PDF4LHC meeting, CERN, June 26, 2015

Final combination of META2.0 parton distributions

Jun Gao, Joey Huston, Pavel Nadolsky

META2.0, status update, 2015-05-26

- Finished the META2.0 combination based on the **official CT14 (arXiv:1506. 07443)**, MMHT'14, and NNPDF3.0; NLO & NNLO
- Verified stability w.r.t. to the choice of the functional form of meta-parametrizations (**Bernstein polynomials** in v. 2)
- Selected an optimal number of eigenvector error sets (**60**) for the final combination
- Completed **benchmark comparisons** for LHC observables
- Beta development of a **public Mathematica module MP4LHC** for meta-analysis: improvements in user interface, platform independence, mostly focusing on smooth compilation with LHAPDF6+BOOST library

META parameters of PDFs (arXiv:1401.0013)

- The core idea of the meta-analysis approach is to cast all input PDFs into a shared parametric representation.
- META parameters can be selected in many ways
  - ***Our approach:*** By fitting $f_i(x, Q)$ by flexible functions $F_i(\{a\}; x, Q)$, such as those based on smooth polynomials
  - ***MC2Hessian approach:*** By treating the PDF values themselves as parameters, $f_i(x_j, Q_l) \equiv f_{ijl}$;
    - is close, or even equivalent, to using step functions of x as the expansion basis

Confirming the absence of the functional bias

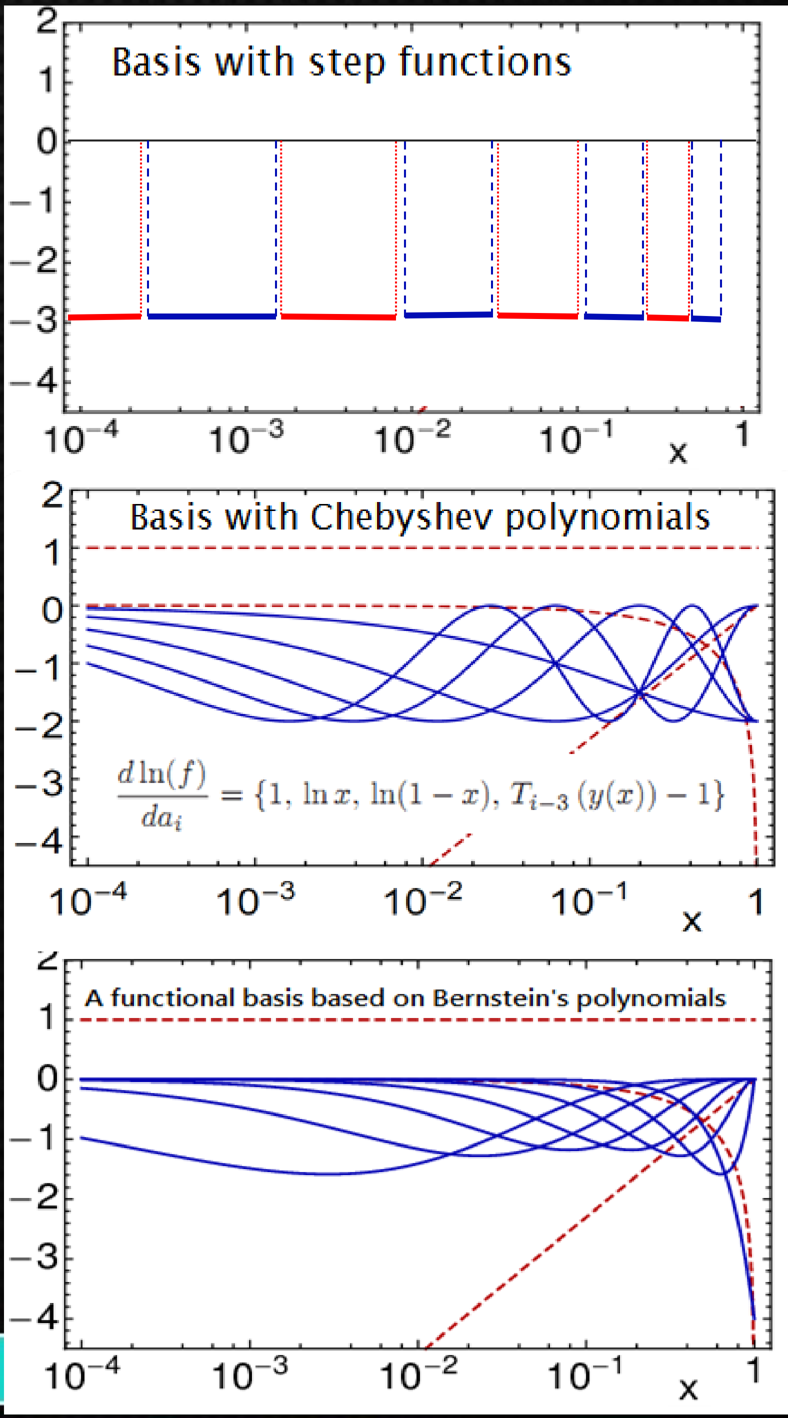
- We verified that the combination is not biased by the choice of parametrization with several tests
  - By noticing that different functional forms (e.g., based on Chebyshev and Bernstein polynomials) produce compatible PDF uncertainties and correlations
  - By checking that dependence on the number of eigenvector sets is weak

Which basis functions to use?

A step function is constant in one x interval; no correlations between adjacent x regions; many error sets needed for combination (**100-200 sets** (?); compare to 180 error sets of the related **mc2hessian** method)

A Chebyshev polynomial (**META1.0**) has multiple peaks: redundant correlations between far x regions, **~80-100 error sets**

A Bernstein polynomial (**META2.0**) has a single peak only; correlations are present only with neighboring x regions; **~40-60 error sets**



Reduction of the error PDFs

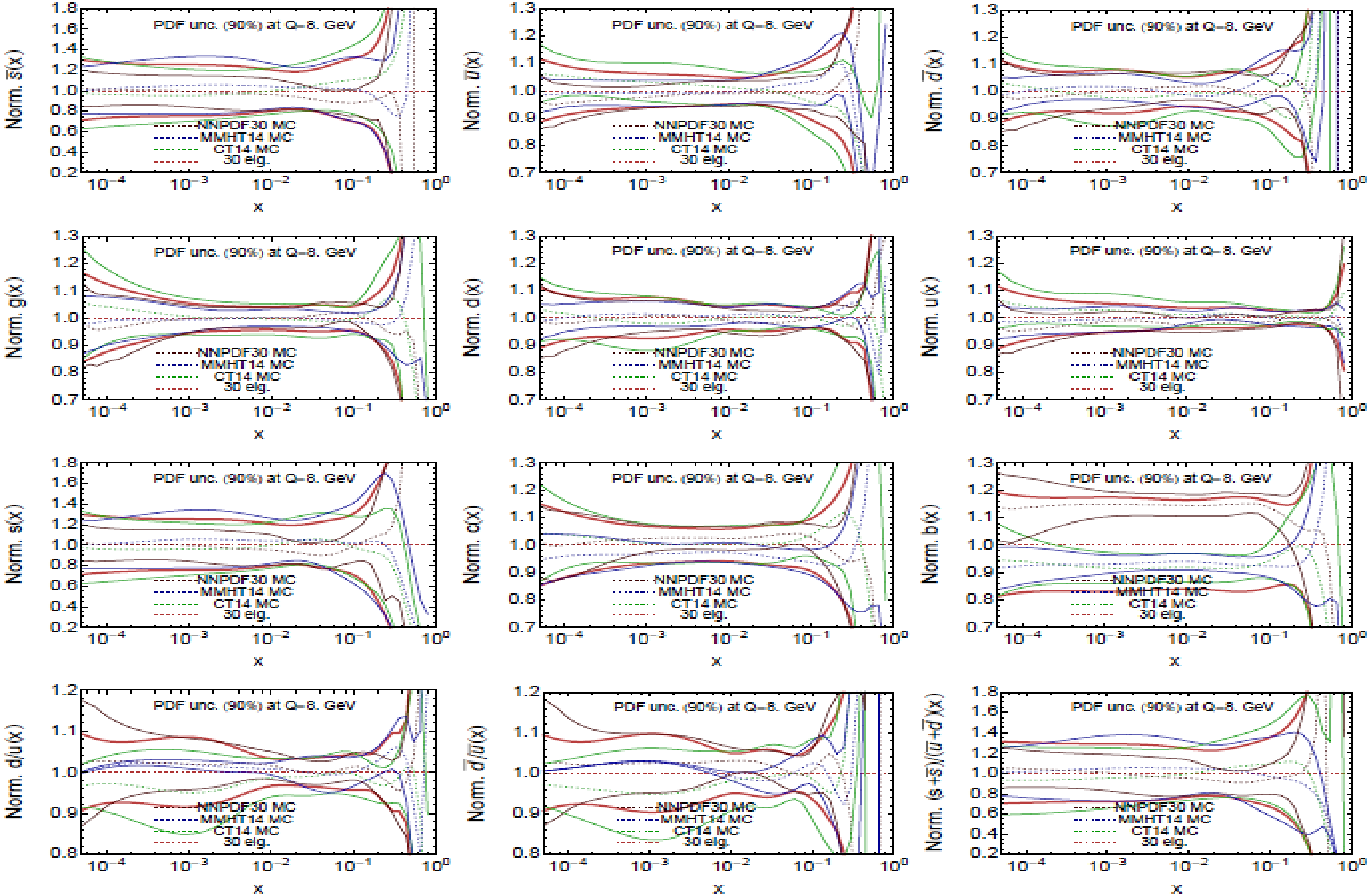
The number of final error PDFs is much smaller than in the input ensembles

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

⇒ 600 MC replicas for reconstructing the combined probability distribution

⇒60 Hessian META sets for most LHC applications (**general-purpose** ensembles META2.0; 40 or 100-member sets can be also made available)

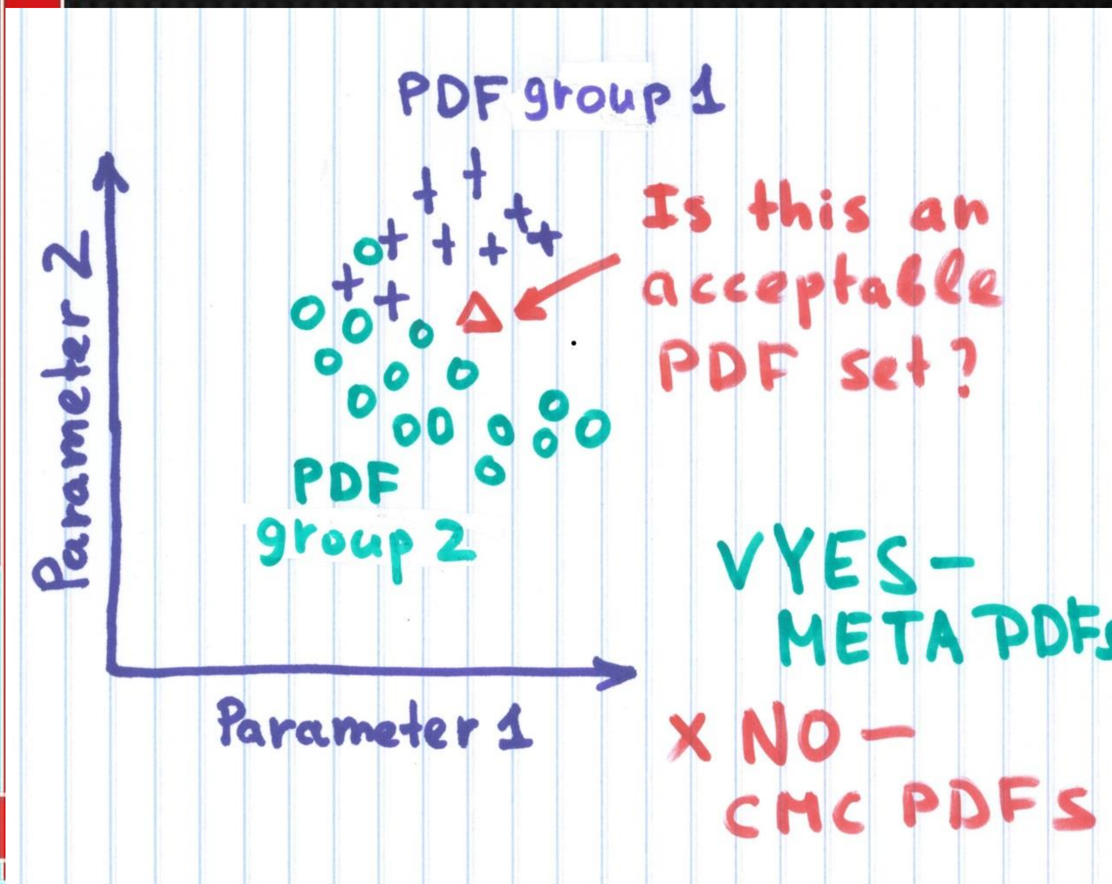
⇒ 20 META sets for LHC Higgs production observables (**reduced ensemble** META LHCH, obtained using the method of data set diagonalization)



META 2.0 PDFs utilizes the Gaussian approximation. The META60 ensemble "averages out" non-Gaussian features of input PDFs and their ratios from CT, MMHT, NNPDF MC sets

Gaussian confidence regions are convex

Gaussian error propagation implies that a linear combination of solutions from the input PDF groups is also an allowed solution.



This may be a reasonable assumption in many practical situations, e.g., when the input PDF groups do not sample all possible parameter space.

This feature is absent in the combination with MC replicas

Choice of the initial scale Q_0

- **Initial scale for META2.0, $\mathbf{Q_0 = 8}$ GeV**, is selected sufficiently above the bottom mass. Below Q_0, different heavy-quark schemes must be used in hard cross sections for CT, MMHT, NNPDF PDFs $\Rightarrow$ The user must be made aware they cannot be naively combined.
- At $Q > Q_0$, META PDFs can be used with zero-mass hard cross sections, 5 active flavors. Individual heavy-quark schemes are unnecessary. PDFs can be combined.

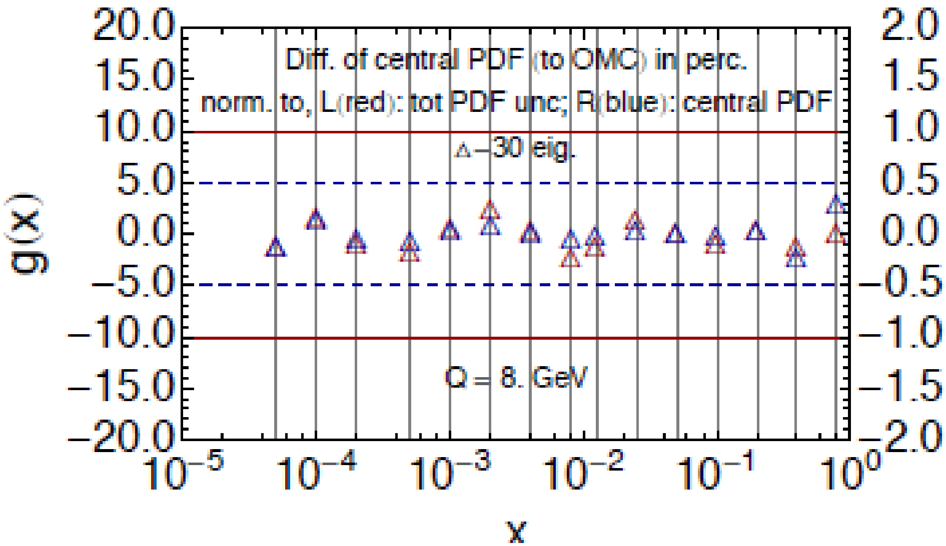
Mathematica module MP4LHC

- Implements all necessary functions to perform META analysis, data set diagonalization, etc. within ≈ 1 day
- **IMPORTANT:** Mathematica finds **all** eigenvalues of the Hessian matrix H_{ij} with high accuracy. Eigenvalues of H_{ij} for a typical PDF set span up to 10 orders of magnitude. Common diagonalization codes can lose precision dramatically. For CTEQ Hessian analysis, Pumplin had to revise CERN MINUIT to evaluate small eigenvalues, prevent wrong solutions for poorly constrained eigenvector sets.
- MP4LHC utilizes versatile *Mathematica* methods for singular value decomposition of H_{ij}. It can achieve essentially arbitrary accuracy for any reasonable number of parameters

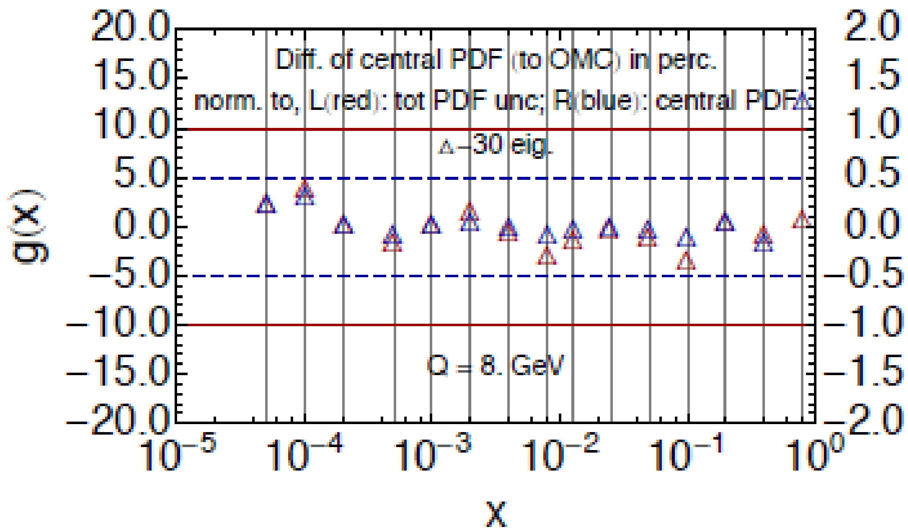
META2 validation: examples of comparisons

Differences between META60 and MC600

For central PDFs



For PDF uncertainties

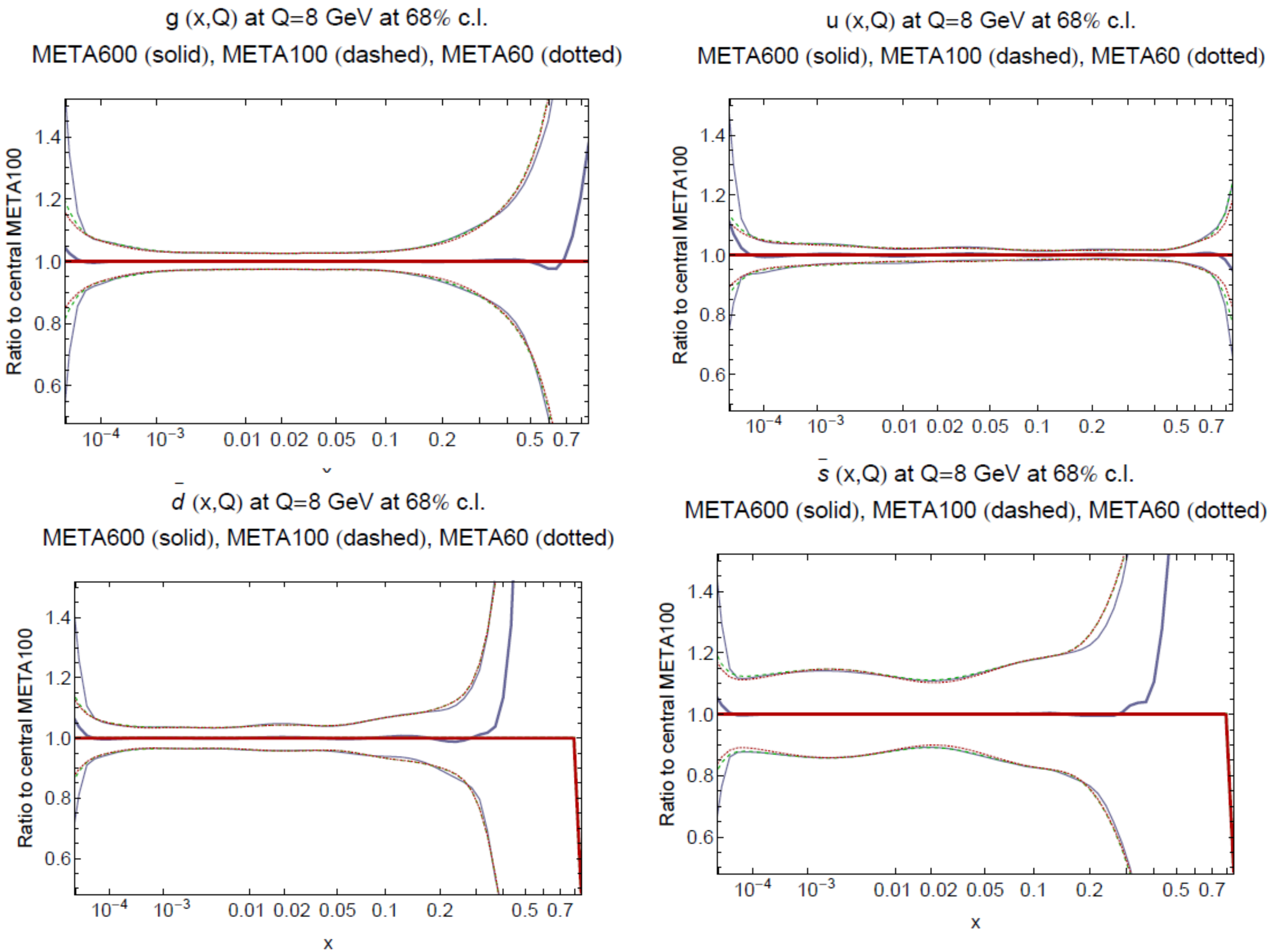


Left axis, red triangles: as percentages of PDF uncertainty

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

More comparisons in backup slides

META2 validation: 600 → 100 → 60 error sets

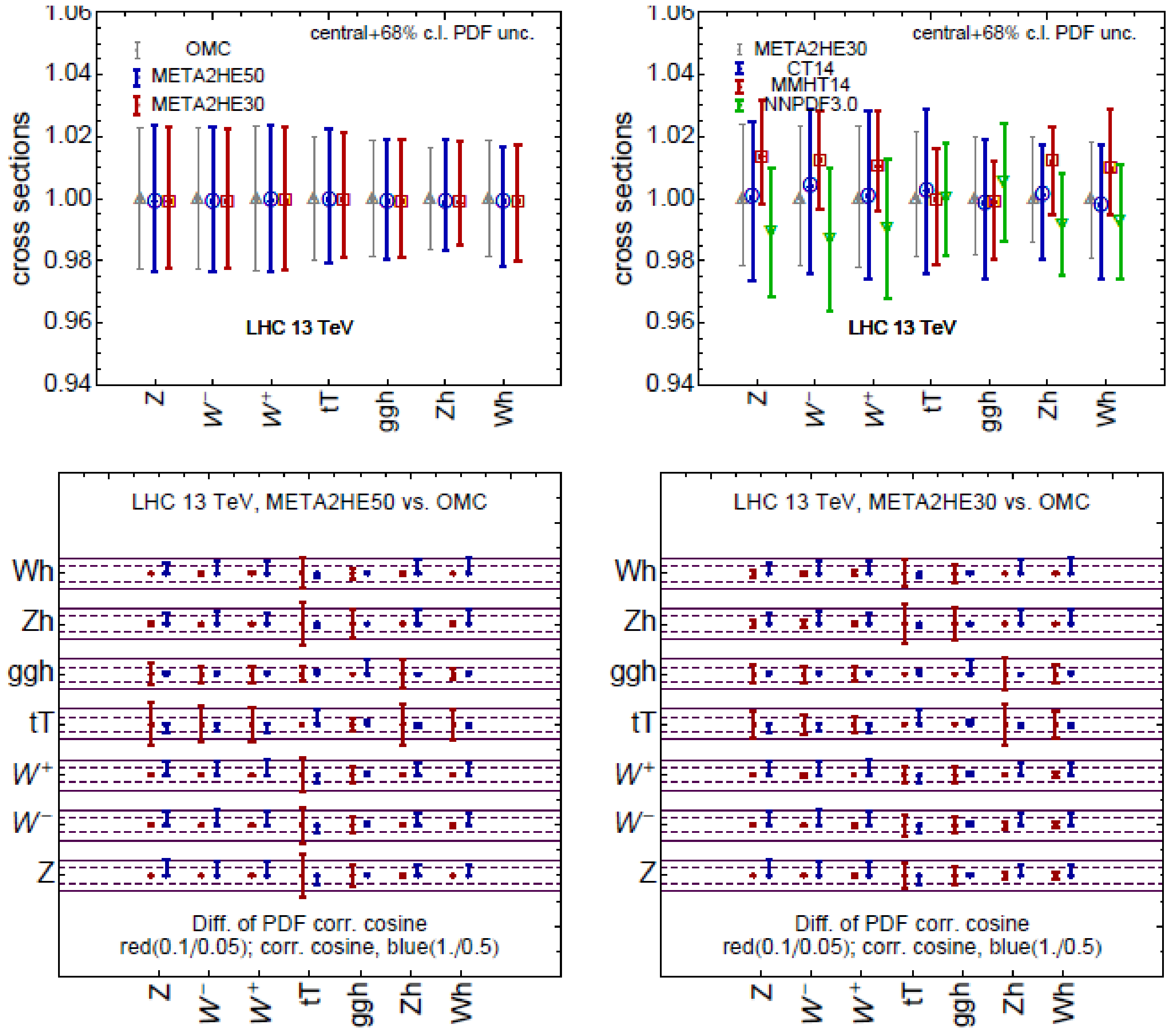


Dependence on the number of eigenvectors is **weak**

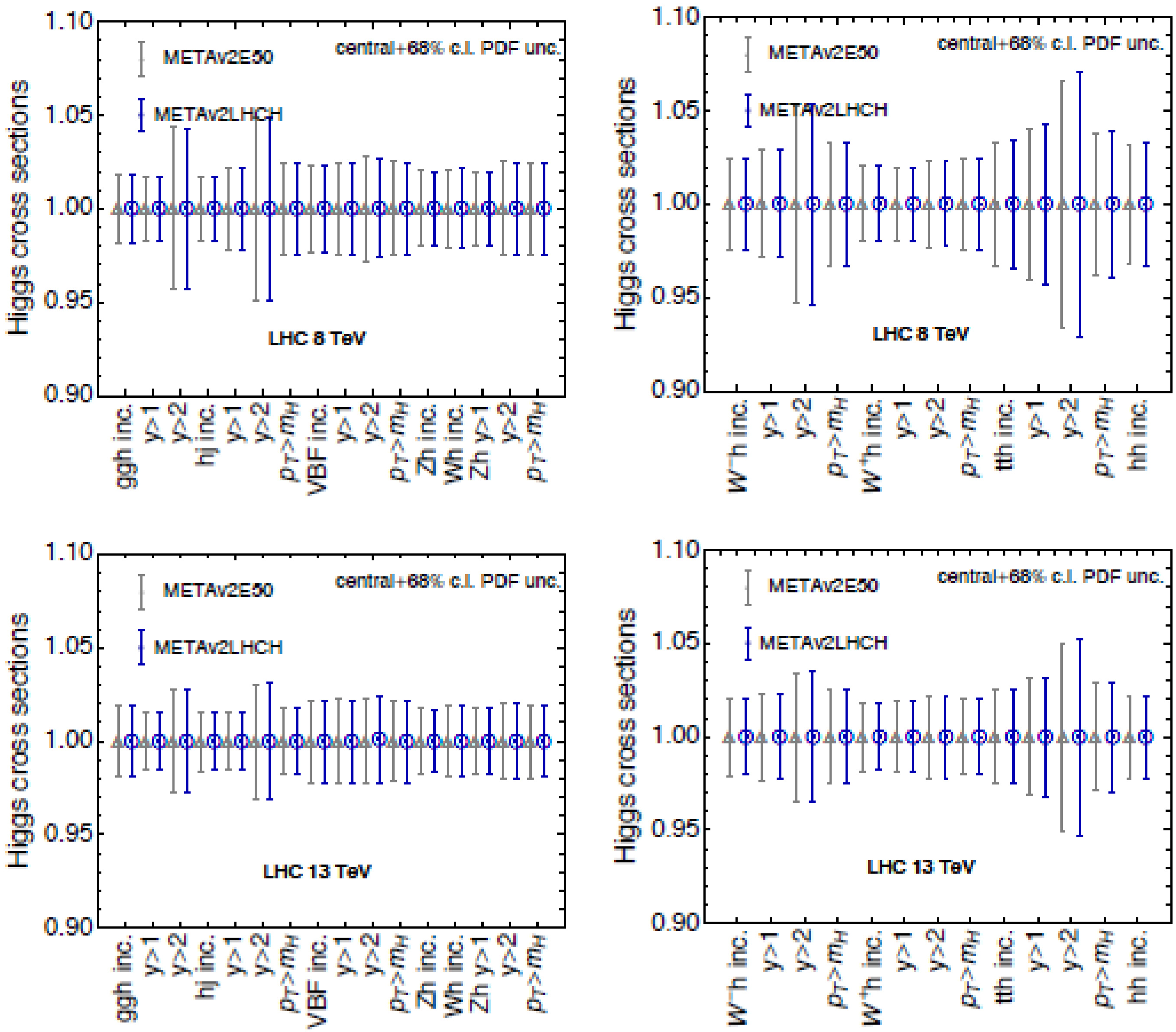
META2.0 validation: LHC benchmark cross sections, correlations

Total cross sections (up) and correlations (down) from META600 (=OMC) are well reproduced by META60 (=META2E30)

Blue pts: $cos\varphi$
Red pts: $\delta\cos\varphi$

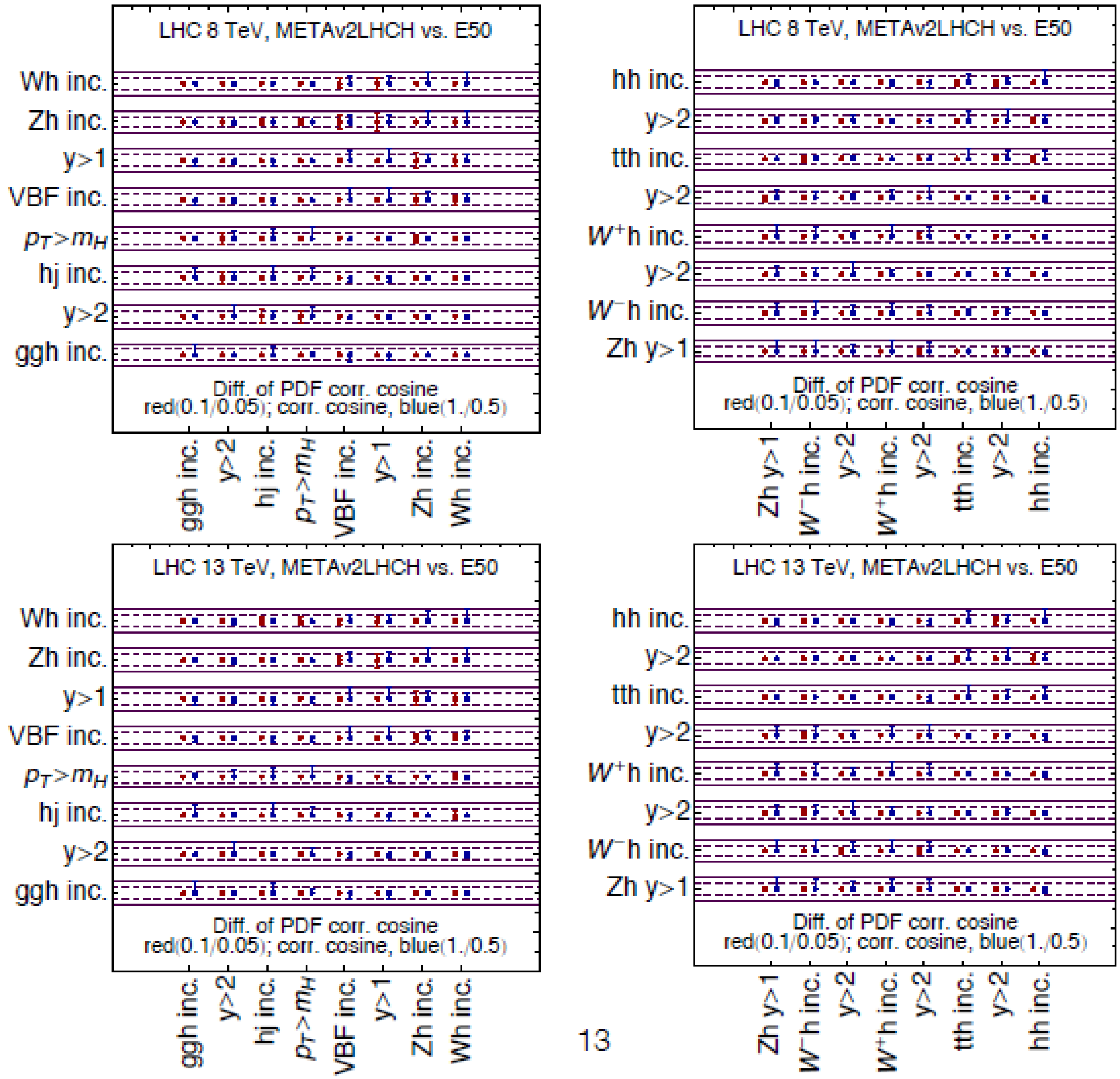


META2LHCH: combined PDFs for LHC Higgs observables 1

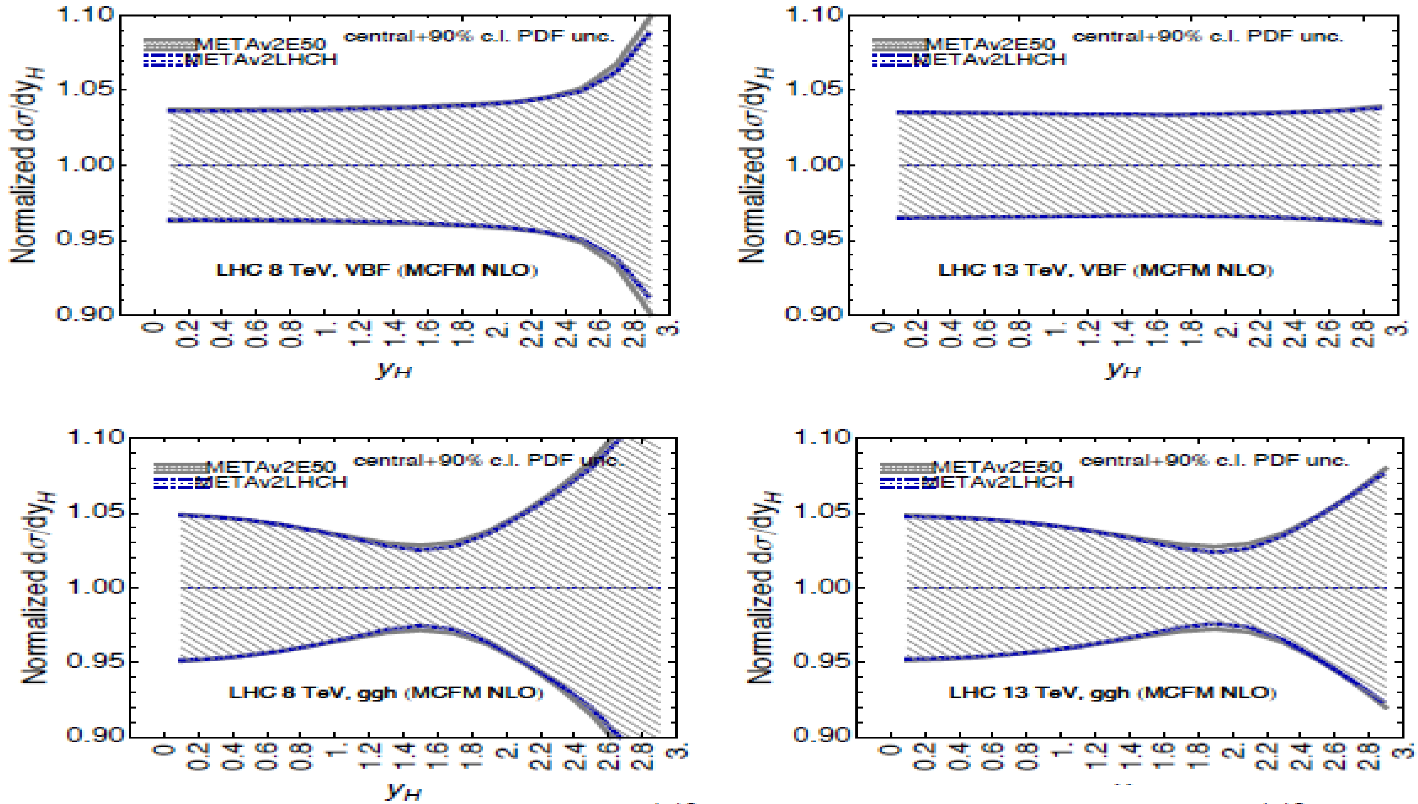


A 20-member ensemble to reproduce PDF uncertainties and correlations for ~60 observables in various Higgs production processes, including large rapidities

META2LHCH for Higgs observables 2: PDF correlations

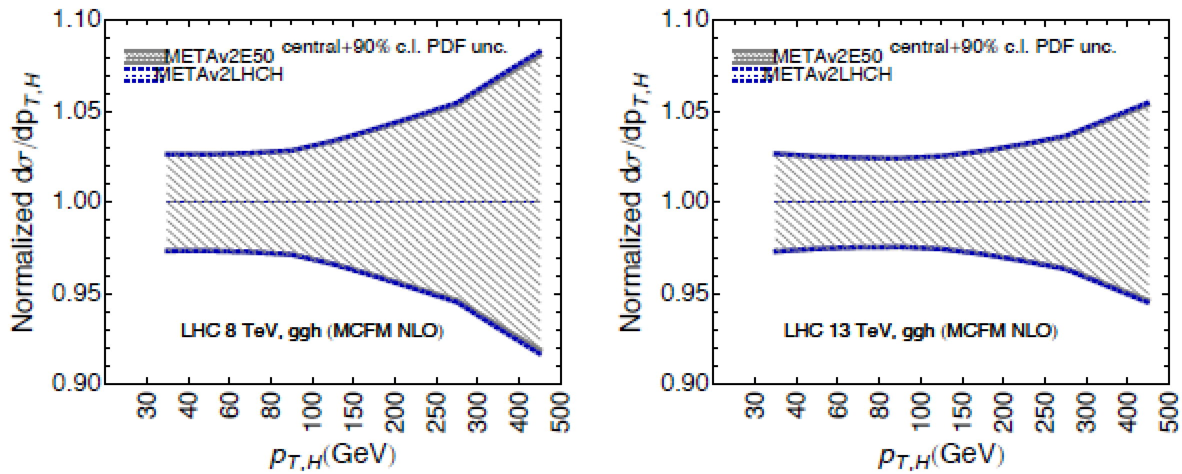


META2LHCH 3: Higgs rapidity and p_T distributions



Higgs rapidity

Higgs transverse momentum



Combination methods, side-by-side
DRAFT: please help to fill in

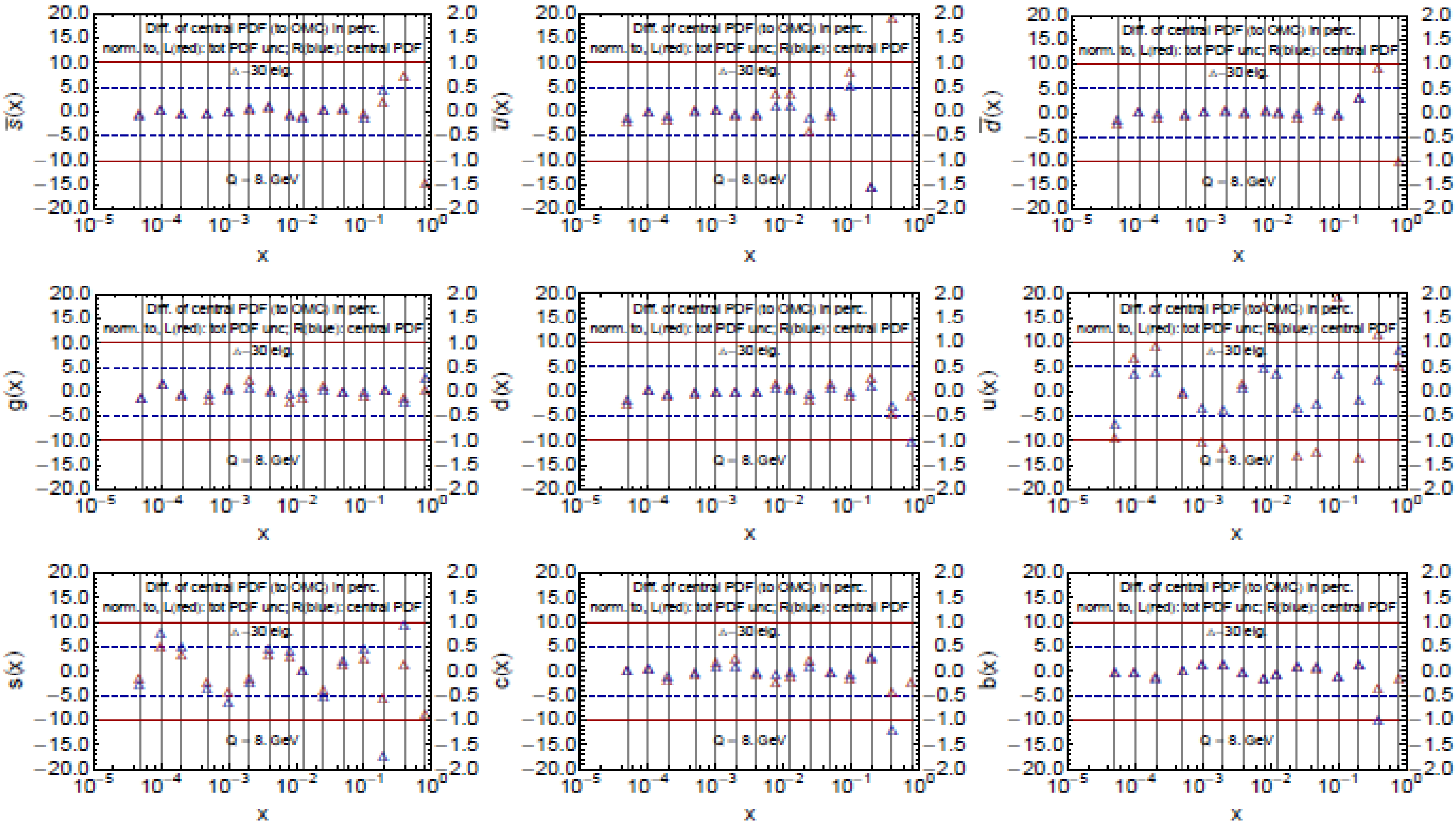
	META2.0	MC2Hessian	Compressed MC
Number of error sets	61 or 21 (LHCH)	181-201	40
Hessian approximation	Yes	Yes	No
Initial scale	8 GeV	2 GeV (?)	From initial PDFs
Parametrization basis	Bernstein	X-space	N/A
Minimization	Analytic	Genetic algorithm (what kind ?)	Genetic algorithm (what kind ?)
Numeric realization	C++, Mathematica	C++, Python	C++
PDF errors	Reproduced	Reproduced	Reproduced
PDF correlations	Better reproduced in data regions	Reproduced (how?)	Better reproduced in extrapolation regions (?)

The META analysis is a mature, validated approach

- Has been evolving for 2.5 years. Exists in several realizations.
- A general and intuitive framework. Implemented in a public Mathematica module MP4LHC
- The PDF parameter space of all input ensembles is visualized explicitly. A functional basis (Bernstein polynomials) with negligible bias is available.
- Data combination procedures familiar from PDG can be applied to each meta-PDF parameter
- Asymmetric Hessian errors can be computed, similarly 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 20 META –LHCH error PDFs.
- Future prospects: extension of fitted regions to smaller/larger x, diagonalization for non-Higgs LHC observables, inclusion of non-global ensembles, …

Back-up slides

META2 validation 1: central PDFs

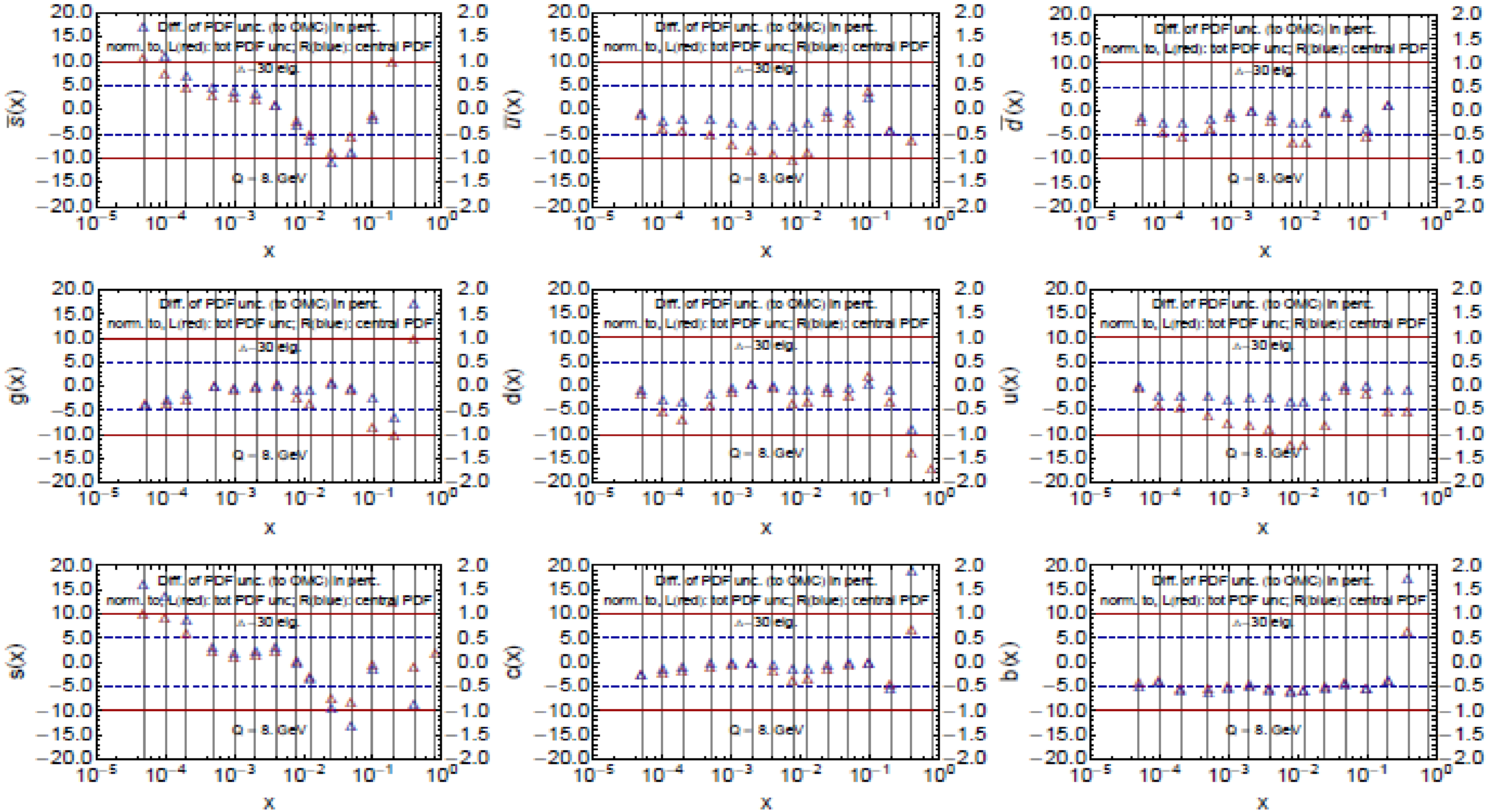


Differences in <u>central PDFs</u> between META60 and MC600.

Left axis, red triangles: as percentages of PDF uncertainty

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

META2 validation 2: PDF errors

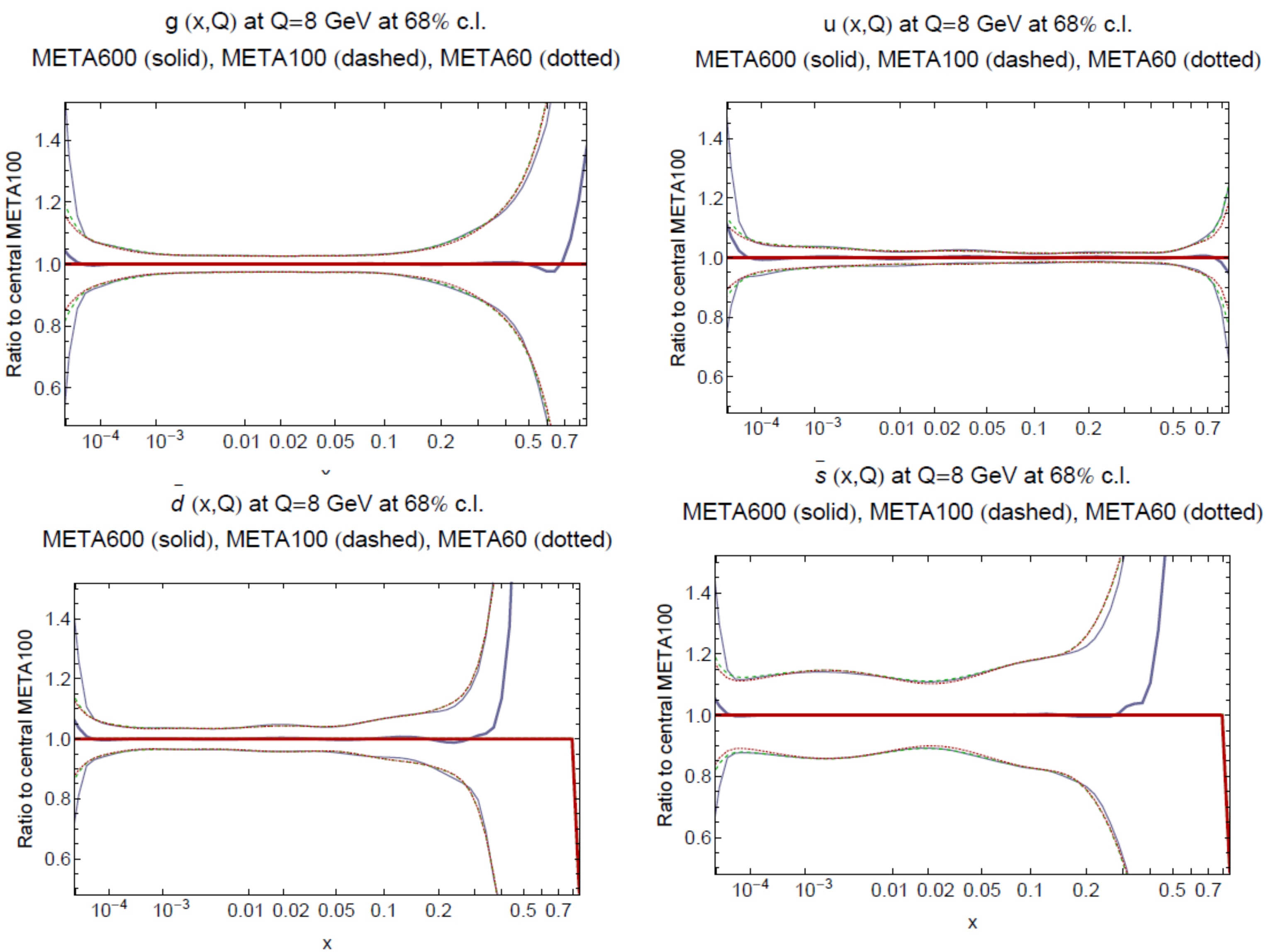


Differences of <u>PDF uncertainties</u> between META60 and MC600.

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

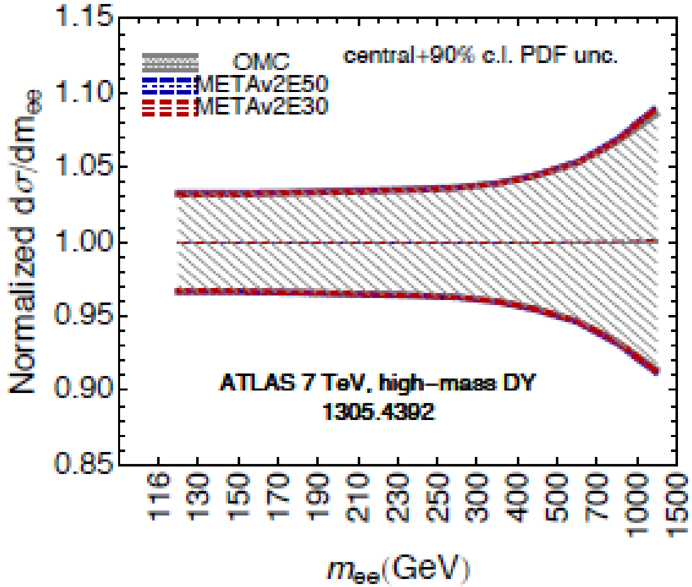
Right axis, blue triangles: as percentages of the central PDF

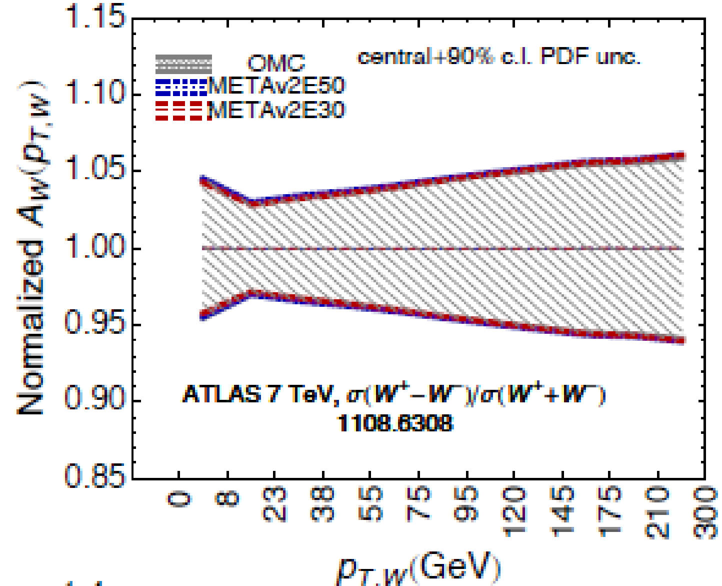
META2 validation 3: 600 → 100 → 60 error sets

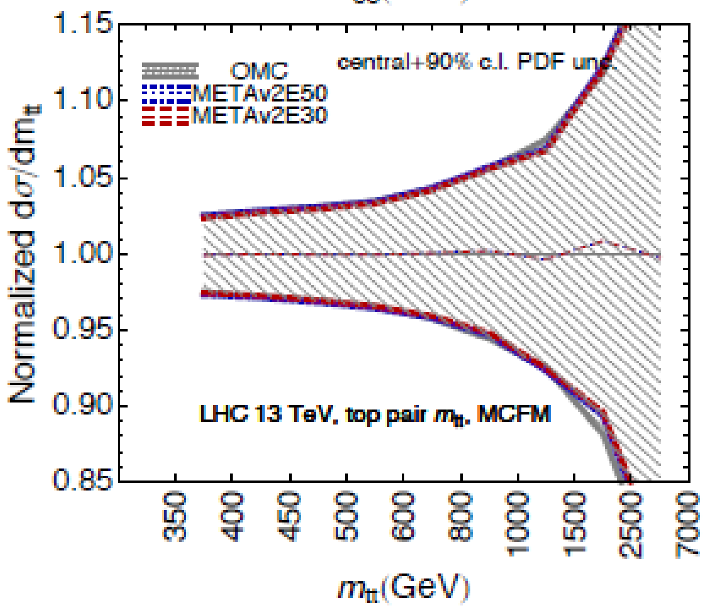


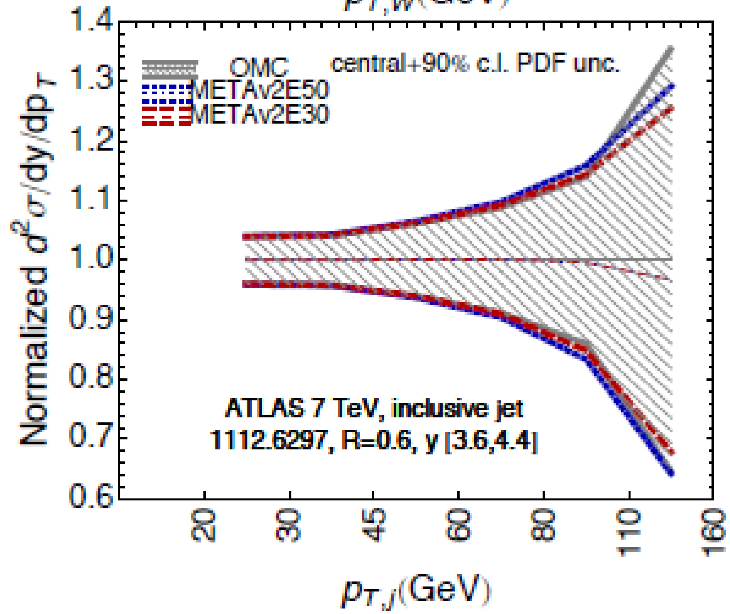
Dependence on the number of eigenvectors is **weak**

META2.0 validation 5: differential cross sections

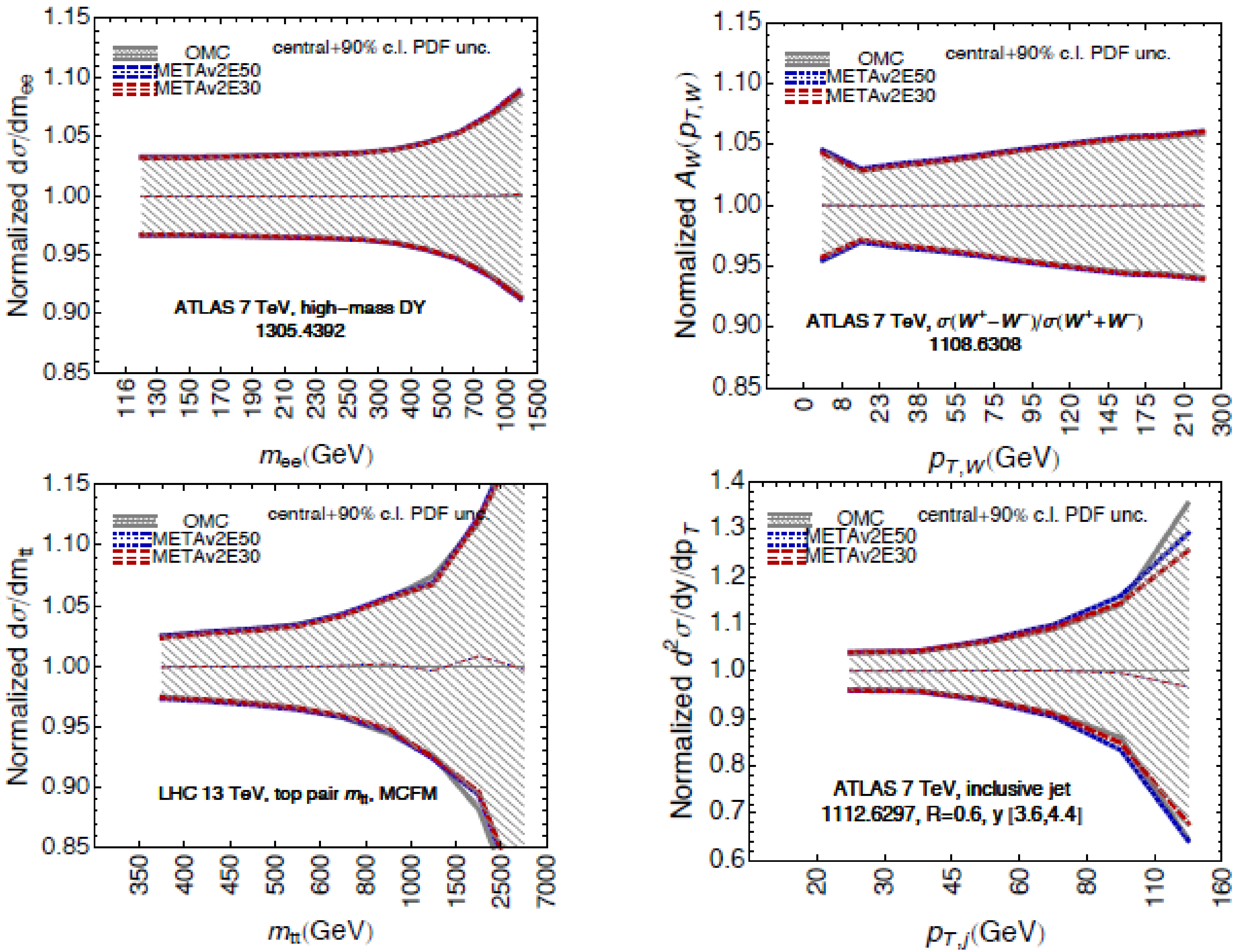




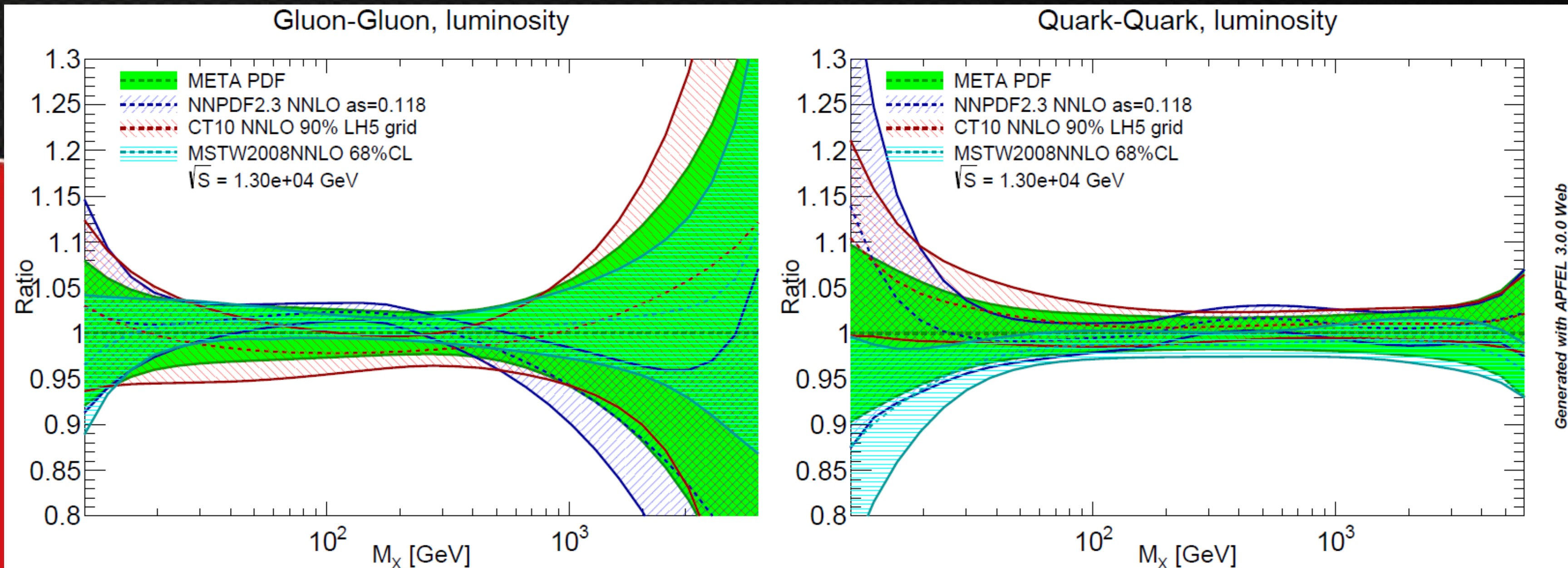


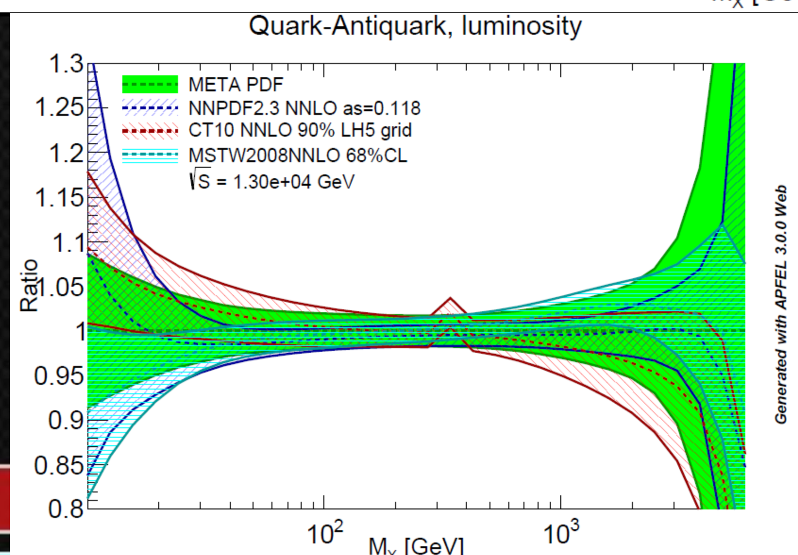


META2.0 validation 6: differential cross sections



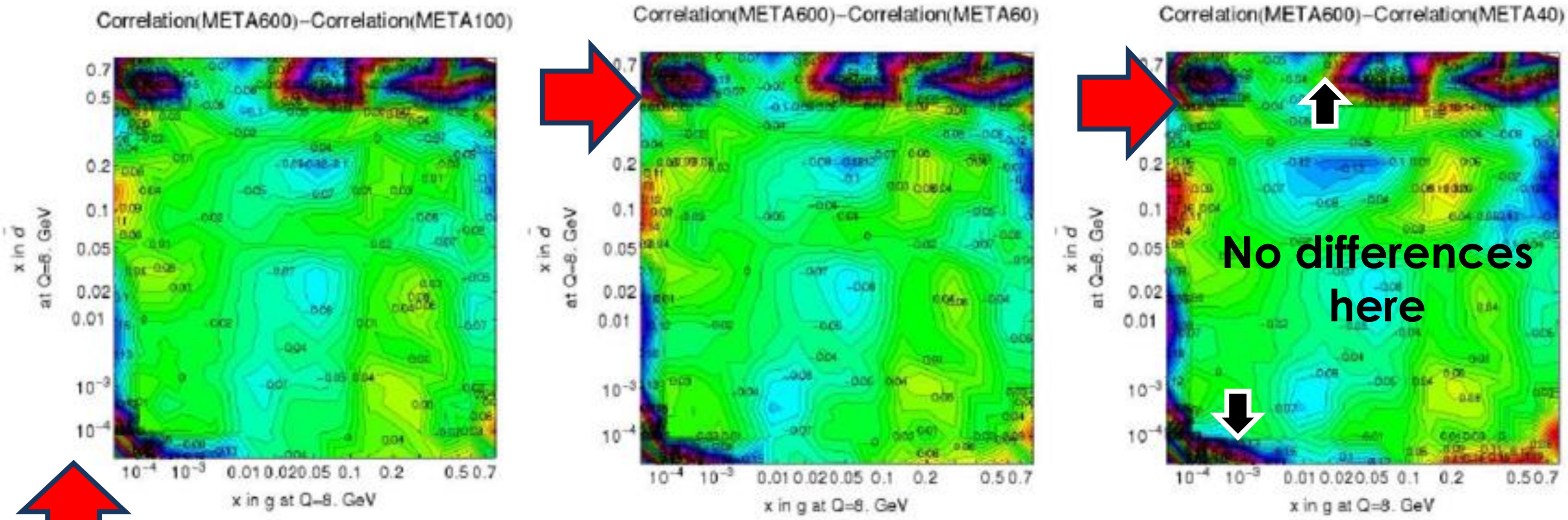
Some parton luminosities

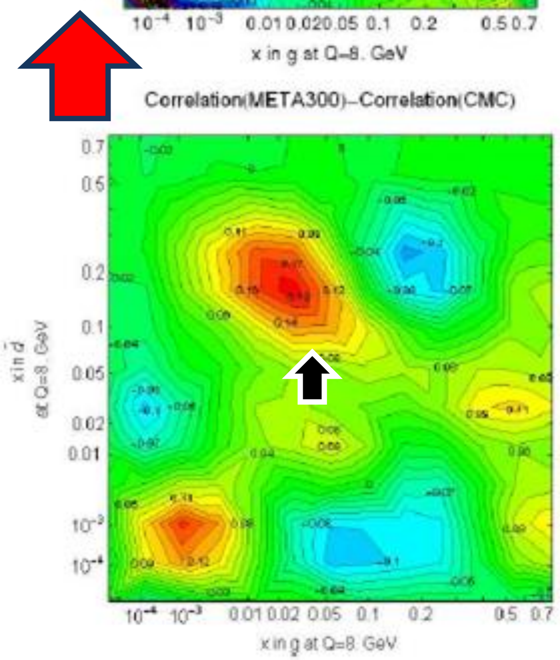


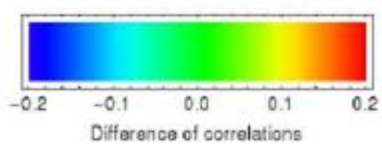


Plots are made with APFEL WEB (apfel.mi.infn.it; Carrazza et al., [1410.5456](1410.5456))

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

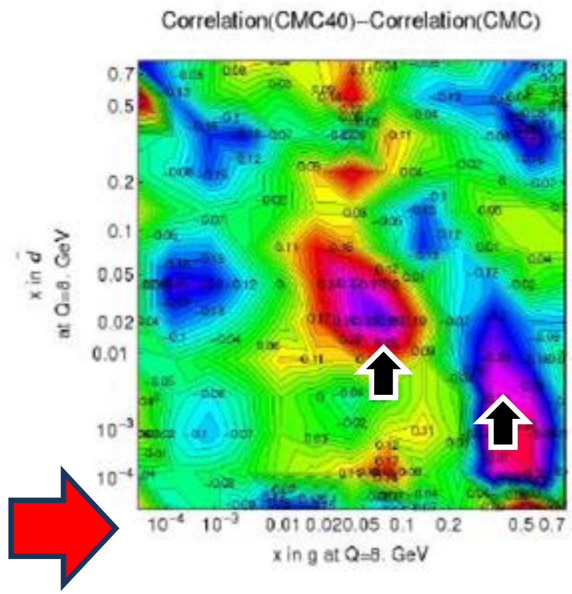






Upper row: META600->60->40

Lower row: CMC300->CMC40



PRELIMINARY