The PDF4LHC Recommendations for Run II

Based on J. **Butterworth et al, arXiv:1510.03865**
Journal of Physics G 43 (2016) 023001

Juan Rojo
STFC Rutherford Fellow
Rudolf Peierls Center for Theoretical Physics
University of Oxford

DIS2016 Workshop
DESY, Hamburg, 12/04/2016

Why do we need a recommendation?

Higgs production cross-section in gluon fusion at the LHC is a crucial process to test the Higgs sector: the **higher the theory accuracy**, the more powerful our **discovery power for New Physics** is

Higgs couplings may indicate new physics: a few percent precision is a good target

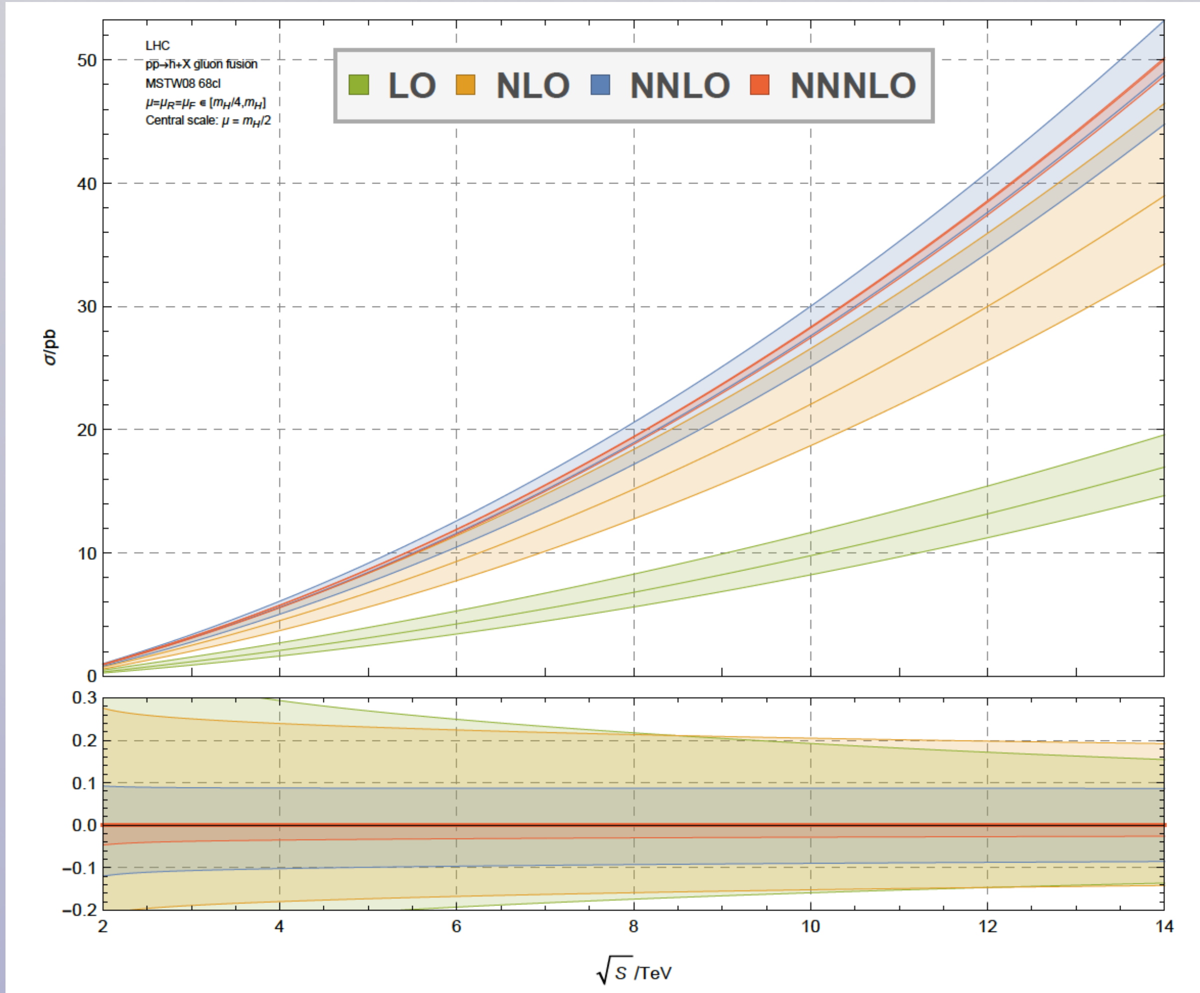
Higgs Snowmass report (arXiv:1310.8361)
Deviation from SM due to particles with M=1 TeV

Model	κ_V	κ_b	κ_γ
Singlet Mixing	$\sim 6\%$	$\sim 6\%$	$\sim 6\%$
2HDM	$\sim 1\%$	$\sim 10\%$	$\sim 1\%$
Decoupling MSSM	$\sim -0.0013\%$	$\sim 1.6\%$	$\sim -.4\%$
Composite	$\sim -3\%$	$\sim -(3-9)\%$	$\sim -9\%$
Top Partner	$\sim -2\%$	$\sim -2\%$	$\sim +1\%$

Future LHC data: measure H couplings at 2-8% level (cf 20-50% today), and to access rare decays such as $H \to \mu\mu$

Why do we need a recommendation?

Higgs production cross-section in gluon fusion at the LHC is a crucial process to test the Higgs sector: the **higher the theory accuracy**, the more powerful our **discovery power for New Physics** is



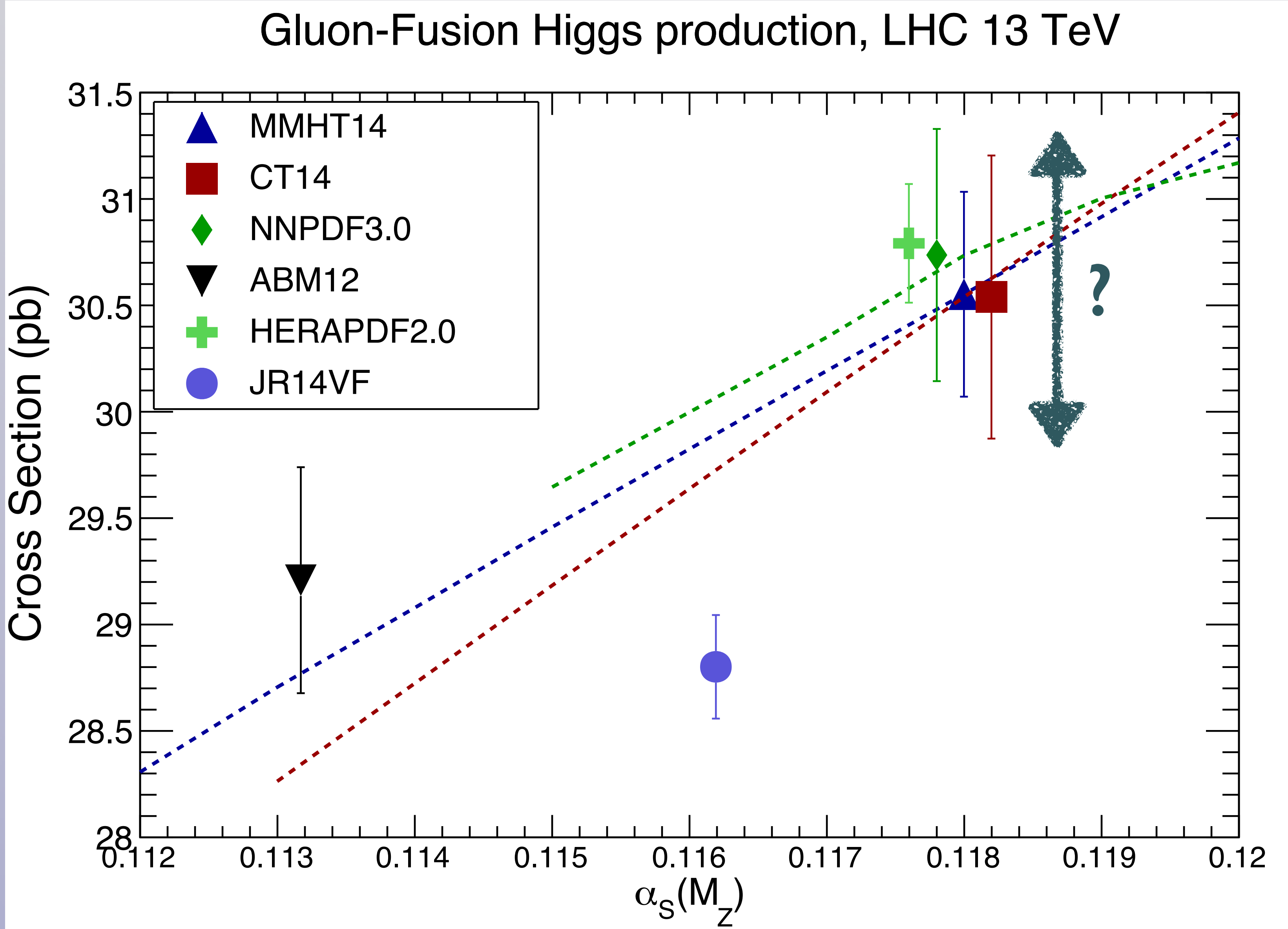
- Impressive progress in perturbative calculations! This process is known at **N3LO**, scale errors down **δσ≃3%!**

- Does this mean that we can **probe BSM physics** at this level of precision?

- This will depend on the size of other theory uncertainties, first and foremost the **PDF and strong coupling uncertainties**

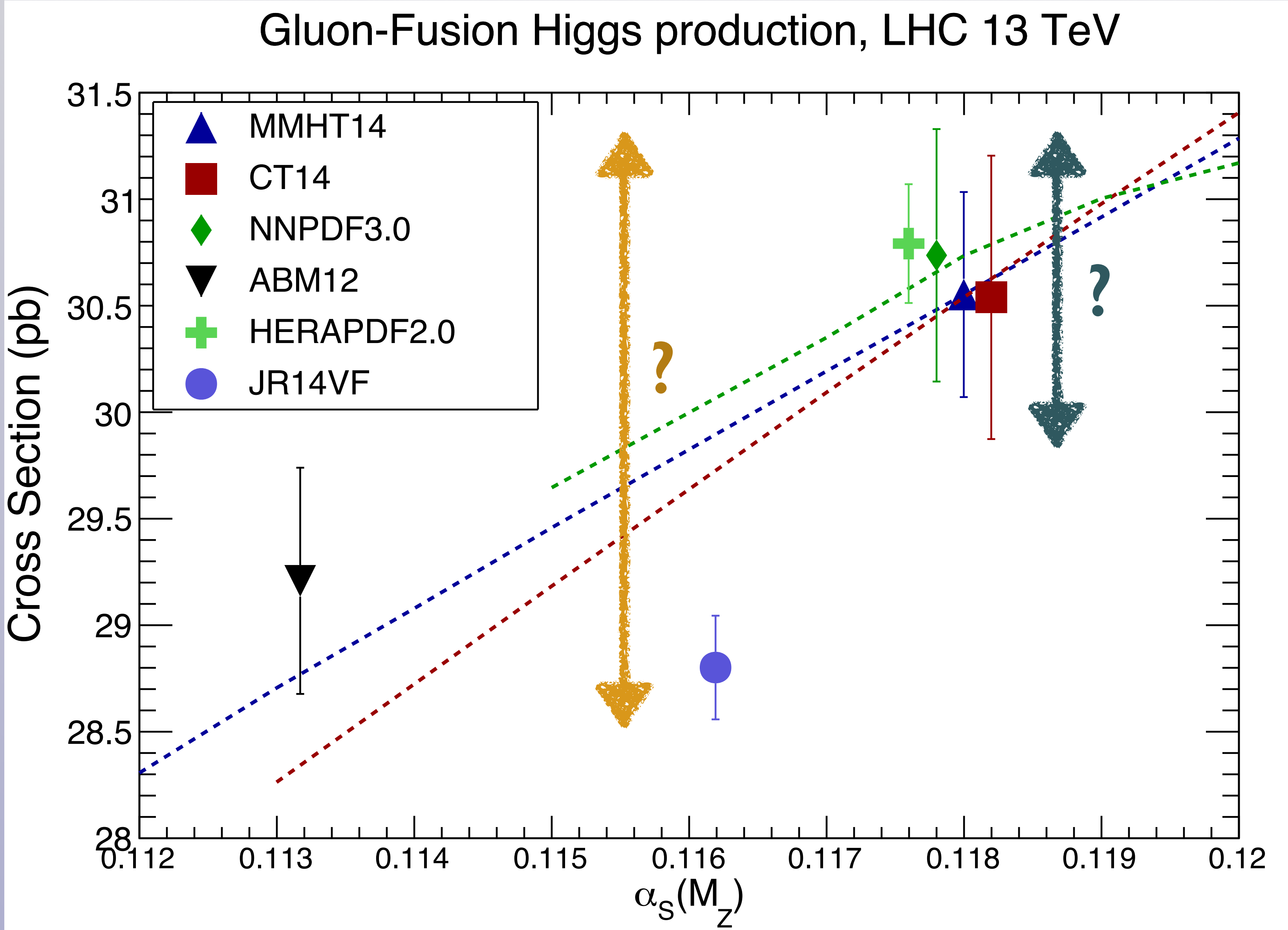
Why do we need a recommendation?

A well-defined PDF recommendation is mandatory for compare data with theory for Higgs physics



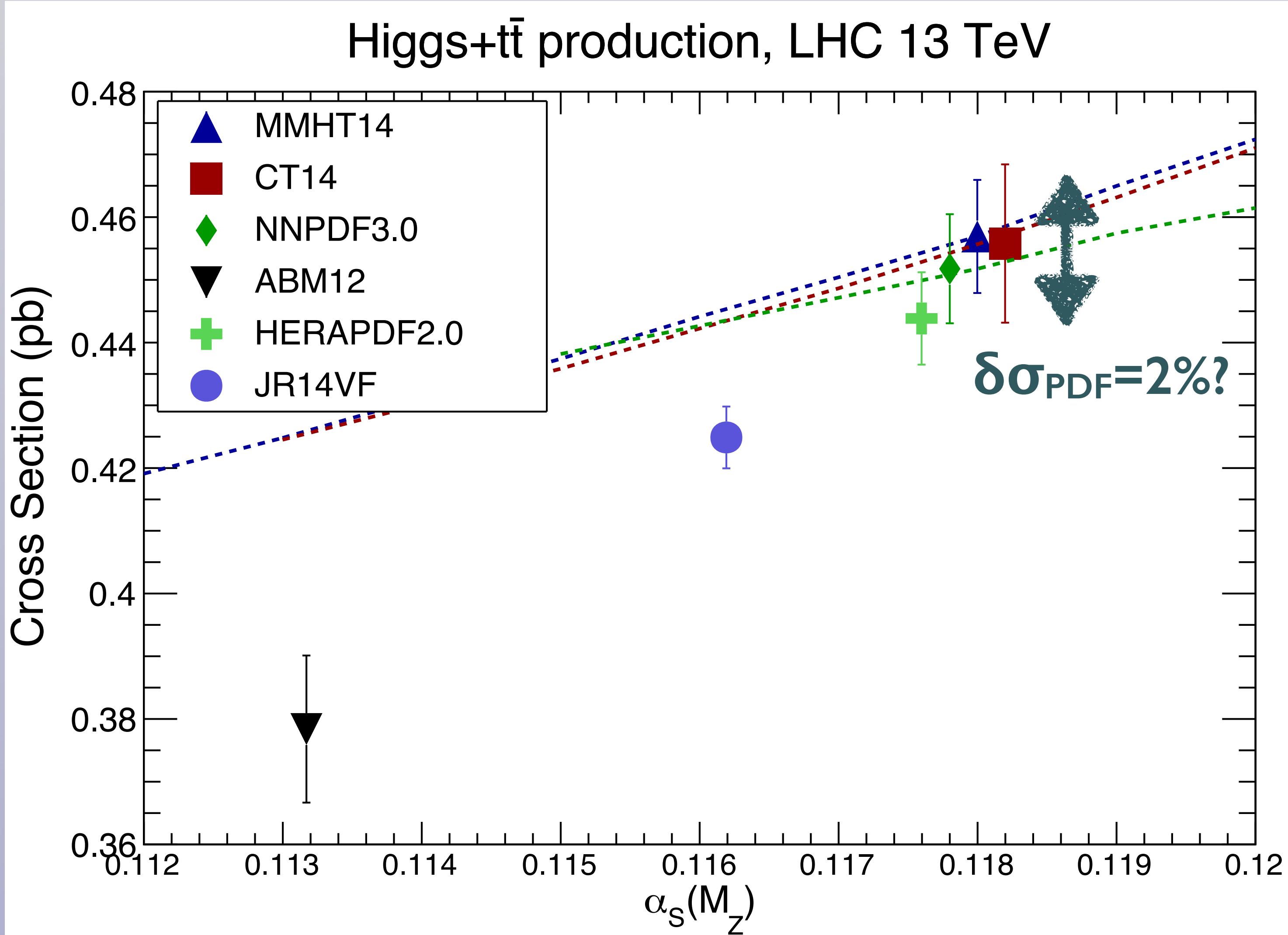
Why do we need a recommendation?

A well-defined PDF recommendation is mandatory for compare data with theory for Higgs physics



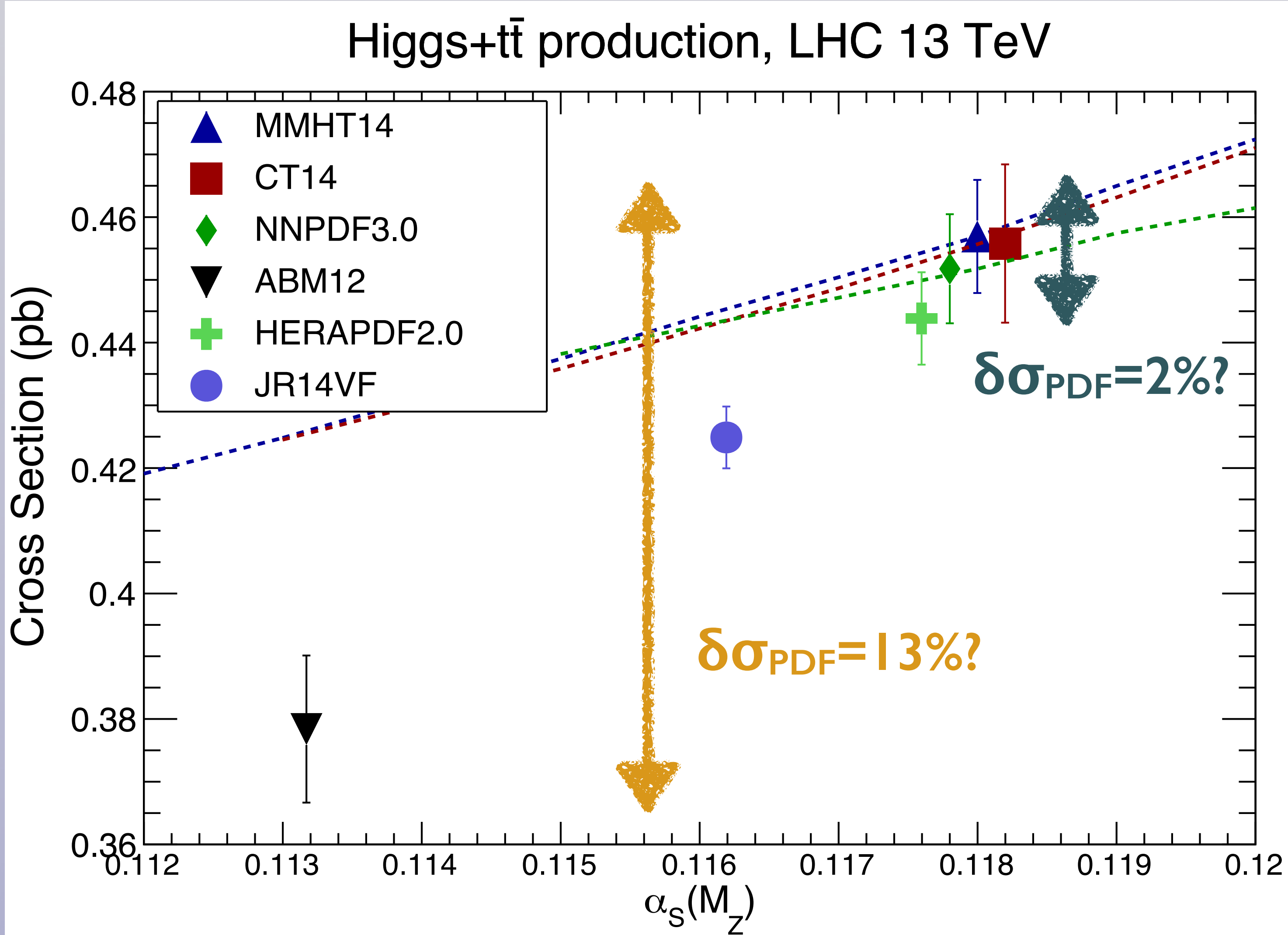
Why do we need a recommendation?

Depending on how the **total PDF uncertainty on a given cross-section** is defined, results can differ by large amount, dramatically **degrading the BSM discovery potential of the LHC**



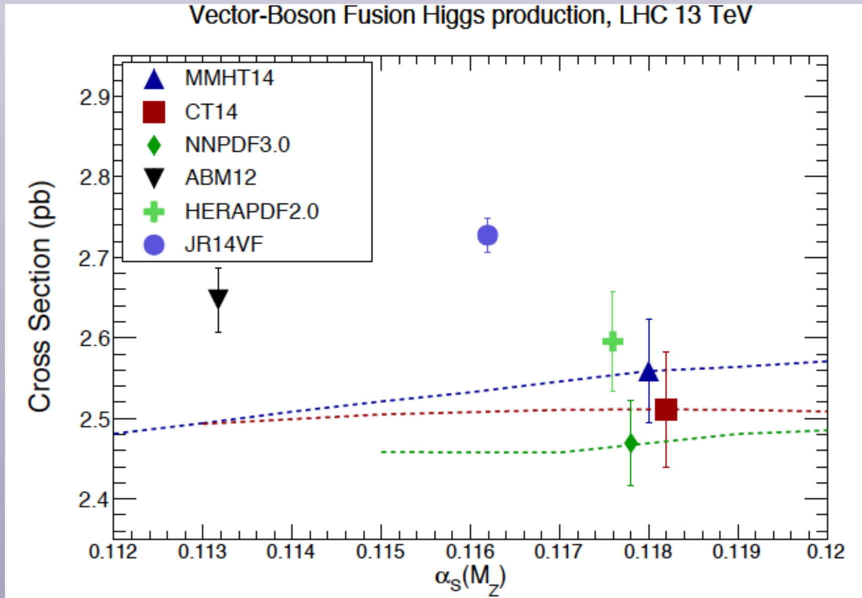
Why do we need a recommendation?

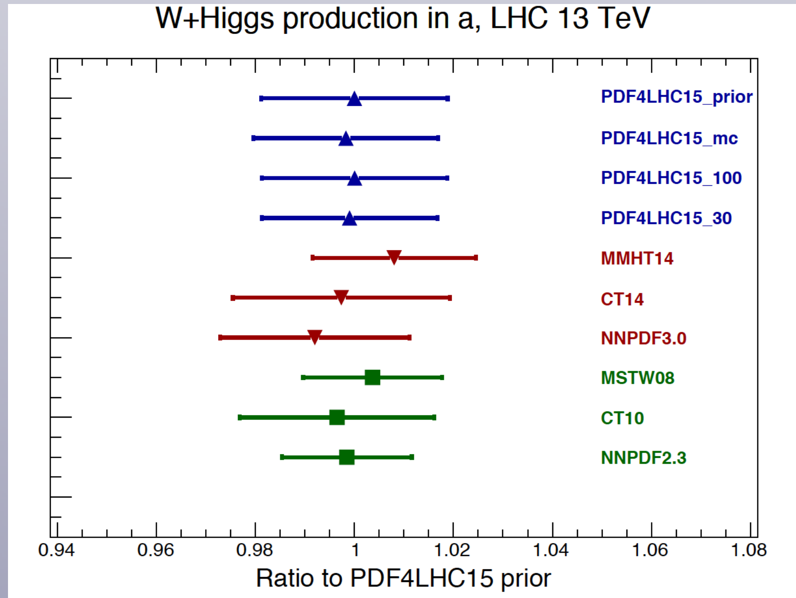
Depending on how the **total PDF uncertainty on a given cross-section** is defined, results can differ by large amount, dramatically **degrading the BSM discovery potential of the LHC**



Why do we need a recommendation?

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_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
CT14	0.09	-0.32	-0.44	-0.26	-0.03	-0.18
MMHT14	0.90	-0.22	-0.52	0.08	-0.18	-0.33
NNPDF3.0	0.90	-0.17	-0.21	0.18	0.52	0.49





A recommendation is required to establish a framework for the **consistent calculation of PDF uncertainties in different LHC processes** and **their correlation**, *i.e.*, in the case of Higgs production

Roadmap to PDF4LHC15

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

Hessian reductions

CMC-PDFs

MC2H

SM-PDFs

Combination tailored to **specific processes**

Meta-PDFs

Meta-H PDFs

Input to the PDF4LHC15 combination

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.

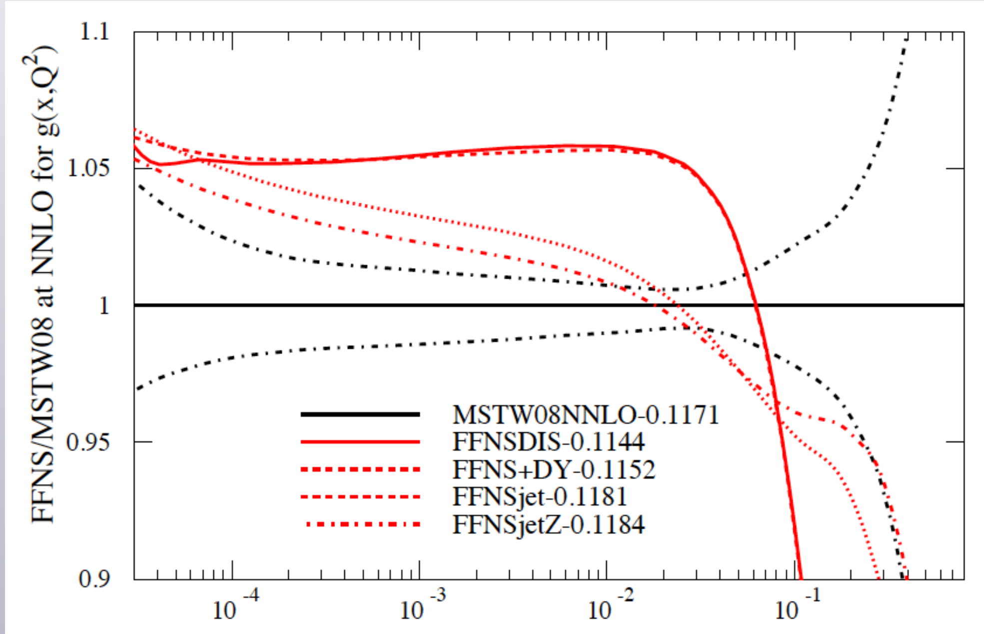
3. The central value of $\alpha_s(m_Z^2)$ *should be fixed at an agreed common value*, consistent with the PDG world-average [107]. This value is currently chosen to be $\alpha_s(m_Z^2) = 0.118$ at both NLO and NNLO.[2] For the computation of α_s uncertainties, two additional PDF members corresponding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

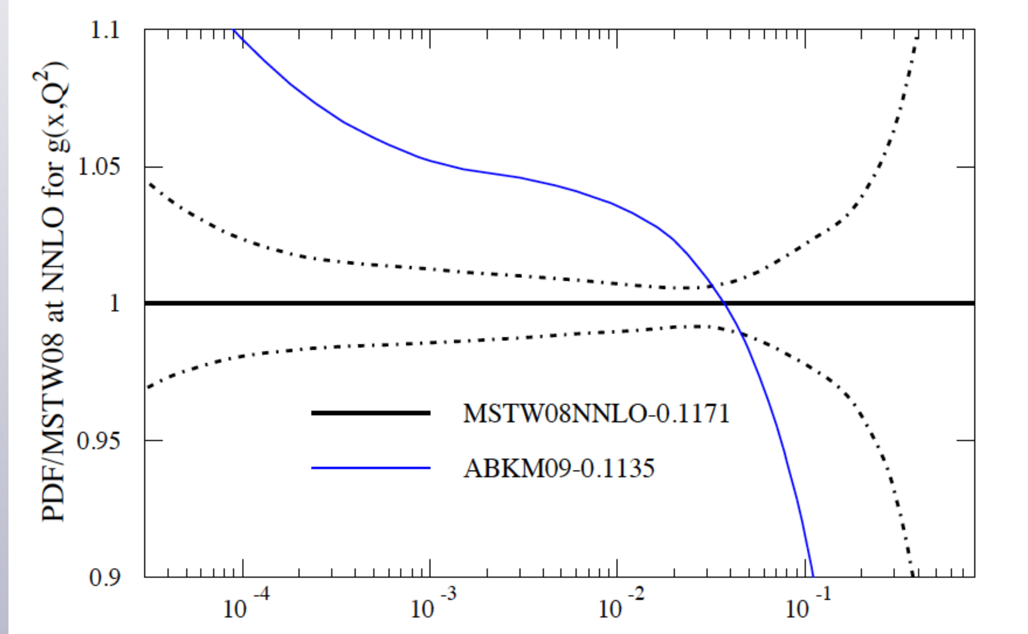
   The input values of m_c and m_b should be compatible with their world-average values; either pole or $\overline{\mathrm{MS}}$ masses are accepted.

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

FFN vs GM-VFN

Thorne, arXiv:1402.3536





- The use of **FFN heavy quark scheme** as opposed to **GM-VFN** leads to a **suppression (enhancement) of the large-x (small-x) gluon**, which is made more prominent if a **DIS-only** dataset is used
- The fit quality to the **high-Q2 inclusive HERA data** is worse in the FFN than in the GM-VFN. Explained by the slow perturbative convergence of the FFN scheme in certain *x* regimes (*ie* intermediate *x*)
- The **spread between FFN and GM-VFN fits** is not a genuine theoretical uncertainty
- Similar results obtained by NNPDF, *arXiv:1303.1189*

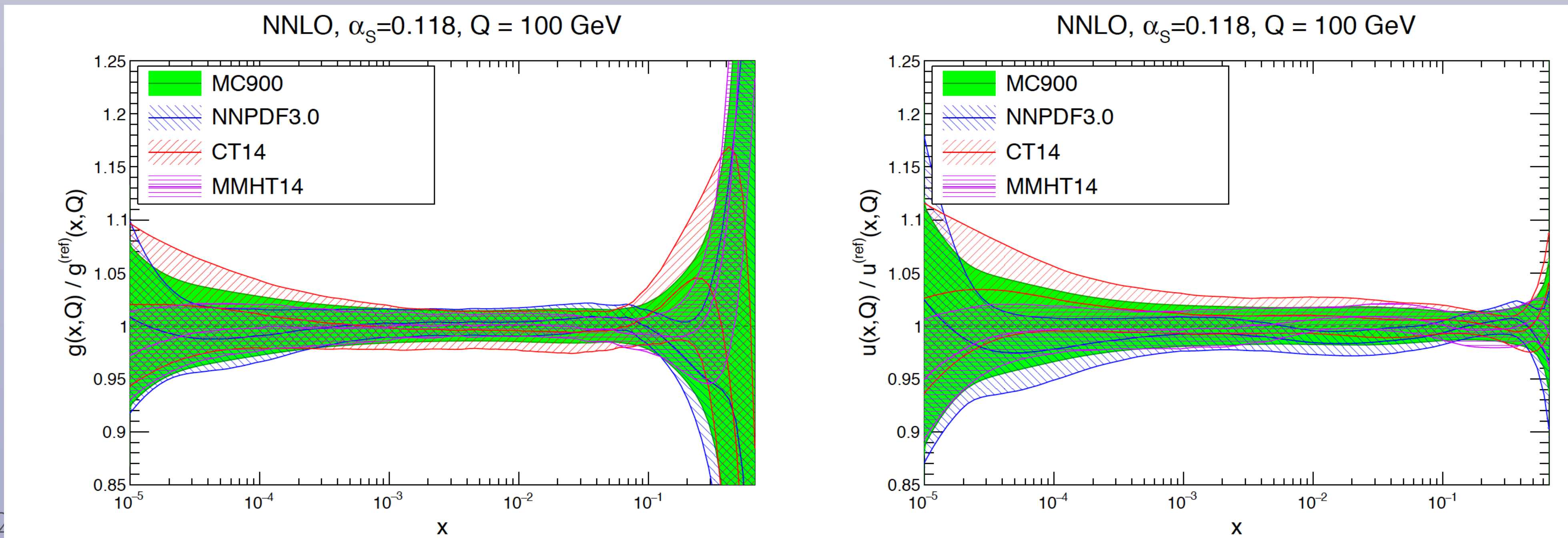
Monte Carlo combination

- First of all, select the **PDF sets that enter the combination**. Must be reasonably consistent among them for a meaningful combination. The 2015 combination is based on **CT14, MMHT14** and **NNPDF3.0**
- Transform the **Hessian PDF sets into their Monte Carlo representation:**

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

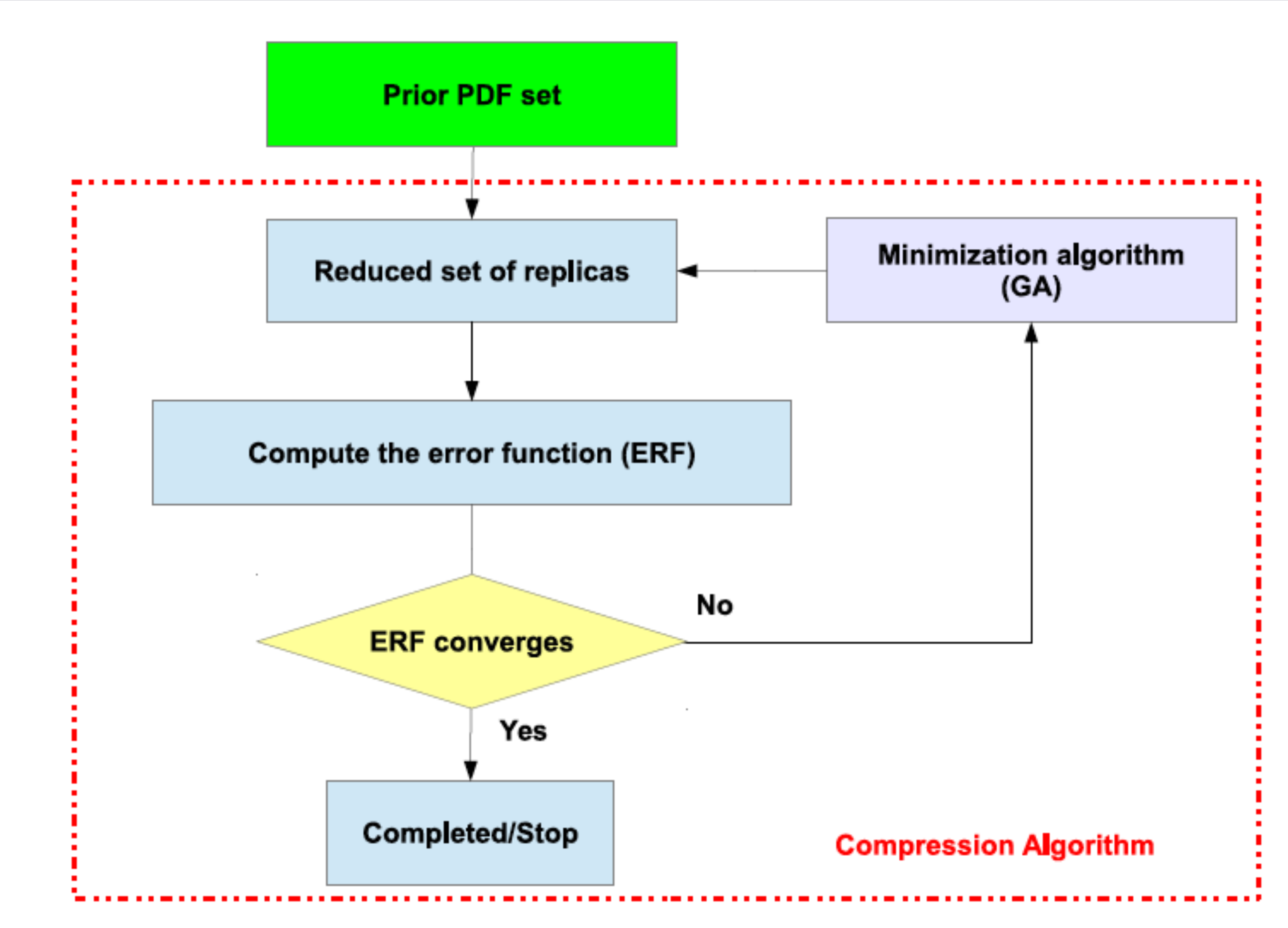
Watt and Thorne, arXiv:1205.4024
Forte, arXiv:1011.5247
Forte and Watt, arXiv:1301.6754

- 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 compared to the **original PDF4LHC envelope,** but now with the statistically correct treatment of outliers



Monte Carlo reduction: CMC-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



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

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

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

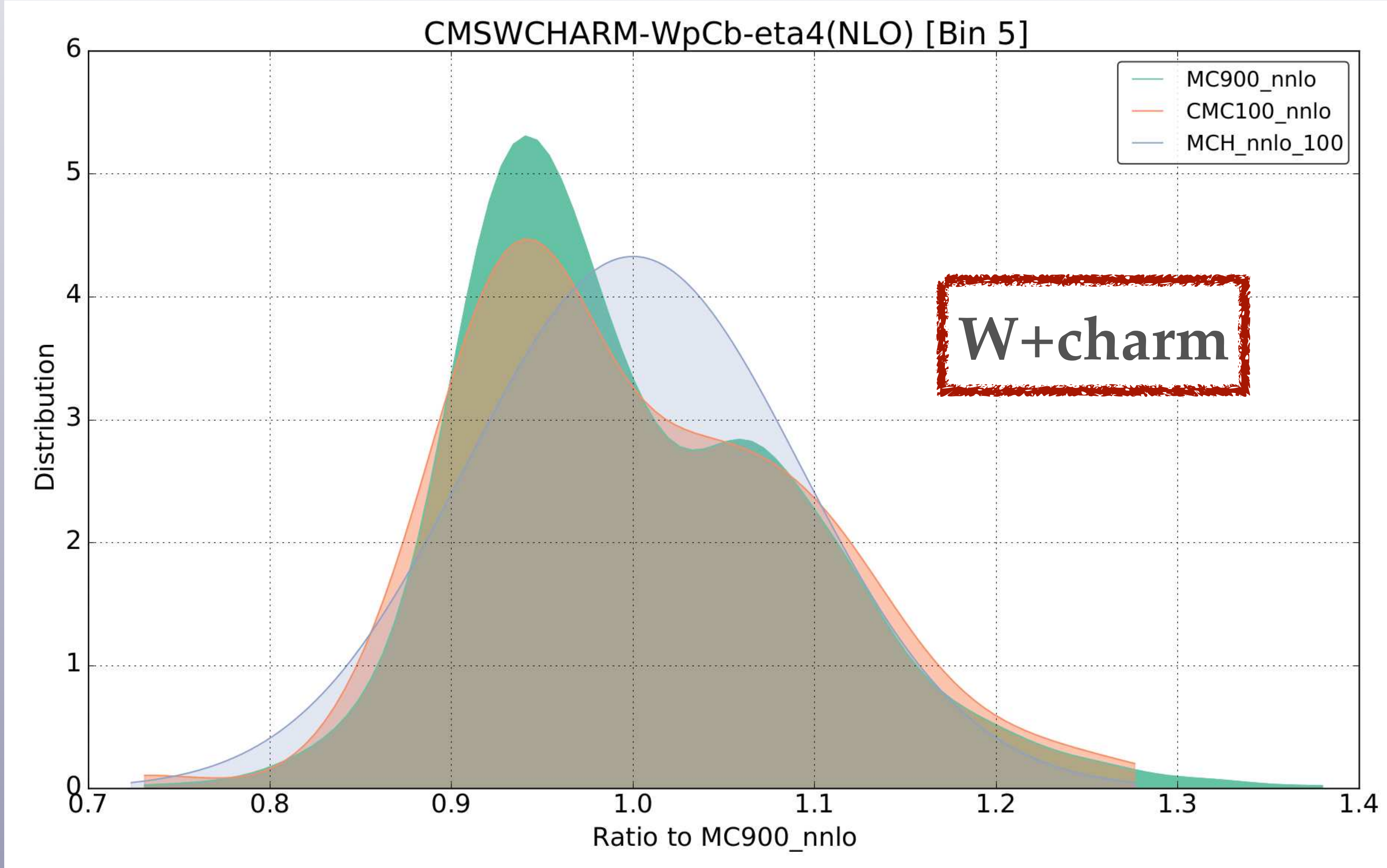
Similar for **variances, correlations, skewness, kurtosis** and the **Kolmogorov** distance

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

Carrazza, Latorre, J. R., Watt, arXiv:1504.06459

Non-Gaussian features in LHC cross-sections

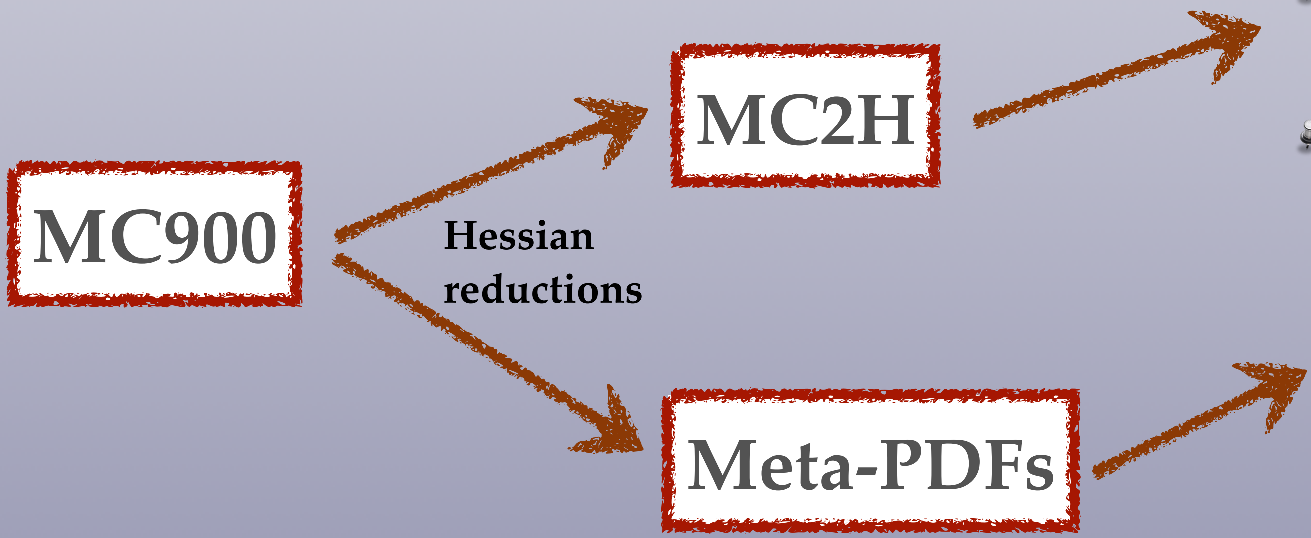
CMC100 reproduces correctly **non-Gaussian features in MC900**, such as **double-hump distribution**



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**

Carrazza, Forte, Kassabov, Latorre and J. R., arXiv:1505.06736

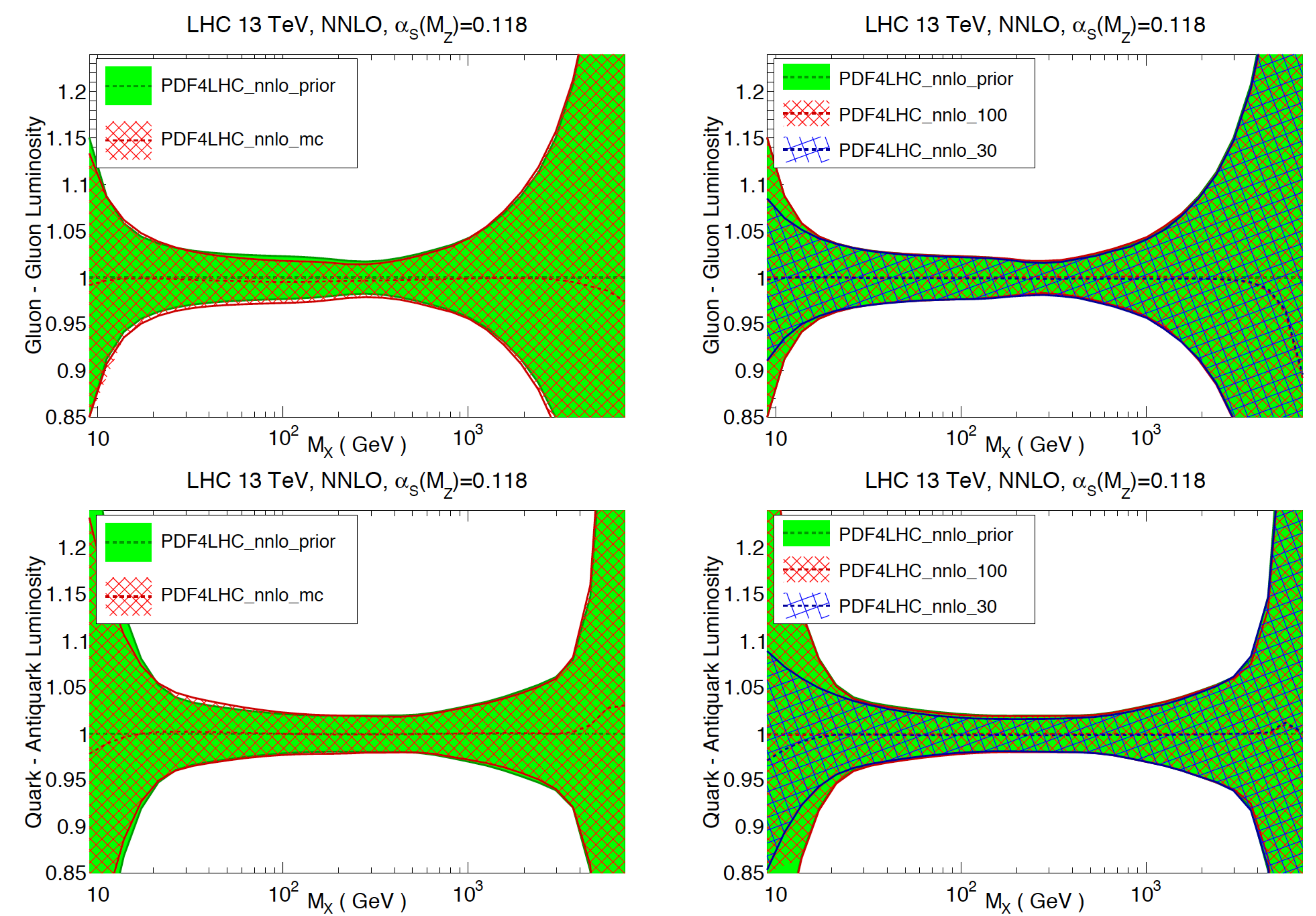


MC2H:
- **MC replicas** as expansion basis
- **Principal Component Analysis (PCA)** to reproduce PDF covariance matrix with arbitrary precision
- Standard linear algebra problem: **Singular Value Decomposition**

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

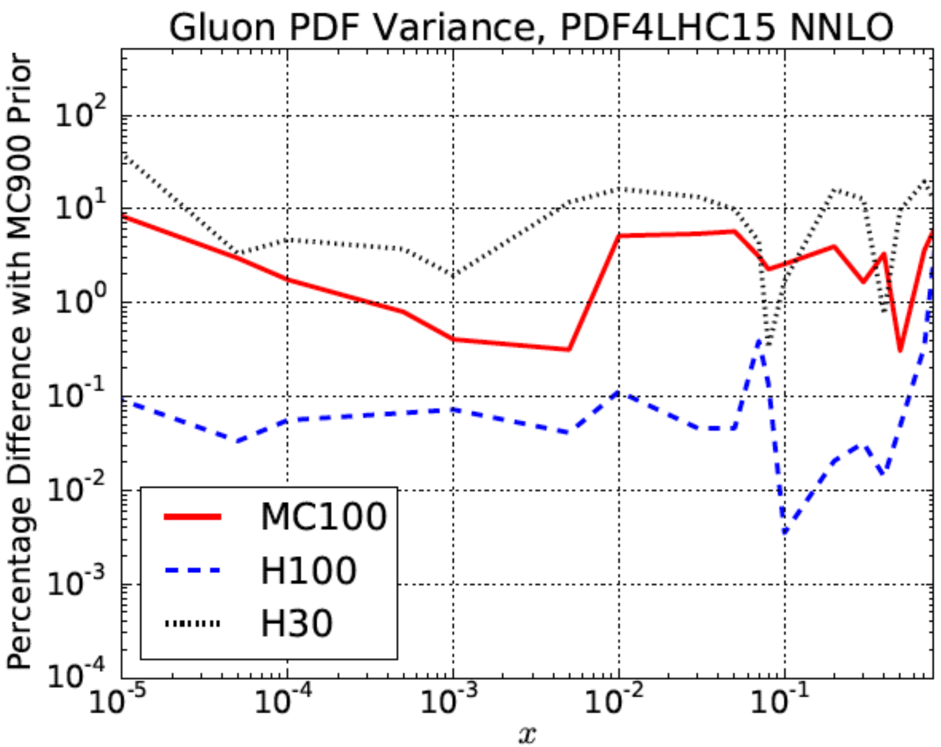
Gao and Nadolsky, arXiv:1401.0013

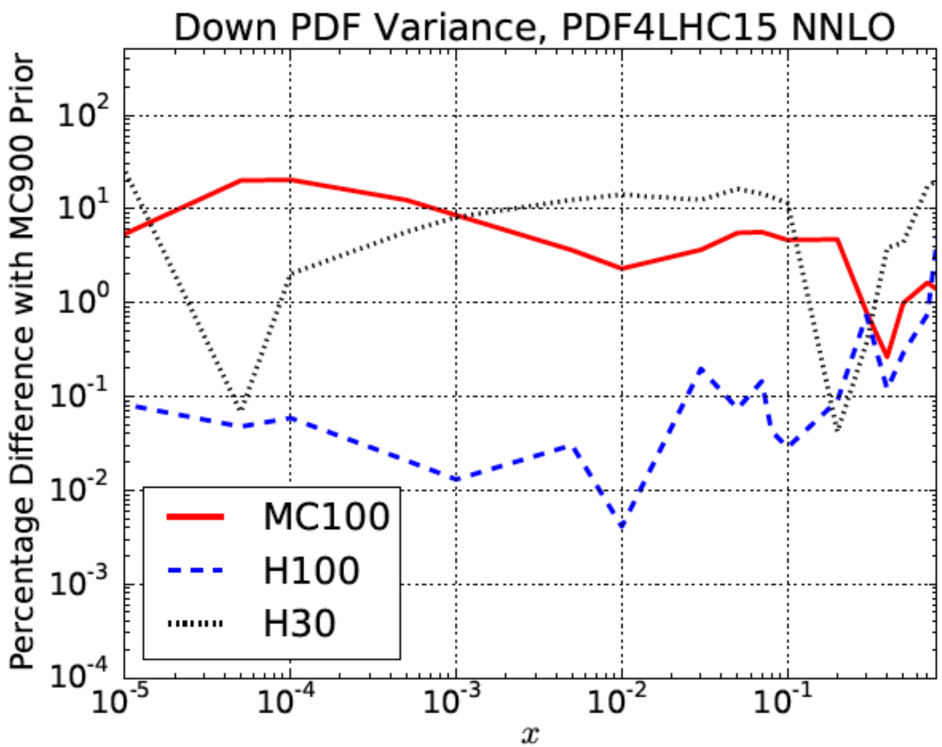
Comparison of reduction algorithms



Comparison of reduction algorithms

- The three reduction methods **successfully reproduce the variances** (PDF uncertainties) of the prior PDF4LHC15 combination
- The **highest accuracy** is achieved for the **PDF4LHC15_100**, with then similar performance of PDF4LHC_30 and PDF4LHC_mc





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**

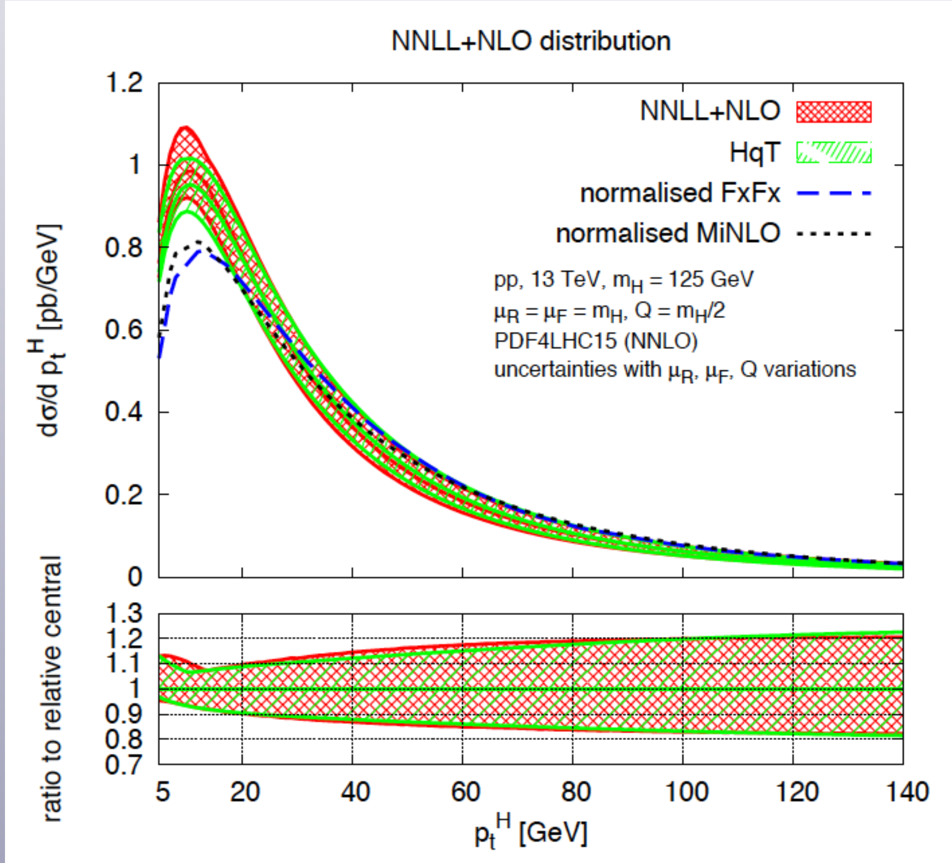
   Recommendation: Use the `PDF4LHC15_100` sets.

   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.

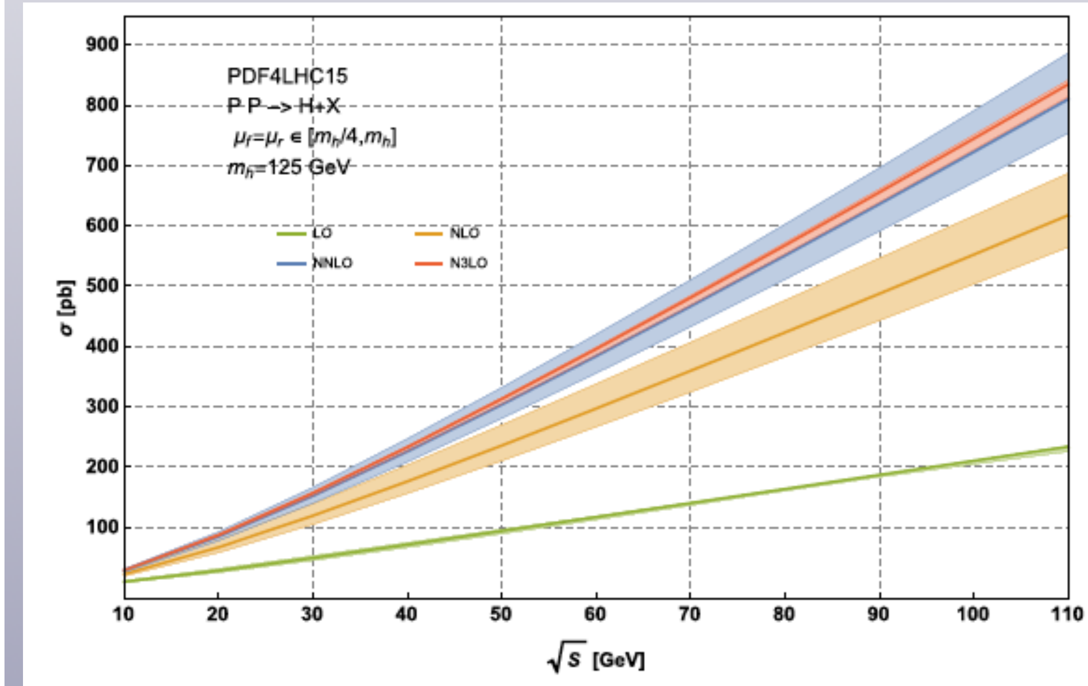
Towards HXSWG Yellow Report IV

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

NNLL+NLO Higgs p_T



N3LO ggH inclusive cross-section



The task of the report is to provide the **theory predictions for Higgs analysis at the LHC Run II**

An alternative proposal

A recent paper suggests **alternative procedures for PDF usage at the LHC:**

Recommendations for PDF usage in LHC predictions

arXiv:1603.08906, see also Ringaile's talk

A. Accardi[a,b], S. Alekhin[c,d], J. Blümlein[e], M.V. Garzelli[c], K. Lipka[f], W. Melnitchouk[b], S. Moch[c], R. Plačakytė[f], J.F. Owens[g], E. Reya[h], N. Sato[b], A. Vogt[i] and O. Zenaiev[f]

(i) Precision theory predictions

Recommendation: Use the individual PDF sets ABM12, CJ15, CT14, JR14, HERA-PDF2.0, MMHT14 and NNPDF3.0 (or as many as possible), together with the respective uncertainties for the chosen PDF set, the strong coupling $\alpha_s(M_Z)$ and the heavy quark masses m_c, m_b and m_t.

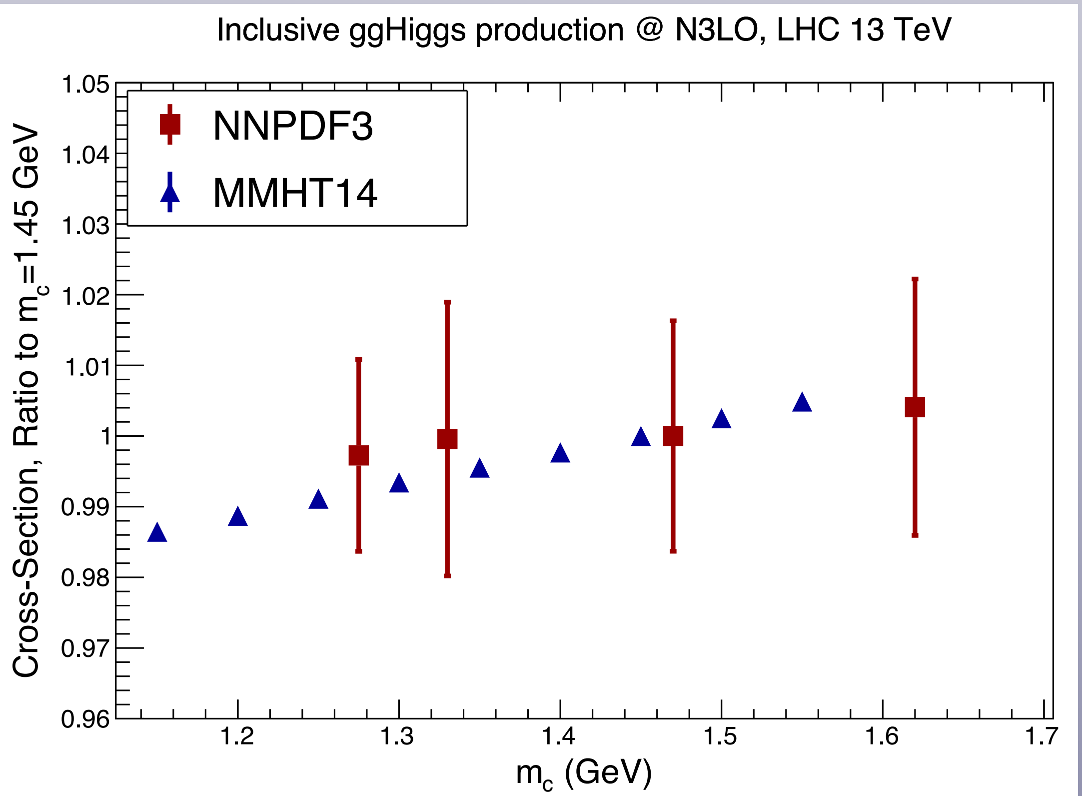
- Do we have any reason to **disregard experimental data**, including LHC constraints on PDFs?
- Should we mix **state-of-the art PDF sets** with **PDFs obtained several years ago**?
- Is it justified to believe that the **PDG average of $\alpha_S(M_Z)$ is not correct**?

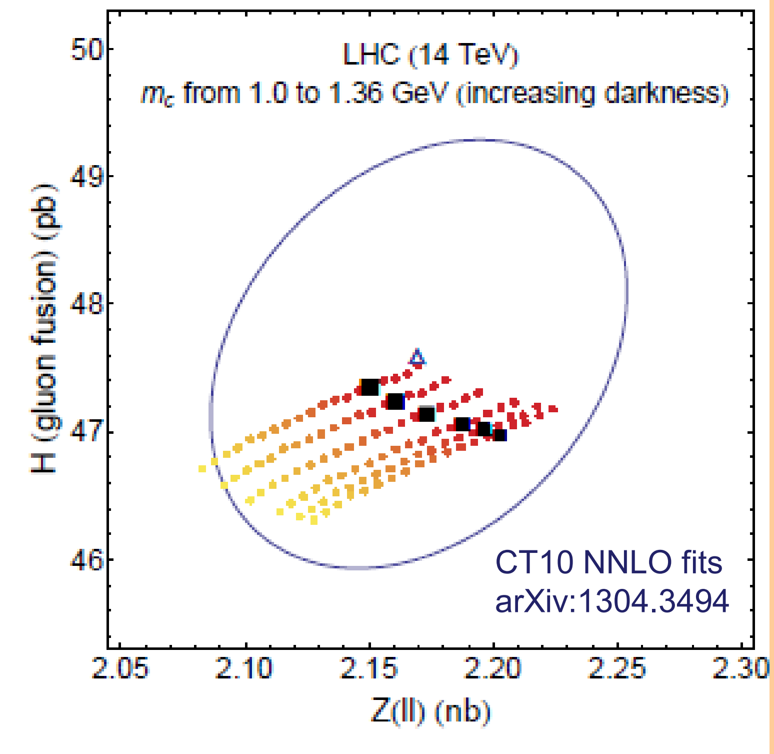
This would tend to **redundant calculations with many PDF sets** and to **blow up PDF uncertainties in precision physics and BSM searches**

Is the charm mass a tuning parameter?

compatible with the world average of the PDG [39]. Thus, in some PDF fits, the numerical value of the charm-quark mass effectively takes over the role of a "tuning" parameter for the Higgs cross section.

- Are global PDF fits really so **sensitive to the value of the charm mass?**
- When **m_c is varied in a wide range**, changes in NNPDF, MMHT and CT Higgs cross-sections are **within their PDF uncertainties**
- The GM-VFN schemes used by the three groups **extensively benchmarked** up to NNLO

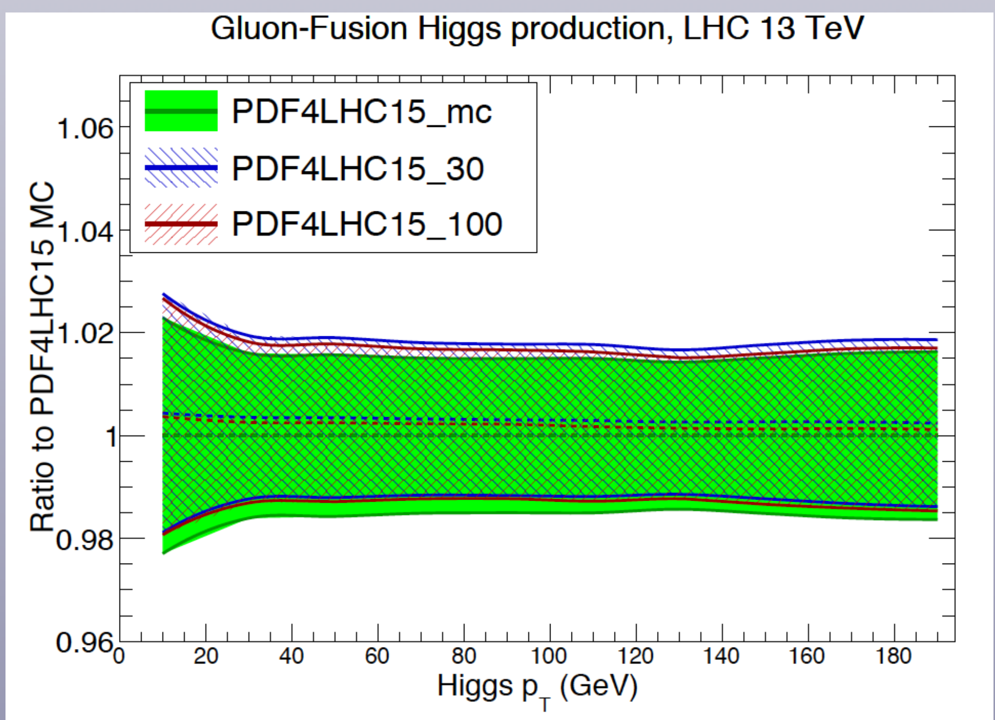




σ_{tot} for $m_c(m_c) = 1 - 1.36$ GeV and matching parameter λ varied independently, $Q_0 = 1$ GeV. Black boxes are for $m_c(m_c) = 1.28$ GeV (close to world average), for the explored λ. The error ellipse is for nominal 90% C.L. @ $Q_0 = 1.3$ GeV.

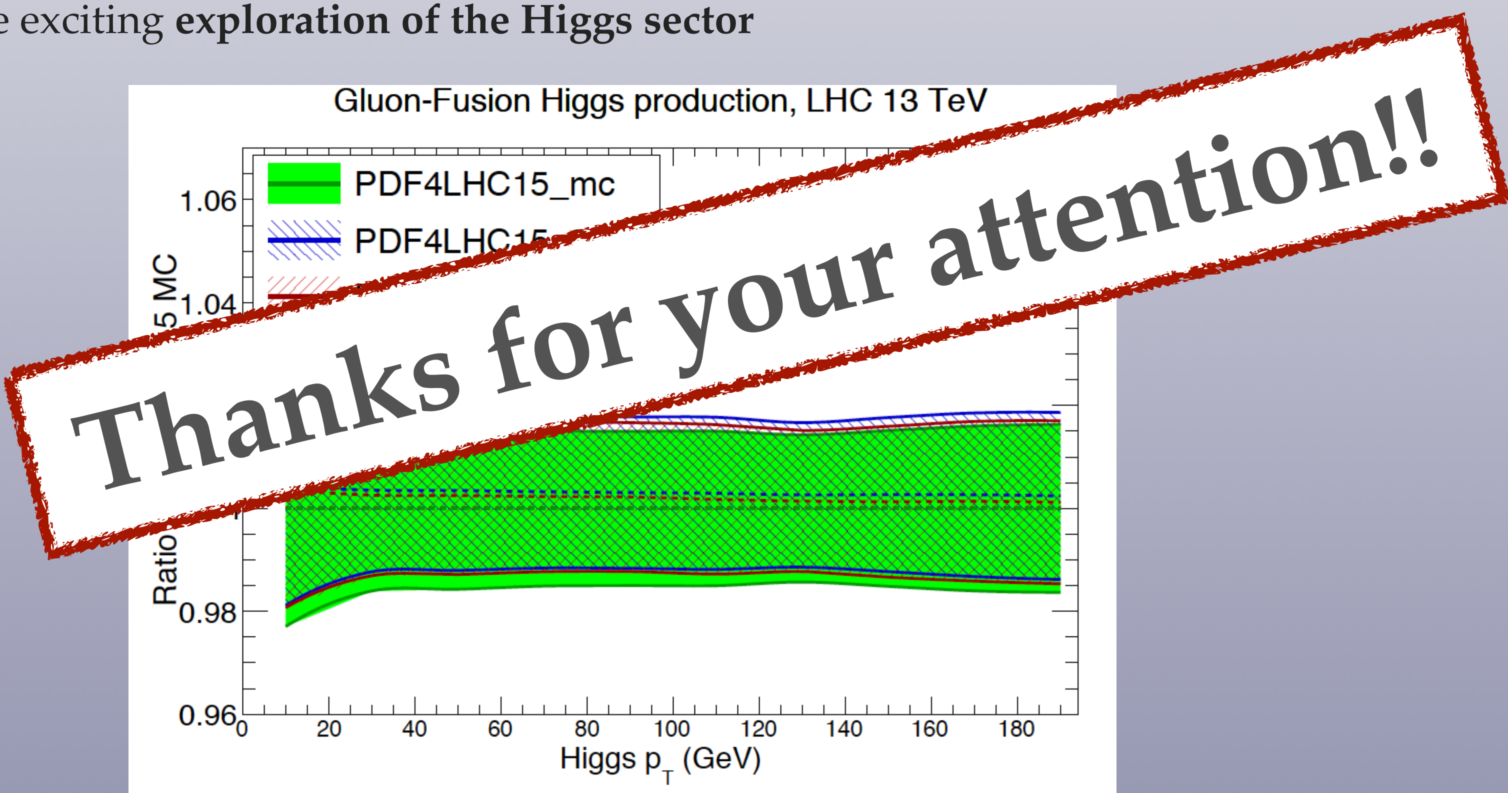
The PDF4LHC15 recommendations

- The extensive discussions within the **PDF4LHC working group** during the last two years have led to the **2015 recommendations for PDF usage at Run II**
- These are based on a set of tools to **combine PDF sets** and reduce the resulting number of eigenvectors/replicas in order to **streamline the usage of PDF sets** by the LHC experiments
- While there is room for further optimisation, the **general strategy is robust and flexible enough** to accommodate any future PDF4LHC combination
- The **PDF4LHC15 recommendations** are intended for many backbone LHC analyses at Run II, including the exciting **exploration of the Higgs sector**



The PDF4LHC15 recommendations

- The extensive discussions within the **PDF4LHC working group** during the last two years have led to the **2015 recommendations for PDF usage at Run II**
- These are based on a set of tools to **combine PDF sets** and reduce the resulting number of eigenvectors/replicas in order to **streamline the usage of PDF sets** by the LHC experiments
- While there is room for further optimisation, the **general strategy is robust and flexible enough** to accommodate any future PDF4LHC combination
- The **PDF4LHC15 recommendations** are intended for many backbone LHC analyses at Run II, including the exciting **exploration of the Higgs sector**



Additional Material

SM-PDFs

Carrazza, Forte, Kassabov and J.R, arXiv:1602.00005

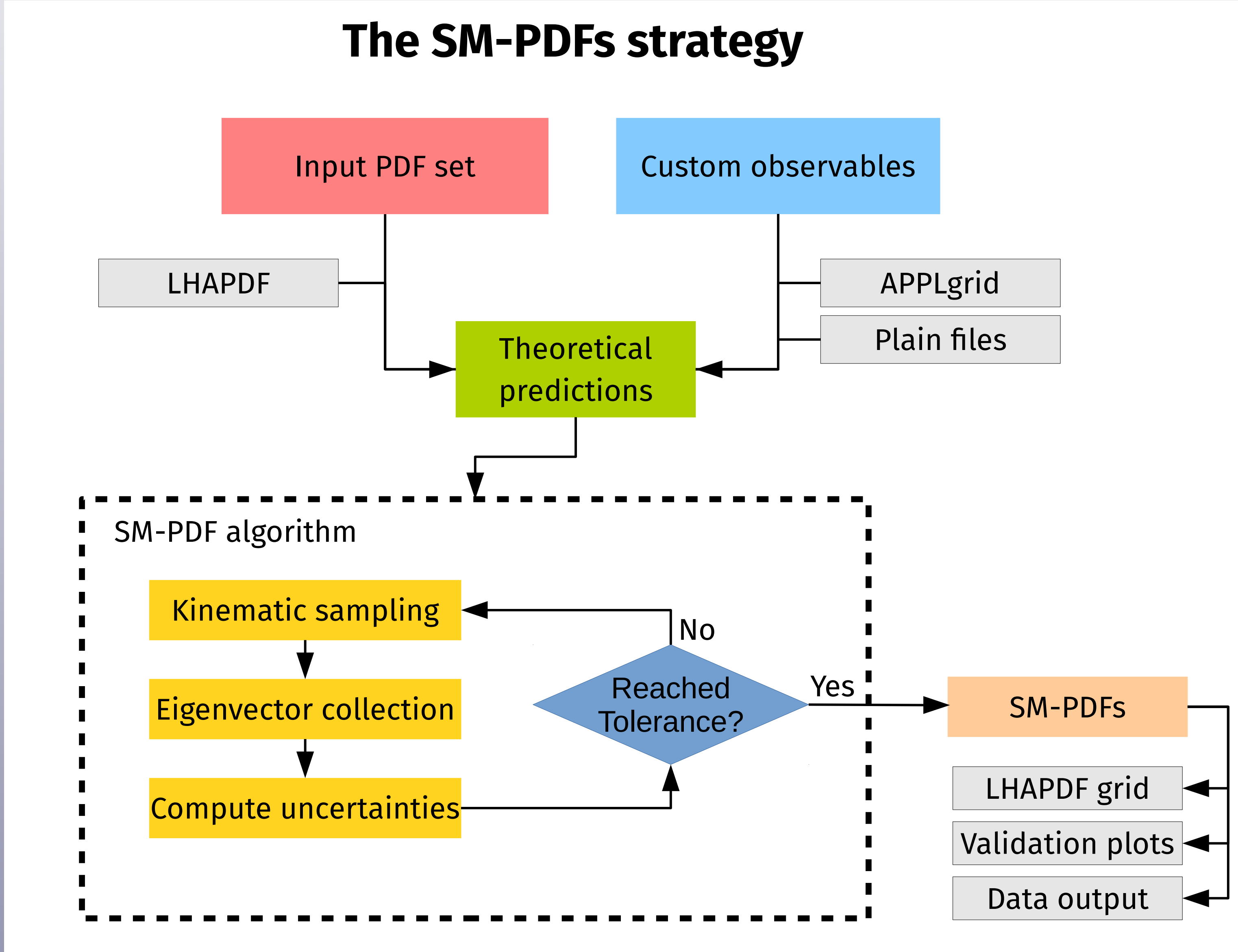
- **PDF sets specialised to a particular set of processes** might be useful to reduce the CPU burden of theory calculations, and facilitate the **treatment of PDF uncertainties and PDF-induced correlations**
- One clear example are **Higgs analysis in the HXS WG**, but similar **Specialized Minimal PDFs (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

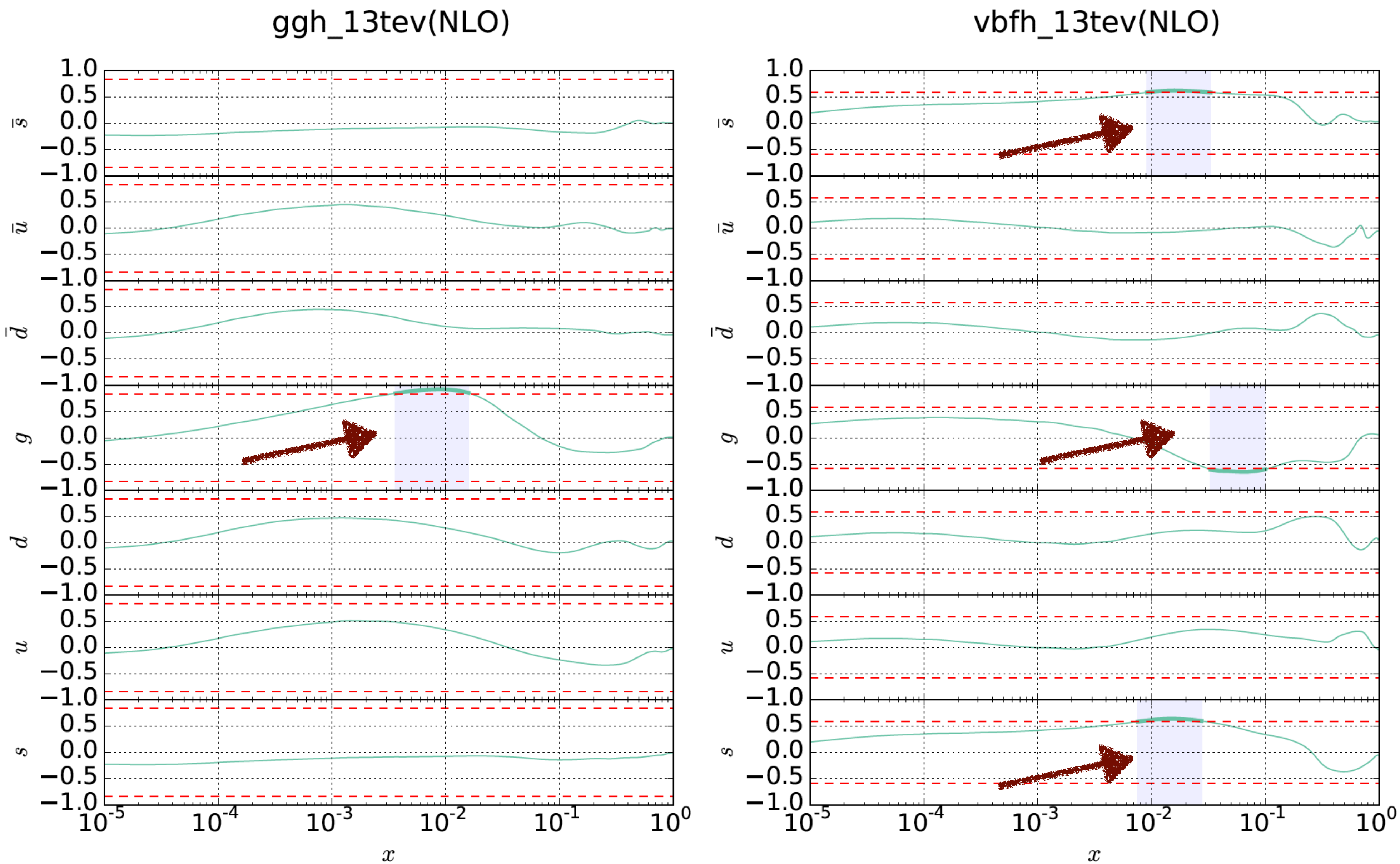
General strategy

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**

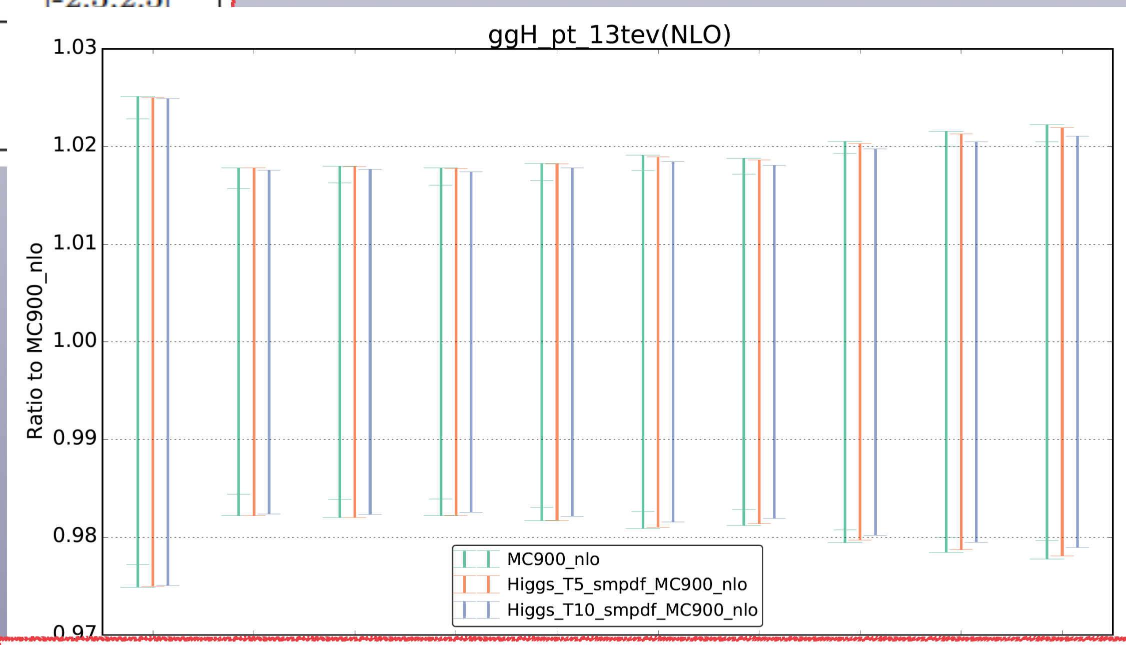


SM-PDFs for Higgs physics

- 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**



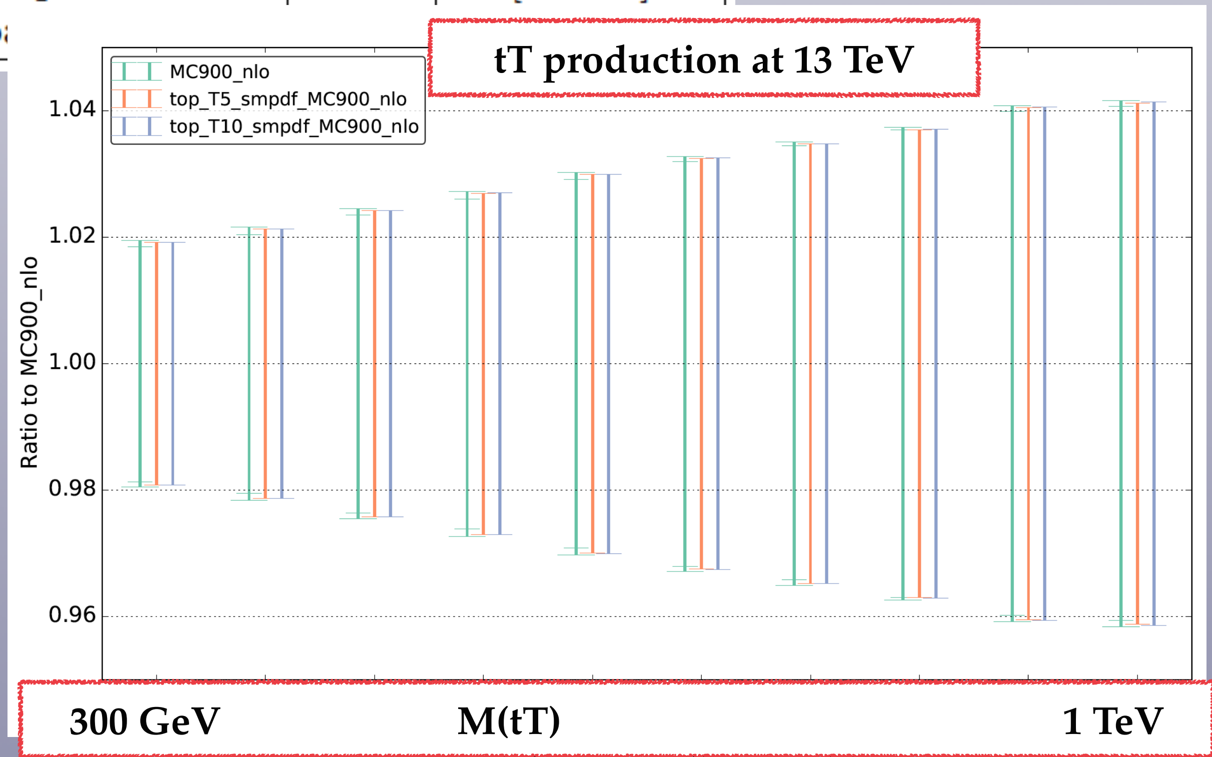
10 GeV **Higgs pT** **200 GeV**

SM-PDFs for Top physics

- Then construct **Specialised 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

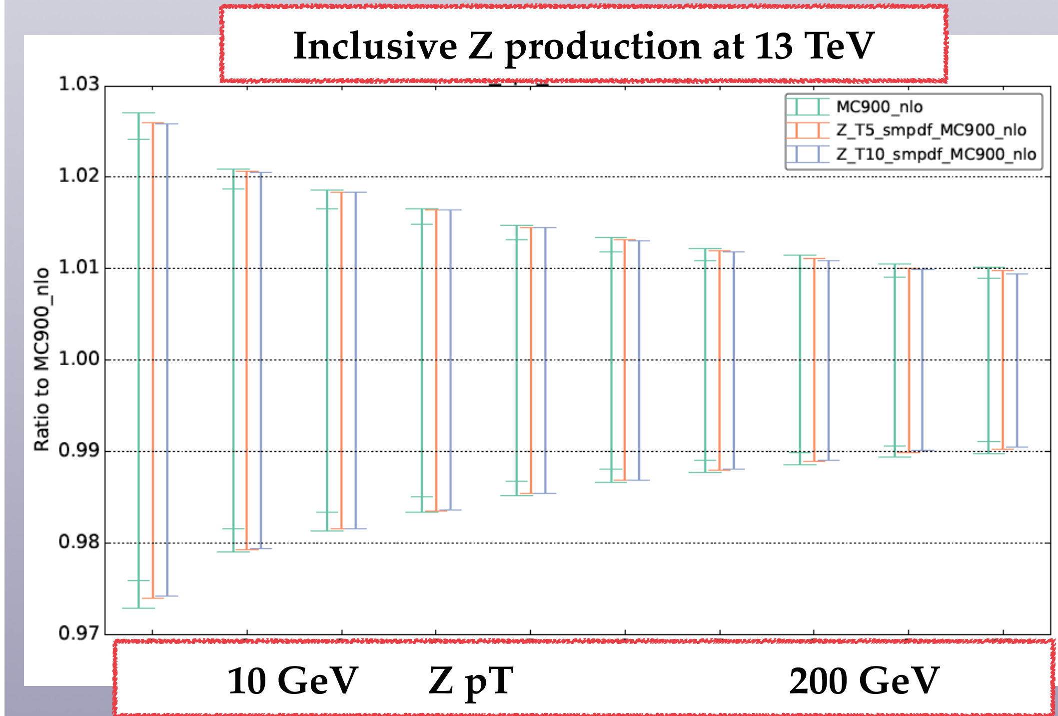
- Then construct S**pecialised Minimal PDFs for W mass determination studies quark physics. Should be useful for ongoing ATLAS and CMS analysis**
- **Included the following set of distributions:**

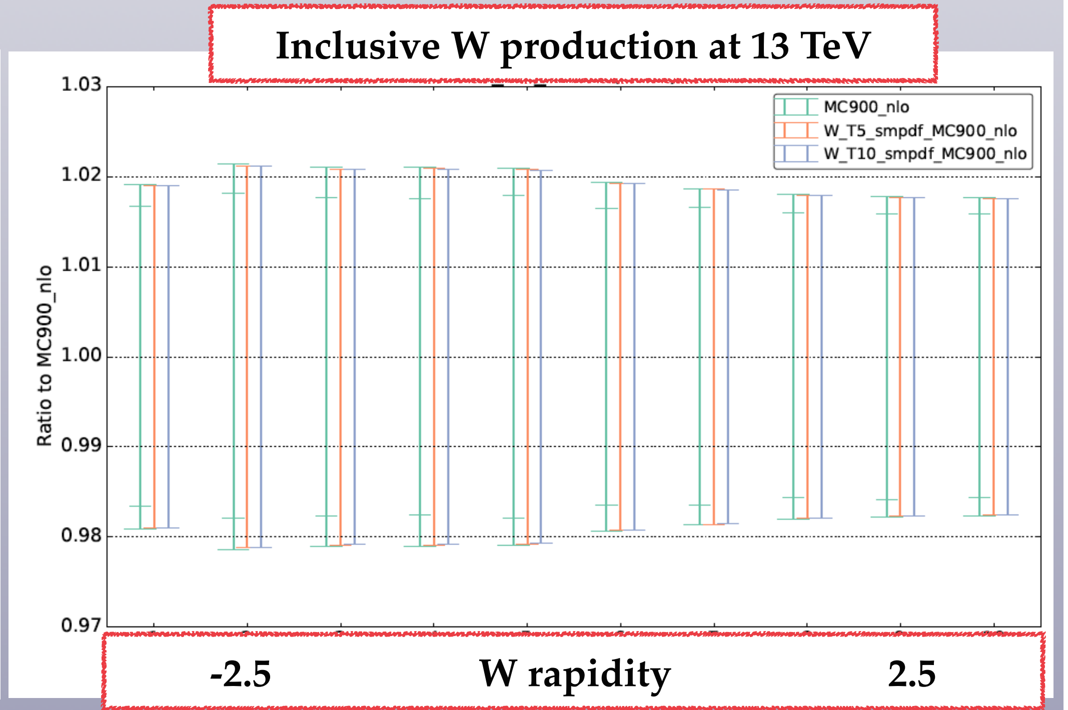
Process	APPLgrid	$\sqrt{s}$ (TeV)	$N_{\rm bins}$	Range	Cuts
Z	z_13tev	13	1	-	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	z_lmpt_13tev	13	10	[0,200] GeV	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	z_lmy_13tev	13	10	[-2.5,2.5]	$p_T(l) \geq 10$ GeV, $\lvert\eta^l\rvert \leq 2.5$
	z_lppt_13tev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

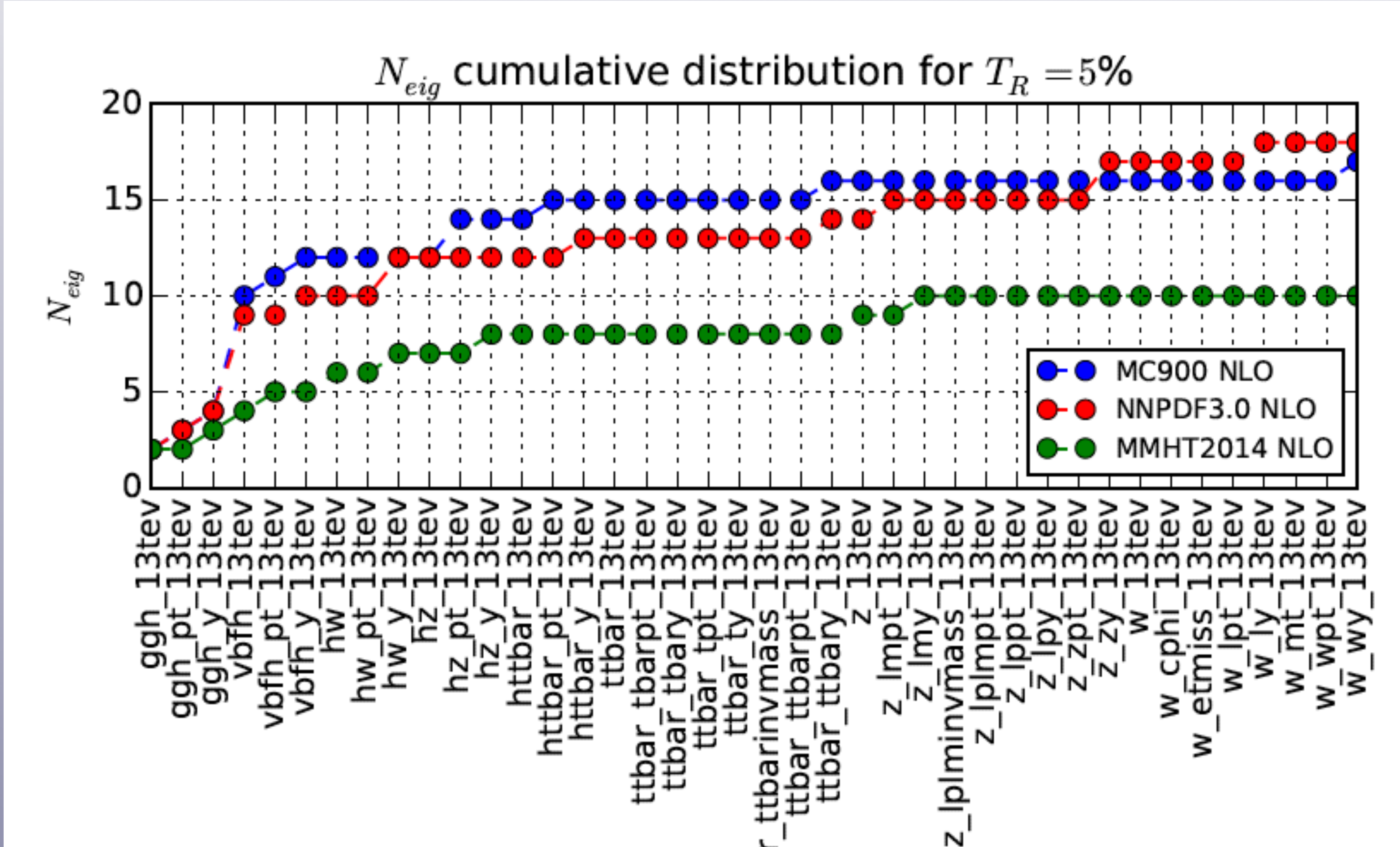
- Then construct **Specialised Minimal PDFs for W mass determination studies. Should be useful for ongoing ATLAS and CMS analysis**
- A somewhat **larger number of eigenvectors selected** here, since W,Z production involve all PDF flavours in a wide range of x),
- Even in this case **with only 8 to 14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



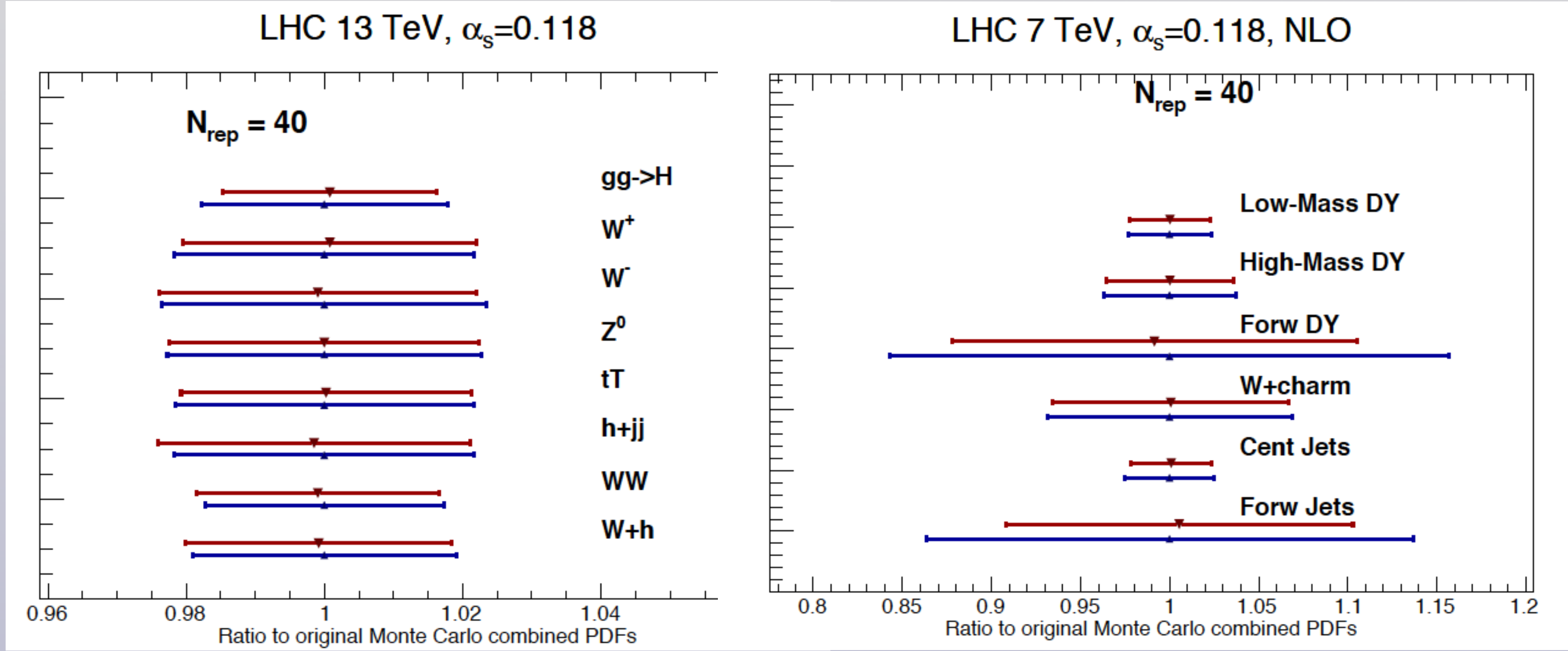


Climbing up the ladder

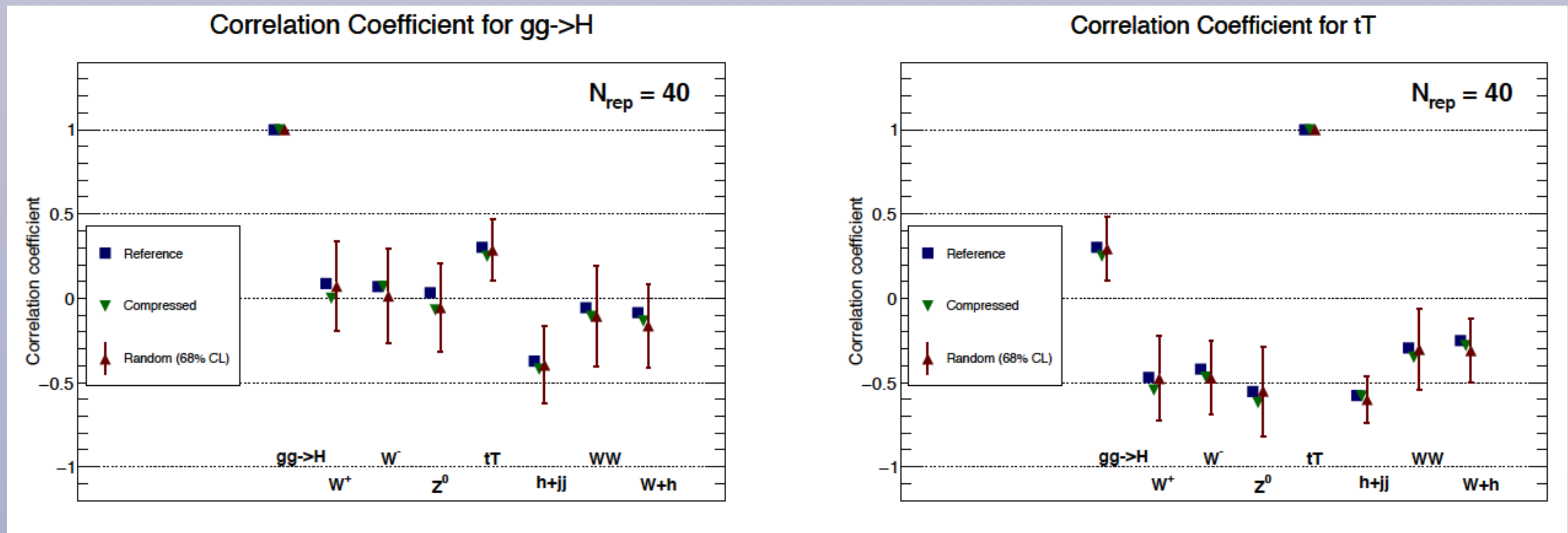
- The **number of eigenvectors** required **increases monotonically** as more and more input cross-sections are included in the procedure
- In the limit where all possible cross-sections are included, the **original MC2H set is reproduced**



LHC Phenomenology

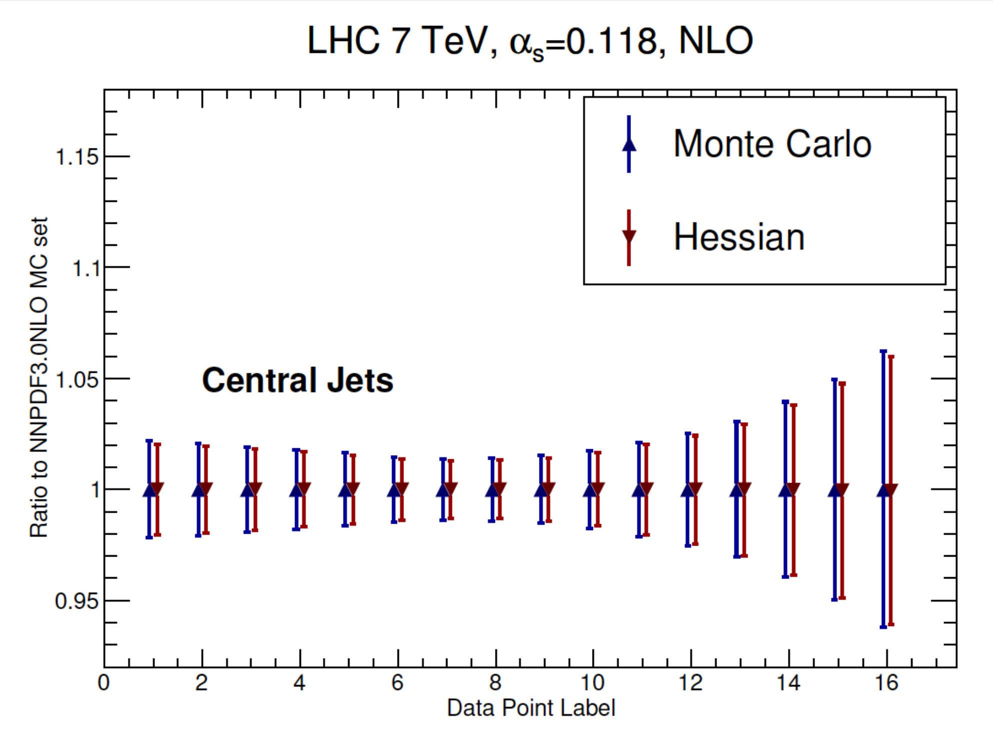


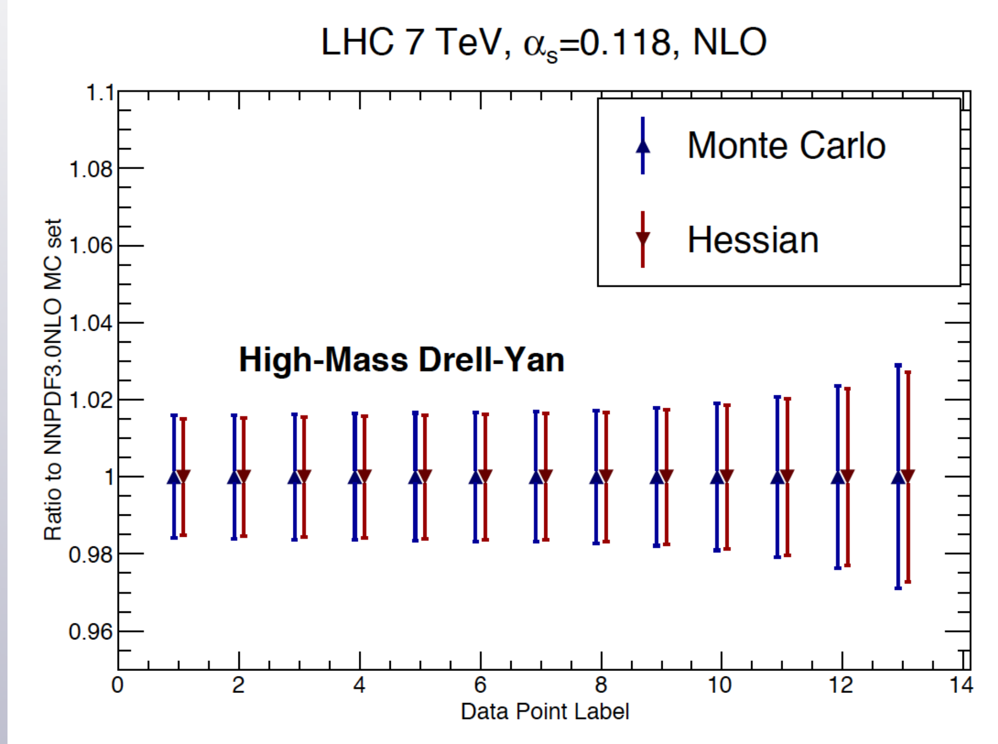
As expected from good agreement at the PDF level, CMC-PDFs also validated for **LHC inclusive cross-sections and diff distributions,** including **correlations**

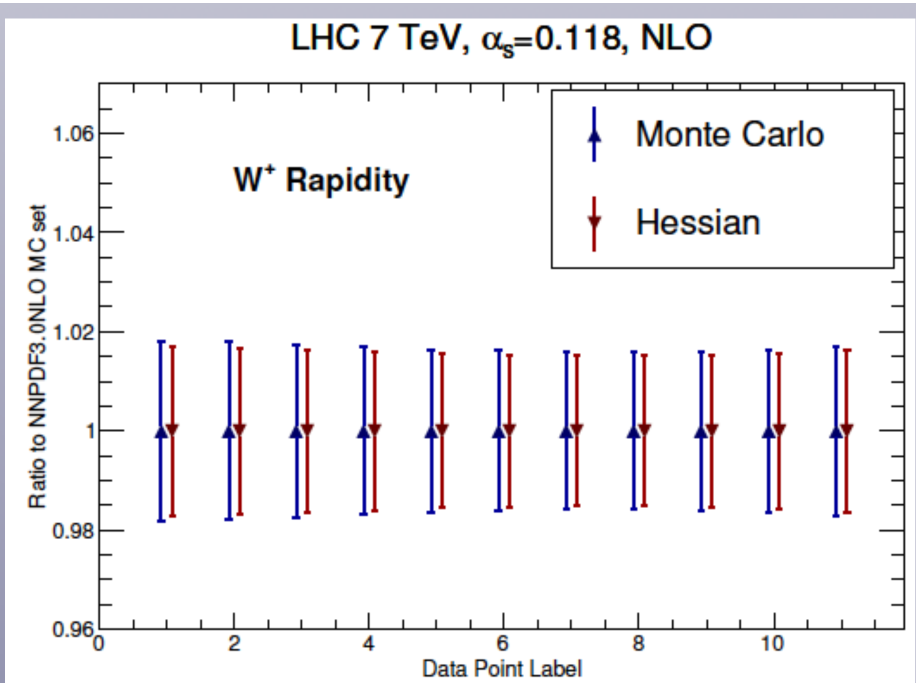


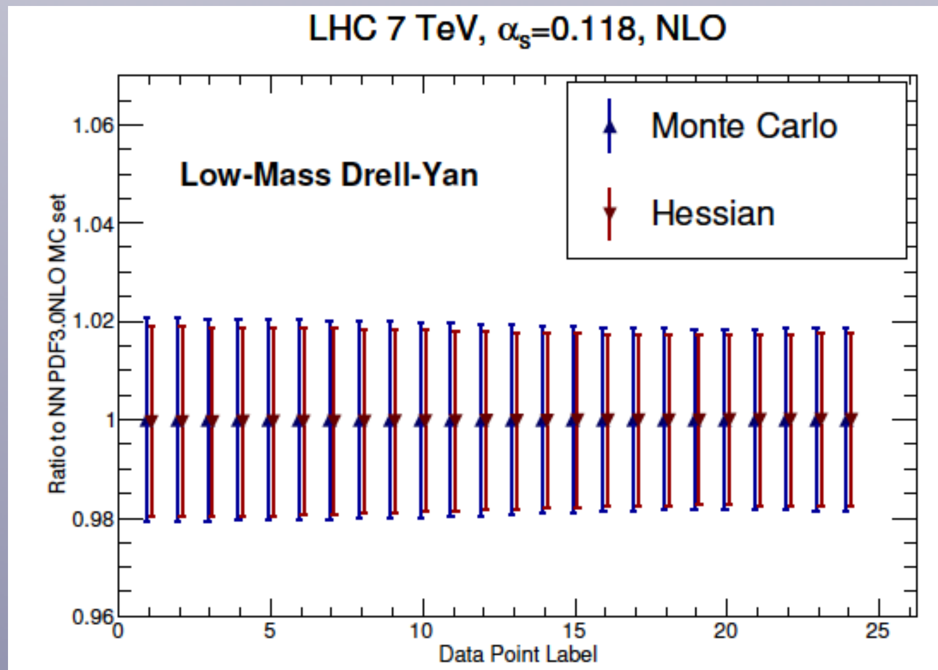
Agreement between MC900 and CMC significantly better than **random selection of same number of replicas**

MC2H LHC phenomenology



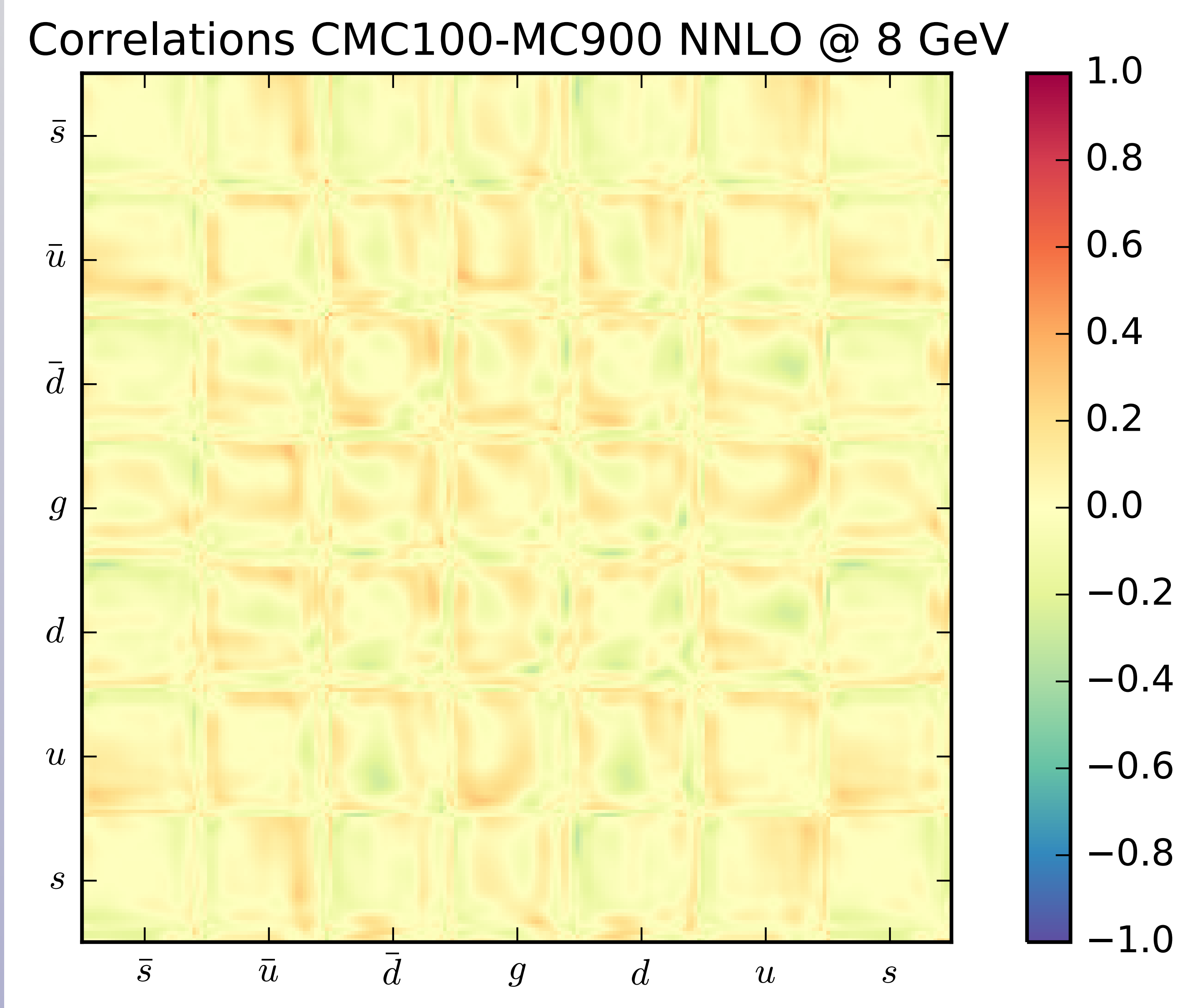






CMC-PDFs

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

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

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

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

$$X = U S V^t$$

Matrix with orthogonal eigenvectors of covariance matrix (U)

Matrix constructed from singular values of X (S)

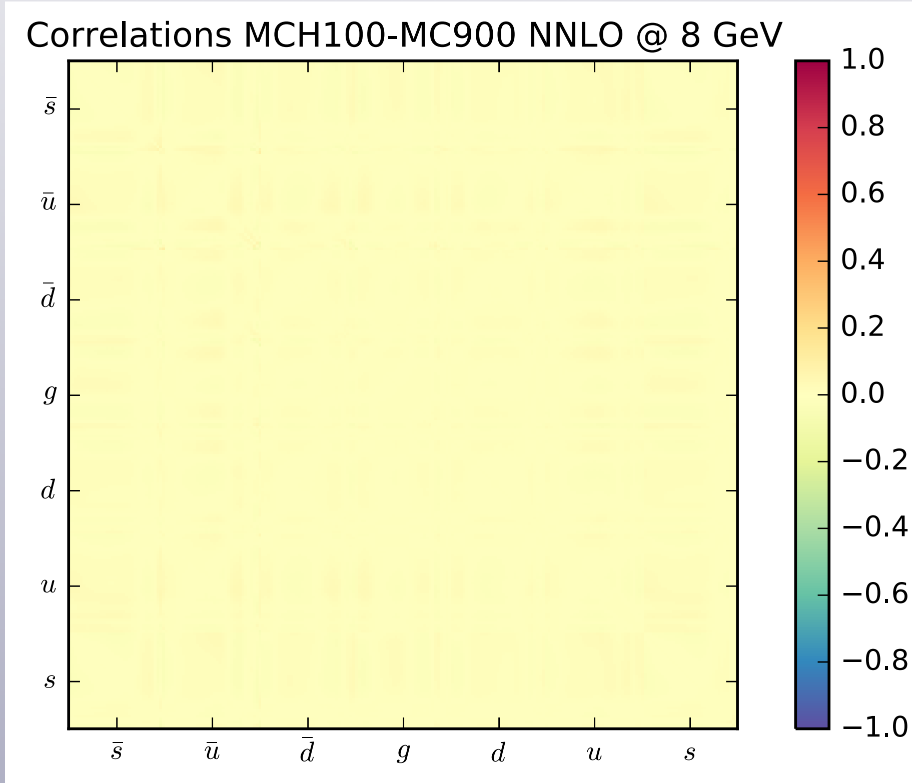
Orthogonal matrix of coefficients (V^t)

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, $\mathbf{N_{eig}=N_{pdf}\,N_x}$, but with SVD methods possible to **optimise the way information is stored** by **reducing $\mathbf{N_{eig}}$** while minimising information loss, ordering eigenvectors.
- Use **Principal Value Decomposition** to reduce $\mathbf{N_{eig}}$ 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 $\mathbf{N_{eig} = 100}$
- Basic assumption: reproduce the entire PDF covariance matrix, even in regions where the **underlying probability distribution is non-Gaussian**

Correlation matrix

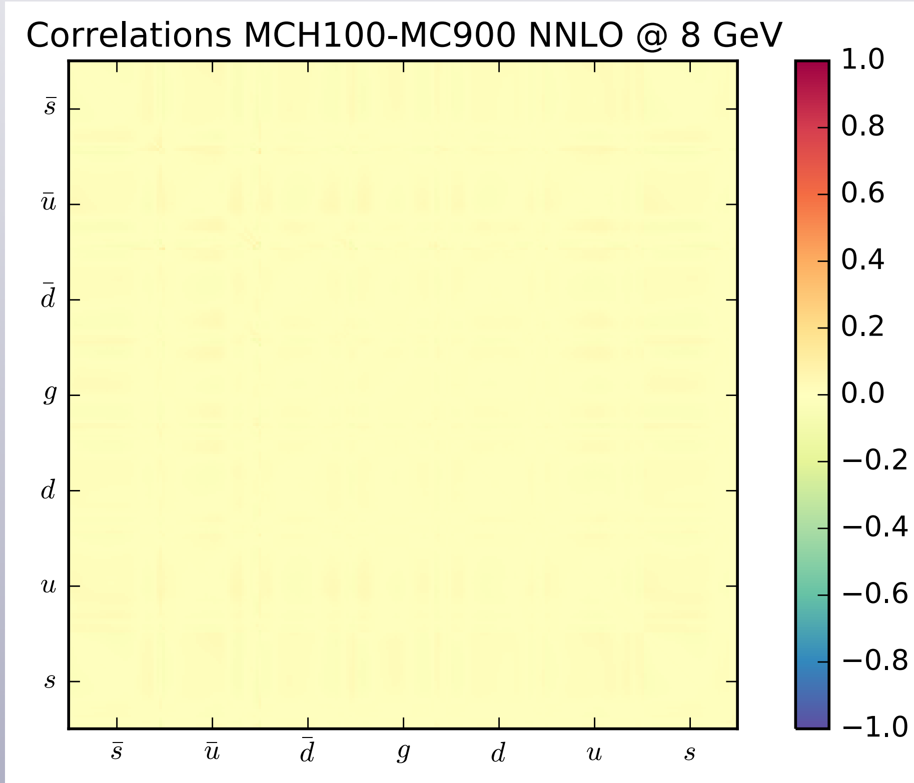
Since the method is essentially exact, one expects that the **PDF correlation matrix** is **perfectly reproduced** (modulo the eigenvector reduction step)

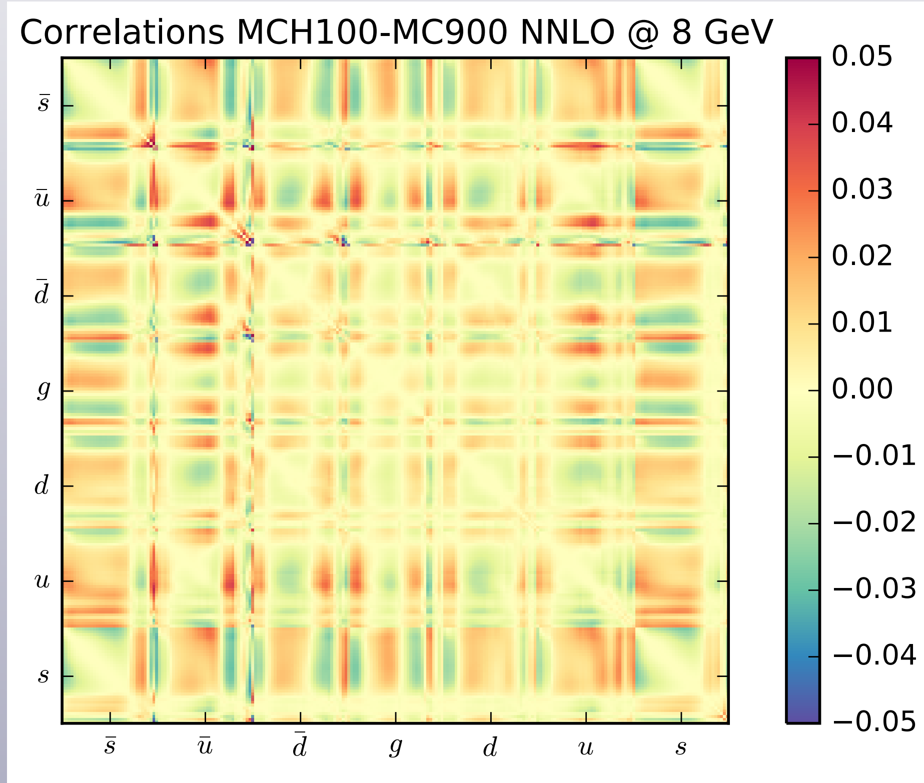


Perfect?

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!

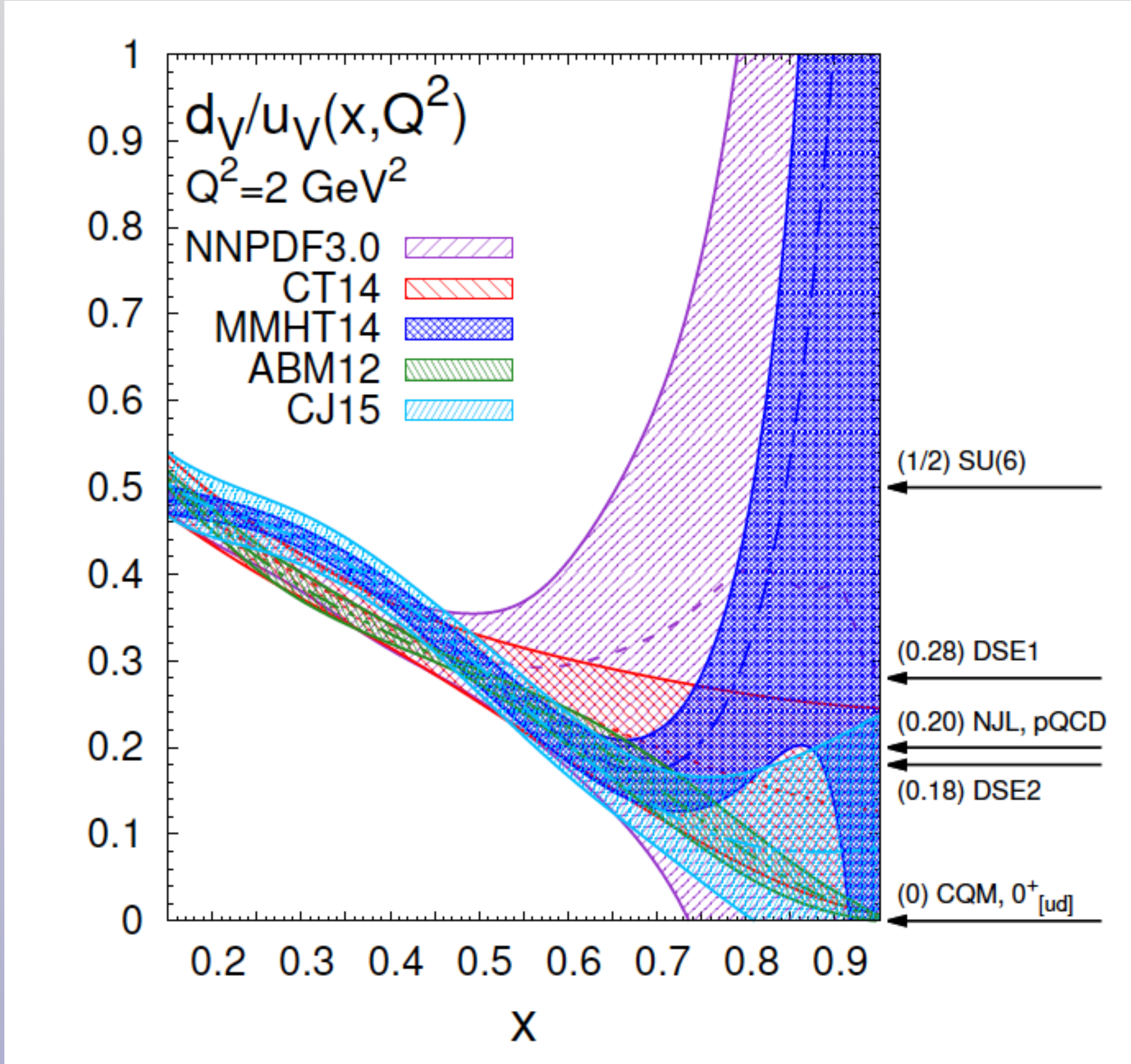
Tiny residual differences at the level of few perfect of the correlation coefficient at most, **irrelevant for LHC phenomenology**

PDFs in BSM searches

- **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,** and missing correlated systematics in many experiments.

- Useful for specific cross-checks, but many caveats needed

- Moreover, **CJ12 results consistent with the PDF4LHC15 uncertainties,** for instance well within the **NNPDF3.0 error band for d/u**



MC2H

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

MMHT14

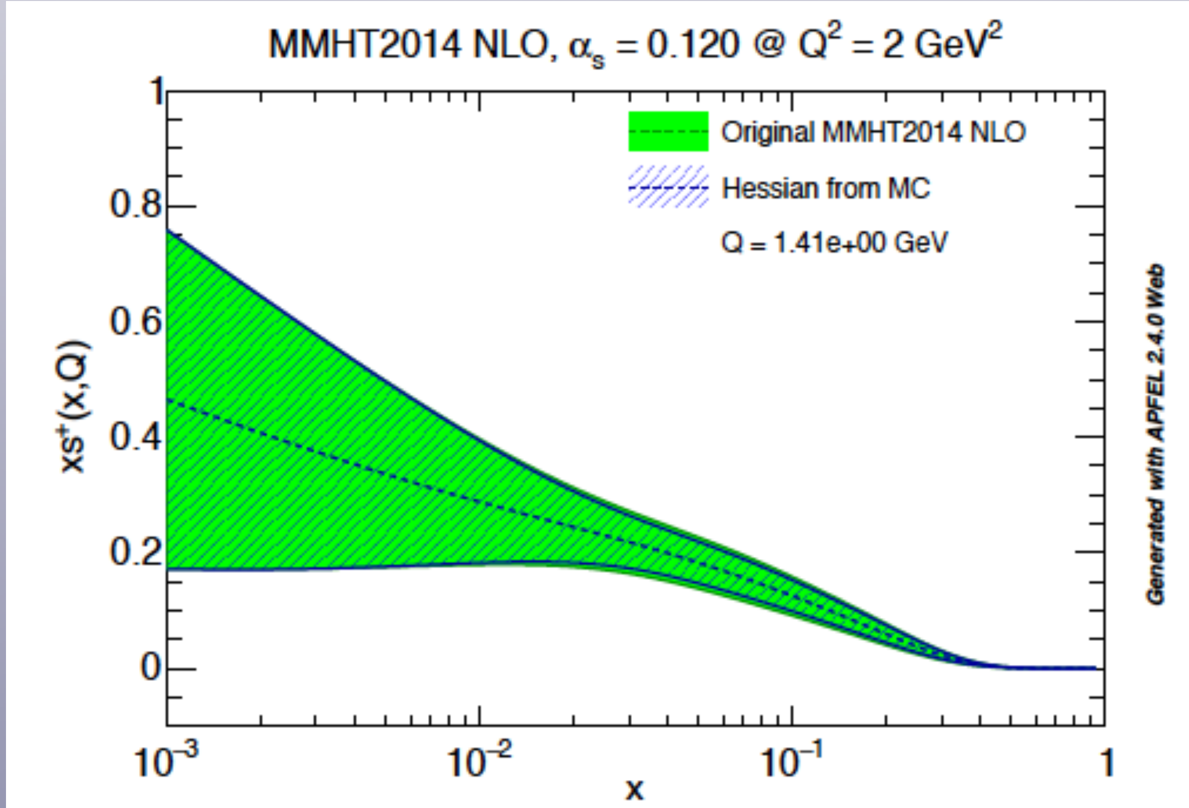
Hessian in space of fit parameters

MMHT14-MC

Monte Carlo representation

MMHT14-MC2H

Hessian in space of MC replicas



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]*

CMC-PDFs

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