PDF4LHC recommendations for LHC Run-2

Pavel Nadolsky

Southern Methodist University

With Joey Huston (MSU) and Jun Gao (ANL)

for the PDF4LHC working group

OUTP-15-17P
SMU-HEP-15-12
TIF-UNIMI-2015-14
LCTS/2015-27

PDF4LHC recommendations for LHC Run II

Jon Butterworth[1], Stefano Carrazza[2], Amanda Cooper-Sarkar[3], Albert De Roeck[4,5], Joël Feltesse[6], Stefano Forte[2], Jun Gao[7], Sasha Glazov[8], Joey Huston[9], Zahari Kassabov[2,10], Ronan McNulty[11], Andreas Morsch[4], Pavel Nadolsky[12], Voica Radescu[13], Juan Rojo[14] and Robert Thorne[1].

[1] *Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK.*
[2] *TIF Lab, Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano, Via Celoria 16, I-20133 Milano, Italy*
[3] *Particle Physics, Department of Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK.*
[4] *PH Department, CERN, CH-1211 Geneva 23, Switzerland*
[5] *Antwerp University, B2610 Wilrijk, Belgium*
[6] *CEA, DSM/IRFU, CE-Saclay, Gif-sur-Yvette, France*
[7] *High Energy Physics Division, Argonne National Laboratory, Argonne, Illinois 60439, U.S.A.*
[8] *Deutsches Elektronen-Synchrotron (DESY), Notkestrasse 85, D-22607 Hamburg, Germany.*
[9] *Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824 U.S.A.*
[10] *Dipartimento di Fisica, Università di Torino and INFN, Sezione Via Pietro Giuria 1, I-10125 Torino, Italy*
[11] *School of Physics, University College Dublin Science Centre UCD Belfeld, Dublin 4, Ireland*
[12] *Department of Physics, Southern Methodist University, Dallas, TX 75*
[13] *Physikalisches Institut, Universität Heidelberg, Heidelberg, Ge*
[14] *Rudolf Peierls Centre for Theoretical Physics, 1 Keble Ro University of Oxford, OX1 3NP Oxford, UK*

A major revision of the previous PDF4LHC recommendation in arxiv:1101.0538, arXiv:1211.5142

Abstract

PDF4LHC publication, topics

1. **Review of updates on PDFs from various groups**

NNLO Global PDF sets: CT14, MMHT'14, NNPDF3

PDFs using other methodologies: ABM'12, CJ15, HERAPDF2.0

→

2. Average PDF sets by PDF4LHC group: PDF4LHC15_30, _100, _MC

Criteria for combination

3. Recommendation on selecting PDF sets for various LHC applications

New physics searches

Precision tests of SM and PDFs

Monte-Carlo simulations

Acceptance estimates

2012→2015: Agreement between NNLO PDFs greatly improved

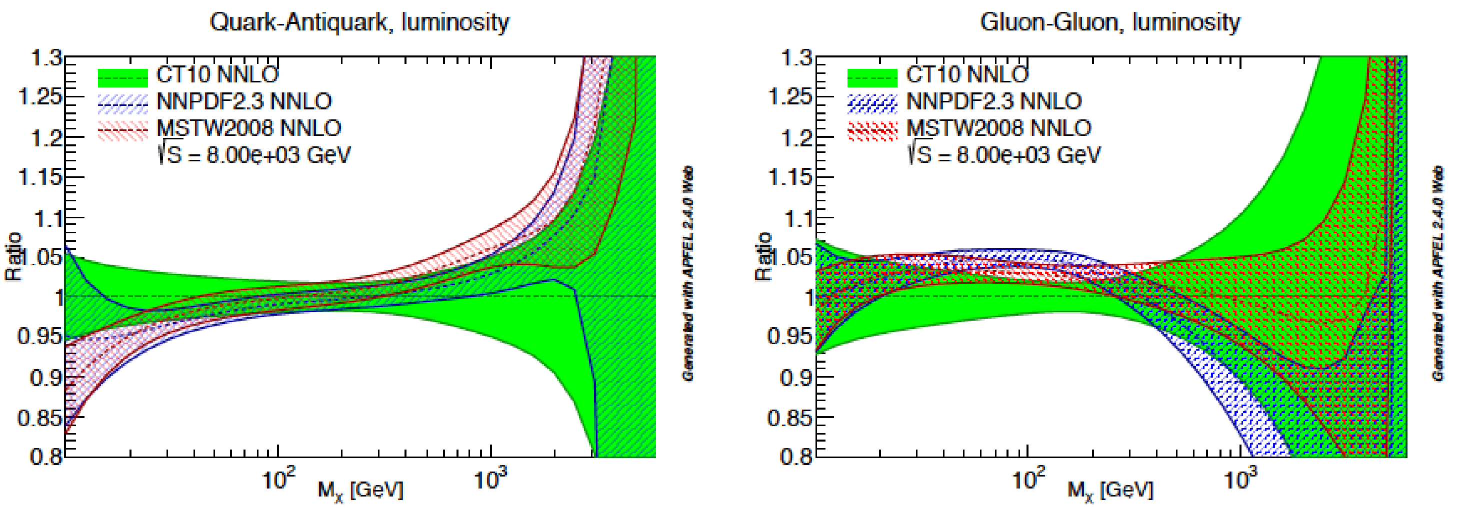
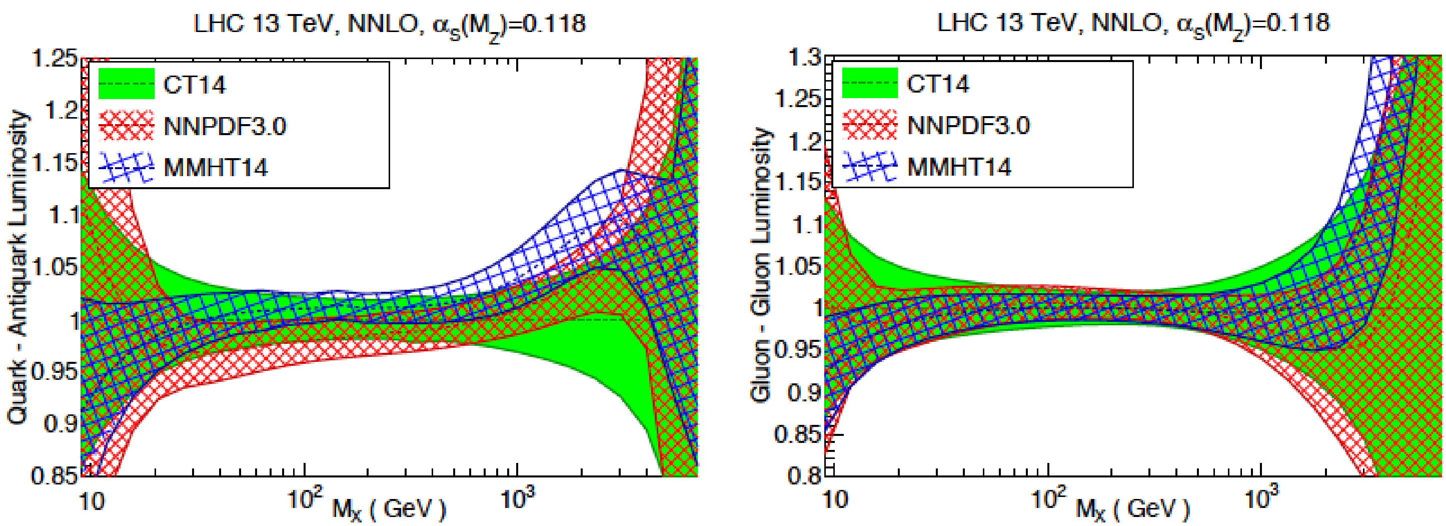


Figure 1: Comparison of the $q\bar{q}$ (left) and gg (right) PDF luminosities at the LHC 8 TeV for CT10, MSTW2008 and NNPDF2.3. Results are shown normalized to the central value of CT10.



Note in particular the changes in the gg luminosity, especially important in the Higgs mass region

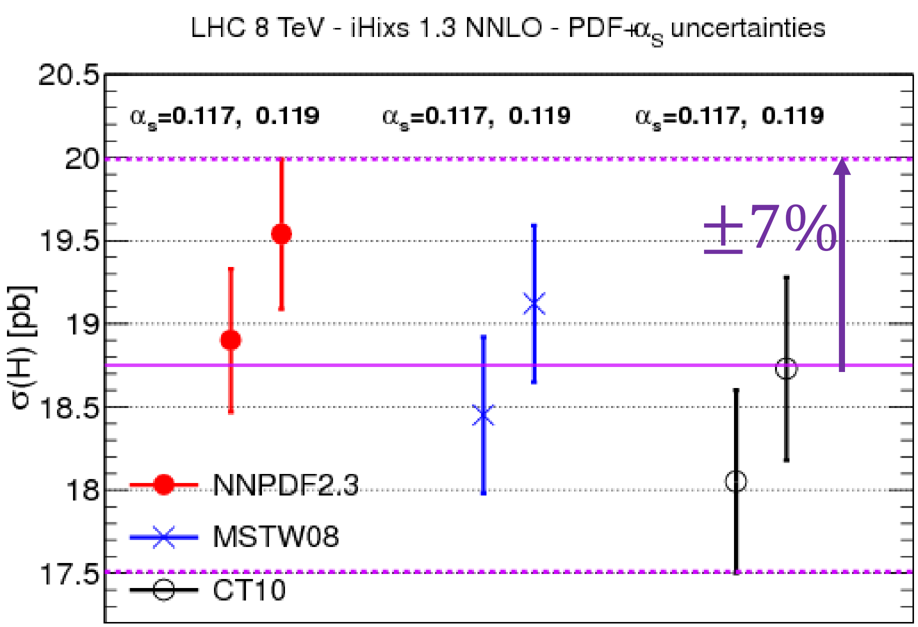
LHC data has been added for all 3 new PDFs, but most of change is due to changes in formalisms

Note also differences in high mass region remain

$gg \rightarrow$ SM Higgs

$2012\!: \delta_{PDF} \approx 7\%$

$2015\!: \delta_{PDF} \approx 2-3\%$



R. Ball et al., **arXiv:1211.5142**

$\sigma(gg \rightarrow H^0)$ at NNLO

	CT14	MMHT2014	NNPDF3.0
8 TeV	18.66 pb -2.2% +2.0%	18.65 pb -1.9% +1.4%	18.77 pb -1.8% +1.8%
13 TeV	42.68 pb -2.4% +2.0%	42.70 pb -1.8% +1.3%	42.97 pb -1.9% +1.9%

J.Huston, PDF4LHC, April 2015

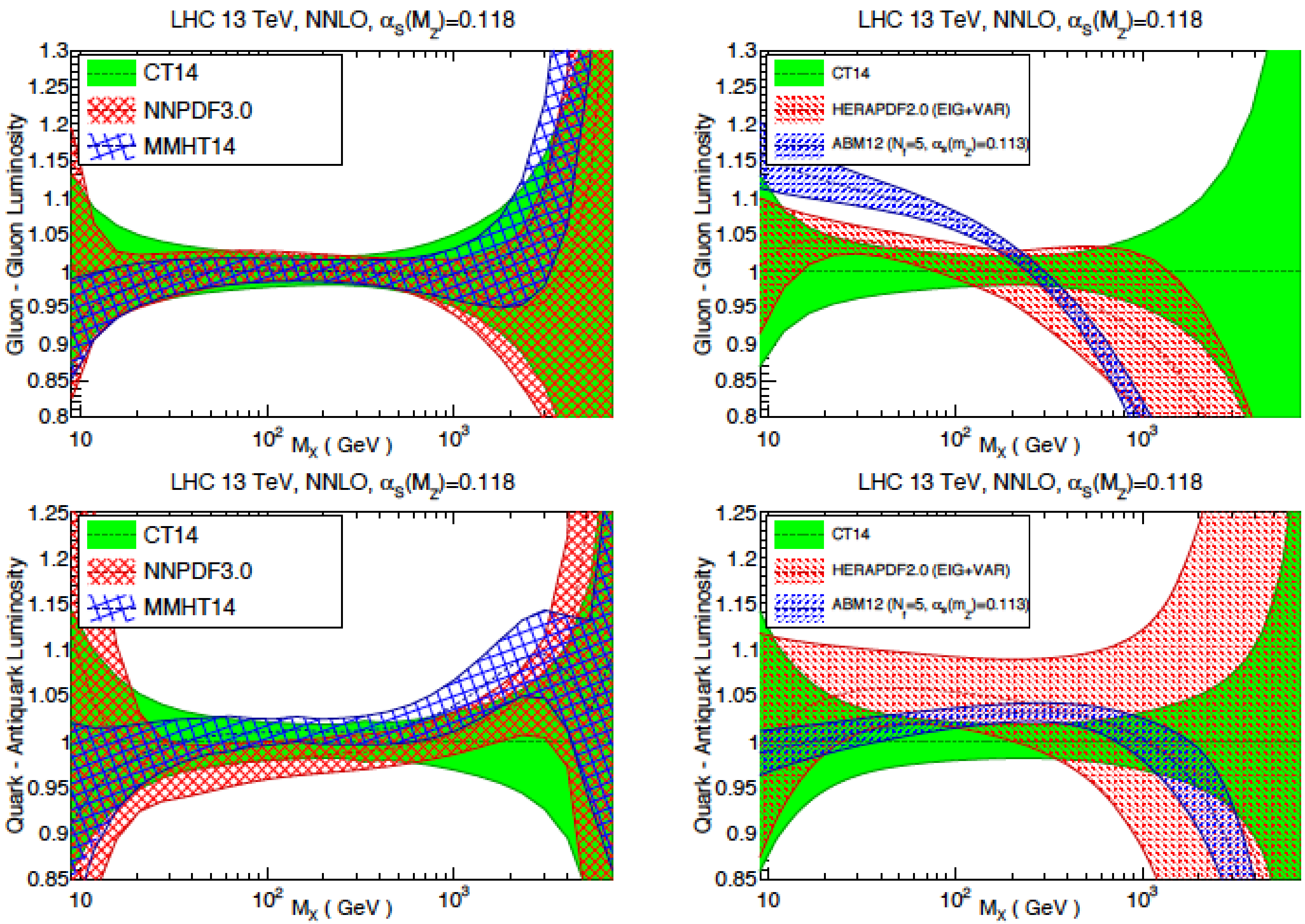
Disagreement in central values

Good agreement of central values

N3LO scale dependence on σ_H is <3%

Similar agreement for $t\bar{t}$ cross sections

Other new sets published as well



behavior for HERAPDF2.0 and ABM12 somewhat different

HERAPDF2.0 uncertainties tend to be larger

Figure 5: Comparison of the gluon-gluon (upper plots) and quark-antiquark (lower plots) PDF luminosities from the CT14, MMHT14 and NNNPDF3.0 NNLO sets (left plots) and from the NNPDF3.0, ABM12 and HERAPDF2.0 NNLO sets (right plots), for a center-of-mass energy of 13 TeV, as a function of the invariant mass of the final state M_X.

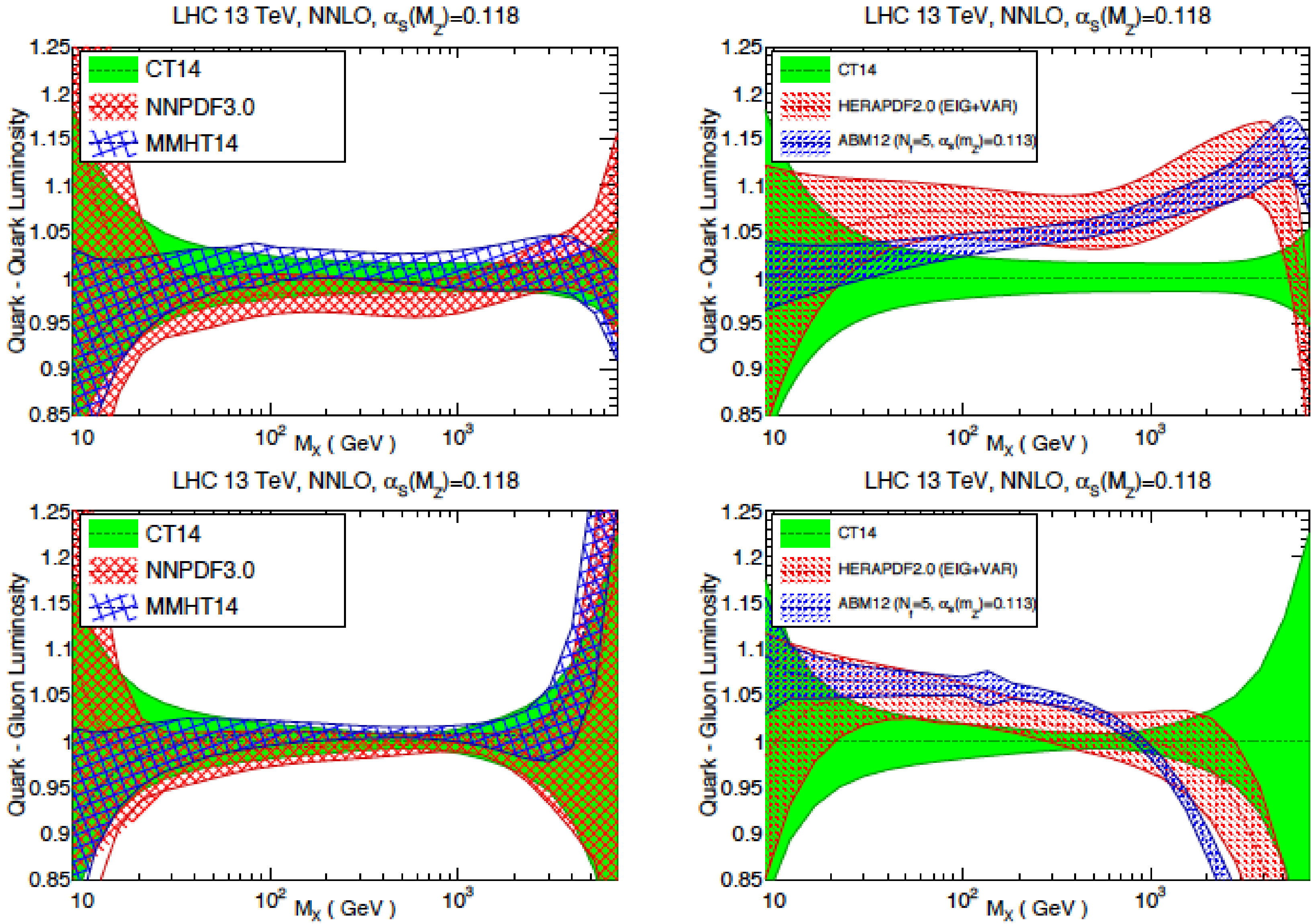


Figure 6: Same as Fig. 5 for the quark-quark (upper plots) and the quark-gluon (lower plots) PDF luminosities.

Three main uses of PDFs at the LHC

1. Assessment of the *total uncertainty on a cross section* based on the available knowledge of PDFs, *e.g.*, when computing the cross section for a process that has not been measured yet (such as supersymmetric particle production cross-sections), or for estimating acceptance corrections on a given observable. This is also the case of the measurements that aim to verify overall, but not detailed, consistency with Standard Model expectations, such as when comparing theory with Higgs measurements.

2. Assessment of the accuracy of the *PDF sets themselves* or of related Standard Model parameters, typically done by comparing theoretical predictions using individual PDF sets to the most precise data available.

3. Input to the *Monte Carlo event generators* used to generate large MC samples for LHC data analysis.

For 2), compute cross sections with individual PDF sets.

For 1), the PDF uncertainty based on the totality of available PDF sets must be estimated. Need to estimate the combined PDF error.

For 3), may want to use **an average of various PDF sets**.

Averaging of PDF ensembles

The 2012 recommendation estimated the combined uncertainty as an envelope of **cross sections** for 3 PDF sets; the envelope was overly sensitive to outliers

Recently, several methods for combination (averaging) of **PDFs** (before computing cross sections) were developed. Criteria allowing the combination were outlined.

Combination workflow:

1. Generate 900 MC replicas from all input ensembles (currently CT14, MMHT14, NNPDF3.0) using Thorne-Watt procedure
   - Other PDF sets can be added in the future if they satisfy the listed criteria
2. Reduce the number of final replicas from 900 to 100 or 30 by keeping most relevant PDF combinations

Which PDFs can be “easily” combined?

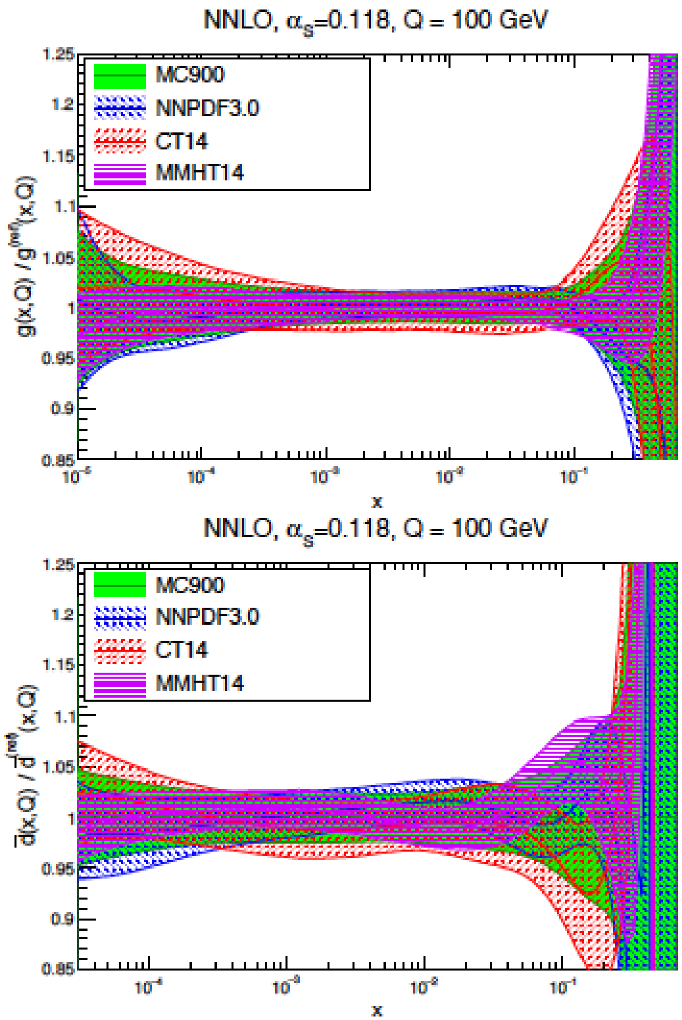
1. The PDF sets to be combined should be *based on a global dataset*, including a large number of datasets of diverse types (deep-inelastic scattering, vector boson and jet production, ...) from fixed-target and colliders experiments (HERA, LHC, Tevatron).

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.

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 the experimental data, uncertainties associated with incompatibility of the fitted experiments, procedural uncertainties such as those related to the functional form of PDFs, the handling of systematic errors, etc. Sets entering the combination must account for these through suitable methods, such as separate estimates for additional model and parametrization components of the PDF uncertainty [9], tolerance [6, 10], or closure tests [11].

MC900

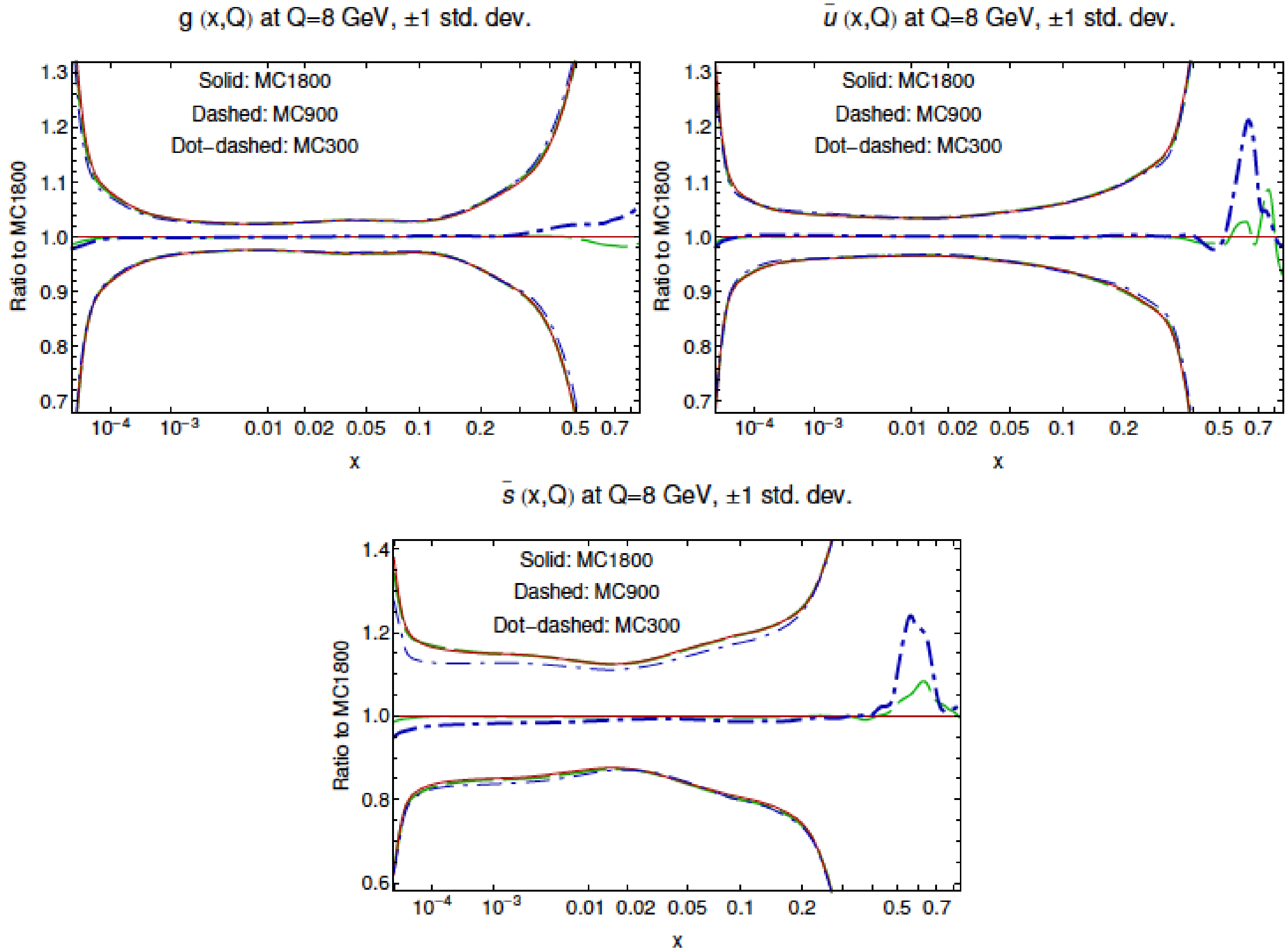


Note that MC900 is not the envelope of the 3 PDF error bands

The PDF error bands themselves are similar for the precision physics region, but not for low mass/high mass

Figure 8: Comparison of the MC900 PDFs with the sets that enter the combination: CT14, MMHT14 and NNPDF3.0 at NNLO. We show the gluon and the up, anti-down and strange quarks at $Q = 100$ GeV. Results are normalized to the central value of MC900.

Why 900 Monte-Carlo replicas?



900 replicas seems enough. Increasing the # of replicas to 1800 does not significantly change the average PDF uncertainty

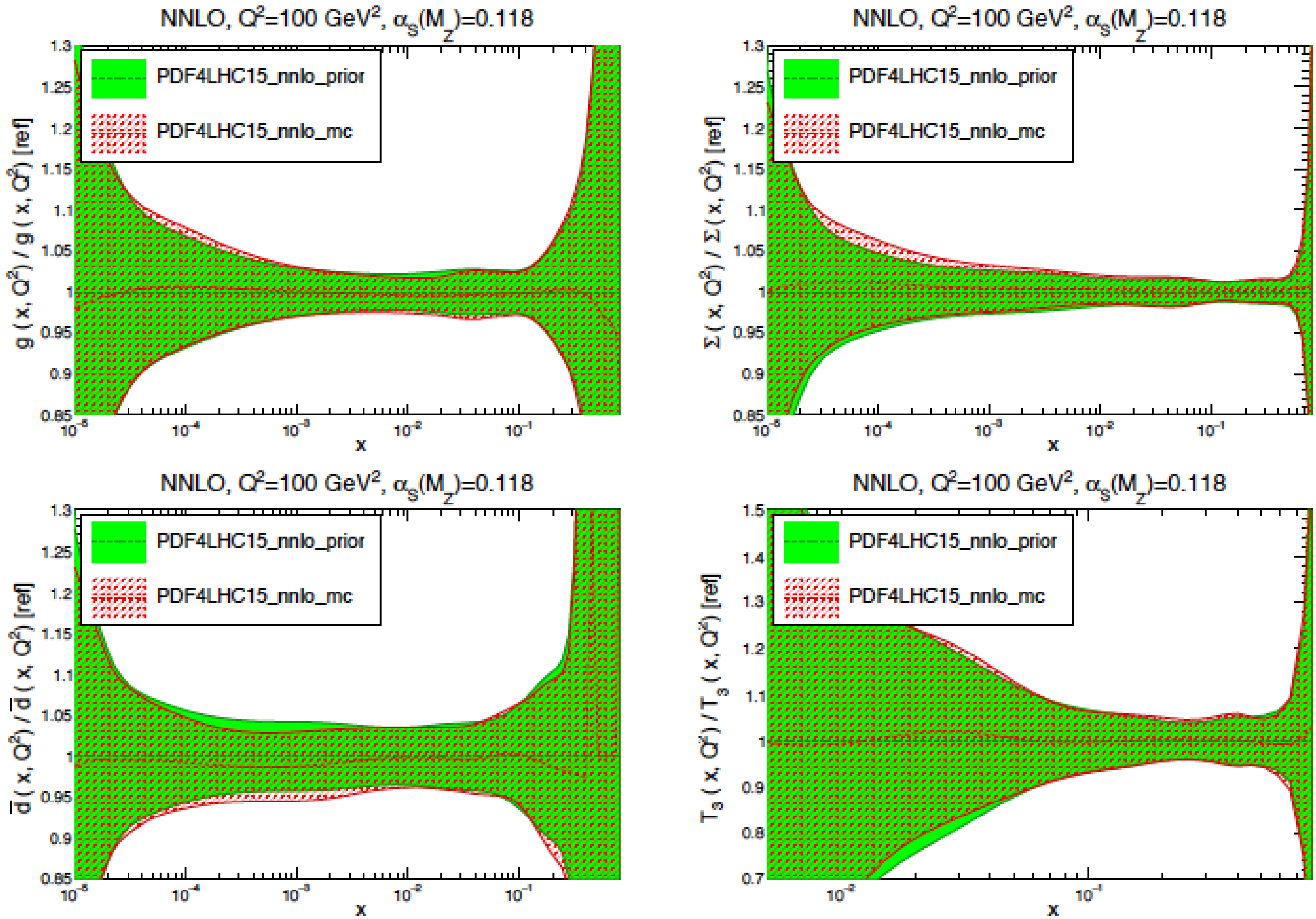
->MC900
or
PDF4LHC_prior

Figure 7: Comparison of central values and uncertainties for the MC combination of CT14, MMHT14 and NNPDF3.0 for different values of $N_{\rm rep}$, 300, 600 and 900, denoted by MC300, MC900 and MC1800 respectively.

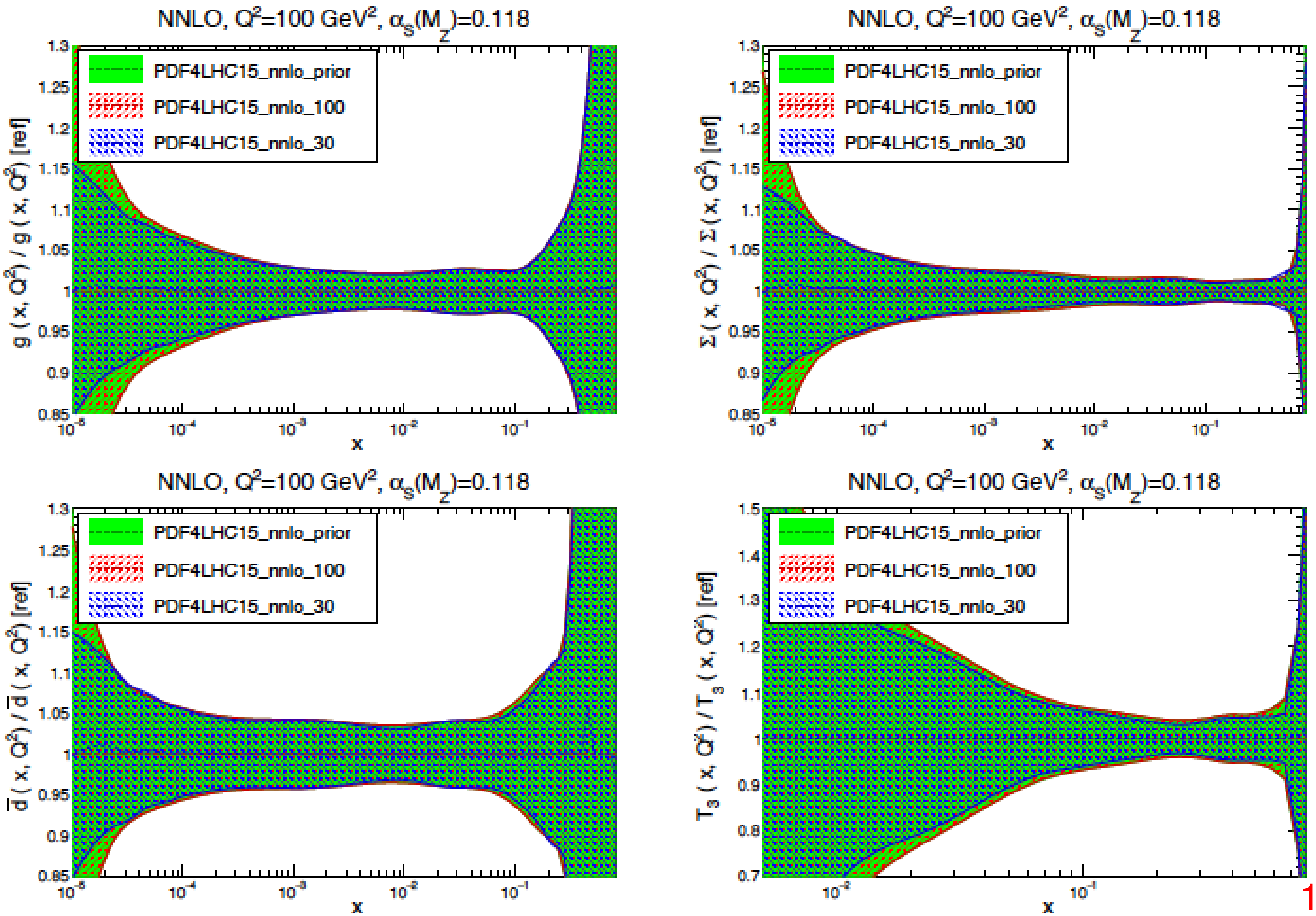
Reduced sets

- 900 error PDFs are too much for general use
- We would like to reduce this number while still maintaining as much information on the uncertainties and on correlations between PDF uncertainties as possible
- We have settled on 3 techniques/outputs
  - Compressed Monte Carlo PDFs (PDF4LHC15_nnlo(nlo)_mc)
    - 100 PDF error sets; preserve non-Gaussian errors
  - META Hessian PDFs (PDF4LHC15_nnlo(nlo)_30
    - 30 PDF error sets using METAPDF technique; Gaussian (symmetric) errors
  - MCH Hessian PDFs (PDF4lhc15_nnlo(nlo)_100
    - 100 PDF error sets using MCH technique; Gaussian (symmetric errors)
- The META technique is able to more efficiently reproduce the uncertainties when using a limited number (30) of error PDFs
- The MCH technique best reproduces the uncertainties of the 900 MC set prior

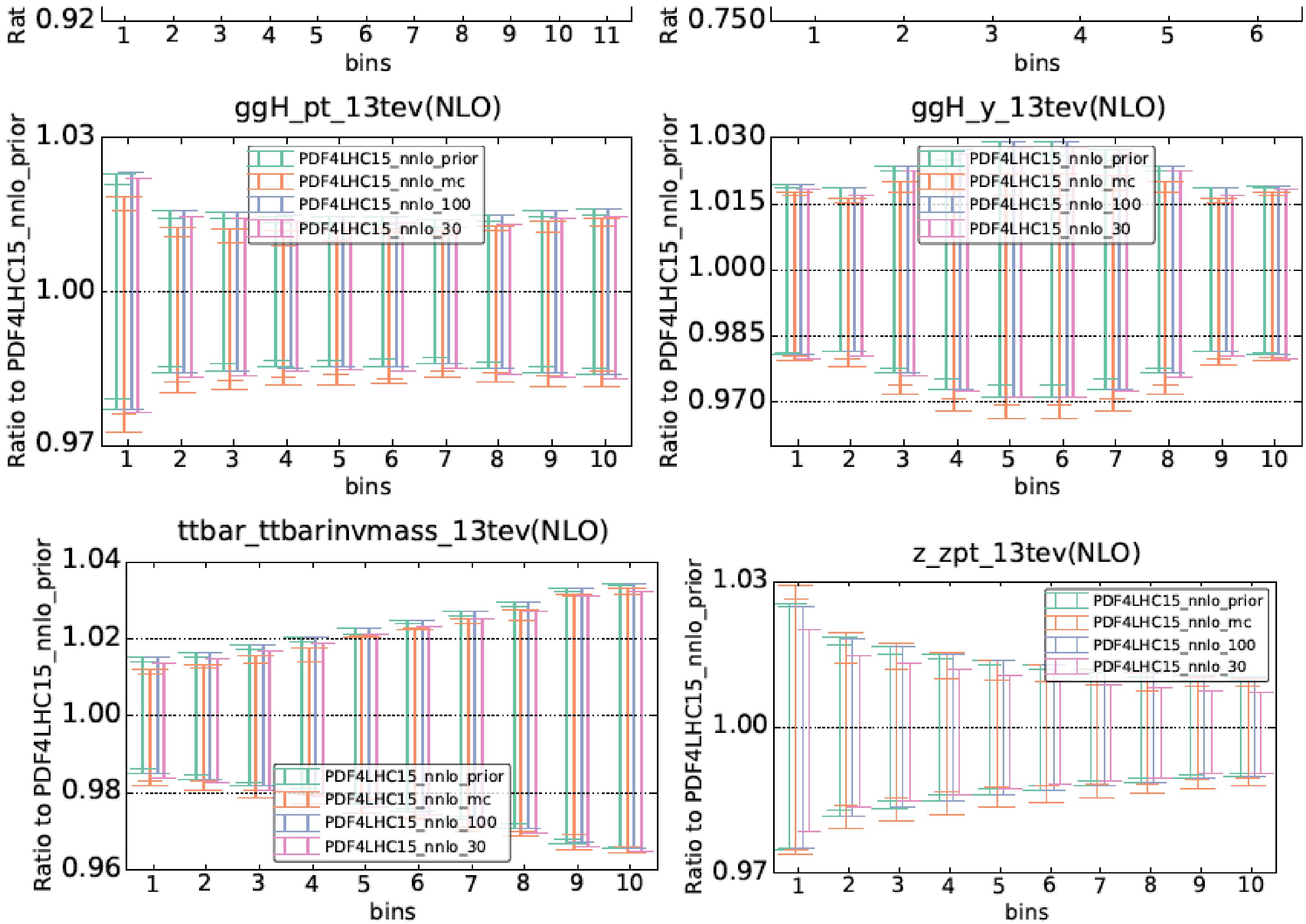
Some comparisons: mc PDFs



Some comparisons: Hessian sets



Application to cross sections



PDF Set	Correlation coefficient					
	Z, W	$Z, t\bar{t}$	Z, ggh	$Z, ht\bar{t}$	Z, hW	Z, hZ
PDF4LHC15_nlo_prior	0.90	-0.60	0.22	-0.64	0.55	0.74
PDF4LHC15_nlo_mc	0.92	-0.49	0.41	-0.58	0.61	0.77
PDF4LHC15_nlo_100	0.92	-0.60	0.23	-0.64	0.57	0.75
PDF4LHC15_nlo_30	0.90	-0.68	0.16	-0.71	0.55	0.76
PDF4LHC15_nnlo_prior	0.89	-0.49	0.08	-0.46	0.56	0.74
PDF4LHC15_nnlo_mc	0.90	-0.44	0.18	-0.42	0.62	0.80
PDF4LHC15_nnlo_100	0.91	-0.48	0.09	-0.46	0.59	0.74
PDF4LHC15_nnlo_30	0.88	-0.63	0.04	-0.61	0.56	0.72

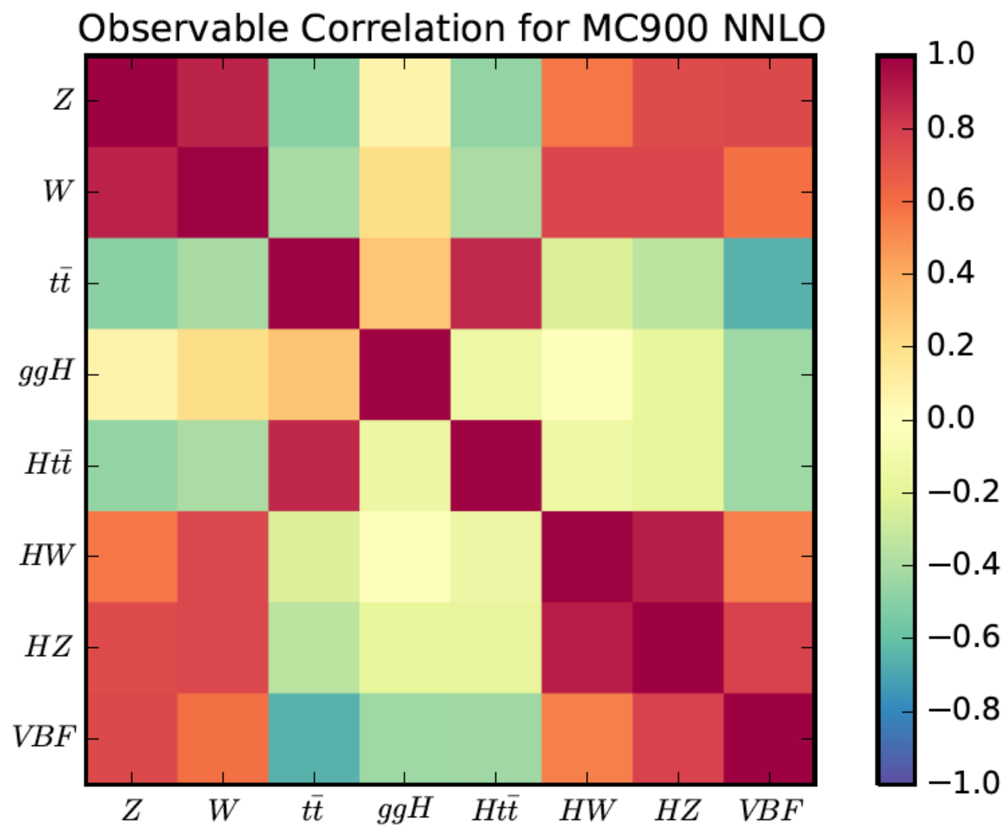
Table 1: Correlation coefficient between the Z production cross-sections and the W, $t\bar{t}$, ggh, $ht\bar{t}$, hW and hZ production cross-sections. The PDF4LHC15 prior is compared to the Monte Carlo and the two Hessian reduced sets, both at NLO and at NNLO.

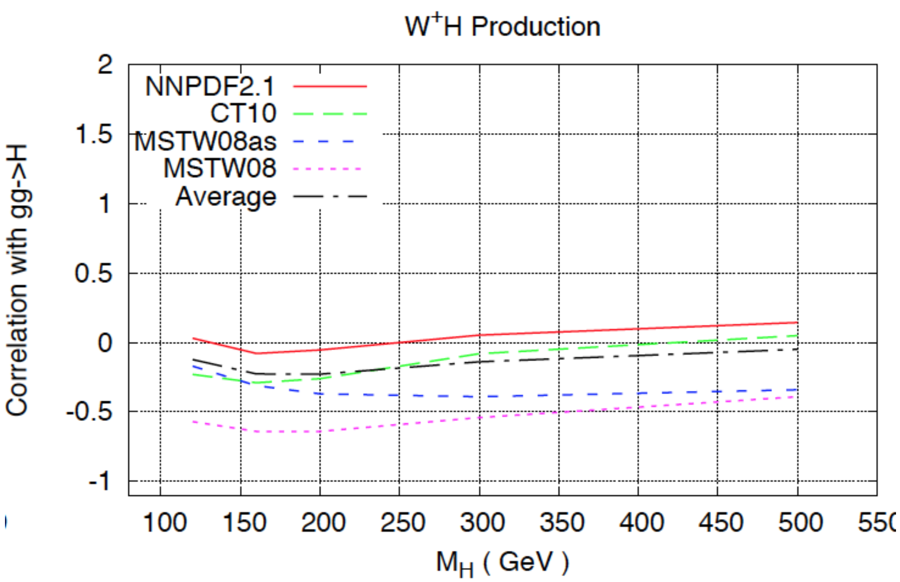
PDF Set	Correlation coefficient					
	$W, t\bar{t}$	W, ggh	$W, ht\bar{t}$	W, hW	W, hZ	$t\bar{t}, ggh$
PDF4LHC15_nlo_prior	-0.46	0.32	-0.51	0.77	0.78	0.27
PDF4LHC15_nlo_mc	-0.35	0.49	-0.46	0.81	0.80	0.27
PDF4LHC15_nlo_100	-0.47	0.32	-0.52	0.77	0.79	0.27
PDF4LHC15_nlo_30	-0.52	0.28	-0.56	0.79	0.81	0.32
PDF4LHC15_nnlo_prior	-0.40	0.20	-0.40	0.76	0.77	0.30
PDF4LHC15_nnlo_mc	-0.44	0.26	-0.42	0.81	0.82	0.32
PDF4LHC15_nnlo_100	-0.40	0.20	-0.40	0.76	0.77	0.30
PDF4LHC15_nnlo_30	-0.47	0.19	-0.47	0.77	0.76	0.31

Table 2: Same as Table 1 for the correlation coefficient of additional pairs of LHC inclusive cross-sections.

Correlations

probably only 1 digit for the correlations is significant, so plot like on right is more relevant





correlations can differ significantly for example from the individual PDFs

from YR2

Now on LHAPDF: NLO, NNLO, varied α_s sets $N_f = 5$ and 4 (upcoming)

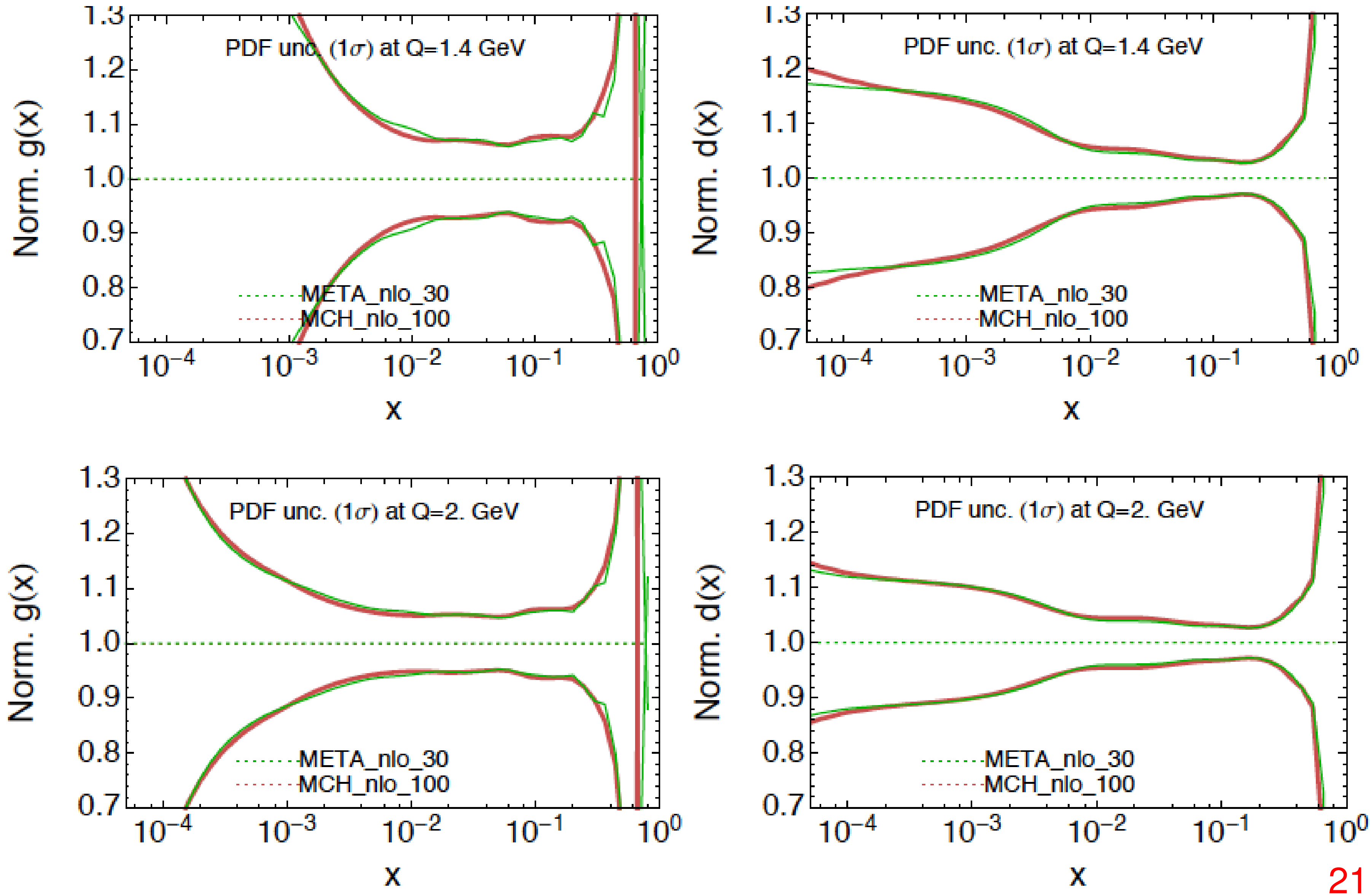
LHAPDF6 grid	Pert order	ErrorType	$N_{\rm mem}$	$\alpha_s(m_Z^2)$
PDF4LHC15_nnlo_mc	NNLO	replicas	100	0.118
PDF4LHC15_nnlo_100	NNLO	symmhessian	100	0.118
PDF4LHC15_nnlo_30	NNLO	symmhessian	30	0.118
PDF4LHC15_nnlo_mc_pdfas	NNLO	replicas+as	102	mem 0:100 $\to$ 0.118 mem 101 $\to$ 0.1165 mem 102 $\to$ 0.1195
PDF4LHC15_nnlo_100_pdfas	NNLO	symmhessian+as	102	mem 0:100 $\to$ 0.118 mem 101 $\to$ 0.1165 mem 102 $\to$ 0.1195
PDF4LHC15_nnlo_30_pdfas	NNLO	symmhessian+as	32	mem 0:30 $\to$ 0.118 mem 31 $\to$ 0.1165 mem 32 $\to$ 0.1195
PDF4LHC15_nnlo_asvar	NNLO	-	1	mem 0 $\to$ 0.1165 mem 1 $\to$ 0.1195

Table 5: Summary of the combined NNLO PDF4LHC15 sets with $n_f^{\rm max} = 5$ that are available from LHAPDF6. The corresponding NLO sets are also available. Members 0 and 1 of PDF4LHC15_nnlo_asvar coincide with members 101 and 102 (31 and 32) of PDF4LHC15_nnlo_mc_pdfas and PDF4LHC15_nnlo_100_pdfas (PDF4LHC15_nnlo_30_pdfas). Recall that in LHAPDF6 there is always a zeroth member, so that the total number of PDF members in a given set is always $N_{\rm mem} + 1$. See text for more details.

Caveat: Heavy flavors and low Q

- $N_f = 5$ averaged PDFs are strictly valid at $Q^2 \gg m_b^2$, where they should be used with $N_f = 5$ hard cross sections
- At $Q \leq m_b$, each PDF group assumes a different heavy-quark mass scheme/heavy quark mass; at $Q < m_b$ **the averaged PDFs should not be naively used in fixed-order calculations, but can be used with parton showering programs if some accuracy loss is tolerated**
- With this understanding, the _30, _100, and _MC PDFs are provided for $Q > 1.4$, 1, and 1 GeV.
- The _30 PDFs have been parametrized at Q_o=8 GeV and extended down to Q=1.4 GeV by backward DGLAP evolution.The resulting _30 PDFs match up well with the _100 set for Q<8 GeV

_30 and _100 sets at $Q = 1.4$ and 2 GeV



PDF4LHC NLO sets with $N_f = 4$

- Made available for NLO calculations in the 4-flavor scheme
- $\alpha_s^{N_f=5}(M_Z) \approx 0.118 \rightarrow \alpha_s^{N_f=4}(M_Z) \approx 0.113$. For other μ values, $\alpha_s^{N_f=4}(\mu)$ is found from $\alpha_s^{N_f=5}(\mu)$ via 4-loop relations by Chetyrkin et al., hep-ph/0512060

LHAPDF6 grid	Pert order	ErrorType	N_{mem}	$\alpha_s^{(n_f=5)}(m_Z^2)$
PDF4LHC15_nlo_nf4_100	NLO	symmhessian	100	0.118
PDF4LHC15_nlo_nf4_30	NLO	symmhessian	30	0.118
PDF4LHC15_nlo_nf4_100_pdfas	NLO	symmhessian+as	102	mem 0:100 $\rightarrow$ 0.118
				mem 101 $\rightarrow$ 0.1165
				mem 102 $\rightarrow$ 0.1195
PDF4LHC15_nlo_nf4_30_pdfas	NLO	symmhessian+as	32	mem 0:30 $\rightarrow$ 0.118
				mem 31 $\rightarrow$ 0.1165
				mem 32 $\rightarrow$ 0.1195
PDF4LHC15_nlo_nf4_asvar	NLO	-	1	mem 0 $\rightarrow$ 0.1165
				mem 1 $\rightarrow$ 0.1195

Table 6: Same as Table 5 for the combined PDF4LHC15 sets in the $n_f = 4$ scheme. We indicate the value of $\alpha_s^{(n_f=5)}(m_Z^2)$ in the $n_f = 5$ scheme, the actual value in the $n_f = 4$ scheme is substantially smaller, see text.

Recommendations

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.

   Rationale: Measurements such as jet production, vector-boson single and pair production, or top-quark pair production, have the power to constraining PDFs, and this is best utilized and illustrated by comparing with many individual sets.

   As a rule of thumb, *any measurement that potentially can be included in PDF fits* falls in this category.

   The same recommendation applies to the *extraction of precision SM parameters*, such as the strong coupling $\alpha_s(m_Z^2)$ [75,124], the W mass M_W [125], and the top quark mass m_t [126] which are directly correlated to the PDFs used in the extraction.

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.

   Rationale: In many situations, PDF uncertainties may affect the extraction of physics parameters. From the point of view of the statistical analysis, it might be useful in some cases to *limit the number of error PDFs* that need to be included in such analyses. In these cases, use of the `PDF4LHC15_30` sets may be most suitable.

   In addition, the calculation of *acceptances, efficiencies or extrapolation factors* are affected by the corresponding PDF uncertainty. These quantities are only a moderate correction to the measured cross-section, and thus a mild loss of accuracy in the determination of PDF uncertainties in these corrections is acceptable, while computational speed can be an issue. In these cases, use of the `PDF4LHC15_30` sets is most suitable.

   However, in the cases when PDF uncertainties turn out to be substantial, we recommend to cross-check the PDF estimate by comparing with the results of the `PDF4LHC15_100` 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.

   Whenever the highest accuracy is desired, the `PDF4LHC15_100` set is most suitable.

Pedagogical text about the PDF use has been added

6.2 Formulae for the calculation of PDF and PDF+α_s uncertainties

For completeness, we also collect in this report the explicit formulae for the calculation of PDF and combined PDF+α_s uncertainties in LHC cross-sections when using the PDF4LHC15 combined sets. Let us assume that we wish to estimate the PDF+α_s uncertainty of given cross-section σ, which could be a total inclusive cross-section or any bin of a differential distribution.

First of all, to compute the PDF uncertainty, one has to evaluate this cross-section $N_{\rm mem}+1$ times, where $N_{\rm mem}$ is the number of error sets (either symmetric eigenvectors or MC replicas) of the specific combined set,

$$\sigma^{(k)}\,, \quad k=0,\ldots,N_{\rm mem}\,, \tag{19}$$

so in particular $N_{\rm mem} = 30$ in `PDF4LHC15_30` and $N_{\rm mem} = 100$ in `PDF4LHC15_100` and `PDF4LHC15_mc`.

PDF uncertainties for Hessian sets. In the case of the Hessian sets, `PDF4LHC15_30` and `PDF4LHC15_100`, the master formula to evaluate the PDF uncertainty is given by

$$\delta^{\rm pdf}\sigma = \sqrt{\sum_{k=1}^{N_{\rm mem}} \left(\sigma^{(k)}-\sigma^{(0)}\right)^2}\,, \tag{20}$$

This uncertainty is to be understood as a 68% confidence level. From this expression it is also easy to determine the contribution of each eigenvector k to the total Hessian PDF uncertainty.

…continues with discussion of MC PDFs

Summary

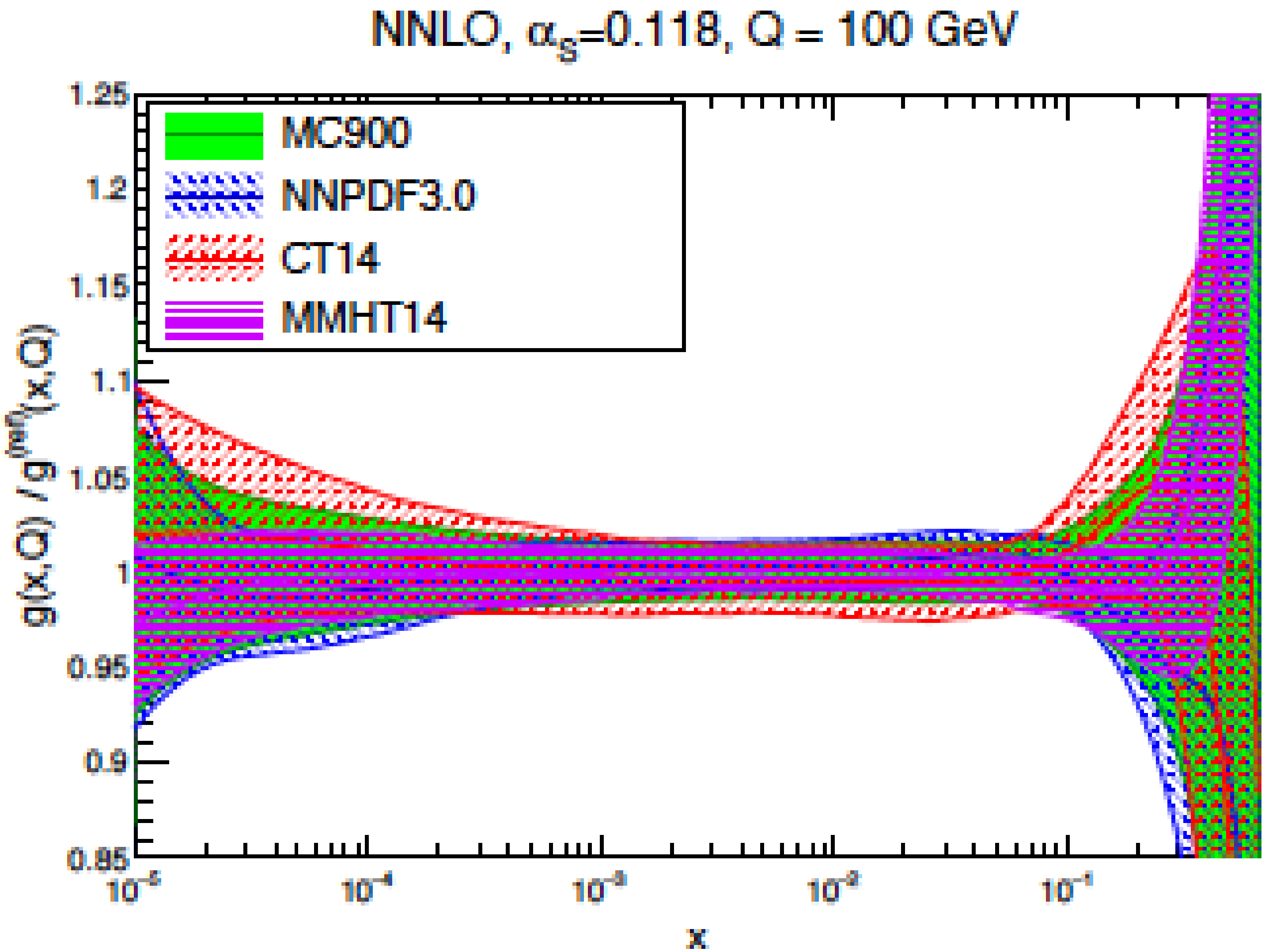
- New PDF4LHC recommendations are based on PDF combinations of CT14, MMHT2014 and NNPDF3.0
- Central PDF and uncertainties derived from 900 MC replicas of error PDFs of above 3 sets
- Three reduction techniques, with either 30 or 100 error PDFs, with uses as discussed previously
- PDF4LHC_XX X sets on LHAPDF:

  $N_f = 5$ NLO and NNLO sets;

  $N_f = 4$ NLO sets;

  member sets with varied a_s
- With this recommendation also comes a new recommendation for the central value of $\alpha_s(M_Z)$ and its uncertainty
  - $\alpha_s(m_Z)$=0.118
  - $\delta\alpha_s(m_Z)$=+/-0.0015 at 68% c.l.

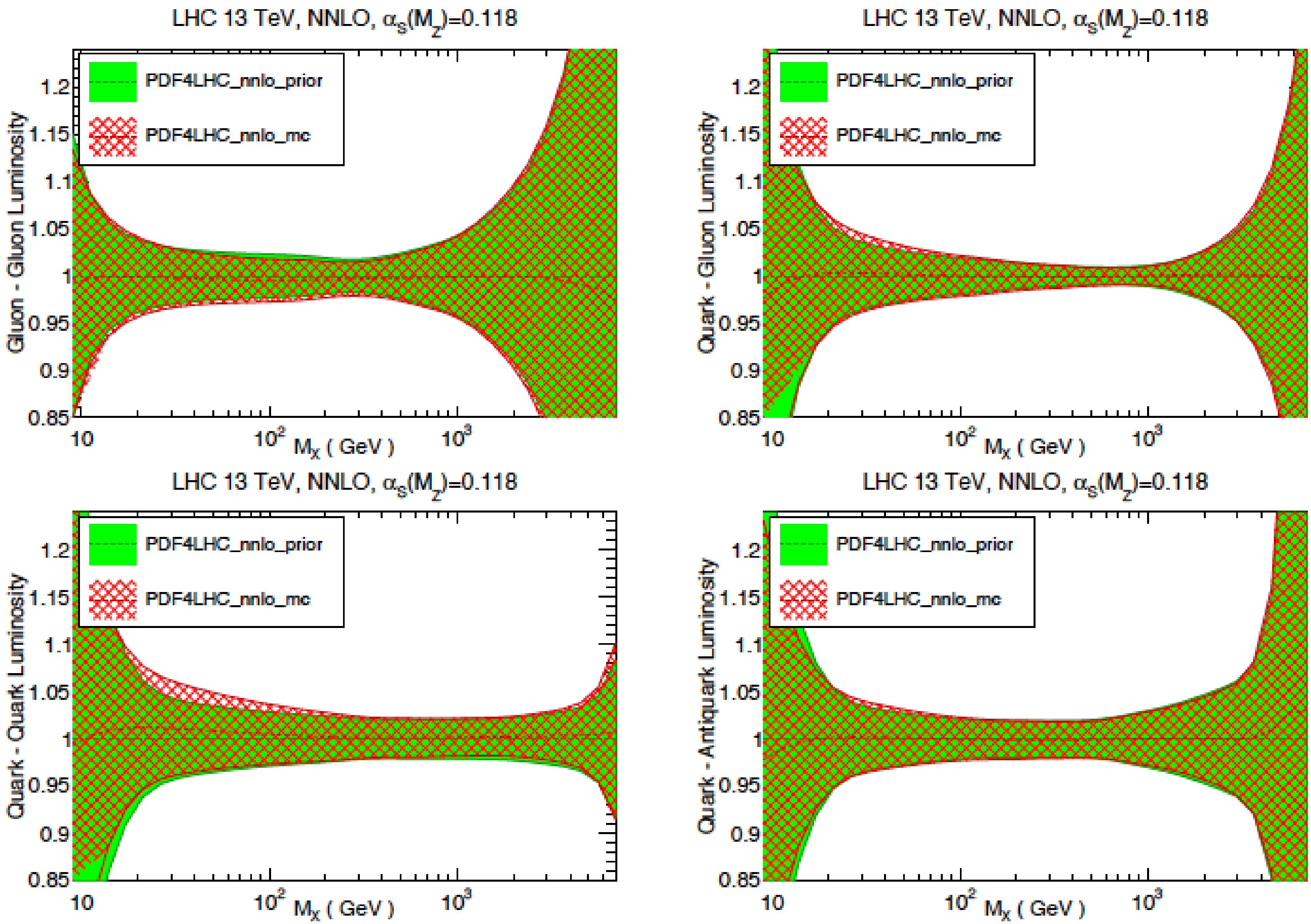
Backup slides



MC900 set does not completely cover/coincide with the uncertainty bands of CT14, NNPDF3.0 or MMHT14, especially at low/high x. However, it's the best approximation we can come up with.

Parton luminosities for combined sets

Some comparisons: mc PDFs



Some comparisons: Hessian sets