Comments on the CMS response to the PDF4LHC15 recommendations

Juan Rojo

STFC Rutherford Fellow,

University of Oxford

CMS PDF Fit Forum

11/02/2016

Disclaimer

Personal comments on the points raised by the CMS reply to the PDF4LHC15 recommendations

Not an expression of the PDF4LHC15 Steering Committee

Comparison with data and theory

1. Comparisons between data and theory for SM measurements

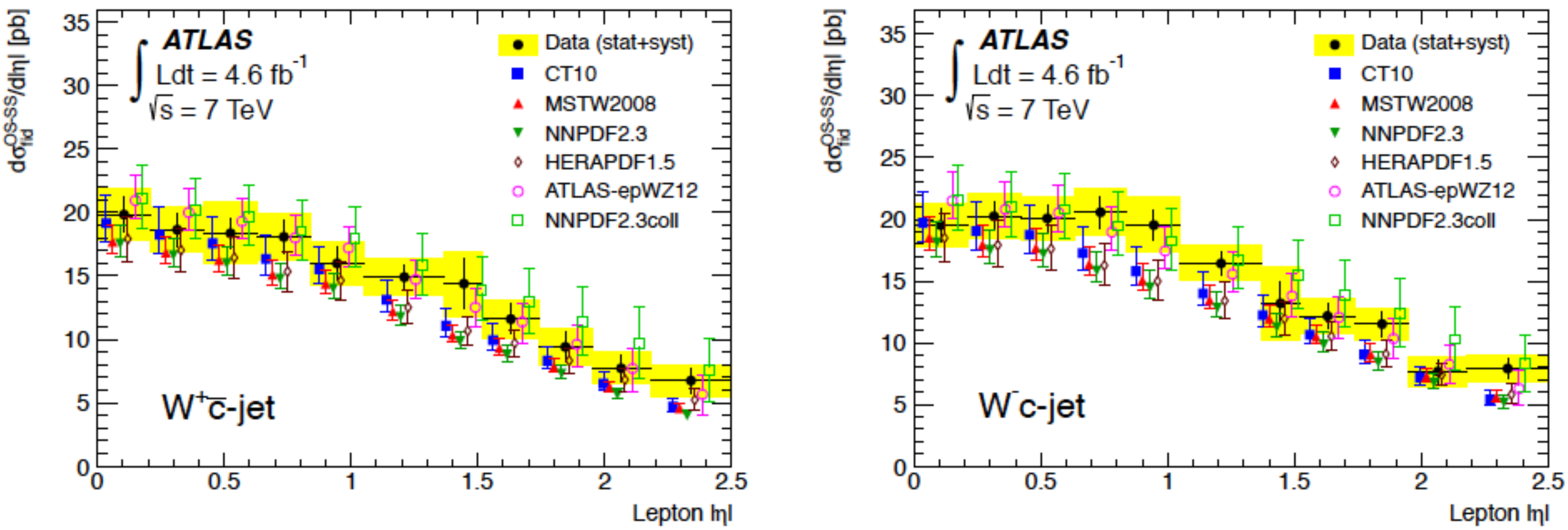
- **Use individual PDF sets, and, in particular, as many of the modern PDF sets [5-11] as possible.**

Observations by us:

- **Clearly, one would do this anyway, no real need for a recommendation ...**
  - **And so: Same recommendation as for Run I.**
  - **But clearly spelled out now to avoid confusion!**
- **With more and more LHC data entering the fits, we will also need PDF sets excluding certain processes to avoid circularity, e.g. no ttbar, no jets, no W (such PDFs are available in some cases already)**
- **For series with ranges in different $\alpha_s(M_Z)$ values it would be interesting to know the change in X^2/n_{dof}**
- **If PDF uncertainties are dominant it might be interesting to check what `PDF4LHC15_100` gives.**
- **?**

Comparison with data and theory

- **PDF sets based on different datasets** already available, and if needed more can be generated upon request
- They can be used **together with the global sets and the combined sets** for data/theory comparisons, depending on the specific measurement



Comparison with data and theory

- The strong coupling constant $\alpha_S(M_Z)$ is an **external parameter** in the global PDF fit, such as the heavy quark masses (or the QED coupling)
- The question of **which value of $\alpha_S(M_Z)$ is preferred in different PDF fits**, and the associated uncertainty, is certainly interesting, but it is separated of ``which is the **most precise value** that should be used for LHC applications'', ***i.e.***, the PDG average
- Consistent with the recommendations of the **HXSWG for SM input parameters**

LHC Higgs Cross Section Working Group*

Internal Note

Standard Model input parameters for Higgs physics

A. Denner, S. Dittmaier, M. Grazzini, R. Harlander,
R. Thorne, M. Spira, M. Steinhauser

Abstract

This note summarises the Standard Model input parameters for Higgs cross section calculations. The same parameters can be used for other SM and BSM processes at the LHC.

Comparison with data and theory

- The strong coupling constant **$\alpha_S(M_Z)$** is an **external parameter** in the global PDF fit, such as the heavy quark masses (or the QED coupling)
- The question of **which value of $\alpha_S(M_Z)$ is preferred in different PDF fits**, and the associated uncertainty, is certainly interesting, but it is separated of ``which is the **most precise value** that should be used for LHC applications'', ***i.e.***, the PDG average
- Consistent with the recommendations of the **HXSWG for SM input parameters**

Concerning the default value for $\alpha_s(m_Z)$ and the estimation of the uncertainties resulting from $\alpha_s(m_Z)$ and the PDFs, one should follow the 2015 PDF4LHC recommendation. This amounts to choose the central value of $\alpha_s(m_Z)$ and the ensuing uncertainty as

$$\alpha_s(m_Z) = 0.118 \pm 0.0015 \,. \tag{10}$$

$$m_c(3 \text{ GeV}) = 0.986 \pm 0.026 \text{ GeV}$$

$$m_b(m_b) = 4.18 \pm 0.03 \text{ GeV}$$

The gauge boson masses and widths from the PDG [1] are

$$m_W = 80.385 \pm 0.015 \text{ GeV}\,, \qquad \Gamma_W = 2.085 \pm 0.042 \text{ GeV}\,,$$
$$m_Z = 91.1876 \pm 0.0021 \text{ GeV}\,, \qquad \Gamma_Z = 2.4952 \pm 0.0023 \text{ GeV}\,.$$

The Fermi constant is

$$G_F = 1.1663787(6) \cdot 10^{-5} \text{ GeV}^{-2}\,.$$

PDFs in BSM searches

2. Searches for Beyond the Standard Model phenomena

- **Use the `PDF4LHC15_mc` sets.**

Observations by us:

- **In particular for searches close to the hadronic threshold where $x \to 1$, it might be interesting to look into the `CTEQ-JLAB` PDFs**
- **To avoid bias from pp or ppbar data in the PDFs, one should also check against PDFs derived without these data.**

PDFs in BSM searches

- The PDF4LHC15 recommendation **(2)** provides a **sensible estimate of PDF errors** which is suitable for general BSM searches

- In case of a **deviation with respect of the SM predictions**, it is essential to **ensure that it is not a PDF effect**, and thus to verify the robustness of such effect with **many different PDF sets**

- However, doing this **in every single search** is maybe a bit too excessive

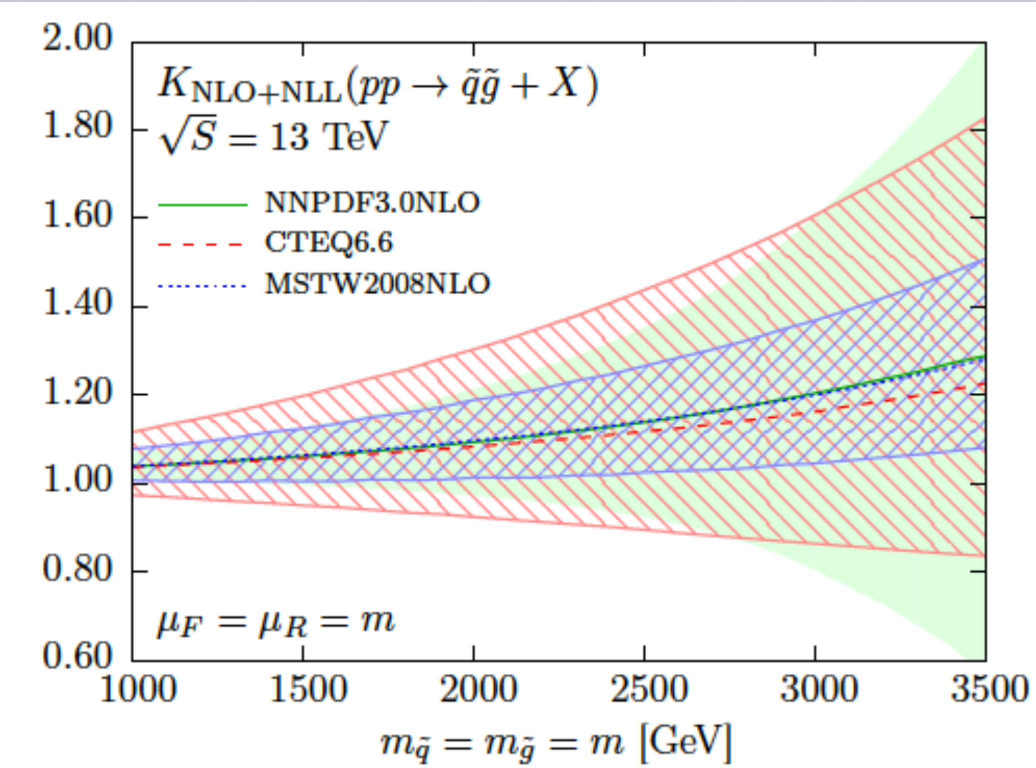
- As before, **non-LHC PDF sets available** and can be used to compare with the results obtained with PDF4LHC15 combined sets or with global sets

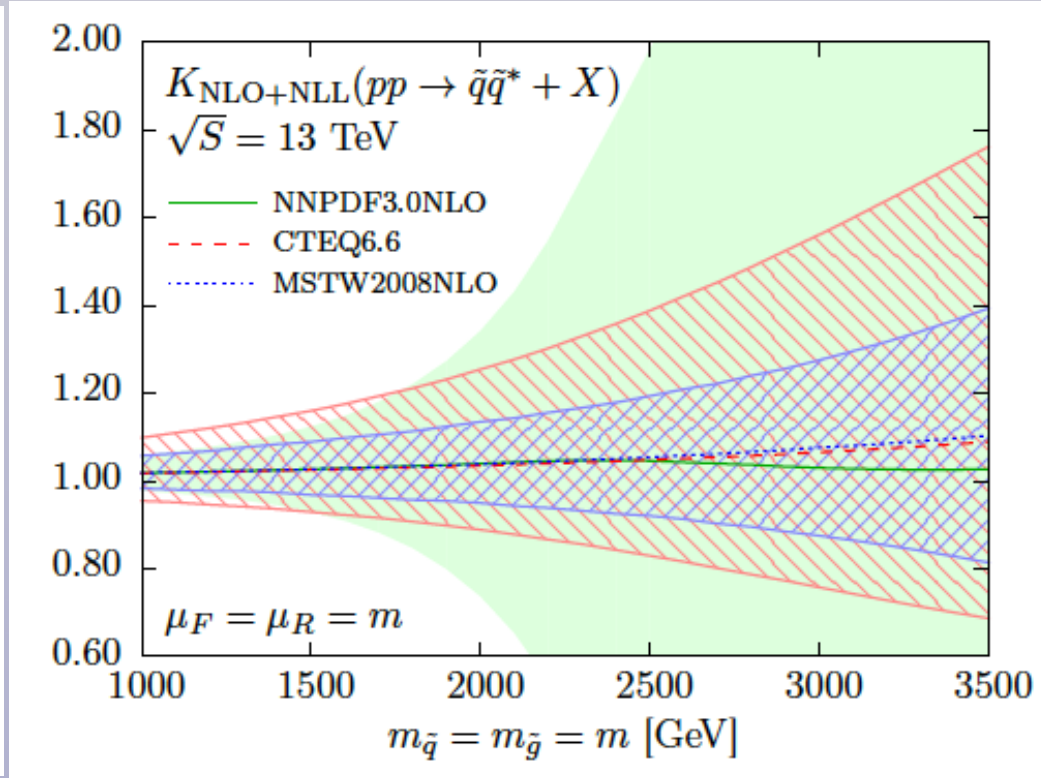
- **CJ PDFs** include **more DIS data at large-x**, but restricted to **NLO** in the **ZM-VFN** scheme. Also, treatment of dynamical higher twists **depends strongly on fit methodology**. Useful for specific cross-checks, but many caveats needed

PDFs in BSM searches

- A **practical recommendation** is needed to answer the question: ``*given my search for high-mass SUSY particles, how I can interpret my results in terms of a mass exclusion region, given TH uncertainties in the SUSY cross-section?"*

- The PDF4LHC15 provides a **sensible, general enough procedure** to answer this question. A purely *agnostic* point of view (define PDF error by envelope of all existing PDF sets) does not seem advisable (nor well defined - where to put cutoff?)





Updated NLO+NLL squark and gluino production at 13 TeV

3. Calculation of PDF uncertainties in situations when computational speed is needed, or a more limited number of error PDFs may be desirable

- **Use the `PDF4LHC15_30` sets.**

Observations by us:

- **There is a nice overview talk from Josh Bendavid from 4. Nov. 2015.**
- **Illustration based on exemplary cross sections instead of parton luminosities is required.**
- **Concerning heavy quark measurements, various mass schemes (RT, FONNL, ACOT) are mixed combining different theory assumptions!?**

Heavy quark schemes

- The **three GM-VFN schemes** used in the CT, MMHT and NNPDF fits are **the same** up to subleading corrections
- Monte Carlo combination method assigns equal weight to each of the different possible treatment of subleading terms: **fully consistent approach**
- Ideally **common values of heavy quark (running) masses** should be used for all sets in the combination
- Combined sets can then be provided for **different mc and mb values**, just as know they are provided for different α_S**(M**$_Z$**)** values
- **Superiority of the GM-VFNS** as compared to FFNS already demonstrated in a number of publications by different groups - objections should be based on identifying specific flaws in existing publications

PDF errors in acceptances

Higgs Cross-Section WG Yellow Report 4

	$\mathcal{A}(gg \to h \to \gamma\gamma)$	$\mathcal{A}(pp \to hW \to hl\nu_l)$
PDF4LHC15_prior	0.728 +- 0.006 (0.9%)	0.7536 +- 0.0014 (0.18%)
PDF4LHC15_mc	0.727 +- 0.006 (0.8%)	0.7538 +- 0.0015 (0.20%)
PDF4LHC15_100	0.728 +- 0.006 (0.9%)	0.7536 +- 0.0013 (0.19%)
PDF4LHC15_30	0.728 +- 0.006 (0.8%)	0.7536 +- 0.0012 (0.16%)
MMHT14	0.728 +- 0.004 (0.6%)	0.7532 +- 0.0012 (0.15%)
CT14	0.725 +- 0.007 (1.0%)	0.7543 +- 0.0014 (0.19%)
NNPDF3.0	0.730 +- 0.005 (0.7%)	0.7534 +- 0.0011 (0.15%)

Table 2: The acceptance corrections $\mathcal{A}$ for Higgs production at the LHC 13 TeV in two different channels with realistic selection cuts, and the corresponding PDF uncertainties. We compare the results of the PDF4LHC15 NNLO prior with the three reduced sets. For completeness, we also show the results for the acceptances computed with the individual PDF sets. See text for more details of the specific selection cuts in each case.

- PDF uncertainties in **acceptances** typically smaller than those in **absolute cross-sections distributions** (due to partial cancellations)
- The use of **PDF4LHC15_30** fully justified in these cases

PDF uncertainties in precision xsecs

<u>4. Calculation of PDF uncertainties in precision observables</u>

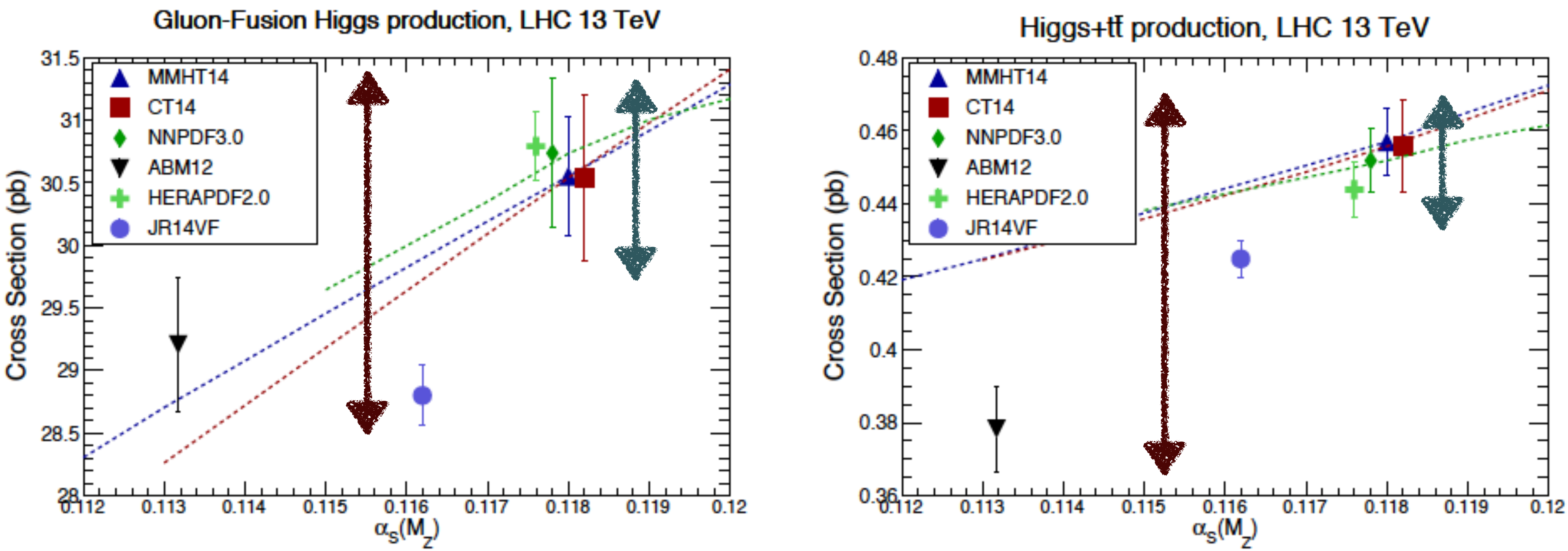
- Use the PDF4LHC15_100 sets.

<u>Observations by us:</u>

- As far as fiducial cross sections are concerned, where extrapolations with additional assumptions or techniques are not required, we do not see why this should be dealt with differently than recommendation 1.

PDF uncertainties in precision xsecs

- What is the **PDF uncertainty to be associated to Higgs production**?
- Crucial to **determine the BSM reach of precision Higgs coupling measurements**
- An *agnostic* point of view concerning PDFs not possible: **need to make choices**!



- What is the **PDF uncertainty to be associated to Higgs production**?
- Crucial to **determine the BSM reach of precision Higgs coupling measurements**
- An *agnostic* point of view concerning PDFs not possible: **need to make choices**!

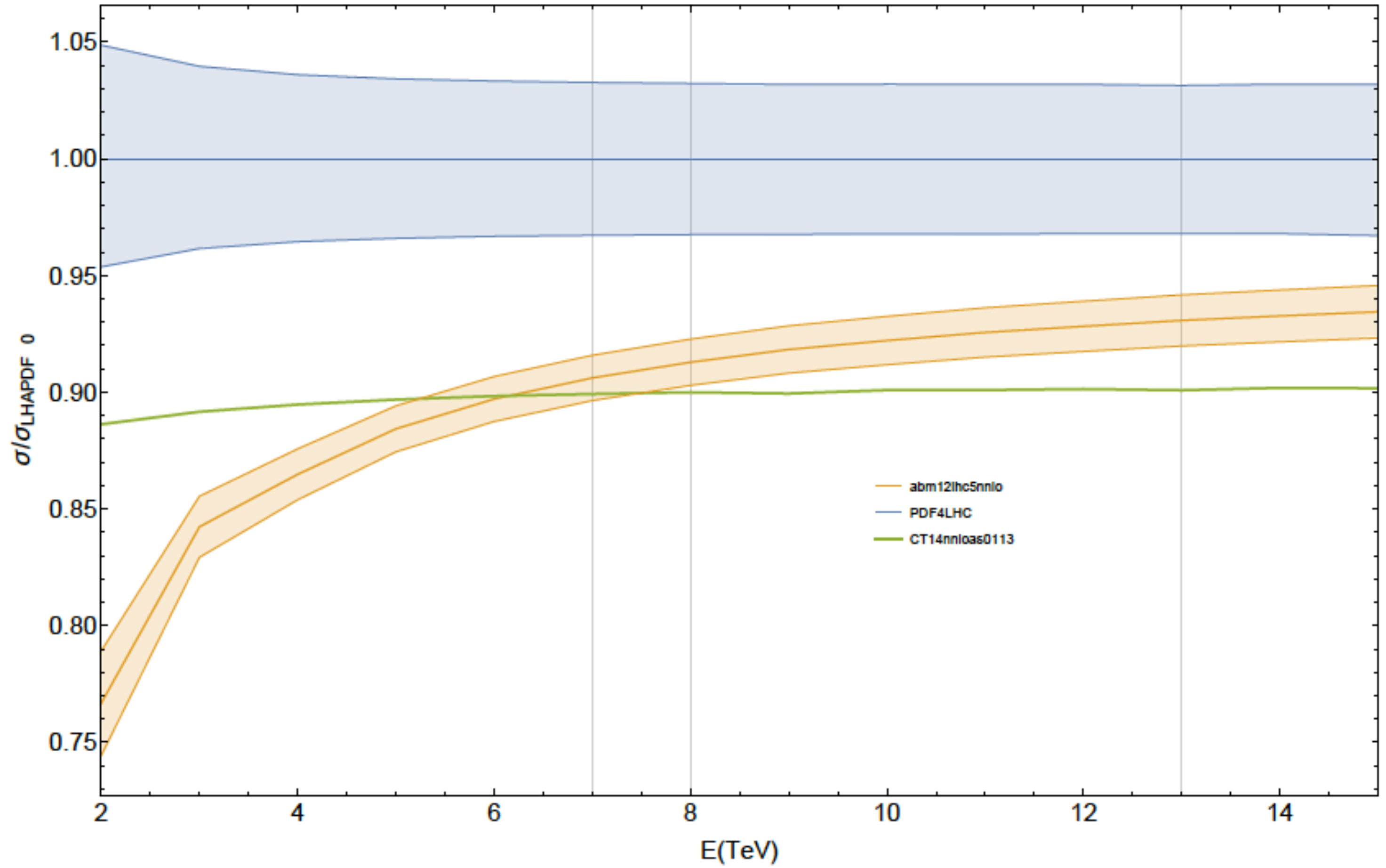
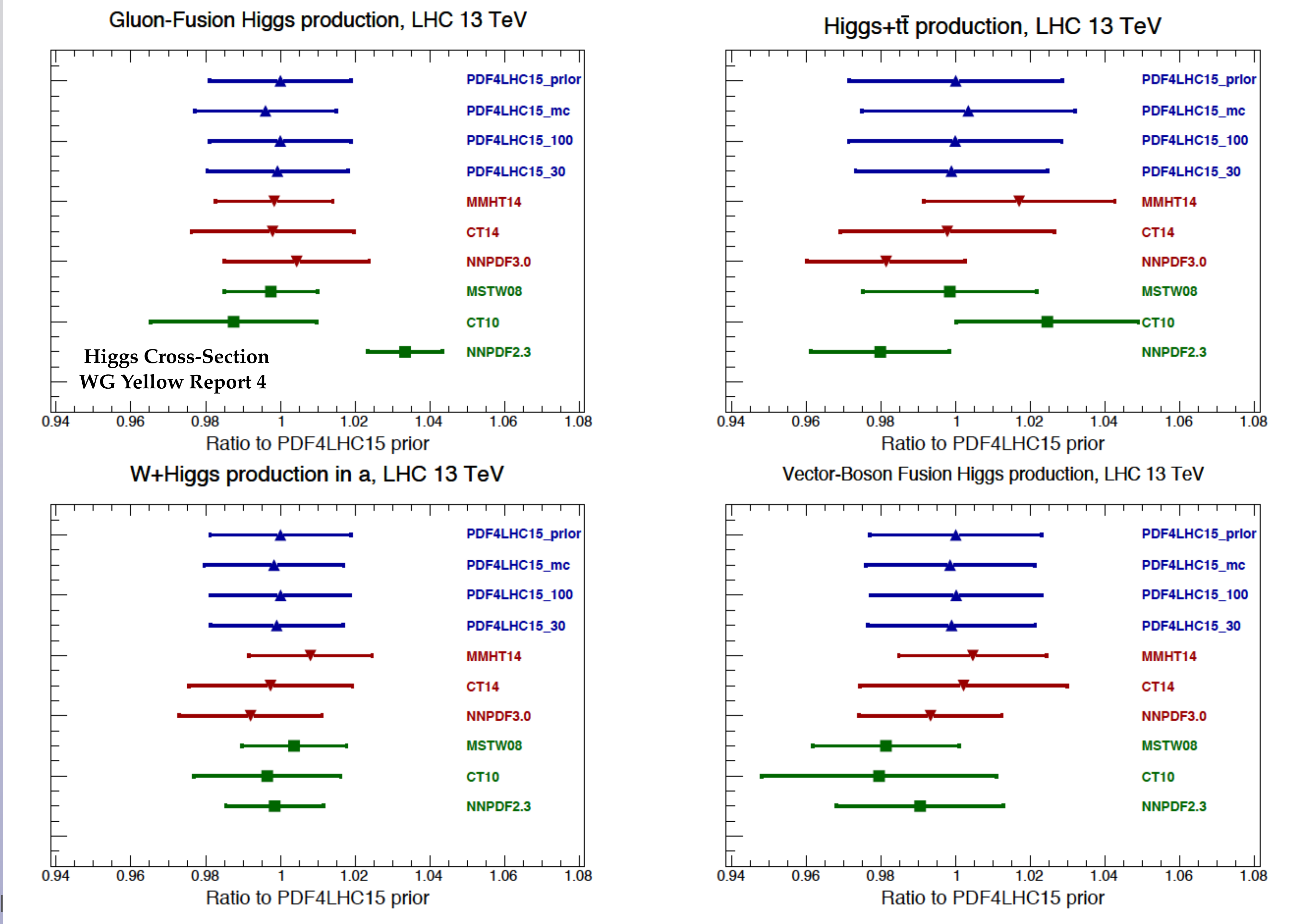


Figure 18: Higgs production cross-section and 68% C.L. PDF+α_s uncertainty from the ABM12 fit and from the CT14 set computed at $\alpha_s = \alpha_s^{ABM}$, normalized by the central value obtained with the PDF4LHC combination.

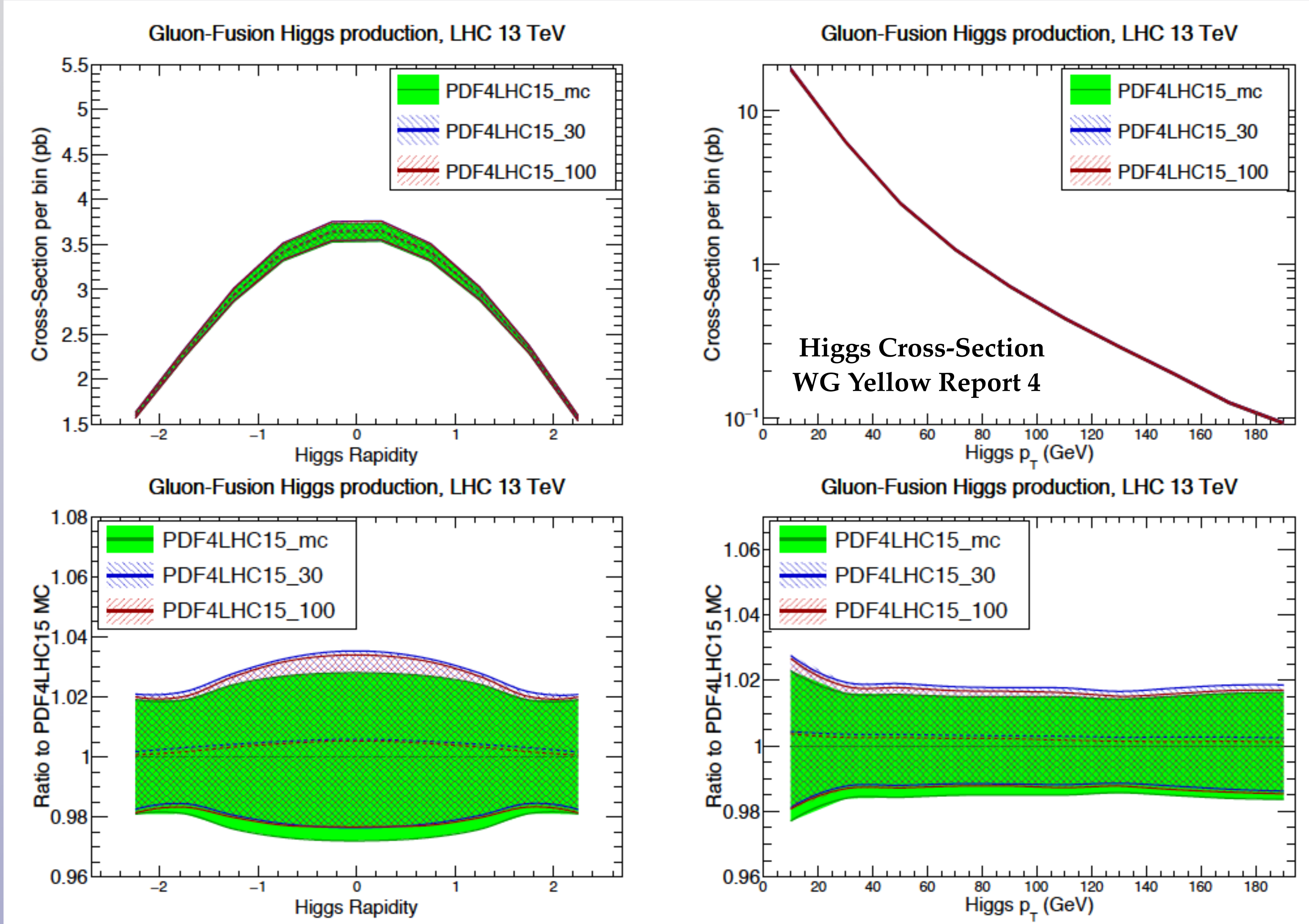
- These are "recommendations" and not something "carved in stone". Each analysis should reflect on their best strategy, of course.
- We heard that some updates are being prepared by the authors.
- LO MC event generators are still in wide-spread use, so it would be nice to have some word on these and LO PDFs as well.

- LO PDFs have associated very **large theory uncertainties**
- The MC combination method is not adequate to LO PDFs
- LO PDFs are used to **tune soft/semi-hard physics in LO MCs**, and the PDF in the shower cannot be modified without having to retune again

Some additional comparisons



Some additional comparisons



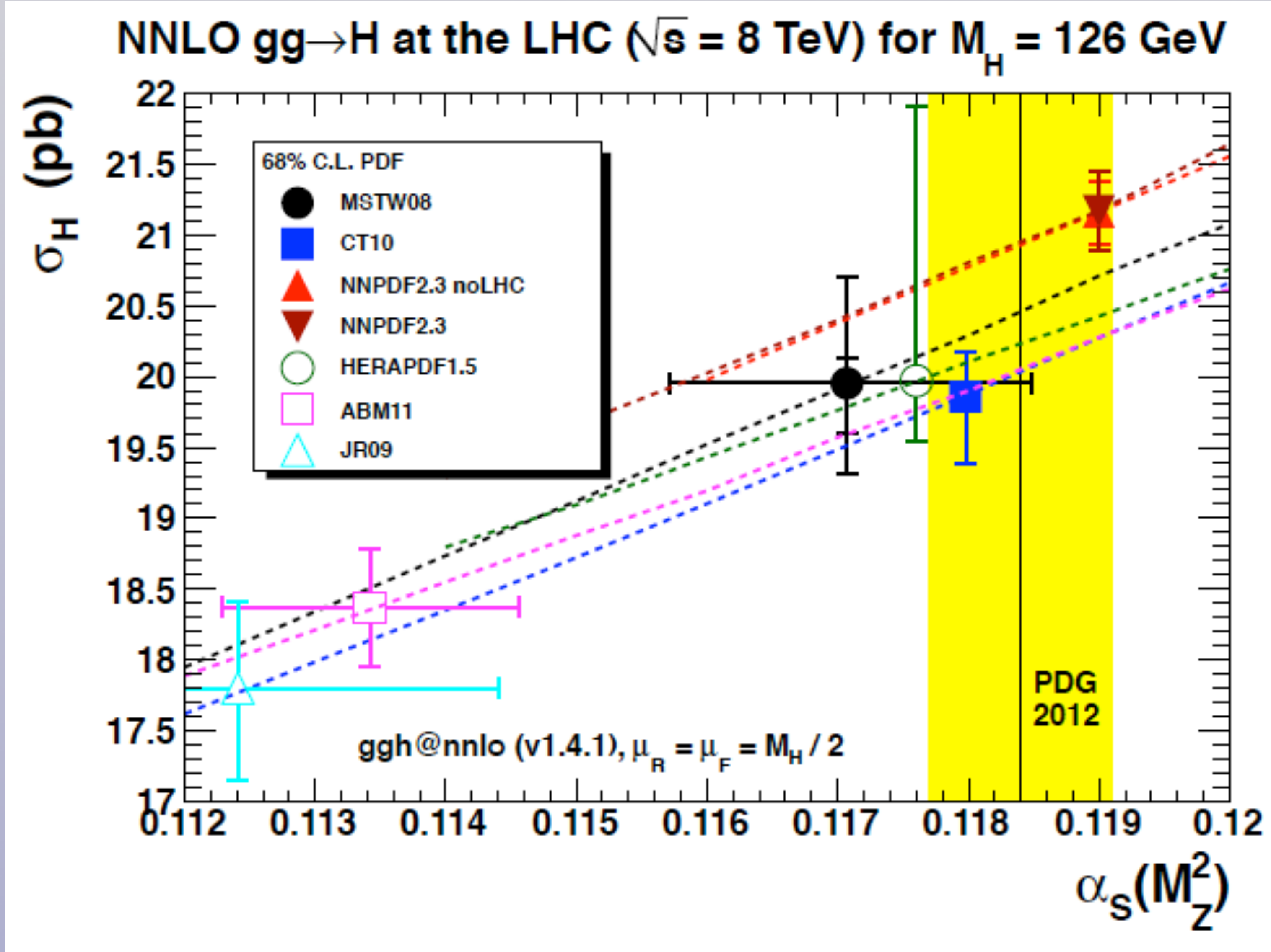
Backup Material

Motivation

What is the **PDF+$\alpha_S(M_Z)$ uncertainty** that should be assigned to the **gluon-fusion Higgs xsec**?

Pressing question since, given precision expected at Run II, **theory uncertainties** can limit our **ability to extract Higgs couplings from LHC data**

6%?



13%?

Motivation

To estimate the **total PDF uncertainty** associated to LHC processes, such as Higgs cross-sections or New Physics searches, a **combination of the results from individual PDF sets** is required

PDF4LHC11 → **Envelope Method**

- Cumbersome to implement: **a posteriori combination** of cross-sections from different PDF sets
- **Statistical interpretation** unclear. Over-conservative: too much weight given to **outliers**

PDF4LHC15 → **Monte Carlo Combination**

- Statistically meaningful combination

MC reduction → **CMC-PDFs**

Hessian reductions → **MC2H**, **Meta-PDFs**

MC2H → **SM-PDFs**

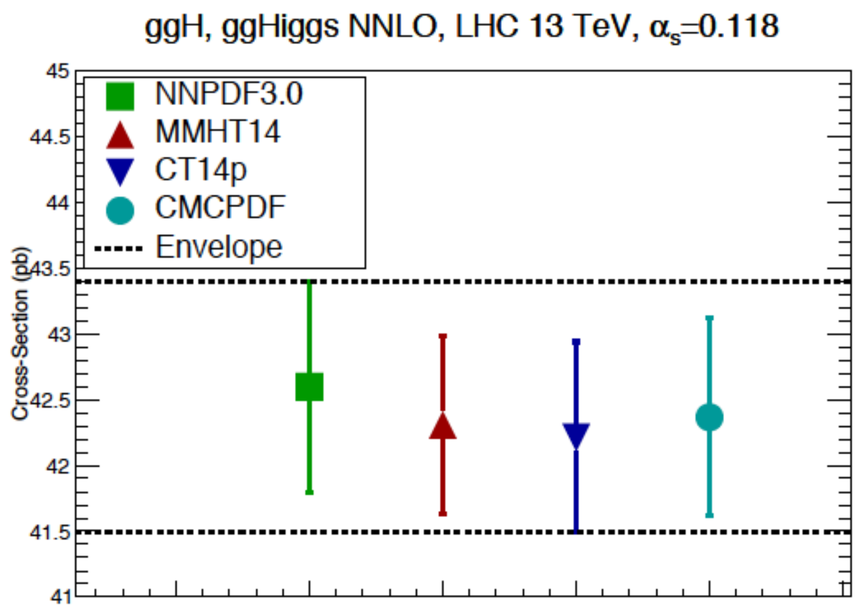
Combination tailored to **specific processes**

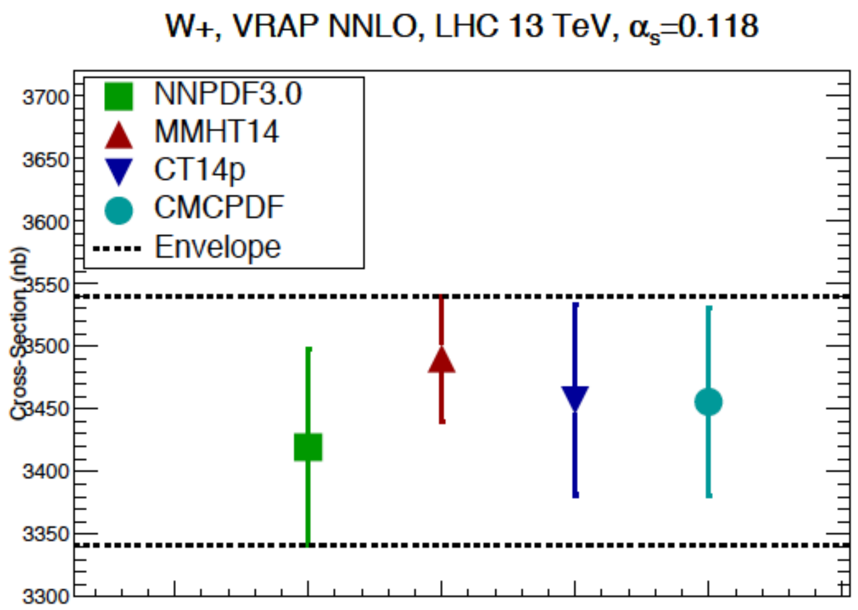
Monte Carlo combination

- First of all, select the **PDF sets that enter the combination**. Must be reasonably consistent among them for a meaningful combination
- Transform the **Hessian PDF sets into their Monte Carlo representation:**

$$F^k = F(q_0) + \frac{1}{2}\sum_{j=1}^{N_{\rm eig}} \left[F(q_j^+) - F(q_j^-)\right] R_j^k \,, \quad k = 1, \ldots, N_{\rm rep} \,,$$

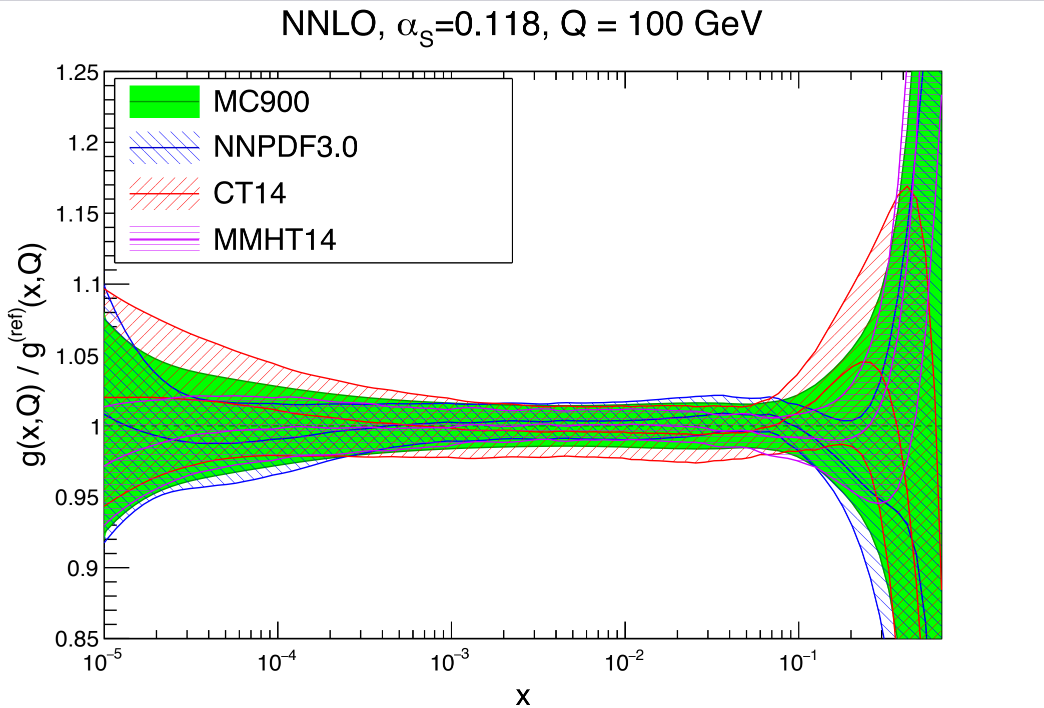
- Assume equal likelihood for all input sets: combine the **same number of replicas**
- The resulting Monte Carlo ensemble has a **robust statistical interpretation.** Similar results, with smaller uncertainties, compared to the **original PDF4LHC envelope** (proper treatment of outliers)

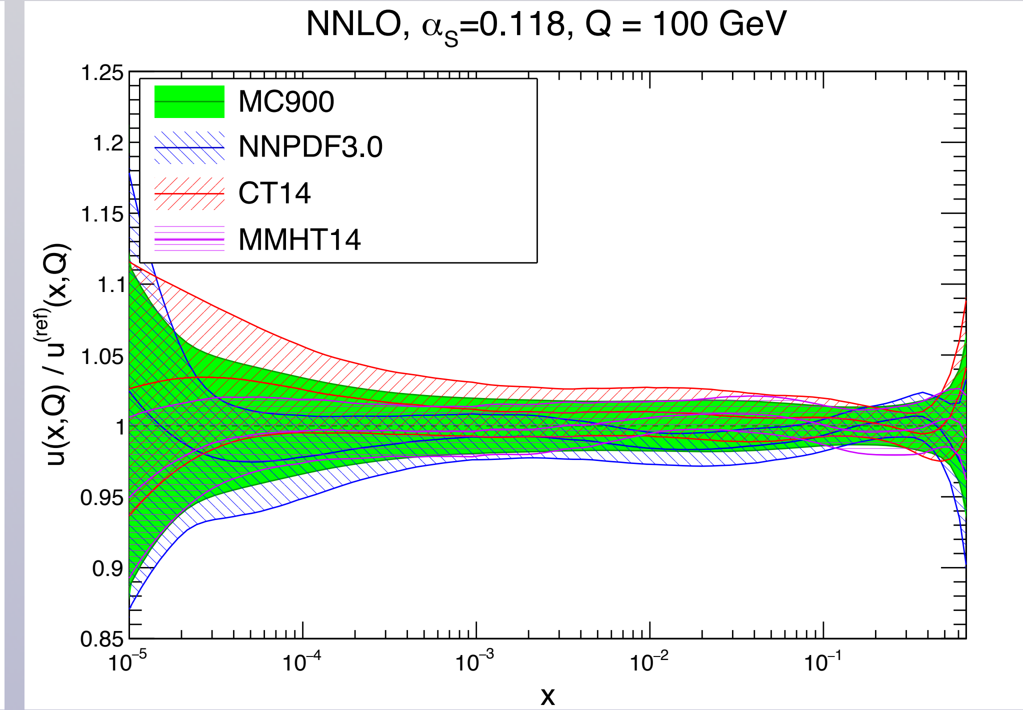




The combined Monte Carlo PDF set

- Combine most recent releases of the three global sets: **NNPDF3.0, CT14, MMHT14**

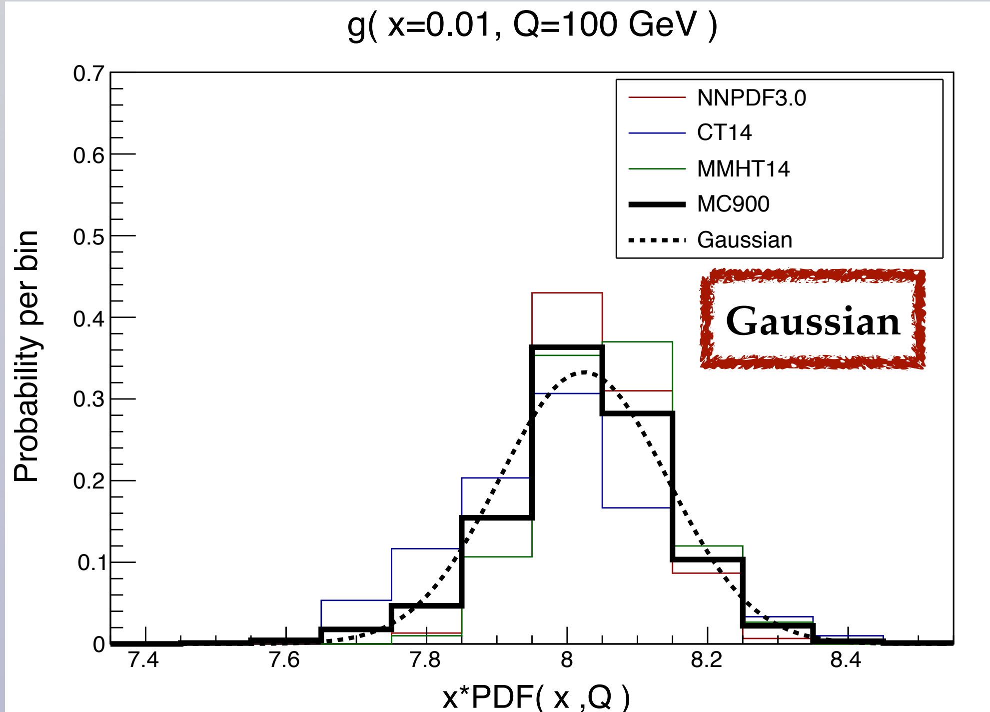


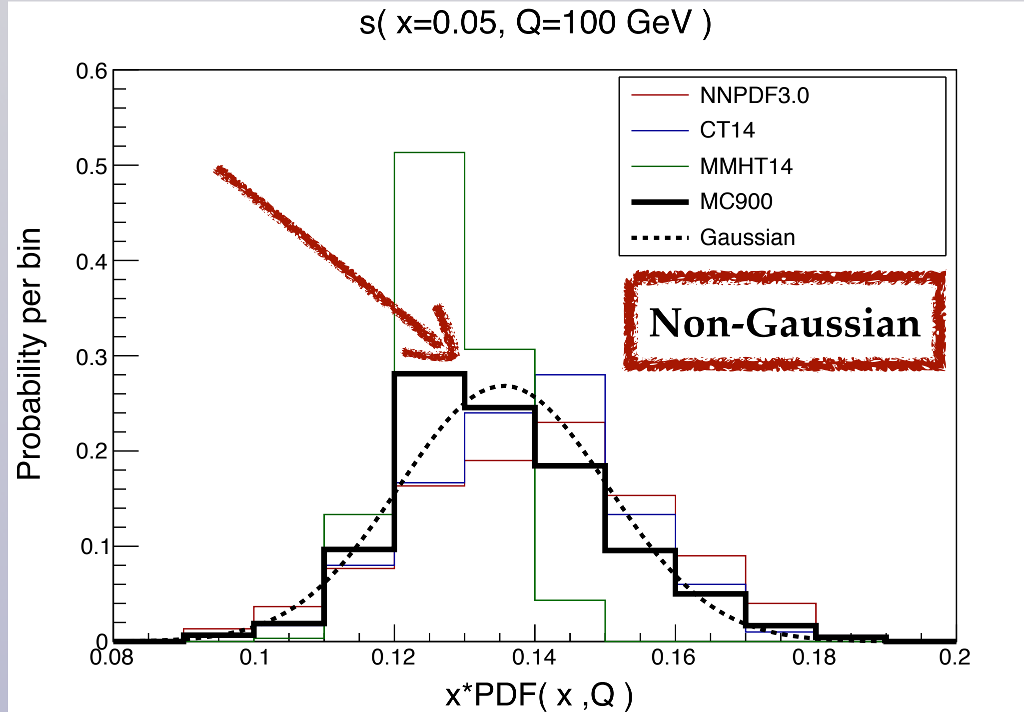


- **900 MC replicas** required to stabilise the combination: these define the **MC900 prior**
- Too many for practical applications: need to find a strategy to **reduce the number of replicas** in which the MC combination is based
- In addition, for many important application, a **Hessian representation** would be required

The combined Monte Carlo PDF set

- Combine most recent releases of the three global sets: **NNPDF3.0, CT14, MMHT14**

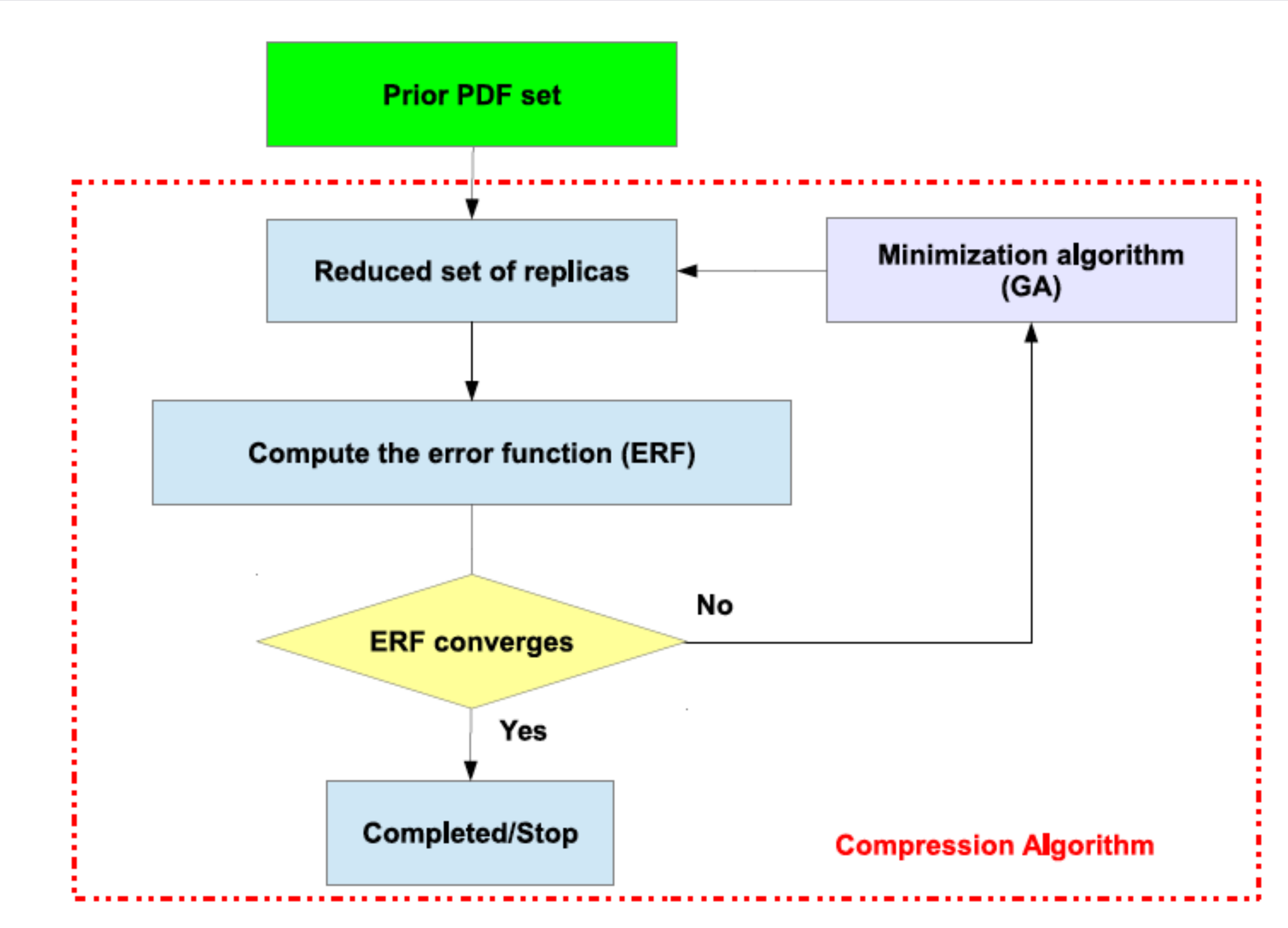




- The **MC combination** is usually **Gaussian** (in the data region) but in many cases **non-Gaussian features** are observed (in extrapolation regions, or for poorly constrained flavors)
- Specially important for **BSM searches**, which rely on PDFs in regions where **PDF errors are large**

Compressed Monte Carlo PDFs

Compress the MC900 prior down to a **smaller number of replicas**, in a way that all the relevant estimators (mean, variances, correlations, higher moments) for the PDFs are fully reproduced



$$\mathrm{ERF} = \sum_k \frac{1}{N_k} \sum_i \left(\frac{C_i^{(k)} - O_i^{(k)}}{O_i^{(k)}} \right)^2 ,$$

$$ERF_{CV} = \frac{1}{N_{\rm CV}} \sum_{i=-n_f}^{n_f} \sum_{j=1}^{N_x} \left(\frac{f_i^{\rm CV}(x_j,Q) - g_i^{\rm CV}(x_j,Q)}{g_i^{\rm CV}(x_j,Q)} \right)^2$$

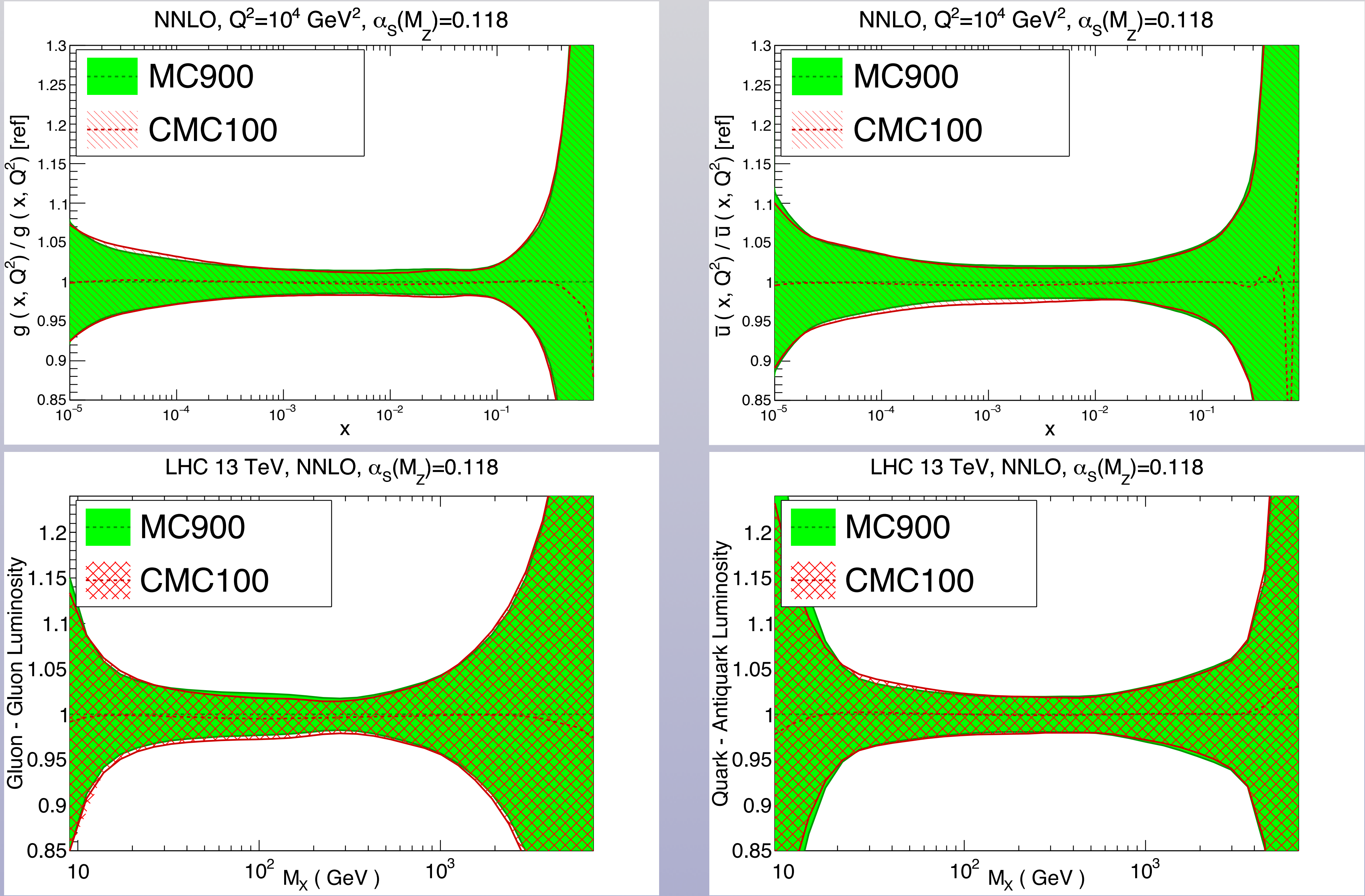
$$f_i^{\rm CV}(x_j,Q) = \frac{1}{N_{\rm rep}} \sum_{r=1}^{N_{\rm rep}} f_i^r(x_j,Q)$$

Similar for **variances, correlations, skewness, kurtosis** and the **Kolmogorov** distance

Minimise **distance between prior and compressed sets** using Genetic Algorithms applied to an **Error Function,** which ensures the reproduction of all statistical properties of the prior

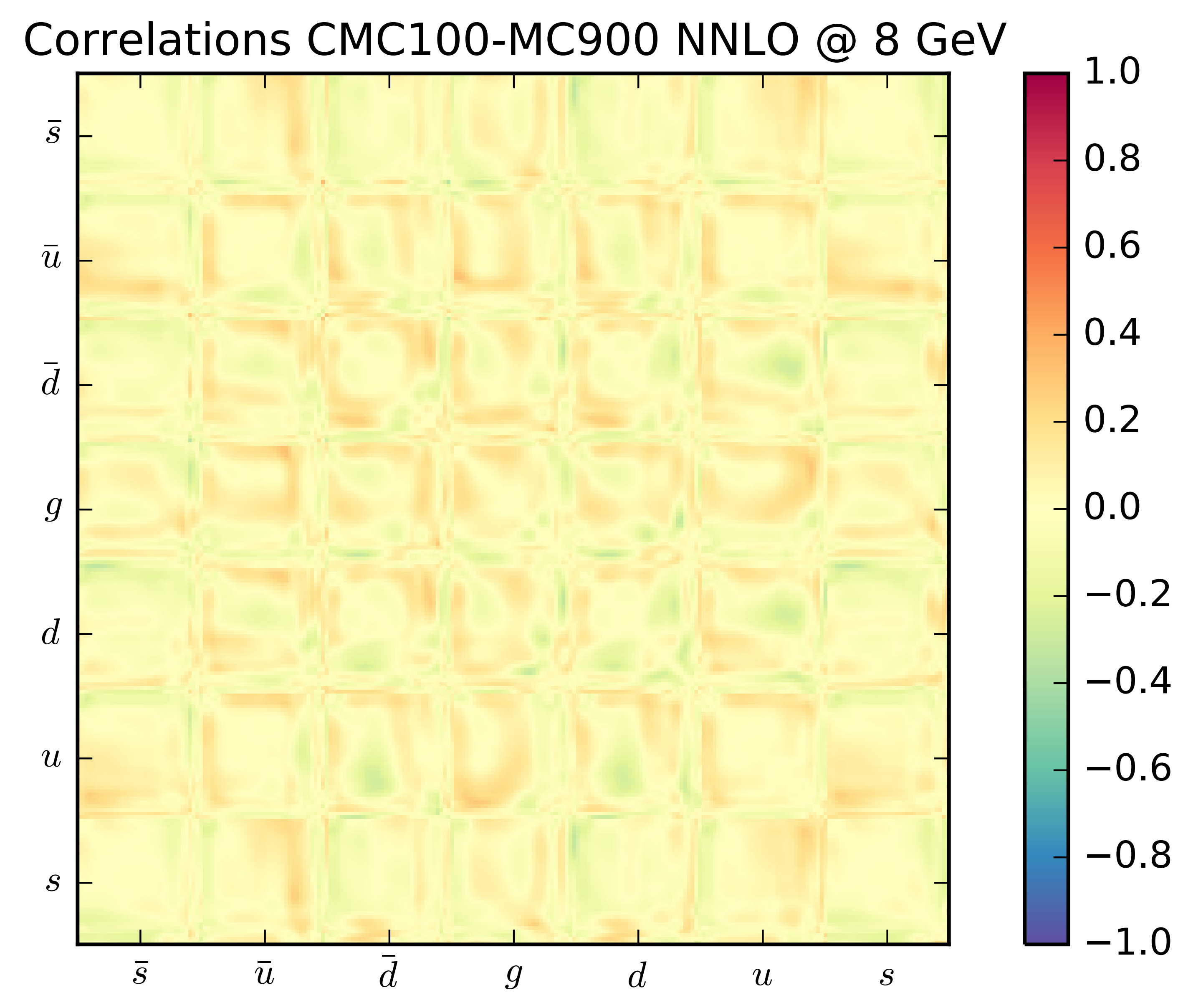
CMC-PDFs

Good agreement between **MC900** and **CMC** from a number of compressed replicas $\mathbf{N_{rep} > 50}$



CMC-PDFs

Reasonable agreement as well for the **correlations between different PDF flavours**

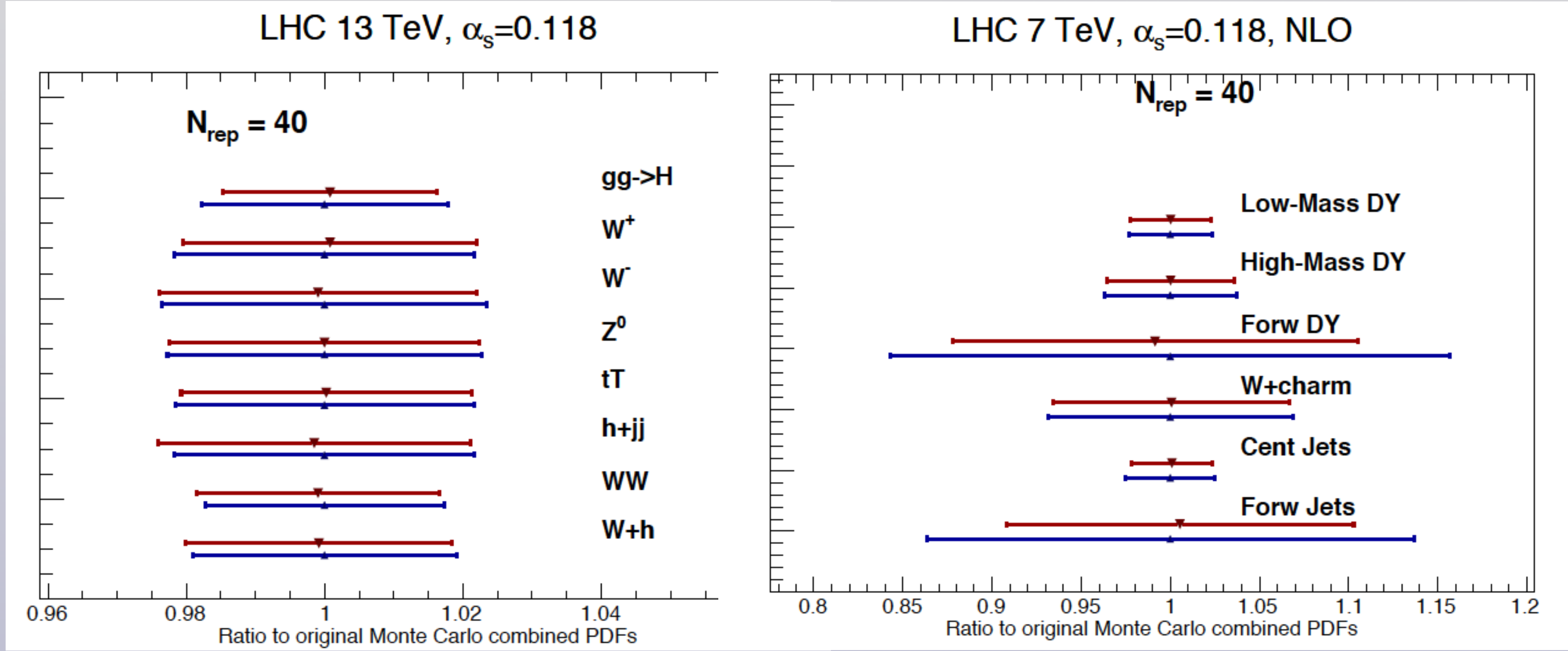


In this contour plot, the **differences in the correlation coefficients of MC900 and CMC100** between different PDF flavours are represented

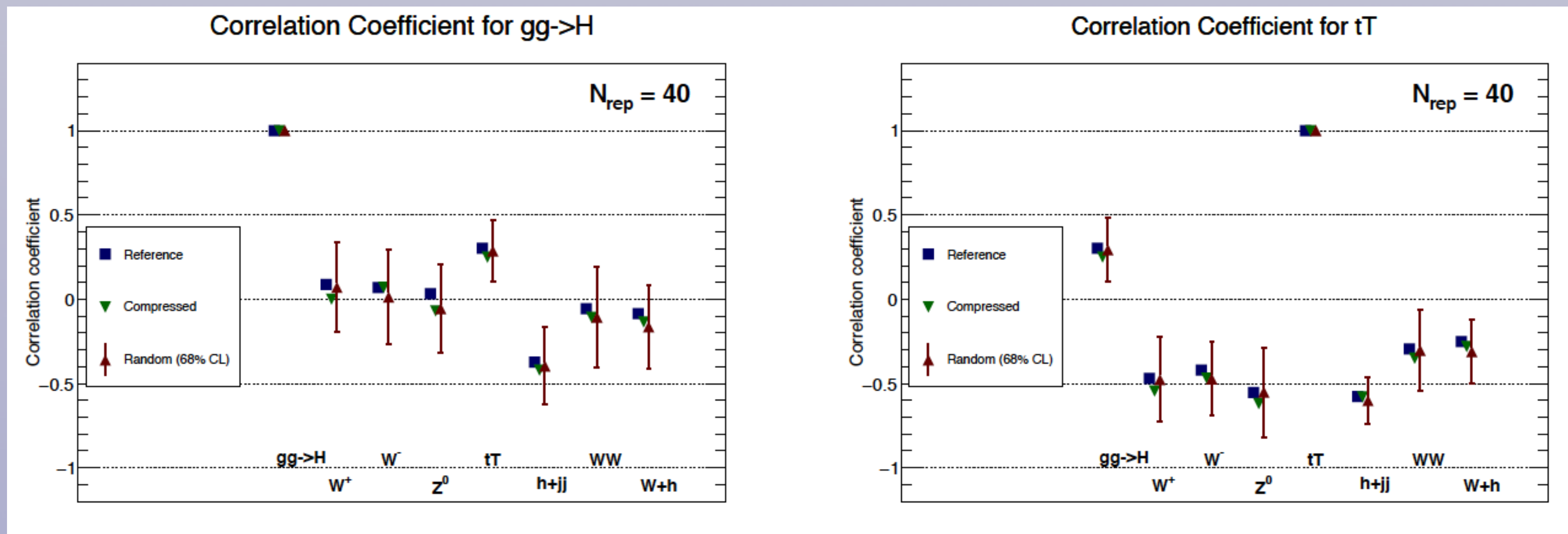
Correlation coefficients are computed at **Q = 8 GeV**, and in the **range of x of [10^{-5},0.9]** logarithmically spaced

LHC Phenomenology

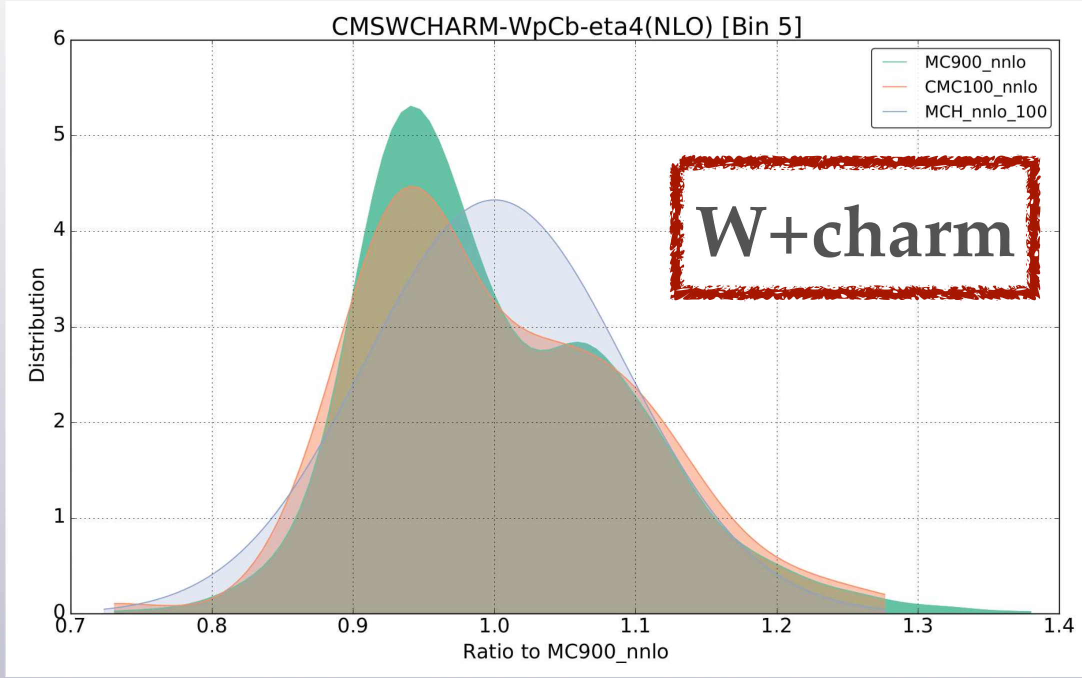
As expected from good agreement at the PDF level, CMC-PDFs also validated for **LHC inclusive cross-sections and diff distributions,** including **correlations**

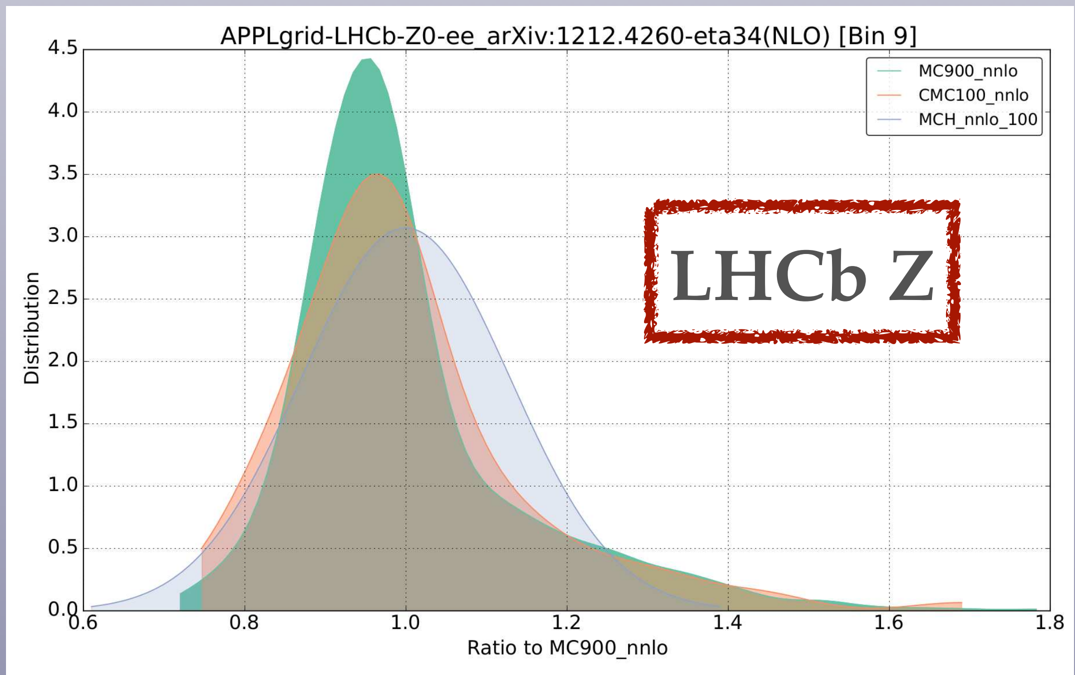


Agreement between MC900 and CMC significantly better than **random selection of same number of replicas**



Non-Gaussian features in LHC cross-sections





CMC100 reproduces properly **non-Gaussian features in MC900**, such as a **double-hump distribution** (in W+charm) and a **skewed distribution** (for forward Z)

Any **Hessian reduction fails** by construction to reproduce such features

Motivation for Hessian reduction strategies

- In regions where **MC900** is approximately Gaussian, a **Hessian representation** is by definition more efficient than **CMC-PDFs**, since it needs to reproduce **only central values and covariances**

- Moreover, a Hessian representation has several advantages, such as the use of **PDF uncertainties as nuisance parameters**, the applicability of **Hessian profiling**, or the **possibility of further reductions** when applied to specific processes

- Two Hessian reduction methods have been developed. Essentially common idea, differ in the choice of **linear expansion basis**

MC900 → **Hessian reductions** → **MC2H**, **Meta-PDFs**

MC2H →

- **MC replicas** as expansion basis
- **Principal Component Analysis (PCA)** to reproduce PDF covariance matrix with arbitrary precision
- Standard linear algebra problem: **Singular Value Decomposition**

Meta-PDFs →

- **MC replicas** are fitted to a common **meta-parametrization**
- **MC900** defines the **68% CL of the parameters** of the fitted function form
- Linear expansion in these fit parameters

MC2H(PCA)

- Define a **matrix for the deviations wrt central value** of all replicas of **MC900**

$$X_{lk}(Q) = f_{\alpha}^{(k)}(x_i,Q) - f_{\alpha}^{(0)}(x_i,Q)\,, \qquad l \equiv N_x(\alpha-1)+i$$

where **i** runs over a grid of **Nx** points in **x, alpha** is the PDF flavour, and **k** runs over the **Nrep** replicas

- The covariance matrix of **MC900** is then defined as

$$\mathrm{cov}_{ll'}(Q) = \frac{1}{N_{\rm rep}-1}\sum_{k}^{N_{\rm rep}} X_{lk}X^t_{kl'} \qquad \mathrm{cov}(Q) = \frac{1}{N_{\rm rep}-1}XX^t$$

- It is now possible to construct a representation of this covariance matrix as a linear combination of MC replicas, using **Singular Value Decomposition**

$$X = USV^t$$

Matrix with orthogonal eigenvectors of covariance matrix

Matrix constructed from singular values of X

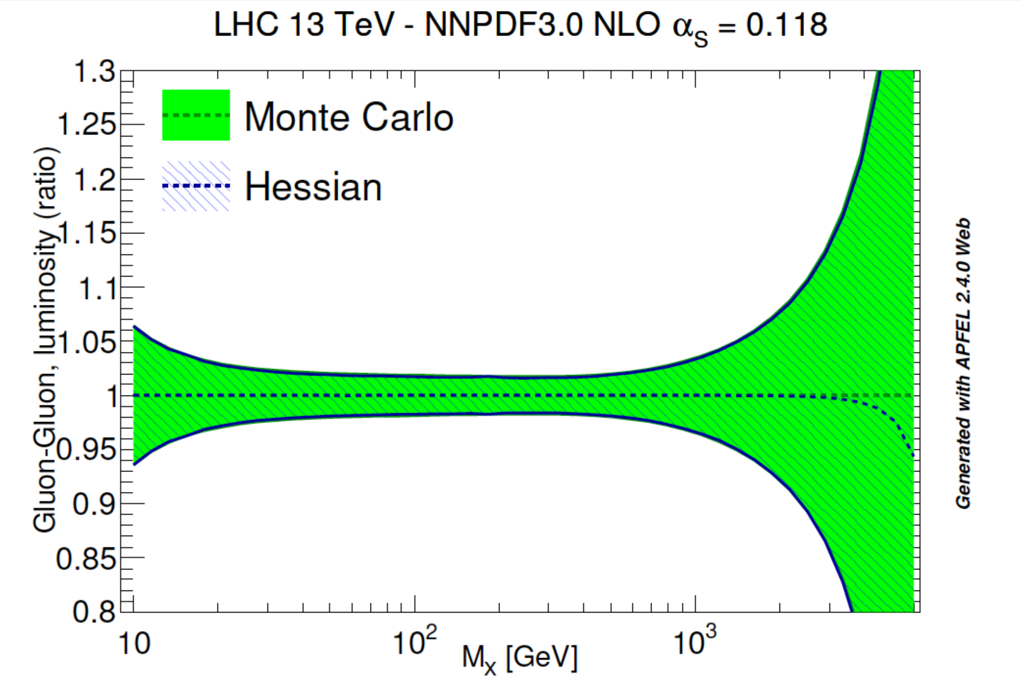
Orthogonal matrix of coefficients

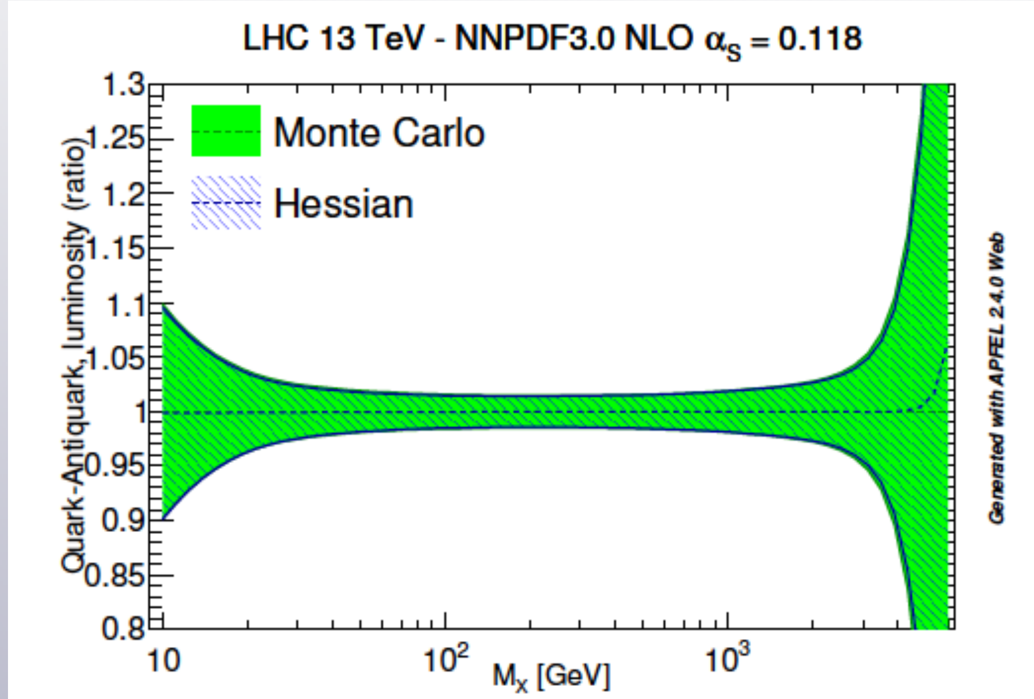
MC2H(PCA)

- It can then be shown that the matrix $Z=XV$ contains the **Hessian parameter**s which yield a **representation of the PDF covariance matrix** as linear combination of MC replicas
- This construction is **purely analytical,** without any numerical approximation involved
- Thus MC2H(PCA) achieves an **exact** representation of the PDF covariance matrix: by construction, this is a **perfect Hessian representation**
- Number of eigenvectors very large, $N_{eig}=N_{pdf}\,N_x$, but with SVD methods possible to **optimise the way information is stored** by **reducing N_{eig}** while minimising information loss, ordering eigenvectors.
- Use **Principal Value Decomposition** to reduce N_{eig} in a perfectly controlled manner (select coeff with largest singular values) so that results are stable within some fixed tolerance
- Optimal number (without compromising the accuracy of the method) is around $N_{eig} = 100$
- Basic assumption: reproduce the entire PDF covariance matrix, even in regions where the **underlying probability distribution is non-Gaussian**

Results: NNPDF3.0 and MMHT14

- The MC2H algorithm succeeds in **producing a Hessian representation of a native MC set,** NNPDF3.0





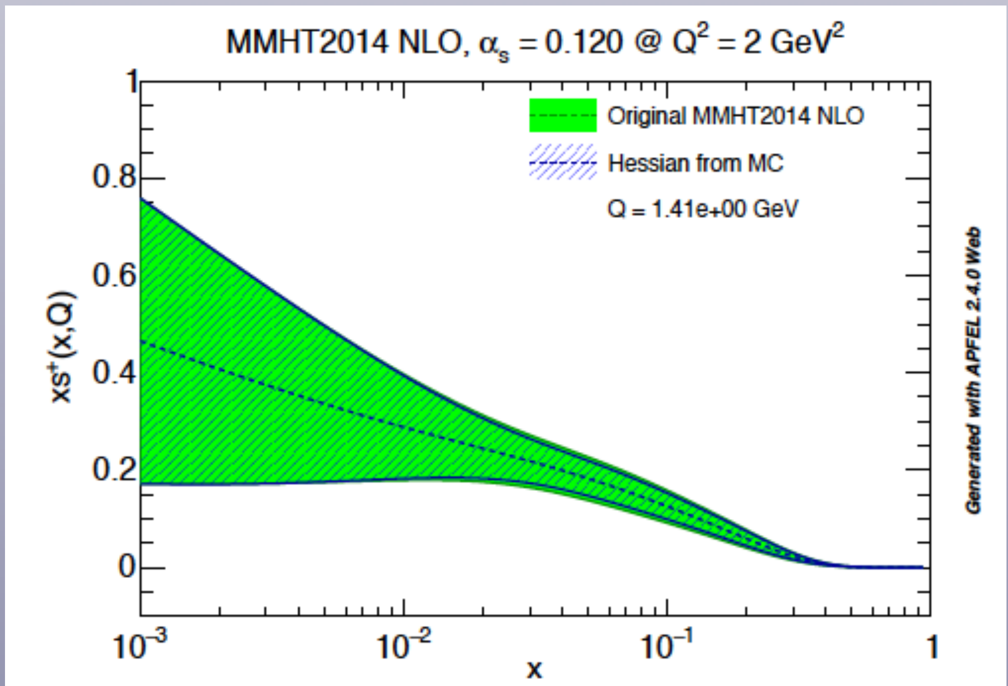
Results: NNPDF3.0 and MMHT14

- The MC2H algorithm succeeds in **producing a Hessian representation of a native MC set,** NNPDF3.0

MMHT14
Hessian in space of fit parameters

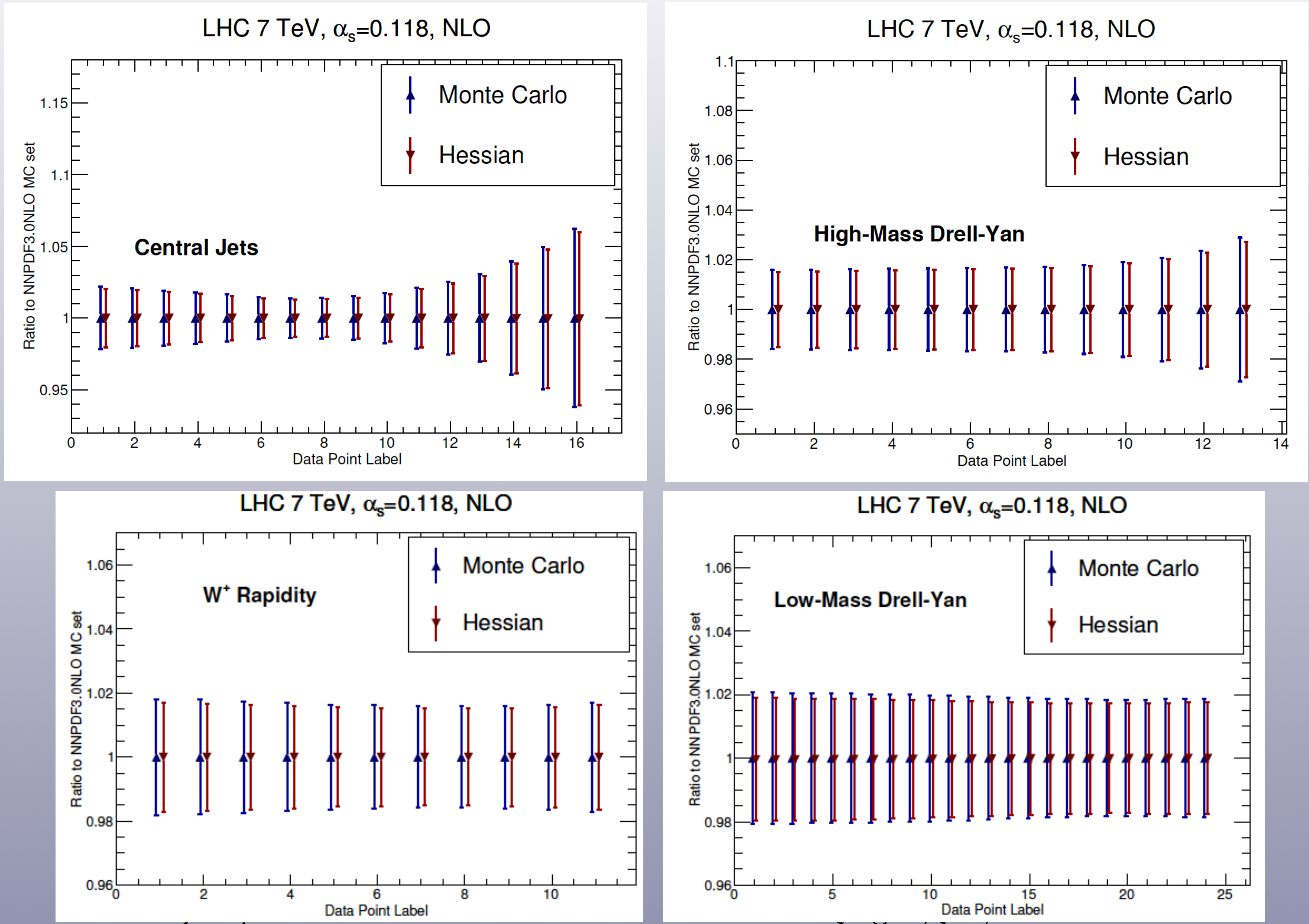
MMHT14-MC
Monte Carlo representation

MMHT14-MC2H
Hessian in space of MC replicas



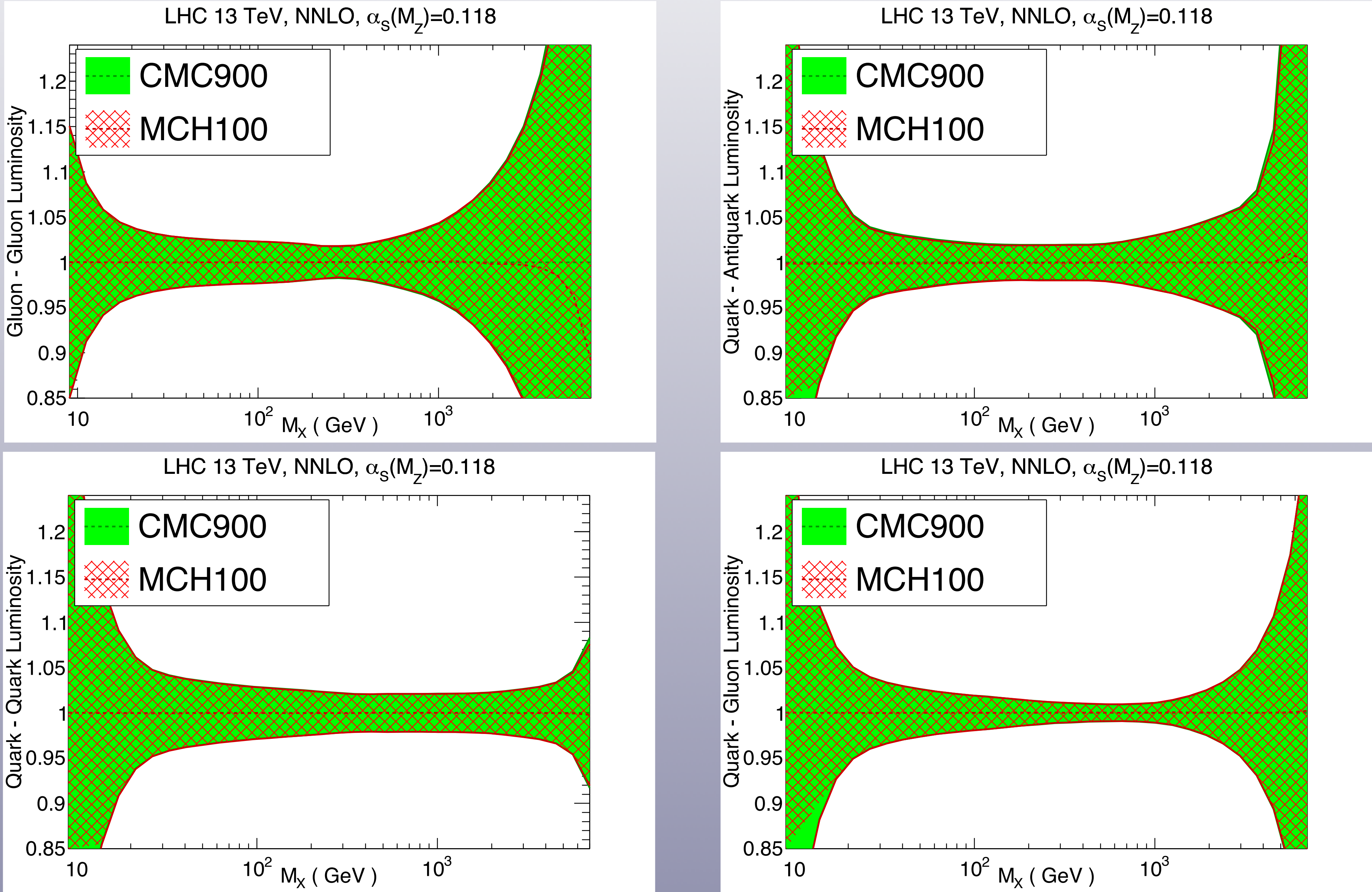
- Also verified that we can produce a **Hessian representation of a native Hessian set, MMH14,** via an **intermediate Monte Carlo representation**

MC2H LHC phenomenology



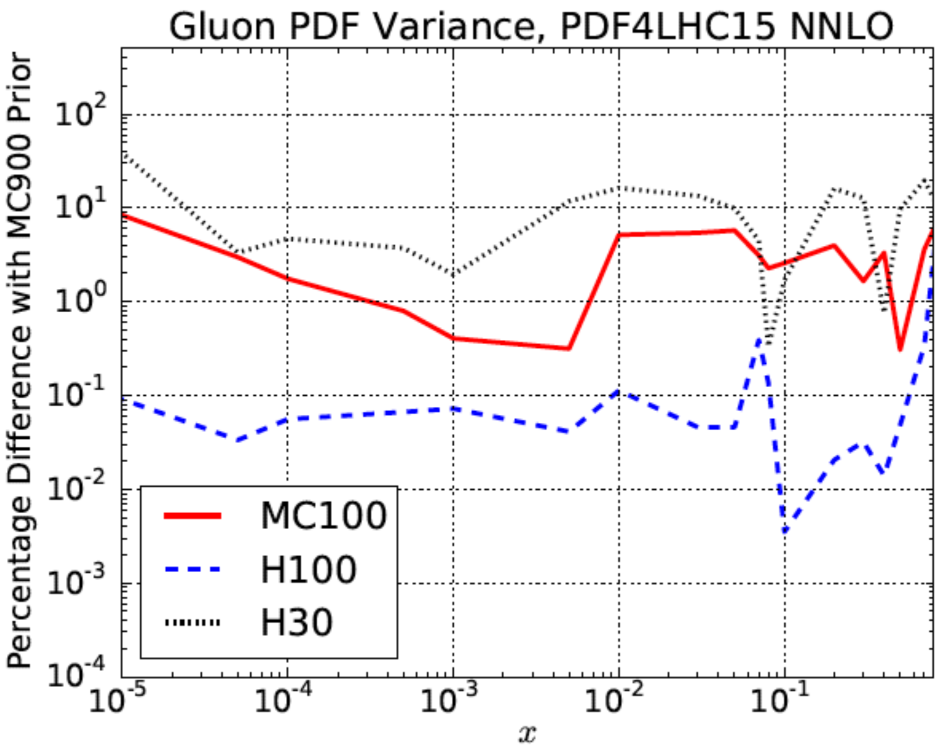
MC2H from MC900

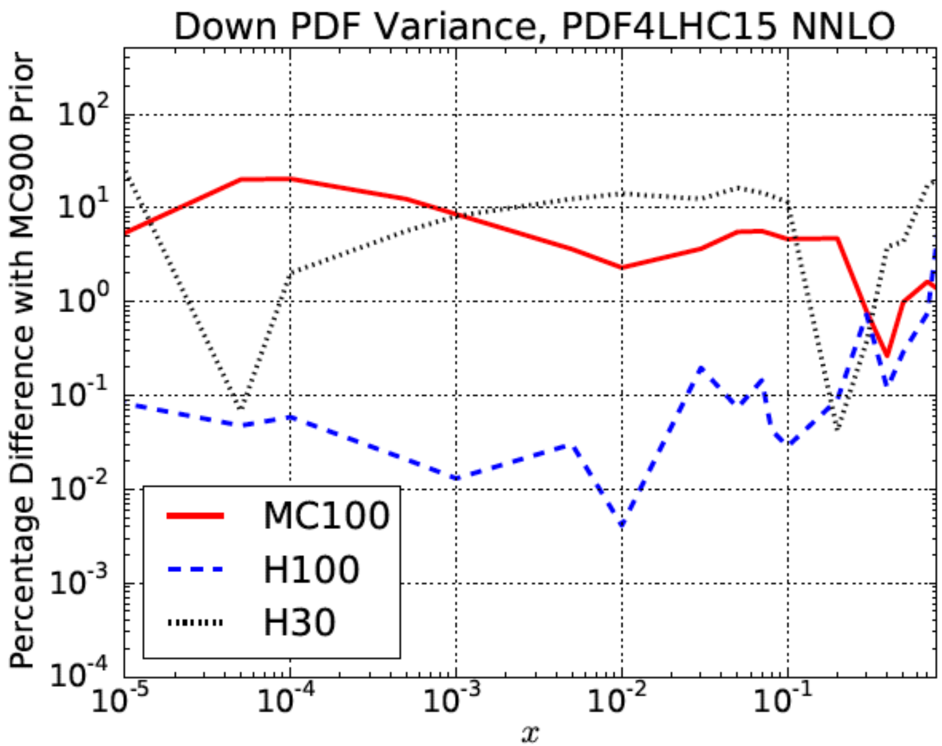
- A Hessian representation of the **MC900 PDF combination** has been constructed using **MC2H(PCA)**
- **Virtually perfect agreement** obtained for PDFs and luminosities as compared with MC900



Reproducing PDF uncertainties

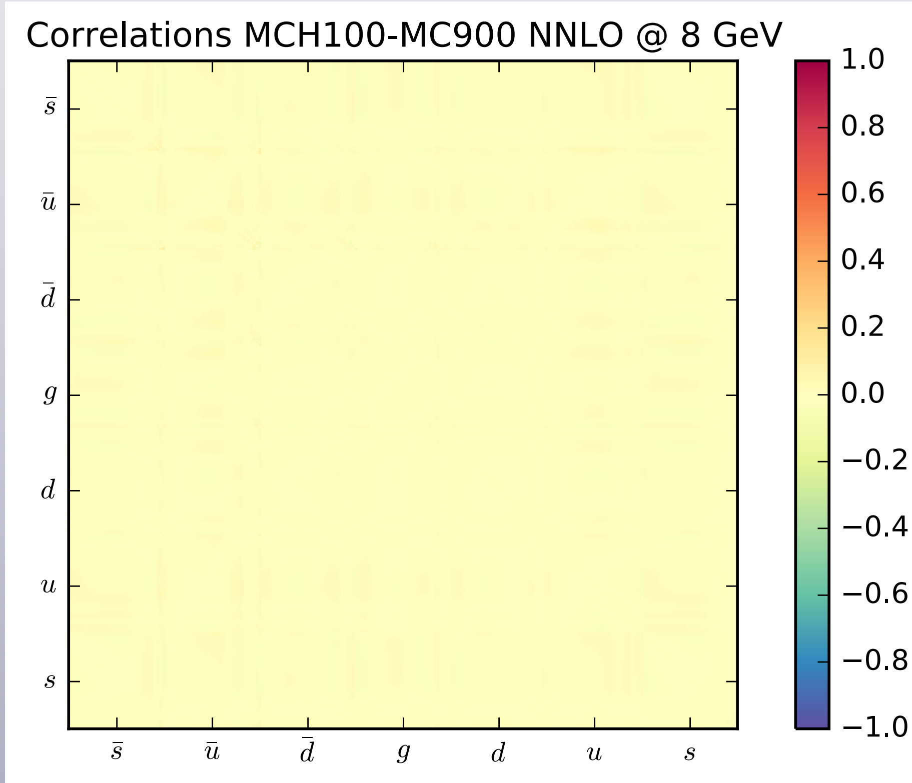
- **Variances** can be arbitrarily well reproduced in the MC2H method
- **Agreement with prior at < 0.1% level** for all PDF combinations in the complete x range
- Similar performance exhibited between CMC and H30





Correlation matrix

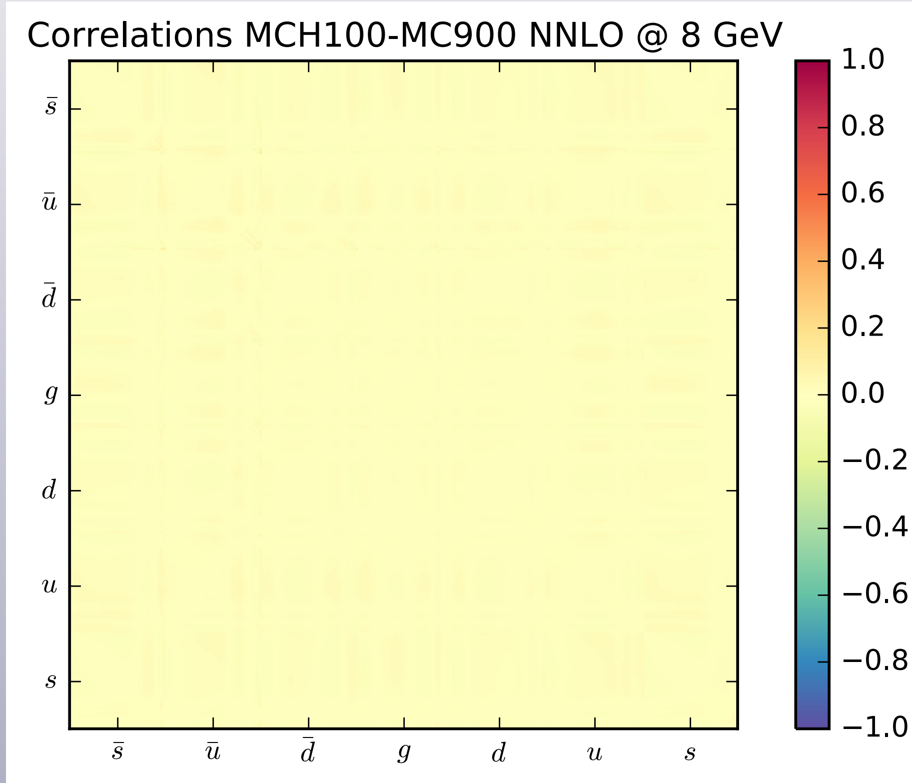
Since the method is essentially exact, one expects that the **PDF correlation matrix** is **perfectly reproduced** (modulo the eigenvector reduction step)

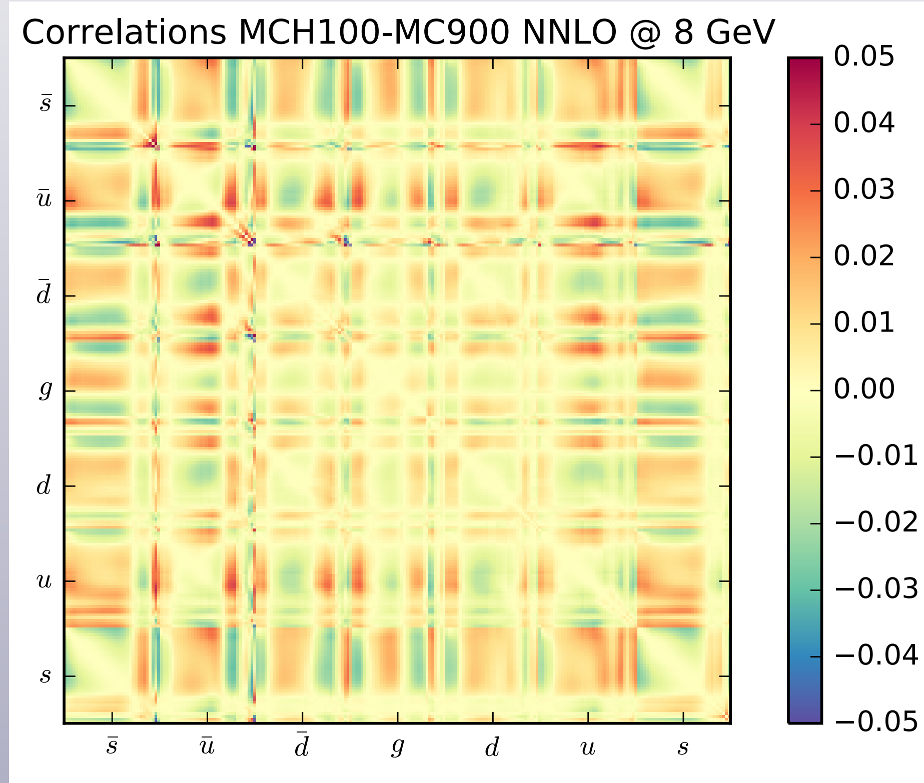


Perfect?

Correlation matrix

Since the method is essentially exact, one expects that the **PDF correlation matrix** is **perfectly reproduced** (modulo the eigenvector reduction step)





Perfect?

Zooming in ...

Almost!

Tiny residual differences at the level of few perfect of the correlation coefficient at most, **irrelevant for LHC phenomenology**

SMPDF

Specialised Minimal PDFs

Carrazza, Forte, Kassabov and J.R, in preparation

Strategy

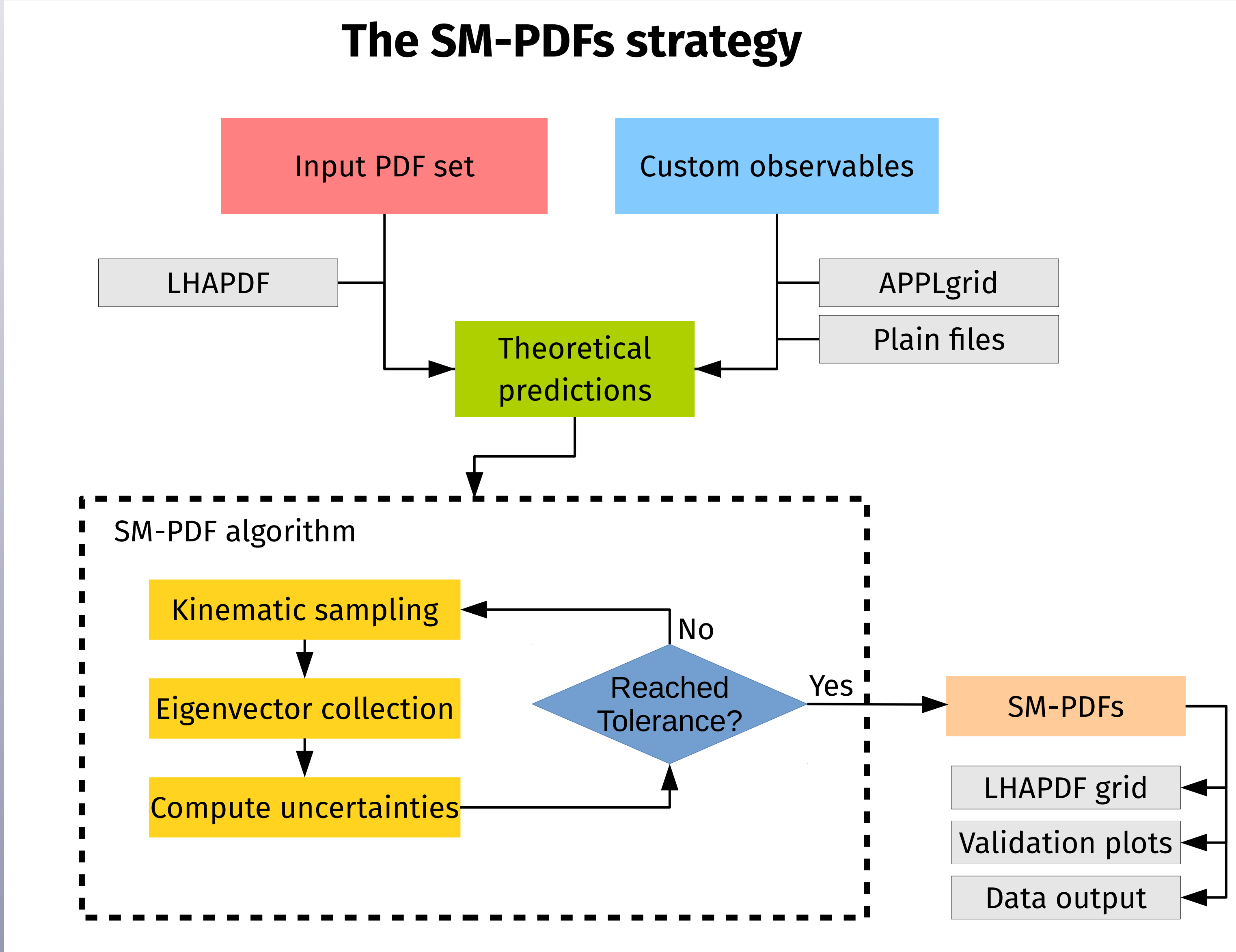
- **PDF sets specialised to a particular set of processes** might be useful to reduce the CPU burden of theory calculations, and facilitate the **treatment of PDF uncertainties and PDF-induced correlations**
- One clear example are **Higgs analysis in the HXS WG**, but similar **Specialized Minimal PDFs (SM-PDFs)** can be useful in the **TOP4LHC WG (**if restricted to top physics**)** and for the **W mass measurement (**if restricted to W, Z production**)**
- We have developed SM-PDFs that basically achieve the same result as the MC2H PDFs but restricted to a **particular subset of LHC cross-sections**
- Same method as MC2H, but different **choice of matrix** X used to construct the PDF covariance matrix, restricted to **values of** x **and PDF flavours relevant** for a given set of processes
- **SM-PDFs** can be **recombined** among them -> Adiabatically, the f**ull MC2H set is recovered**

Number of N_{eig} in the SM-PDF sets forr various processes for a given Tolerance T

Process	MC900		NNPDF3.0		MMHT14	
	$T_R = 5\%$	$T_R = 10\%$	$T_R = 5\%$	$T_R = 10\%$	$T_R = 5\%$	$T_R = 10\%$
h	15	11	13	8	8	7
$t\bar{t}$	4	4	5	4	3	3
W, Z	14	11	13	8	10	9
H+tt+Z+W	17	14	18	11	10	10

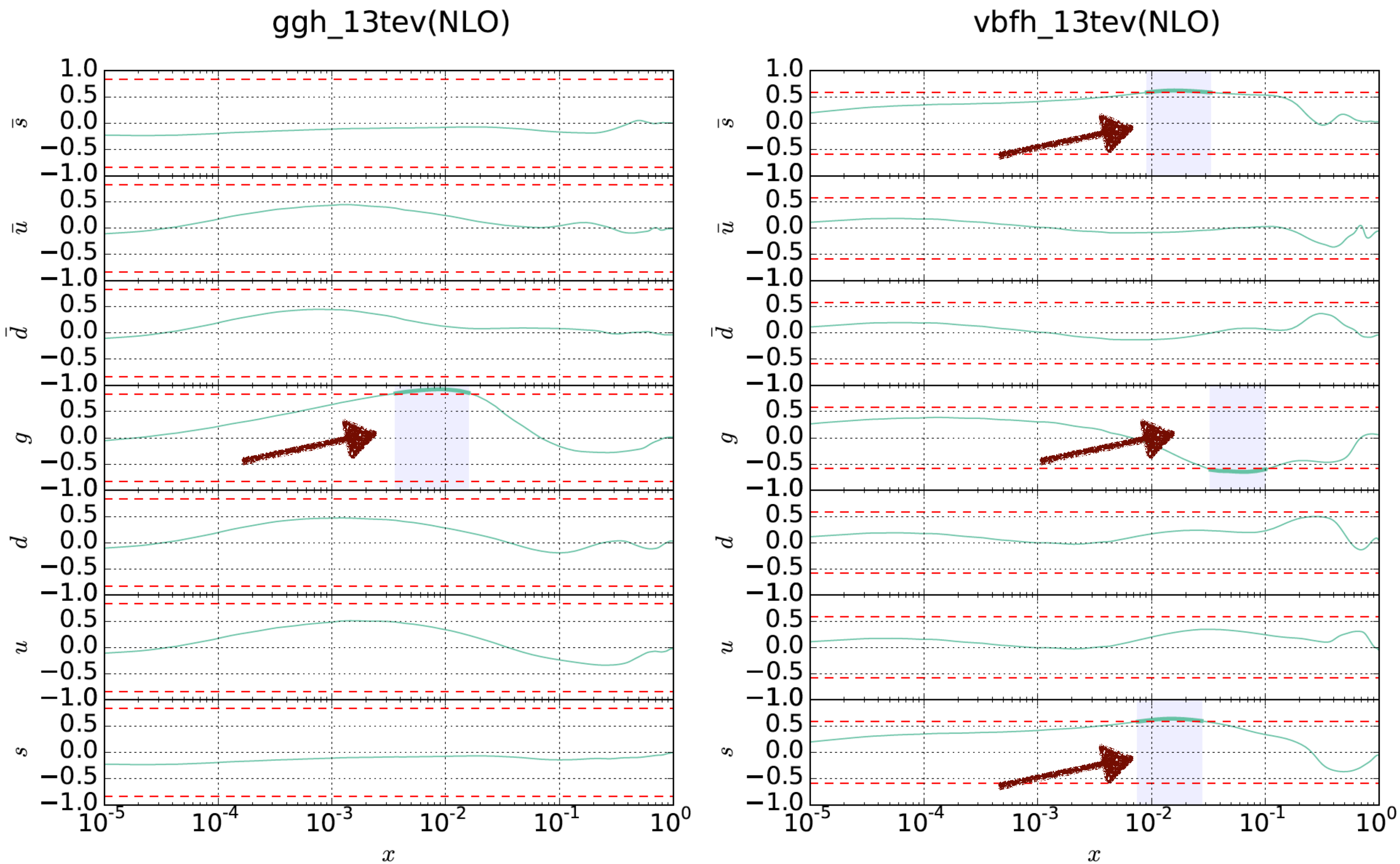
General strategy

Select which **cross-sections** are relevant for some specific application, generate **APPLgrids** for these, determine the region where the **correlations between PDFs and these observables is large**, and perform the **MC2H transformation** only for these, for some choice of **tolerance**



General strategy

Identify, for each input cross-section, t**he PDF flavours and region of Bjorken-x with high correlation**

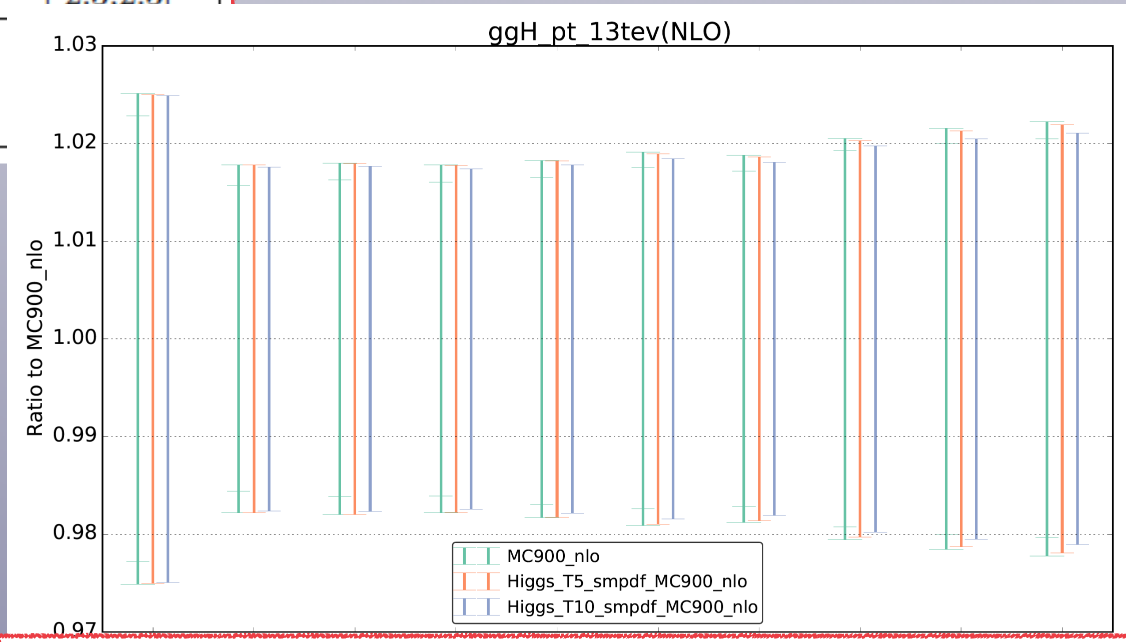


SM-PDFs for Higgs physics

- First exercise: construct **Specialised Minimal PDFs for Higgs physics:**

Input cross-sections for SM-PDFs for Higgs ph				
process	distribution	grid name	N_{bins}	range
$gg \to h$	incl xsec	ggh_13tev	1	-
	$d\sigma/dp_t^h$	ggh_pt_13tev	10	[0,200] GeV
	$d\sigma/dy^h$	ggh_y_13tev	10	[-2.5,2.5]
VBF hjj	incl xsec	vbfh_13tev	1	-
	$d\sigma/dp_t^h$	vbfh_pt_13tev	5	[0,200] GeV
	$d\sigma/dy^h$	vbfh_y_13tev	5	[-2.5,2.5]
hW	incl xsec	hw_13tev	1	-
	$d\sigma/dp_t^h$	hw_pt_13tev	10	[0,200] GeV
	$d\sigma/dy^h$	hw_y_13tev	10	[-2.5,2.5]
hZ	incl xsec	hz_13tev	1	-
	$d\sigma/dp_t^h$	hz_pt_13tev	10	[0,200] GeV
	$d\sigma/dy^h$	hz_y_13tev	10	[-2.5,2.5]
$ht\bar{t}$	incl xsec	httbar_13tev	1	
	$d\sigma/dp_t^h$	httbar_pt_13tev	10	
	$d\sigma/dy^h$	httbar_y_13tev	10	

- Use the **MC900** prior as input
- We check that with **10 (symmetric) eigenvectors** we can **reproduce all Higgs cross-sections and differential distributions**
- This means that of the starting 100 eigenvectors, **90 are not required for Higgs physics**



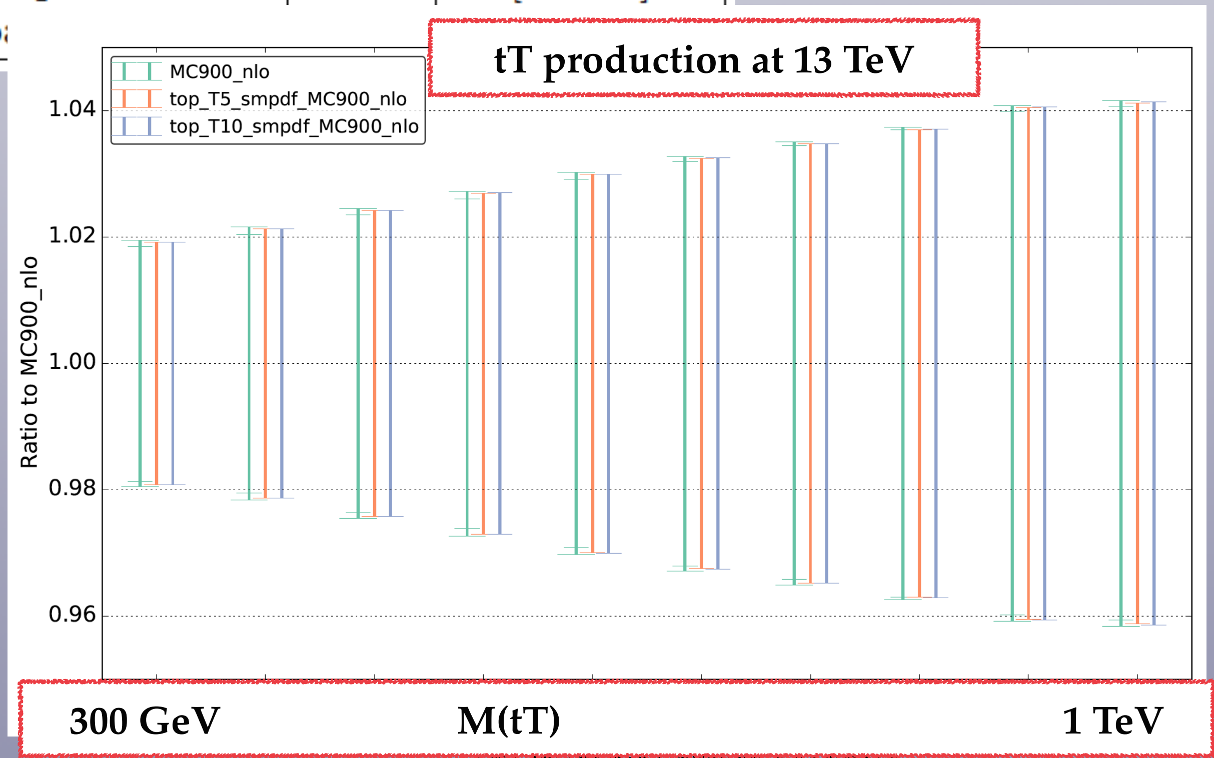
10 GeV **Higgs pT** **200 GeV**

SM-PDFs for Top physics

- Then construct S**pecialised Minimal PDFs for Top quark physics.** Specific request from **TOPLHC WG**.
- **Included the following set of distributions** computed with **aMC@NLO and aMCfast**

Input cross-sections for SM-PDFs for $t\bar{t}$ physics				
process	distribution	grid name	$N_{\rm bins}$	range
$t\bar{t}$	incl xsec	ttbar_13tev	1	-
	$d\sigma/dp_t^{\bar{t}}$	ttbar_tbarpt_13tev	10	[40,400] GeV
	$d\sigma/dy^{\bar{t}}$	ttbar_tbary_13tev	10	[-2.5,2.5]
	$d\sigma/dp_t^{t}$	ttbar_tpt_13tev	10	[40,400] GeV
	$d\sigma/dy^{t}$	ttbar_ty_13tev	10	[-2.5,2.5]
	$d\sigma/dm^{t\bar{t}}$	ttbar_ttbarinvmass_13tev	10	[300,1000]
	$d\sigma/dp_t^{t\bar{t}}$	ttbar_ttbarpt_13tev	10	[20,200]
	$d\sigma/dy^{t\bar{t}}$	ttbar_ttb		

- With **only 4 eigenvectors**, we can reproduce **all relevant distributions in top quark** pair production
- Additional processes (like single top) can be easily added if needed



SM-PDFs for W mass determination

- Then construct S**pecialised Minimal PDFs for W mass determination studies quark physics. Should be useful for ongoing ATLAS and CMS analysis**

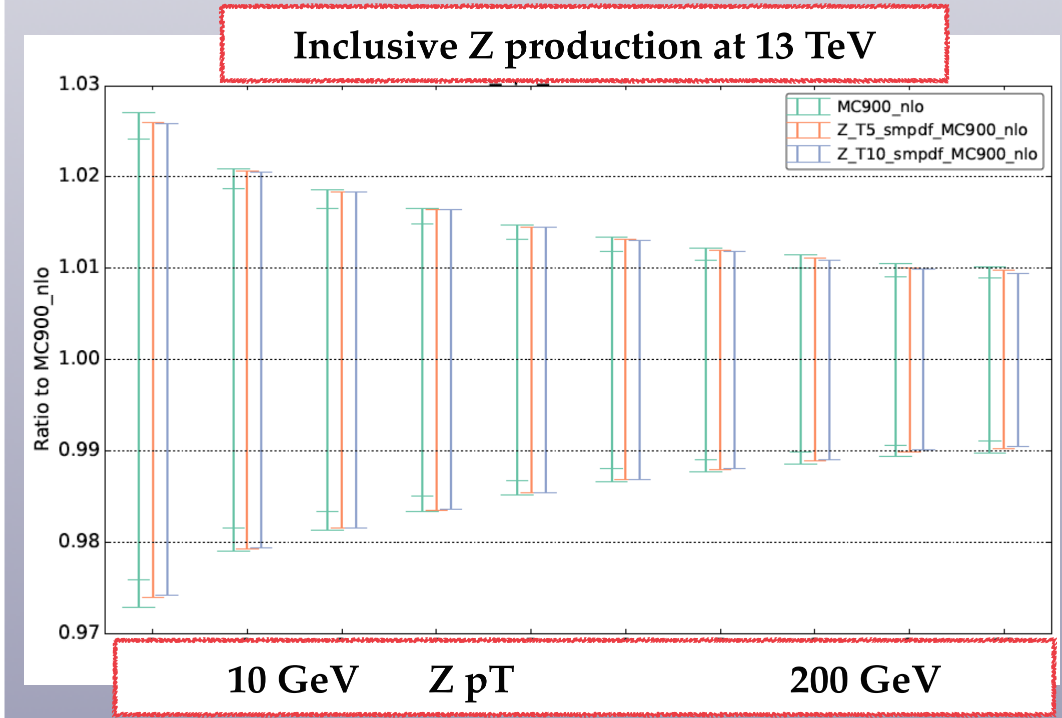
- **Included the following set of distributions:**

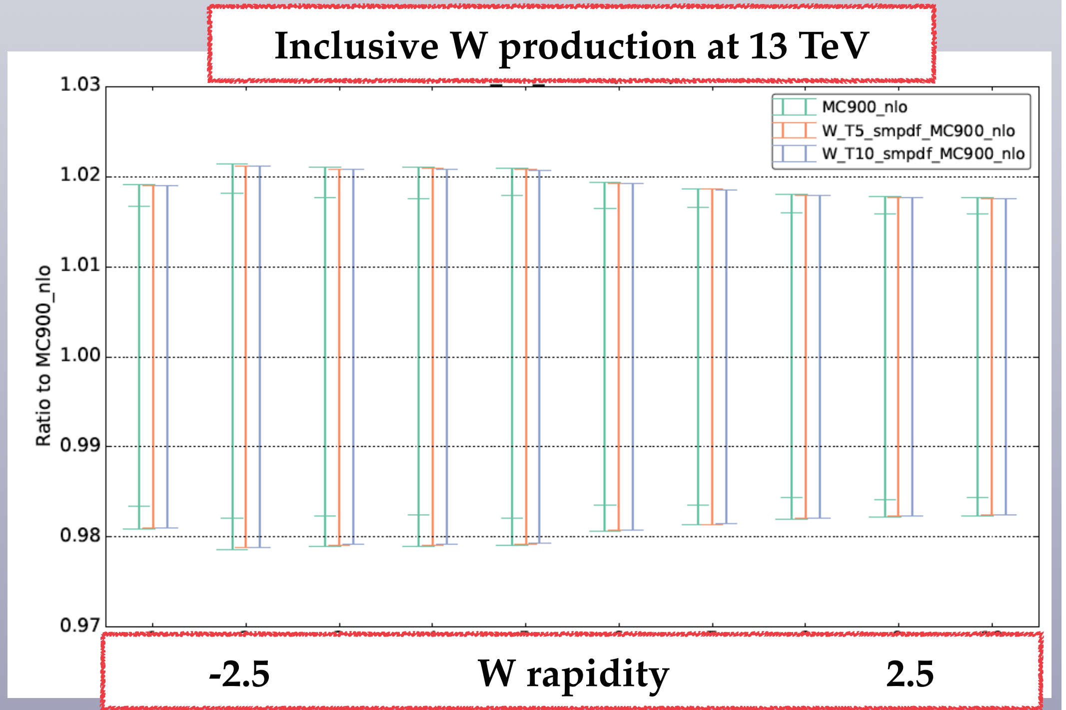
Process	APPLgrid	$\sqrt{s}$ (TeV)	$N_{\rm bins}$	Range	Cuts
Z	z_13tev	13	1	-	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	z_lmpt_13tev	13	10	[0,200] GeV	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	z_lmy_13tev	13	10	[-2.5,2.5]	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	z_lppt_13tev	13	10	[0,200] GeV	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	z_lpy_13tev	13	10	[-2.5,2.5]	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	z_zpt_13tev	13	10	[0,200] GeV	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	z_zy_13tev	13	5	[-4,4]	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	z_lplminvmass_13tev	13	10	[50,130] GeV	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	z_lplmpt_13tev	13	10	[0,200] GeV	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$

Process	APPLgrid	$\sqrt{s}$ (TeV)	$N_{\rm bins}$	Range	Cuts
W	w_13tev	13	1	-	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	w_cphi_13tev	13	10	[-1,1]	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	w_etmiss_13tev	13	10	[0,200] GeV	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	w_lpt_13tev	13	10	[0,200] GeV	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	w_ly_13tev	13	10	[-2.5,2.5]	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	w_mt_13tev	13	10	[0,200] GeV	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	w_wpt_13tev	13	10	[0,200] GeV	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	w_wy_13tev	13	10	[-4,4]	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$

SM-PDFs for W mass determination

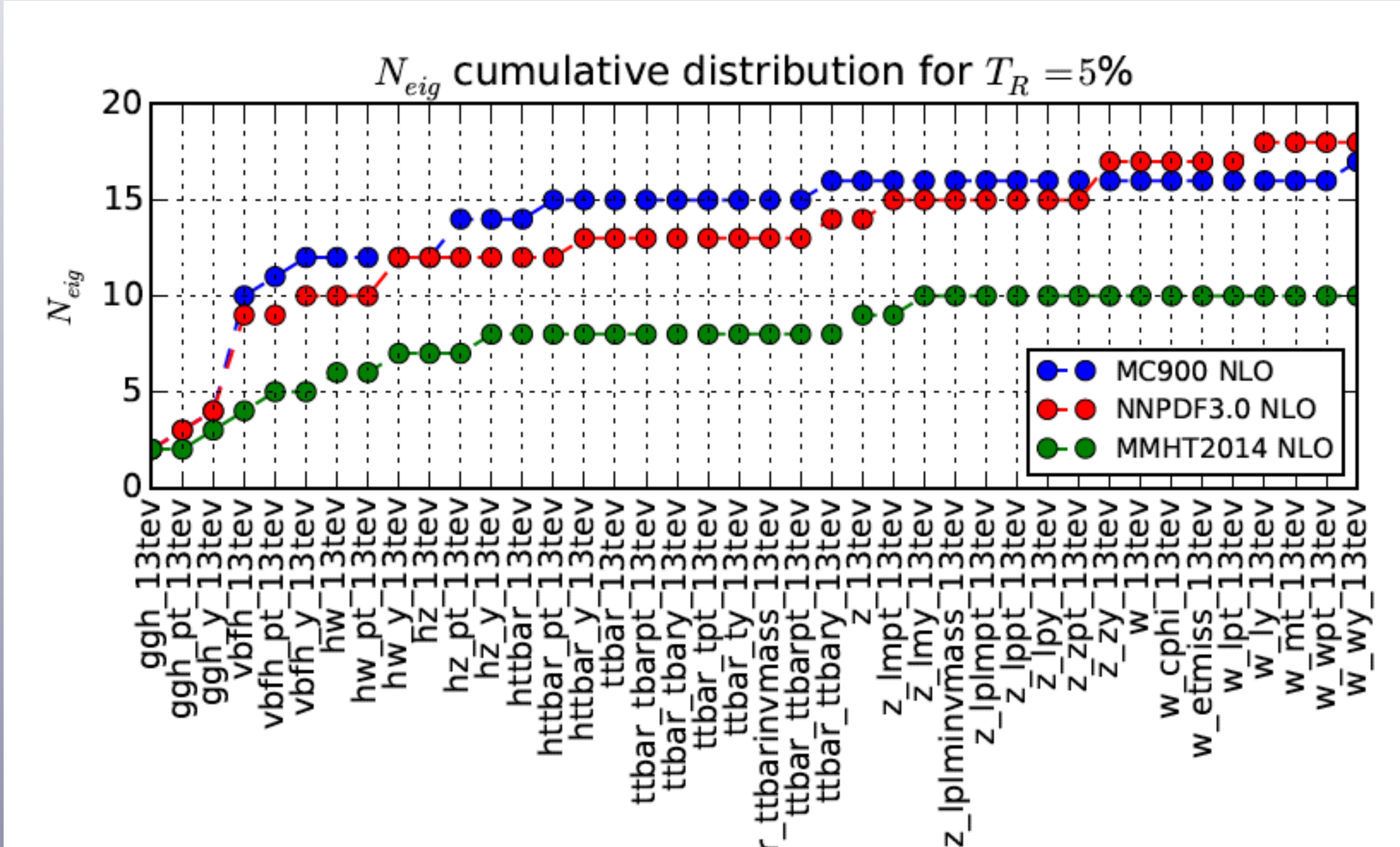
- Then construct **Specialised Minimal PDFs for W mass determination studies. Should be useful for ongoing ATLAS and CMS analysis**
- A somewhat **larger number of eigenvectors selected** here, since W,Z production involve all PDF flavours in a wide range of x),
- Even in this case **with only 8 to 14 eigenvectors** we can reproduce the original **MC900** result





Climbing up the ladder

- The **number of eigenvectors** required **increases monotonically** as more and more input cross-sections are included in the procedure
- In the limit where all possible cross-sections are included, the **original MC2H set is reproduced**



PDF4LHC mandate and activities

The PDF4LHC Working Group has been tasked with:

1. performing benchmark studies of PDFs and of predictions at the LHC, and
2. making recommendations for a standard method of estimating PDF and PDF+$\alpha_s(m_Z^2)$ uncertainties at the LHC through a combination of the results from different individual groups.

This mandate has led to several benchmarking papers [12, 13] and to the 2010 PDF4LHC recommendation [14] which has undergone several intermediate updates, with the last version available (along with a summary of PDF4LHC activities) from the PDF4LHC Working Group website:

`http://www.hep.ucl.ac.uk/pdf4lhc/`.

- Created in 2010 by the CERN directorate
- **Regular PDF4LHC meetings**, discussions between PDF fitters and experimentalists
- **Several benchmarking studies** produced: genuine improved understanding of differences/similarities between PDF fits, with subsequent improved agreement

The PDF4LHC report on PDFs and LHC data: Results from Run I and preparation for Run II

arxiv:1507.0056

Juan Rojo[1], Alberto Accardi[2,3], Richard D. Ball[4,5], Amanda Cooper-Sarkar[6], Albert de Roeck[5,7], Stephen Farry[8], James Ferrando[9], Stefano Forte[10], Jun Gao[11], Lucian Harland-Lang[12], Joey Huston[13], Alexander Glazov[14], Maxime Gouzevitch[15], Claire Gwenlan[6], Katerina Lipka[14], Mykhailo Lisovyi[16], Michelangelo Mangano [5], Pavel Nadolsky[17], Luca Perrozzi[18], Ringaile Plačakytė[14], Voica Radescu[16], Gavin P. Salam[5] and Robert Thorne[12]*

Input to the PDF4LHC15 combination

All PDF sets are equal, but some of them are more equal than others (Orwell, *Animal Farm*)

1. The PDF sets to be combined should be *based on a global dataset*, including a large number of datasets of diverse types (deep-inelastic scattering, vector boson and jet production, ...) from fixed-target and colliders experiments (HERA, LHC, Tevatron).

2. Theoretical hard cross sections for DIS and hadron collider processes should be evaluated up to *two QCD loops in α_s*, in a *general-mass variable-flavor number scheme* with up to $n_f^{\max} = 5$ active quark flavors.[1] Evolution of α_s and PDFs should be performed up to three loops, using public codes such as `HOPPET` [105] or `QCDNUM` [106], or a code benchmarked to these.

3. The central value of $\alpha_s(m_Z^2)$ *should be fixed at an agreed common value*, consistent with the PDG world-average [107]. This value is currently chosen to be $\alpha_s(m_Z^2) = 0.118$ at both NLO and NNLO.[2] For the computation of α_s uncertainties, two additional PDF members corresponding to agreed upper and lower values of $\alpha_s(m_Z^2)$ should also be provided. This uncertainty on $\alpha_s(m_Z^2)$ is currently assumed to be $\delta\alpha_s = 0.0015$, again the same at NLO and NNLO.

   The input values of m_c and m_b should be compatible with their world-average values; either pole or $\overline{\mathrm{MS}}$ masses are accepted.

4. *All known experimental and procedural sources of uncertainty should be properly accounted for*. Specifically, it is now recognized that the PDF uncertainty receives several contributions of comparable importance: the measurement uncertainty propagated from

Usage of the PDF4LHC15 sets

1. **Comparisons between data and theory for Standard Model measurements**

   Recommendations: Use *individual PDF sets*, and, in particular, as many of the modern PDF sets [5–11] as possible.

2. **Searches for Beyond the Standard Model phenomena**

   Recommendations: Use the `PDF4LHC15_mc` sets.

   Rationale: BSM searches, in particular for *new massive particles in the TeV scale*, often require the knowledge of PDFs in regions where available experimental constraints are limited, notably close to the hadronic threshold where $x \to 1$ [127]. In these extreme kinematical regions the PDF uncertainties are large, the *Monte Carlo combination of PDF sets is likely to be non-Gaussian. c.f.* Figs. 10 and 11.

3. **Calculation of PDF uncertainties in situations when computational speed is needed, or a more limited number of error PDFs may be desirable**

   Recommendations: Use the `PDF4LHC15_30` sets.

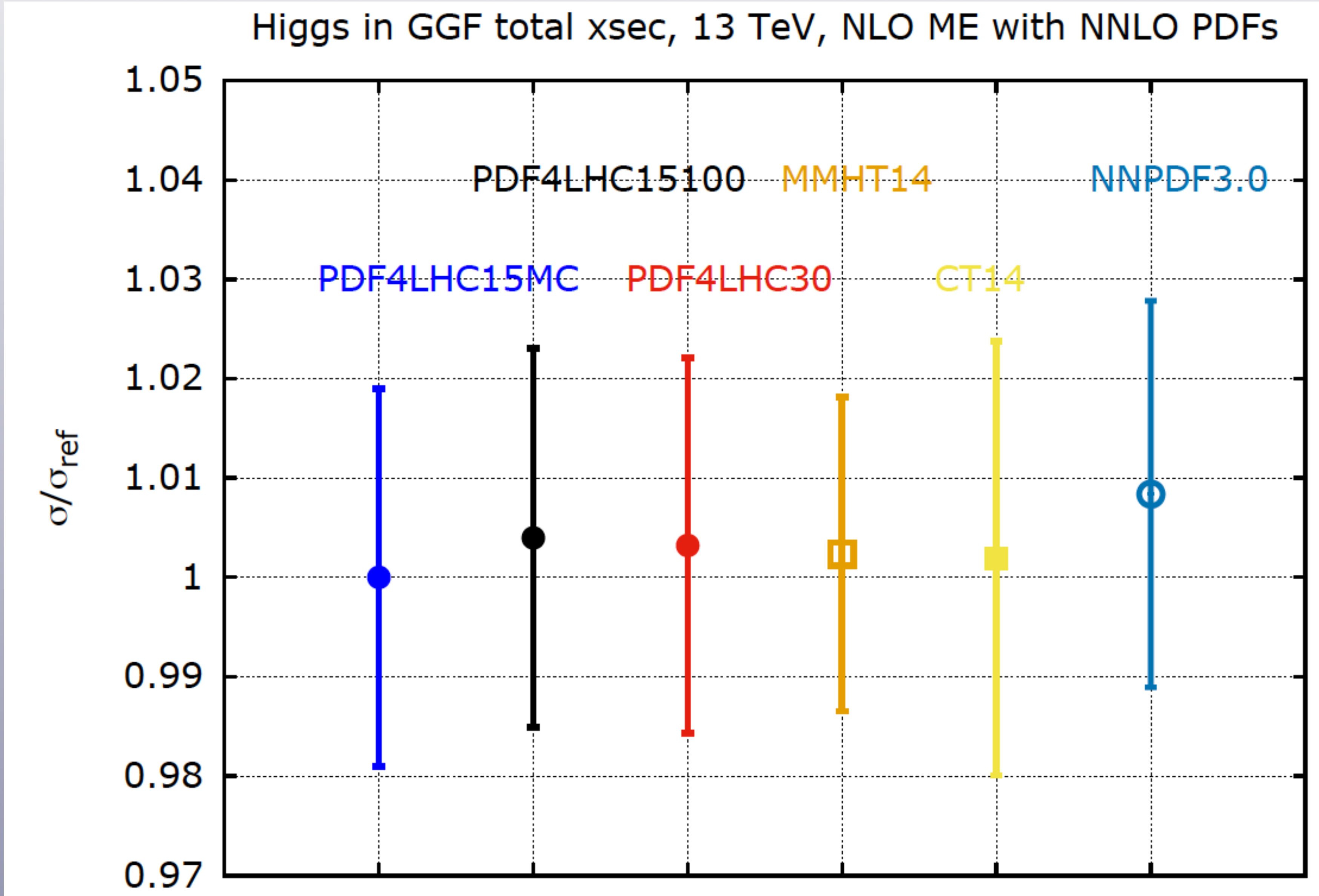
4. **Calculation of PDF uncertainties in precision observables**

   Recommendation: Use the `PDF4LHC15_100` sets.

   Rationale: For several LHC phenomenological applications, the highest accuracy is sought for, with, in some cases, the need to *control PDF uncertainties to the percent level*, as currently allowed by the development of high-order computational techniques in the QCD and electroweak sectors of the Standard Model.

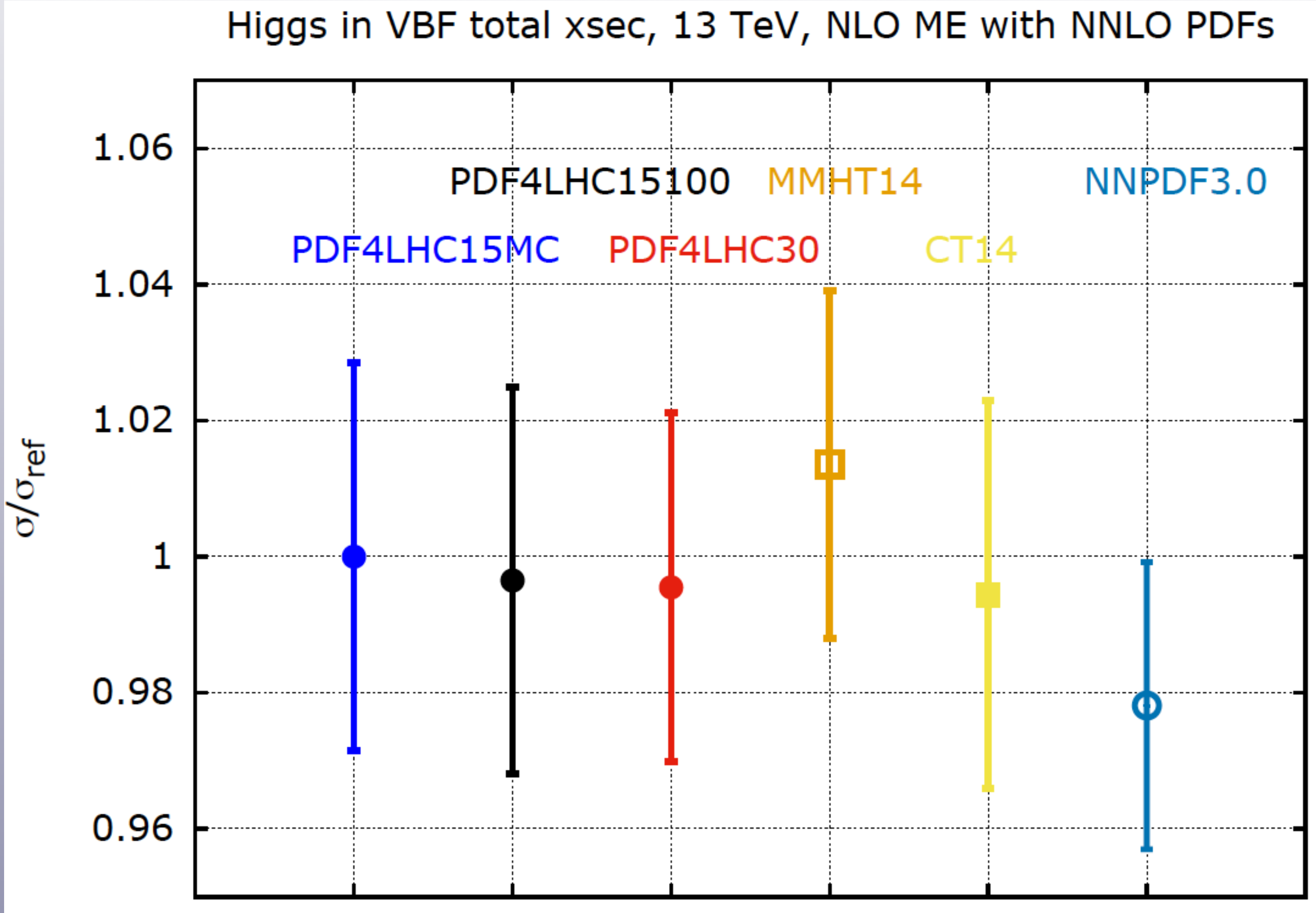
Towards HXSWG Yellow Report IV

The **PDF4LHC15 combined sets** are being used to produce the benchmark cross-sections of **Yellow Report 4** of the **Higgs Cross-Section Working Group**



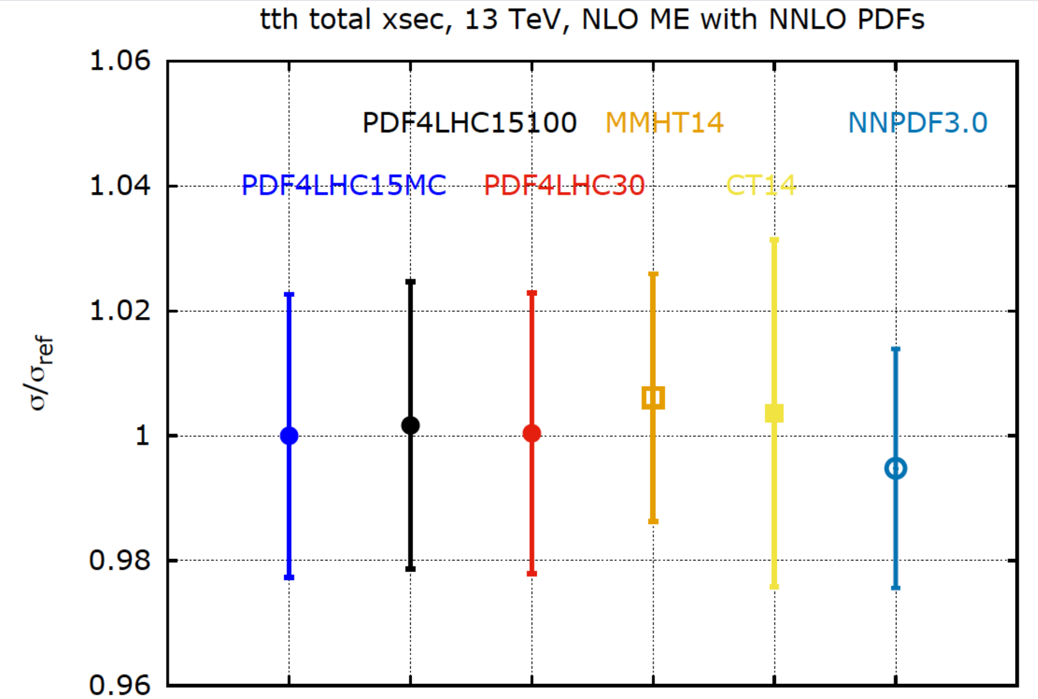
Towards HXSWG Yellow Report IV

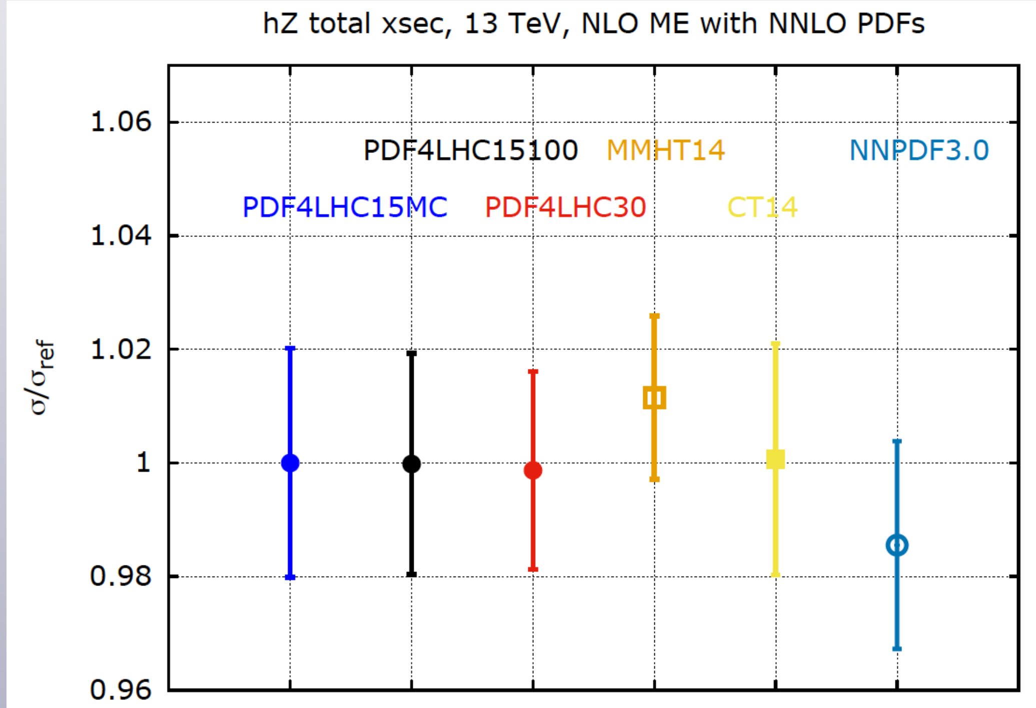
The **PDF4LHC15 combined sets** are being used to produce the benchmark cross-sections of **Yellow Report 4** of the **Higgs Cross-Section Working Group**



Towards HXSWG Yellow Report IV

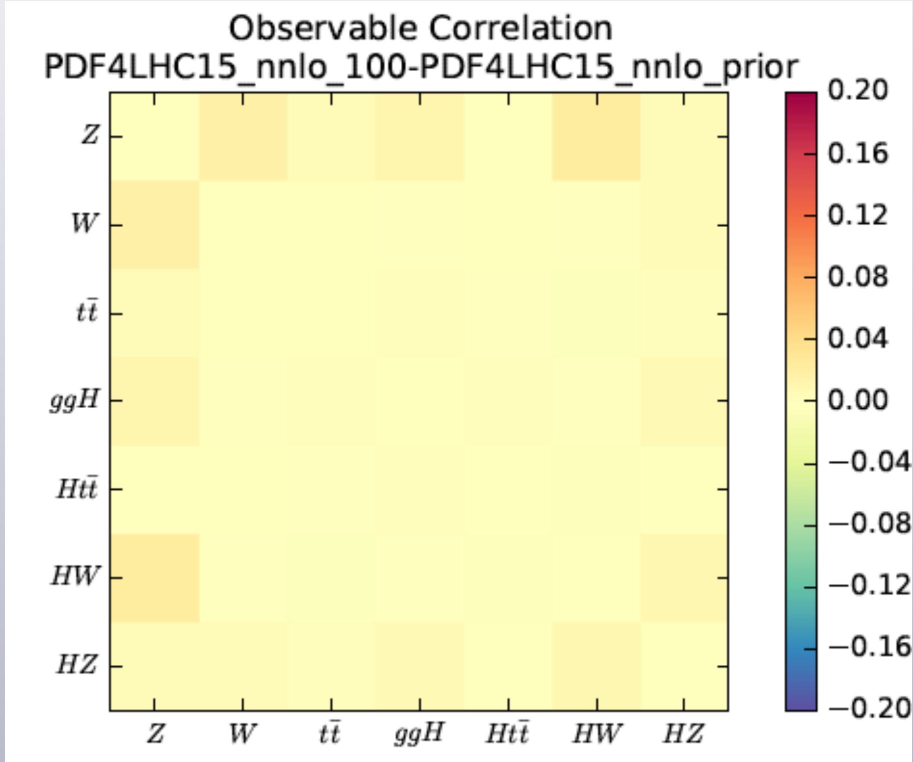
The **PDF4LHC15 combined sets** are being used to produce the benchmark cross-sections of **Yellow Report 4** of the **Higgs Cross-Section Working Group**

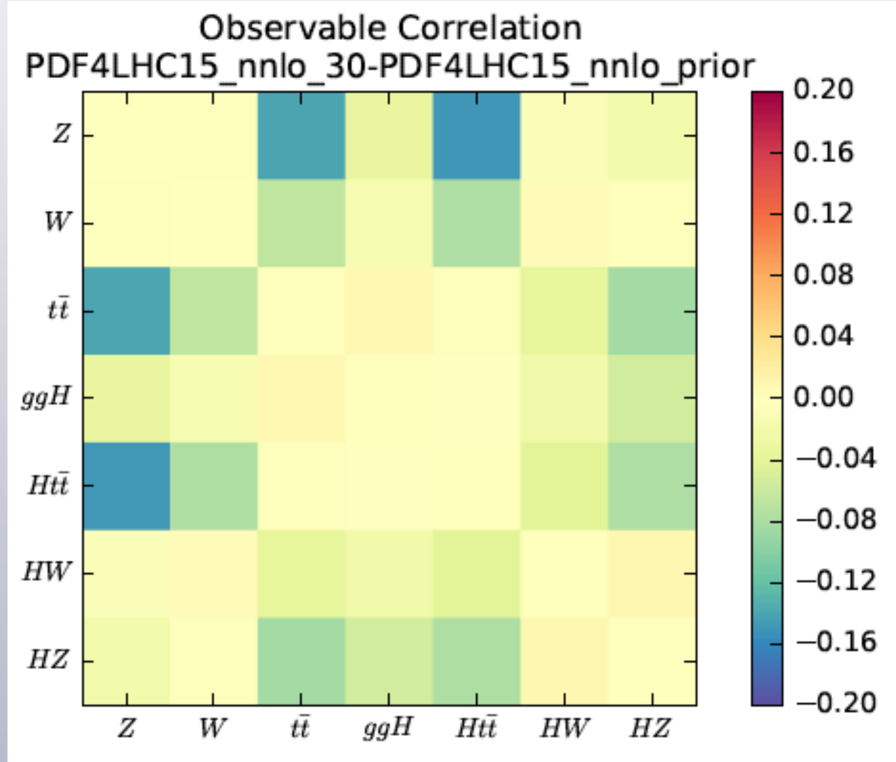




Similar comparisons at the level of **Higgs production and decay differential distributions** are also readily available

Towards HXSWG Yellow Report IV





PDF Set	Correlation coefficient					
	$t\bar{t}, Ht\bar{t}$	$t\bar{t}, hW$	$t\bar{t}, hZ$	$ggh, ht\bar{t}$	ggh, hW	ggh, hZ
PDF4LHC15_nlo_prior	0.93	-0.22	-0.50	-0.02	0.15	0.08
PDF4LHC15_nlo_mc	0.92	-0.14	-0.41	-0.04	0.33	0.27
PDF4LHC15_nlo_100	0.93	-0.22	-0.48	-0.03	0.15	0.08
PDF4LHC15_nlo_30	0.93	-0.25	-0.54	0.02	0.11	-0.01
PDF4LHC15_nnlo_prior	0.87	-0.23	-0.34	-0.13	-0.01	-0.17
PDF4LHC15_nnlo_mc	0.87	-0.27	-0.35	-0.10	0.07	-0.01
PDF4LHC15_nnlo_100	0.87	-0.24	-0.34	-0.13	-0.02	-0.17
PDF4LHC15_nnlo_30	0.87	-0.27	-0.43	-0.13	-0.04	-0.23