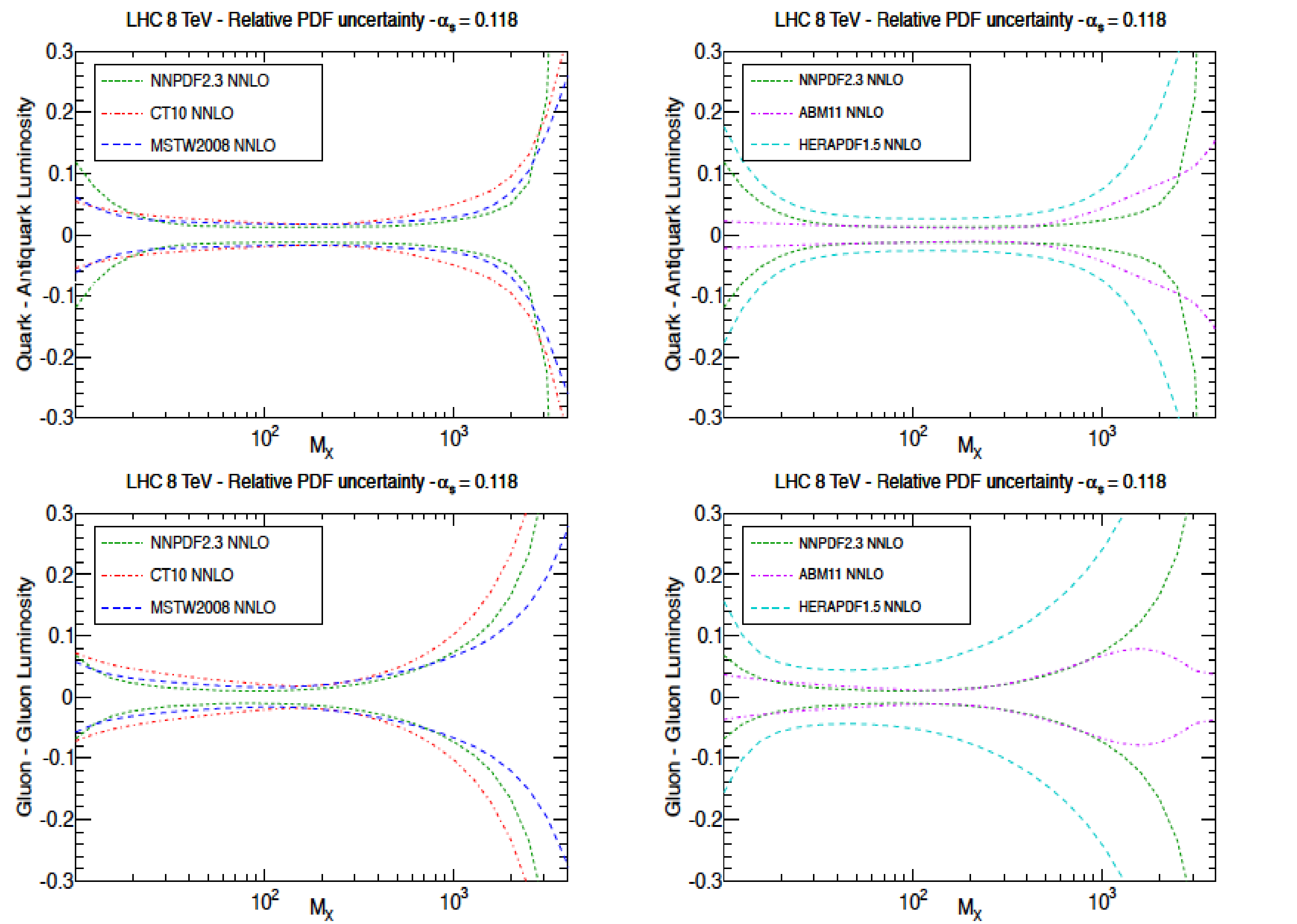


Figure 8: The relative PDF uncertainties in the quark-antiquark luminosity (upper plots) and in the gluon-gluon luminosity (lower plots), for the production of a final state of invariant mass M_X (in GeV) at the LHC 8 TeV. All luminosities are computed at a common value of $\alpha_s = 0.118$.

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

Reduction of the error PDFs

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

In the META1.0 study:
200 CT, MSTW, NNPDF error sets

⇒ 300 MC replicas for reconstructing the combined probability distribution

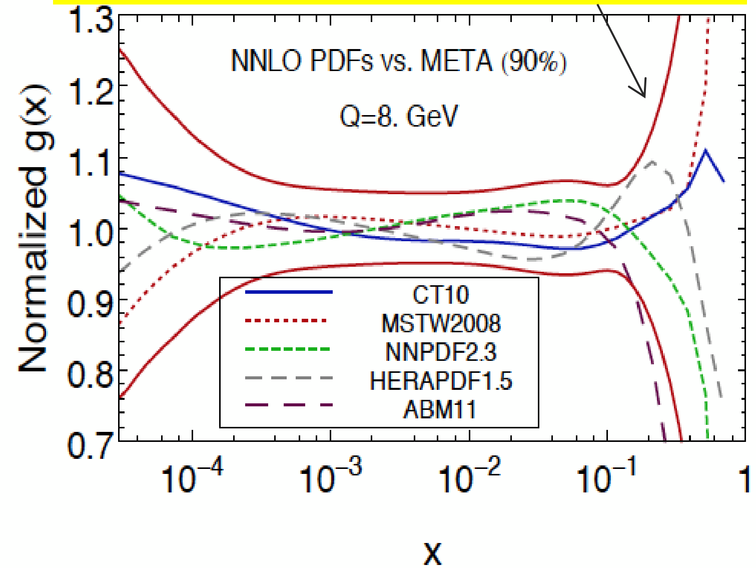
⇒ 100 Hessian META sets for most LHC applications (**general-purpose** ensemble META1.0)

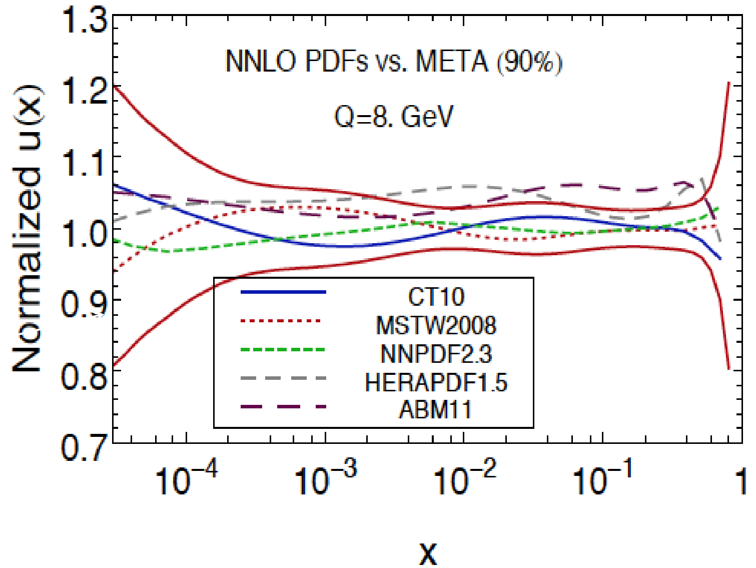
⇒ 13 META sets for LHC Higgs production observables (**reduced ensemble** META LHCH)

General-purpose META PDF ensemble

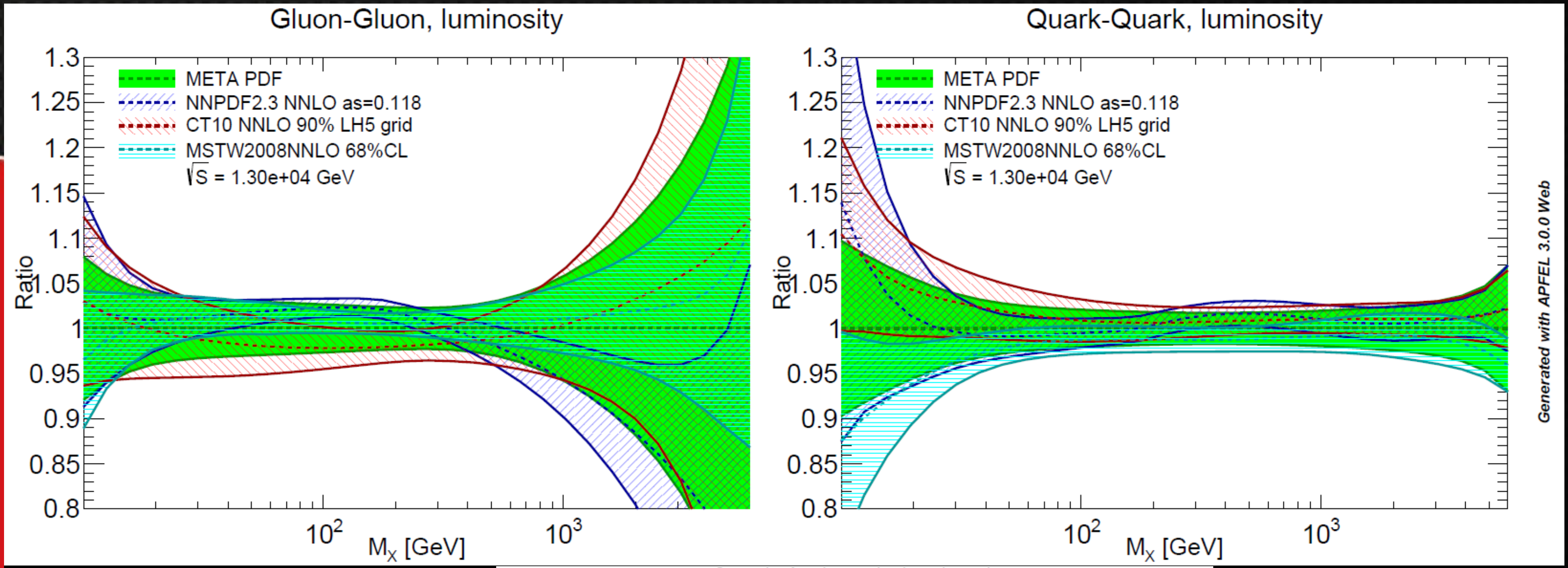
- 50 eigenvectors (100 error sets) provide a very good representation of the PDF uncertainties for all of the 3 PDF error families above
- The META PDFs provide both an average of the chosen central PDFs, as well as a good estimation of the 68% c.l. total PDF uncertainty
- Can re-diagonalize the Hessian matrix to get 1 orthogonal eigenvector to get the α_s uncertainty (H.-L. Lai et al., 1004.4624)

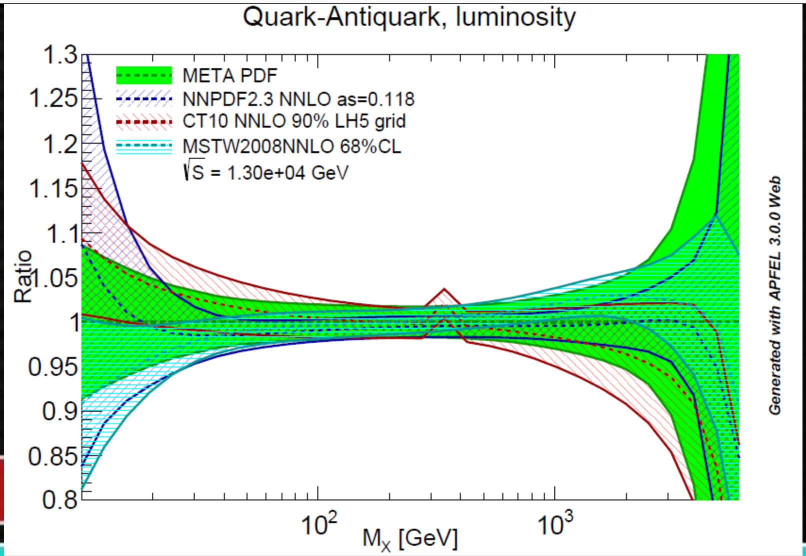
meta-PDF uncertainty band





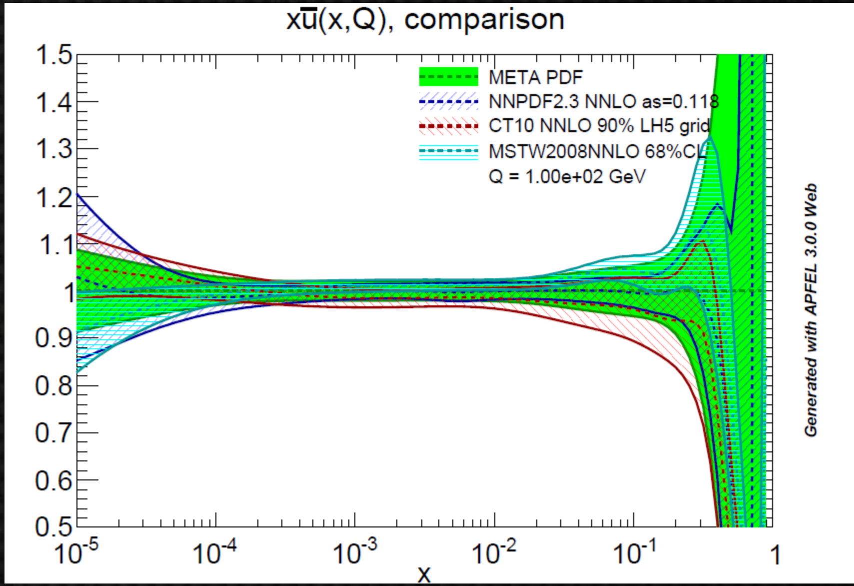
Some parton luminosities

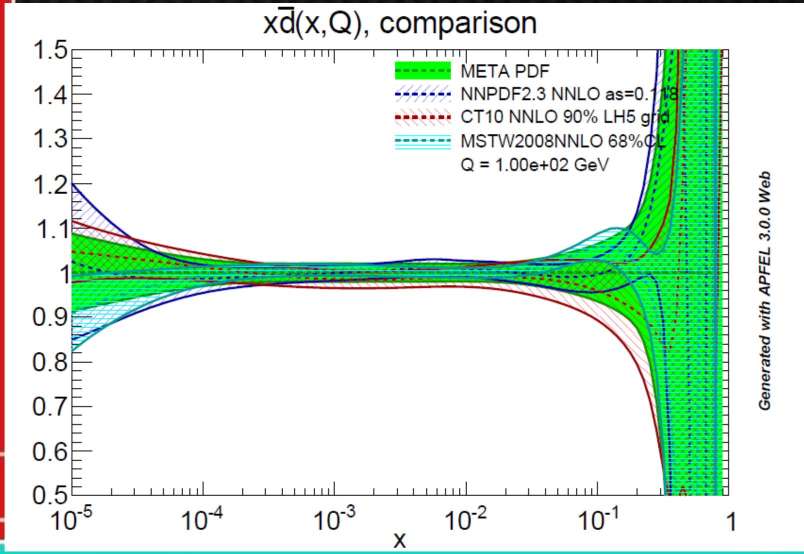


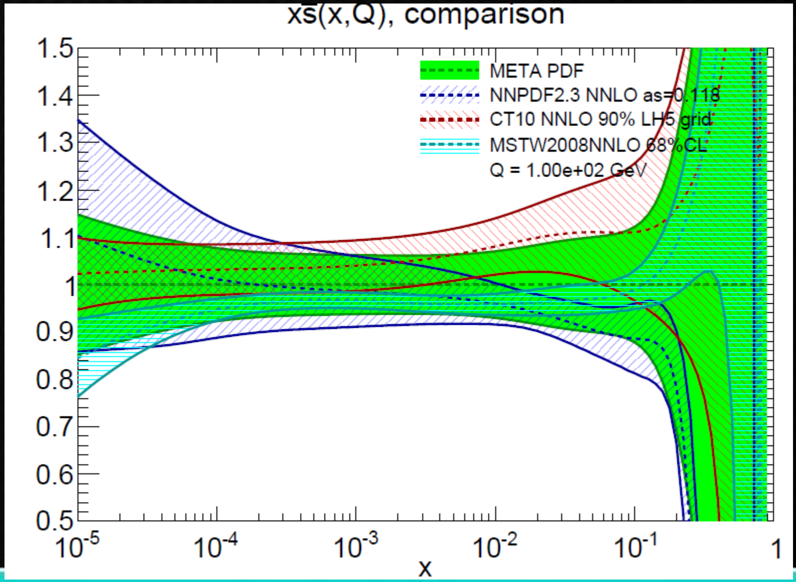


Plots are made with APFEL WEB (apfel.mi.infn.it; Carrazza et al., 1410.5456)

PDFs for sea quarks







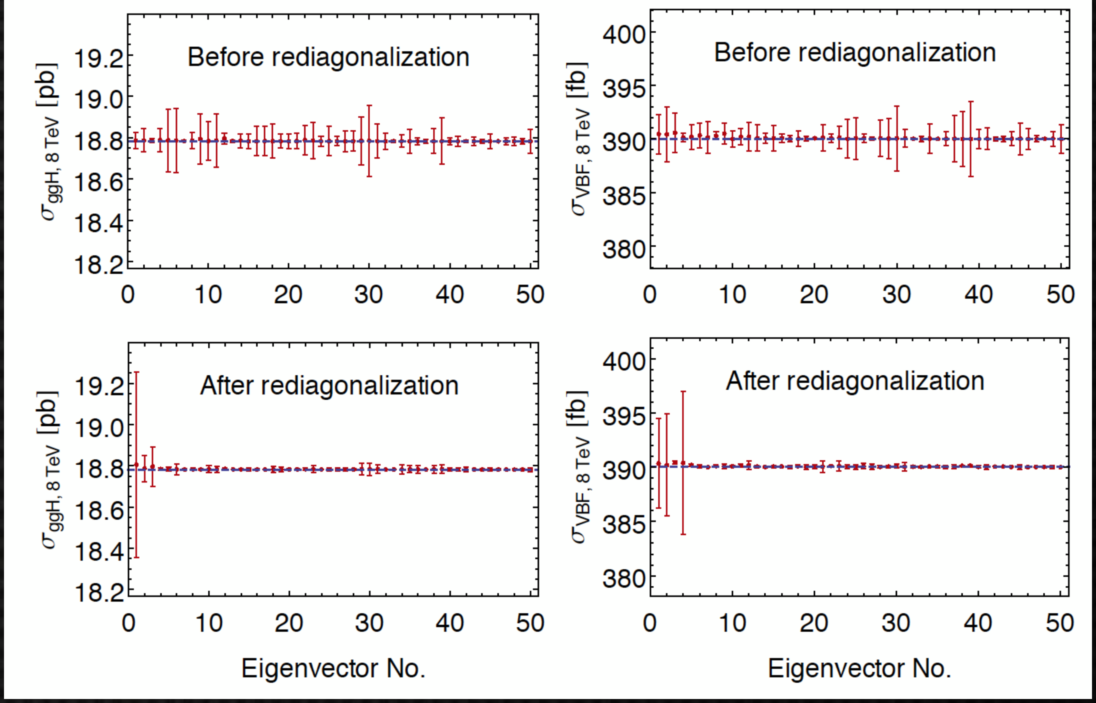
Reduced META ensemble

- 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	—	—	$2m_H$

Data set diagonalization (Pumplin, 0904.2424)

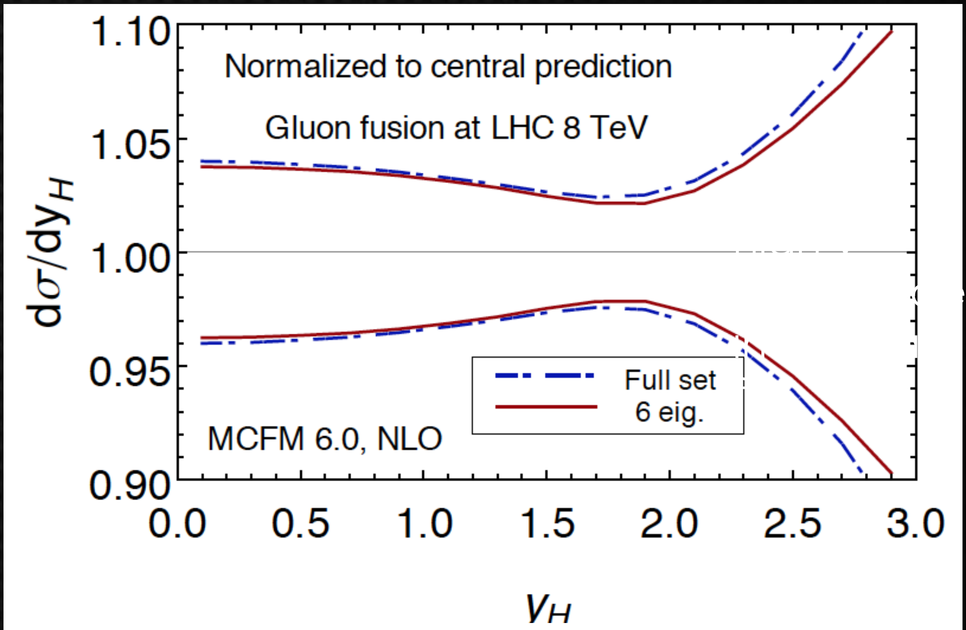
- There are 50 eigenvectors, but can rediagonalize the Hessian matrix to pick out directions important for the Higgs observables listed on previous page; with rotation of basis, 50 important eigenvectors become 6

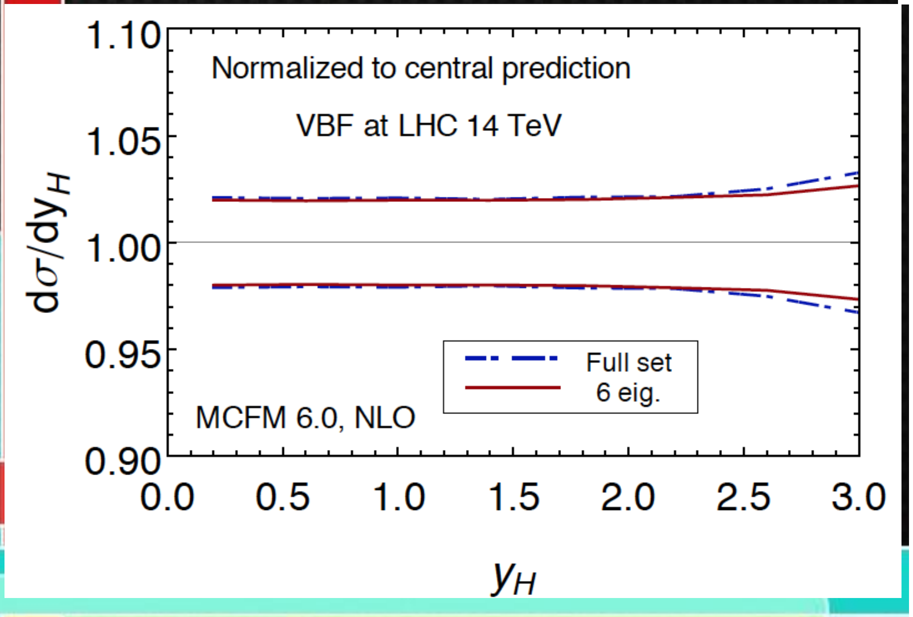


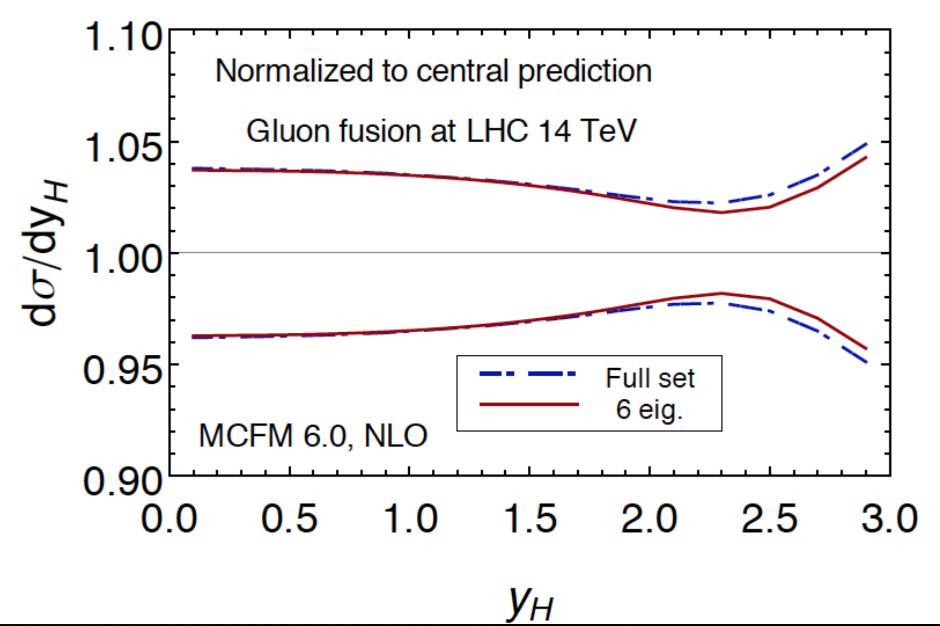
J. Gao,
J. Huston
P. Nadolsky
(in progress)

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

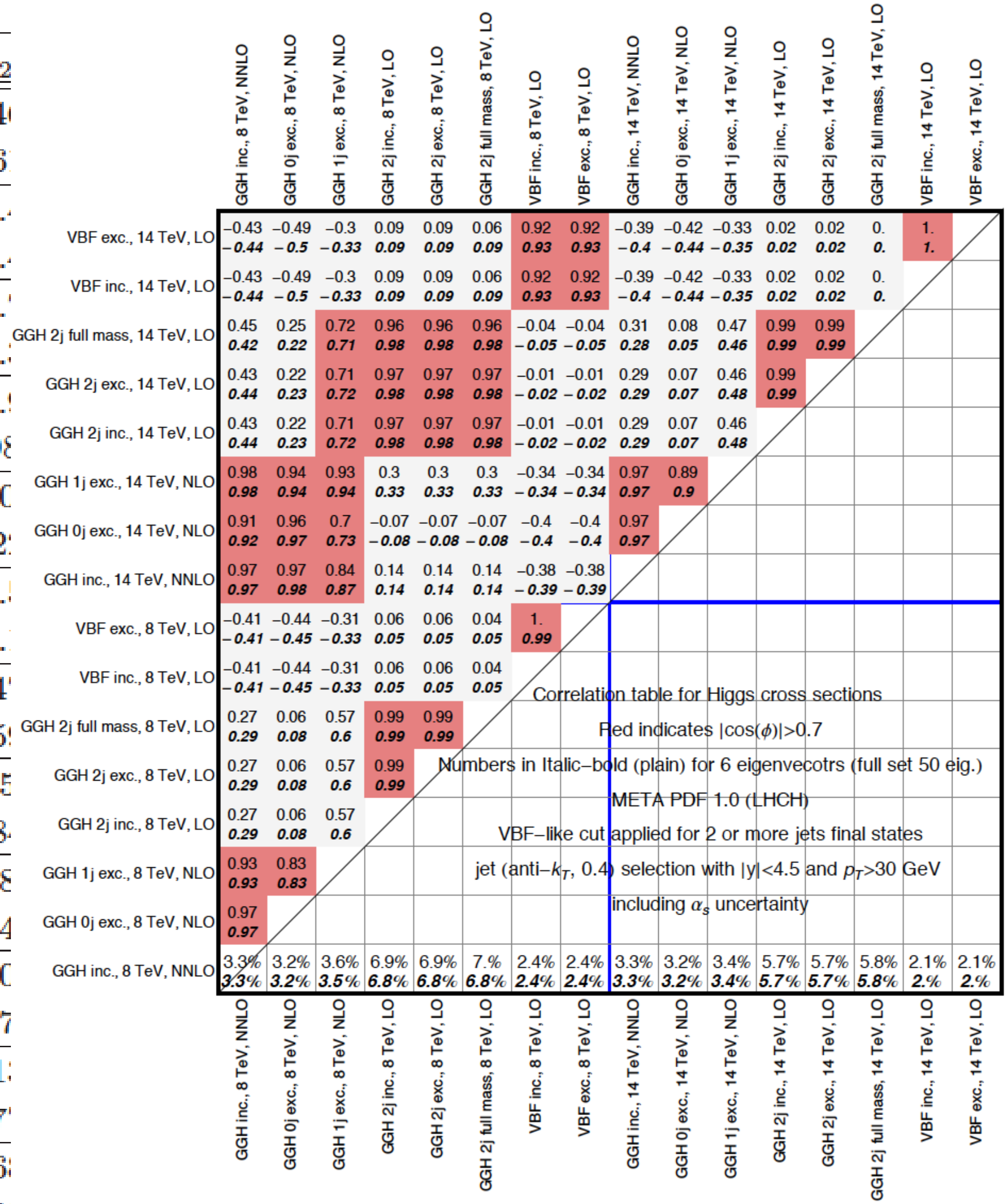
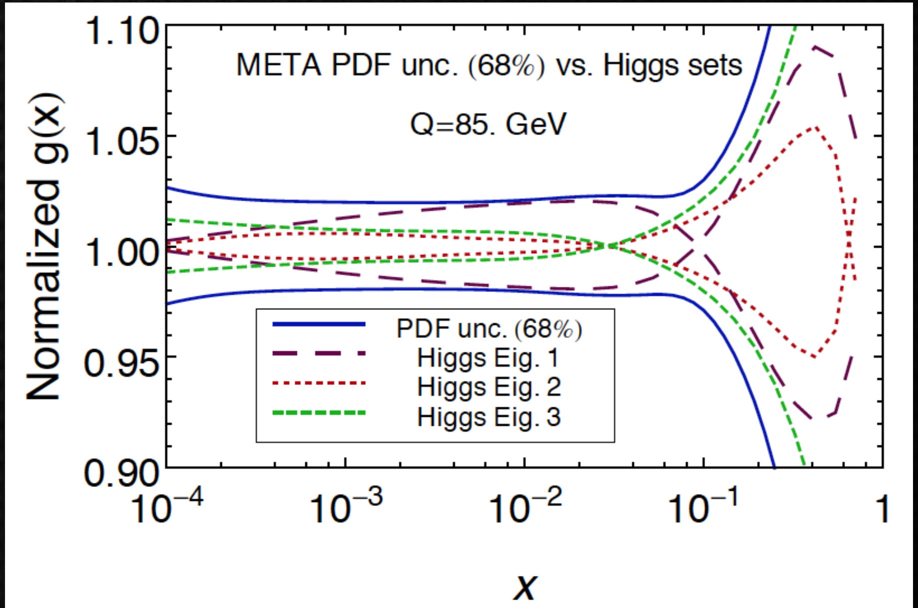


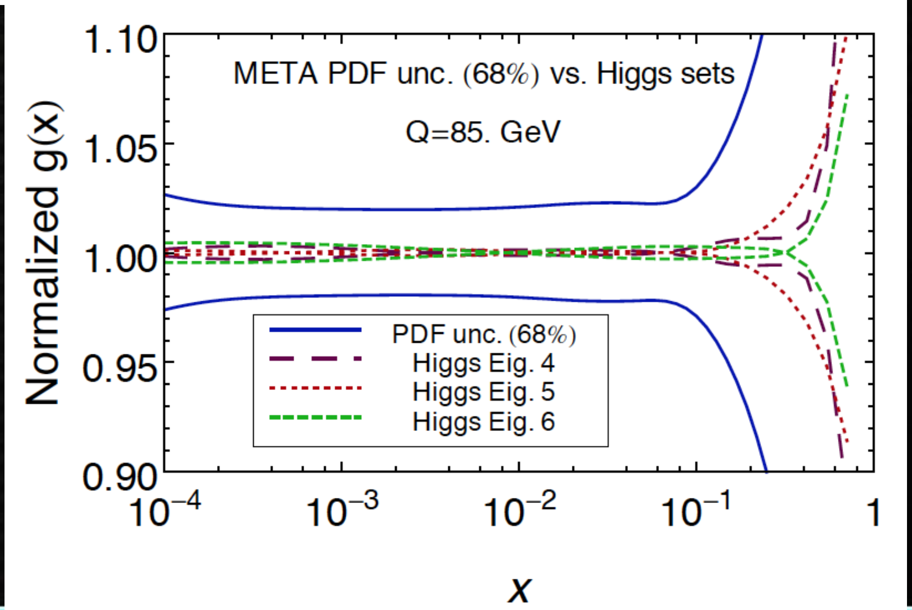
FIG. 7: Same as Fig. 5, with α_s uncertainties included by adding in quadrature.

Re-diagonalized eigenvectors…

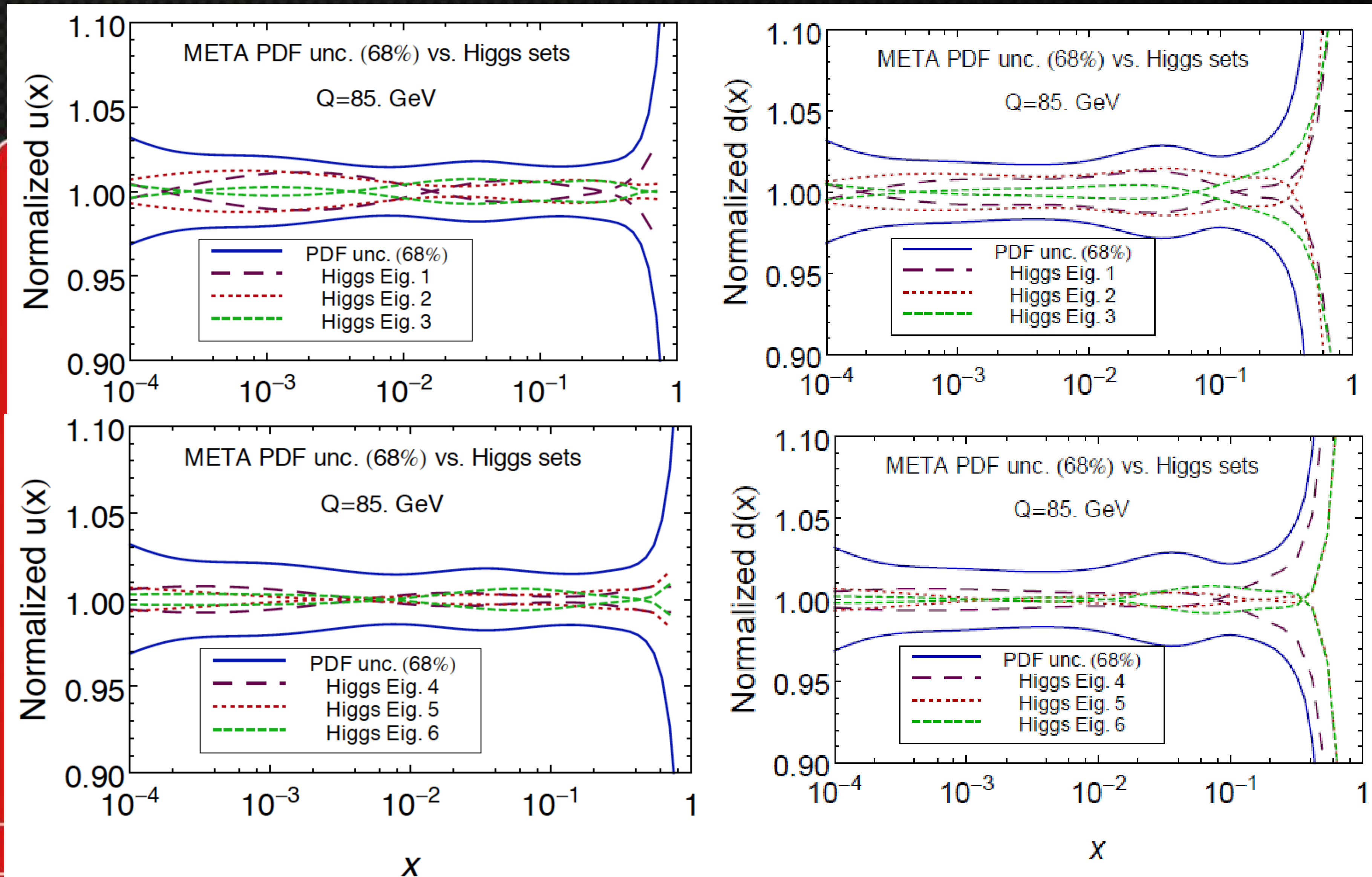
…are associated with the parameter combinations that drive the PDF uncertainty in Higgs, W/Z production at the LHC

- 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

Benefits of the meta-parametrization

- The PDF parameter space of all input ensembles is visualized explicitly.

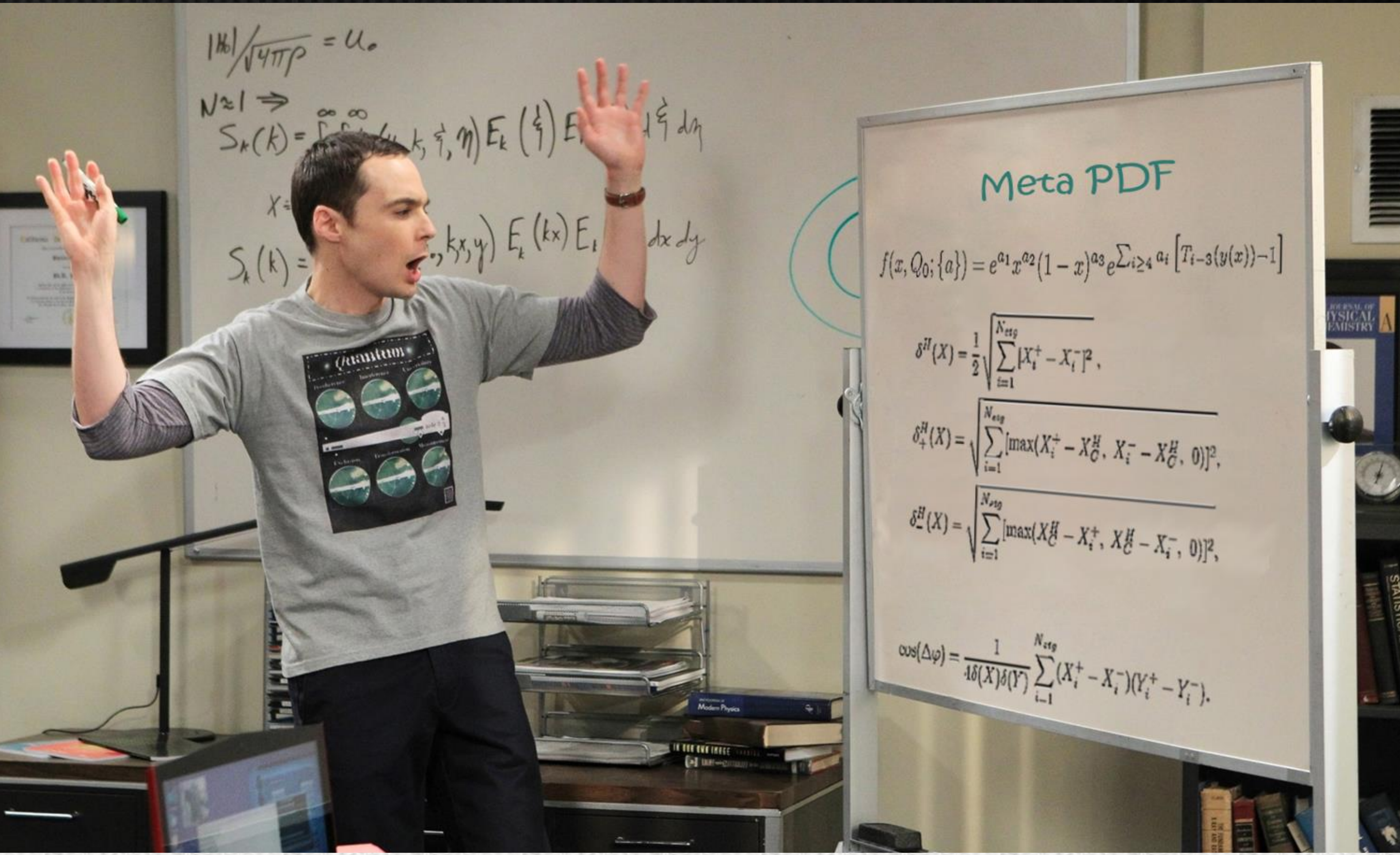
- Data combination procedures familiar from PDG can be applied to each meta-PDF parameter

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

Benefits of the Hessian method

- It is very effective in data reduction, as it makes use of diagonalization of a semipositive-definite Hessian matrix in the Gaussian approximation. [The unweighting of MC replicas is both more detailed and nuanced.]
- Correlations between Higgs signals and backgrounds are reproduced with just 13 META PDFs. It remains to be seen how many MC replicas will be needed to reproduce the correlations in the MC compression approach.

Here is how theorists and experimentalists generally react to our suggestion that 13+2 META PDFs are enough to compute the Higgs PDF+α_s uncertainties :



Jun, Pavel,

Sheldon approves! META PDFs will be on the Big Bang Theory episode on Jan. 29, 2015. Regards, Joey

Implementation of plan

- Apply these ideas to first combine at least CT14, HERAPDFX, MMHT14, NNPDF3.0
- Jun Gao will release a Mathematica notebook that will
  - read input LHAPDF grids for all ensembles
  - fit all grids by a common parametrization
  - produce MC replicas for all ensembles
  - obtain the 68%CL volume for all replicas
  - diagonalize the Hessian matrix for the 68% CL volume
  - perform the data set diagonalization to obtain a reduced Hessian eigenvector set for a given class of observables
- Propose that the above techniques be used by a PDF4LHC working group to obtain META PDFs with both a full set and a reduced set for high Q^2 physics

Back-up slides

22

Meta-parameters of 5 sets and META PDFs

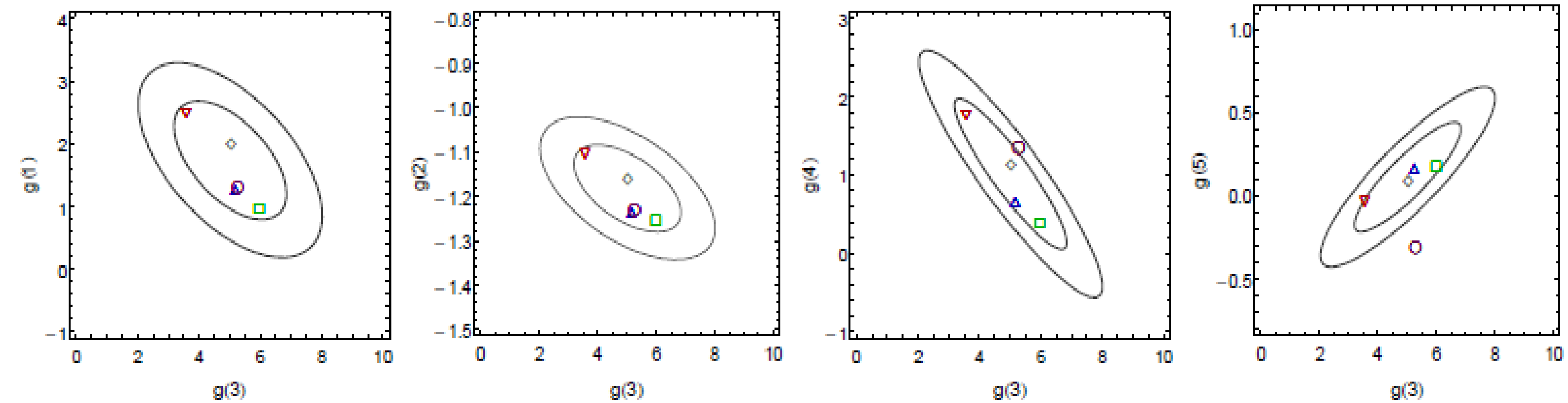


Figure 16: Comparison of META PDF confidence intervals with central NNLO PDFs of the input PDF ensembles in space of meta-parameters a_{1-5} for the gluon PDF. Up triangle, down triangle, square, diamond, and circle correspond to the best-fit PDFs from CT10, MSTW, NNPDF, HERAPDF, and ABM respectively.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

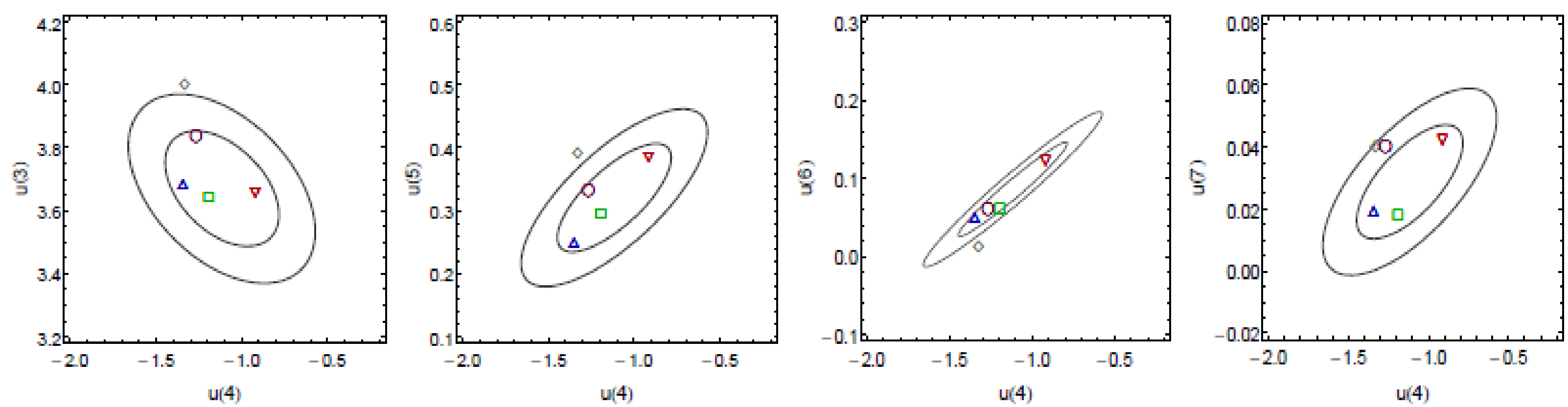


Figure 17: Same as Fig. 16, for a_{3-7} of the u quark PDF.