Published for SISSA by Springer

Received: December 3, 2016
Revised: February 21, 2017
Accepted: March 27, 2017
Published: April 10, 2017

Pinning down the large-x gluon with NNLO top-quark pair differential distributions

Michał Czakon,[a] **Nathan P. Hartland,**[c,d] **Alexander Mitov,**[e] **Emanuele R. Nocera**[b] **and Juan Rojo**[c,d]

[a] *Institut für Theoretische Teilchenphysik und Kosmologie, RWTH Aachen University, D-52056 Aachen, Germany*
[b] *Rudolf Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, OX1 3NP, Oxford, United Kingdom*
[c] *Department of Physics and Astronomy, VU University Amsterdam, De Boelelaan 1081, NL-1081, HV Amsterdam, The Netherlands*
[d] *Nikhef, Science Park 105, NL-1098 XG Amsterdam, The Netherlands*
[e] *Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, United Kingdom*

E-mail: mczakon@physik.rwth-aachen.de, n.p.hartland@vu.nl, adm74@cam.ac.uk, emanuele.nocera@physics.ox.ac.uk, j.rojo@vu.nl

Abstract: Top-quark pair production at the LHC is directly sensitive to the gluon PDF at large x. While total cross-section data is already included in several PDF determinations, differential distributions are not, because the corresponding NNLO calculations have become available only recently. In this work we study the impact on the large-x gluon of top-quark pair differential distributions measured by ATLAS and CMS at $\sqrt{s} = 8\,\text{TeV}$. Our analysis, performed in the NNPDF3.0 framework at NNLO accuracy, allows us to identify the optimal combination of LHC top-quark pair measurements that maximize the constraints on the gluon, as well as to assess the compatibility between ATLAS and CMS data. We find that differential distributions from top-quark pair production provide significant constraints on the large-x gluon, comparable to those obtained from inclusive jet production data, and thus should become an important ingredient for the next generation of global PDF fits.

Keywords: QCD Phenomenology

ArXiv ePrint: 1611.08609

doi:10.1007/JHEP04(2017)044

JHEP04(2017)044

Contents

JHEP04(2017)044

1 Introduction

The accurate determination of the parton distribution functions (PDFs) of the proton [1–4] is an essential requirement for the precision phenomenology program at the Large Hadron Collider (LHC). Traditionally, the bulk of the available experimental information on PDFs came from deep-inelastic scattering (DIS) and fixed-target Drell-Yan (DY) data. In recent years, however, the data from the LHC has provided a wealth of new information on the structure of the proton, see e.g. [5] and references therein. LHC measurements on inclusive electroweak vector boson and jet production are routinely included in most of the modern PDF determinations [6–9]. Furthermore, several dedicated analyses have demonstrated the constraining power of many other LHC processes, including W+charm production [10, 11], the transverse momentum distribution of W and Z bosons [12, 13], prompt photons [14] and charm production in the forward region [15–18].

In the case of top-quark pair production, the next-to-next-to leading order (NNLO) QCD corrections to the total cross-section were computed in 2013 [19–21]. This development allowed for the consistent inclusion of the Tevatron and LHC inclusive top-quark pair measurements into a NNLO global PDF fit [22], which demonstrated how this data could help reducing the rather sizable uncertainty of the gluon PDF for $x \gtrsim 0.1$. With this motivation, top-quark total cross-sections were included in the latest updates of some PDF fits, specifically NNPDF3.0 and MMHT14 (see also the ABM12 fit for related studies).

Last year, the calculation of NNLO corrections to inclusive top-quark pair production was extended to fully differential distributions for stable tops [23–26]. It is therefore natural to investigate how the constraints upon the large-x gluon PDF obtained from inclusive measurements are improved once the additional information contained in the differential distributions is accounted for in a global NNLO analysis (see [27] for a first study based on approximate NNLO). Such a program is enabled by the availability of precision measurements of top-quark pair differential cross-sections at $\sqrt{s} = 8\,\text{TeV}$ from ATLAS [28] and CMS [29], provided with the full information on the breakdown of experimental statistical and systematic uncertainties.

Given that DIS structure functions and DY production provide only a rather loose constraint upon the gluon PDF, particularly at large x, inclusive jet production data has been traditionally used to obtain additional information [30]. While the NLO QCD corrections to jet production at hadron colliders have been available for more than two decades [31–33], the corresponding NNLO corrections (in the leading color approximation) have been computed only very recently [34], building on the partial results of refs. [35, 36]. Since the results of [34] are not yet publicly available, in this work we will not include collider jet data, so that we can use exact NNLO theory for all the processes included in the global PDF fit.

The PDF constraints provided by the ATLAS and CMS top-quark differential distributions at $\sqrt{s} = 8\,\text{TeV}$ will be investigated by means of the NNPDF global analysis framework [6, 37]. For the baseline PDF fit, the input dataset will be largely the same as in NNPDF3.0, with two main differences: the HERA-II structure function data from H1 and ZEUS has been replaced by the final HERA combination [38, 39] and inclusive jet production measurements from CDF, ATLAS and CMS have been excluded. In order to achieve the computational speed required for the PDF fit, we generate theoretical calculations of NLO top-quark pair production with `Sherpa` [40] interfaced to `MCgrid` [41] and dynamical scales as in ref. [24]. These NLO calculations are then supplemented with NNLO/NLO bin-by-bin K-factors consistently derived using the theory settings of [24].

Including the LHC differential distributions from top-quark pair production into the NNPDF global analysis allows us to quantitatively tackle a number of important issues. In particular, we investigate the compatibility between the ATLAS and CMS measurements; how the constraints provided by the differential measurements compare to those obtained from inclusive cross-sections; whether it is advantageous to use normalized or absolute distributions; and which is the optimal combination of LHC top-quark measurements to be included in the global PDF fit. We then demonstrate how differential distributions from top-quark pair production lead to a significant reduction of the gluon PDF uncertainty at large x, and that their impact is comparable to that obtained from inclusive jet measure-

JHEP04(2017)044

ments. The resulting improved gluon will have direct beneficial implications for searches of new physics beyond the Standard Model (BSM) in final states involving top quarks and in general for gluon-initiated processes.

This paper is organized as follows. In section 2 we describe the available LHC top-quark pair production data and the treatment of their experimental uncertainties. In section 3 we discuss the calculation of the NNLO theoretical predictions for top-quark pair differential cross-sections and provide a systematic comparison between them and the LHC data. In section 4 we present NNLO fits including top-quark differential distributions, assess the agreement between data and theory, and discuss their impact in the determination of the large-x gluon. In section 5 we summarize and comment on possible future developments. Further investigations on the compatibility between the ATLAS and CMS data are presented in appendix A.

2 Experimental data

In this section we describe the top-quark pair production data from ATLAS and CMS that will be used as input in the PDF fit. First, we describe the features of the various differential distributions available, including a comparison between absolute and normalized measurements, and then we review the total inclusive cross-sections that will be included alongside the normalized differential distributions.

2.1 Top-quark pair differential distributions from the LHC

In this work we consider the most recent differential cross-section measurements on top-quark pair production at $\sqrt{s} = 8$ TeV from ATLAS [28] and CMS [29] in the lepton+jets final state. These datasets correspond to an integrated luminosity of $20.3\,\text{fb}^{-1}$ and $19.7\,\text{fb}^{-1}$, respectively. In this channel, the $t\bar{t}$ pair is reconstructed from its decays into $W^+bW^-\bar{b}$, with one W boson decaying hadronically and the other into an electron or muon and the associated neutrino. We do not consider earlier measurements at $\sqrt{s} = 7\,\text{TeV}$ [42–45], which are affected by larger uncertainties and are not provided with the full breakdown of systematic error sources.

The ATLAS and CMS top-quark production measurements of refs. [28, 29] are provided in both the fiducial phase space, with observables reconstructed in terms of final-state leptonic and jet variables, and in the full phase space, in terms of the top or top-pair kinematic variables. In our analysis, we are restricted to using the latter as the NNLO calculations are available only for stable top quarks. Ongoing work into extending these calculations to include top-quark decays will eventually overcome this restriction. Among the available distributions, we will focus on the transverse momentum p_T^t and the rapidity y_t of the top quark or antiquark, and on the rapidity $y_{t\bar{t}}$ and the invariant mass $m_{t\bar{t}}$ of the top-quark pair system. We will not consider the transverse momentum of the top-quark pair $p_T^{t\bar{t}}$, for which a complete NNLO theoretical description is not available. The binning and kinematical cuts for each distribution are the same in the ATLAS and CMS measurements, a feature which simplifies the benchmarking of results between the two experiments and their comparison with the theoretical predictions.

JHEP04(2017)044

Exp.	Dataset	Sys. Unc.	$N_{\rm dat}$	Kinematics
ATLAS	ATLAS $d\sigma/dp_T^t$	a	8	$0 < p_T^t < 500\,\text{GeV}$
	ATLAS $d\sigma/d\lvert y_t\rvert$	a	5	$0 < \lvert y_t\rvert < 2.5$
	ATLAS $d\sigma/d\lvert y_{t\bar{t}}\rvert$	a	5	$0 < \lvert y_{t\bar{t}}\rvert < 2.5$
	ATLAS $d\sigma/dm_{t\bar{t}}$	a	7	$345 < m_{t\bar{t}} < 1600\,\text{GeV}$
	ATLAS $(1/\sigma)d\sigma/dp_T^t$		8	$0 < p_T^t < 500\,\text{GeV}$
	ATLAS $(1/\sigma)d\sigma/d\lvert y_t\rvert$		5	$0 < \lvert y_t\rvert < 2.5$
	ATLAS $(1/\sigma)d\sigma/d\lvert y_{t\bar{t}}\rvert$		5	$0 < \lvert y_{t\bar{t}}\rvert < 2.5$
	ATLAS $(1/\sigma)d\sigma/dm_{t\bar{t}}$		7	$345 < m_{t\bar{t}} < 1600\,\text{GeV}$
CMS	CMS $d\sigma/dp_T^t$	b, f	8	$0 < p_T^t < 500\,\text{GeV}$
	CMS $d\sigma/dy_t$	c, f	10	$-2.5 < y_t < 2.5$
	CMS $d\sigma/dy_{t\bar{t}}$	d, f	10	$-2.5 < y_{t\bar{t}} < 2.5$
	CMS $d\sigma/dm_{t\bar{t}}$	e, f	7	$345 < m_{t\bar{t}} < 1600\,\text{GeV}$
	CMS $(1/\sigma)d\sigma/dp_T^t$	b	8	$0 < p_T^t < 500\,\text{GeV}$
	CMS $(1/\sigma)d\sigma/dy_t$	c	10	$-2.5 < y_t < 2.5$
	CMS $(1/\sigma)d\sigma/dy_{t\bar{t}}$	d	10	$-2.5 < y_{t\bar{t}} < 2.5$
	CMS $(1/\sigma)d\sigma/dm_{t\bar{t}}$	e	7	$345 < m_{t\bar{t}} < 1600\,\text{GeV}$

Table 1. The ATLAS and CMS top-quark pair distributions at $\sqrt{s} = 8\,\text{TeV}$ used in this work. For each distribution we indicate the number of data points and their kinematic coverage. In the second column, distributions that are labeled with the same letter have common experimental systematic uncertainties.

In table 1 we summarize the features of each kinematical distribution, indicating whether it is an absolute or a normalized distribution; which of the correlated systematic errors are common; the number of data points $N_{\rm dat}$; and their kinematic coverage. All systematic uncertainties are treated as multiplicative, and absolute distributions share the luminosity uncertainty across each experiment. Moreover, the absolute distributions from CMS also share the same systematic uncertainties of their corresponding normalized distributions (see below). Wherever asymmetric uncertainties are provided, they are symmetrized according to [46]. To avoid double counting, for each experiment only one of the distributions listed in table 1 can be included in a PDF fit, due to the unavailability of the statistical correlations between different distributions within the same experiment. One of the goals of this study is therefore to identify the combination of ATLAS and CMS top-quark pair measurements that maximizes the constraints on the gluon.

In addition to these 8 TeV lepton+jets kinematical distributions, ATLAS and CMS have presented other differential measurements of top-quark pair production. To begin with, differential distributions at $\sqrt{s} = 8\,\text{TeV}$ for the dilepton final state are available from both ATLAS and CMS [29, 47], which in the latter case are also presented in the form of double-differential normalized cross-sections [48]. In addition, measurements of differential distributions of high-p_T boosted top quarks from ATLAS [49] and CMS [50] at

JHEP04(2017)044

$\sqrt{s} = 8\,\text{TeV}$ have also been published, although their interpretation requires an assessment of electroweak corrections [51]. Finally, results on differential distributions at $\sqrt{s} = 13\,\text{TeV}$ in the lepton+jets channel by ATLAS [52] and in the dilepton [48] and lepton+jets channels [53] from CMS are now also available. In this first exploratory work, we concentrate on the most precise data available, the lepton+jets distributions at 8 TeV, but future studies should include also these other available top-quark differential measurements.

2.2 Absolute versus normalized distributions

As indicated in table 1, ATLAS and CMS have presented their measurements of top-quark pair differential distributions in two different ways. In the first case, each distribution is normalized to the sum over the cross-sections in each bin, in a way that it then integrates to one by construction. This procedure is motivated by the partial cancellation of uncertainties, such as the luminosity, that takes place in the ratio. However, some PDF-sensitive information describing the overall normalization of the gluon PDF might be lost in this procedure. In order to compensate for this, the PDF fit should include both total inclusive cross-sections and normalized differential distributions. Typically, the mutual correlation between the two is small and can be neglected.

On the other hand, top-quark pair differential measurements are also provided as absolute distributions. In this case, experimental uncertainties are larger than in the normalized case, though this way one also maintains a handle on the overall magnitude of the gluon. Note that for absolute distributions, the simultaneous inclusion of total and differential measurements would result in a double counting. While constraints arising from the use of either normalized distributions supplemented with total cross-sections or absolute distributions should be equivalent, it turns out that the former are somewhat more stringent than the latter (see section 4).

ATLAS has released measurements for both normalized and absolute distributions, and provided the corresponding full breakdown of systematic uncertainties separately. The former are affected by an additional 2.8% fully correlated uncertainty from the integrated luminosity at 8 TeV. The CMS measurements are available only for the normalized distributions, from which the absolute differential distributions can be reconstructed by means of the corresponding total cross-section measurement [54]. In this procedure, statistical uncertainties from the normalized distribution and the total inclusive cross-section are added in quadrature. Two additional sources of systematics are retained on the absolute differential distribution, which originate respectively from the total systematic and the luminosity uncertainties of the inclusive cross-section.[1]

2.3 Total inclusive cross-section measurements

The LHC measurements of normalized top-quark pair differential distributions benefit from reduced experimental uncertainties as compared to their absolute counterparts, but consequently they might also lose some sensitivity on the overall magnitude of the gluon. It is

[1] We thank the conveners of the CMS Top Quark Physics group for providing us with this recommendation.

JHEP04(2017)044

Exp.	$\sqrt{s}$ [TeV]	Fin. st.	$\mathcal{L}$ [fb^{-1}]	$\sigma^{\rm tot}(t\bar{t})$ [pb]	Ref.
ATLAS	7	l+jets	4.6	182.9 ± 3.1 (stat) ± 4.2 (sys) ± 3.6 (lumi) ± 3.3 (bm)	[55]
	8	**l+jets**	**20.3**	$\mathbf{242.4 \pm 1.7}$ **(stat)** $\pm$ **5.5 (sys)** $\pm$ **7.5 (lumi)** $\pm$ **4.2 (bm)**	[55]
	13	l+jets	3.2	818 ± 8 (stat) ± 27 (sys) ± 19 (lumi) ± 12 (bm)	[56]
CMS	7	l+jets	5.0	173.6 ± 2.1 (stat)$^{+4.5}_{-4.0}$ (sys) ± 3.8 (lumi)	[57]
	8	**l+jets**	**19.7**	$\mathbf{244.9 \pm 1.4}$ **(stat)**$^{\mathbf{+6.3}}_{\mathbf{-5.5}}$ **(sys)** $\pm$ **6.4 (lumi)**	[57]
	13	l+jets	2.2	792 ± 8 (stat) ± 37 (sys) ± 21 (lumi)	[58]

Table 2. Summary of the most precise ATLAS and CMS measurements of the total inclusive $t\bar{t}$ cross-sections at 7, 8 and 13 TeV. We indicate the final state, the integrated luminosity, the breakdown of statistical and systematic uncertainties (where "lumi" stands for the luminosity and "bm" stands for the beam energy). The measurements in boldface are those used in the fits of this work.

therefore important to supplement the normalized distributions included in the PDF fits with the corresponding measurements of the inclusive cross-section in order to obtain a complete picture.

In table 2 we collect the results for the most precise ATLAS and CMS measurements of the total inclusive top-quark pair cross-section at various center-of-mass energies. In each case, we indicate the final state, the integrated luminosity, the value of the total cross-section with the breakdown of statistical and systematic uncertainties (where "lumi" stands for the luminosity and "bm" stands for the beam energy), and the corresponding publication reference. These measurements (with the exception of the 13 TeV measurement) have a total experimental uncertainty of only a few percent. The 8 TeV cross-sections are notably limited by the luminosity uncertainty, which amounts to 2.8% and 2.6% for ATLAS and CMS respectively.

As a general rule, in a global fit it is advantageous to include as many PDF-sensitive observables as possible. In the particular case of fits including top-quark production data, one should then add all the total cross-sections listed in table 2 as well as available measurements in other final states. However, one of the aims of this work is to compare the impact on the large-x gluon of top-quark pair production at 8 TeV, arising from either absolute distributions or from the normalized ones supplemented with the corresponding total cross-sections. To perform such a comparison consistently, we include here only the total cross-sections at 8 TeV from the lepton+jets final state, highlighted in boldface in table 2. Therefore, in the following, whenever one of the ATLAS or CMS normalized differential listed in table 1 is included in the PDF fit, it will always be supplemented by the corresponding 8 TeV total cross-section from table 2.

3 Comparison between NNLO theory and LHC data

In this section, first we provide details on the theory settings used for the calculation of NNLO differential distributions in top-quark pair production. Then we perform a qualitative comparison between the predictions obtained from various NNLO PDF sets and the

JHEP04(2017)044

8 TeV ATLAS and CMS data, for both absolute and normalized distributions. Finally, we quantify these comparisons by means of a χ^2 estimator.

3.1 Differential top-quark pair production at NNLO

The calculation of the NNLO QCD corrections to differential distributions in top-quark pair production has been recently completed [23, 24]. This calculation is however not yet available in a format suitable for its direct inclusion during a global fit, which requires the evaluation of hadronic cross-sections for different input PDFs a large number of times during the minimization procedure. Therefore, in order to include this data into the global NNLO PDF fit, we begin by using fast calculations of NLO matrix elements with NNLO DGLAP evolution and $\alpha_s(Q)$ running. These fast NLO calculations are based upon precomputing the partonic matrix elements in such a way that the standard numerical convolution with generic input PDFs can be reliably approximated by means of interpolation techniques.

There exist two main frameworks for the implementation of this fast convolution procedure, `APPLgrid` [59] and `FastNLO` [60]. In this work we will utilize the former, which has been interfaced to various codes of common use for calculations in PDF fits, such as `NLOjet++` [61], `MCFM` [62], `MadGraph5_aMC@NLO/aMCfast` [63, 64] and `SHERPA` [40]. In particular, here we will use `SHERPA` interfaced to `APPLgrid` using the `MCgrid` code [41] and the `Rivet` [65] analysis package, with `OpenLoops` [66] for the NLO matrix elements. The calculations have been performed with Monte Carlo integration statistics sufficiently large in order to ensure that the residual fluctuations are at the few permille level at most. The NLO `SHERPA/MCgrid` results have been benchmarked with the corresponding calculation using the code of [24], finding excellent agreement for all kinematic distributions.

An important aspect of the NNLO calculation is the choice of central renormalization and factorization scales, μ_R and μ_F. Following ref. [24], the following optimized settings are adopted in this work. For the differential distributions in the top (or anti-top) quark rapidity y_t and in the top-quark pair rapidity $y_{t\bar{t}}$ and invariant mass $m_{t\bar{t}}$ we use

$$\mu_R = \mu_F = \mu = H_T/4\,, \qquad H_T \equiv \sqrt{m_t^2 + \left(p_T^t\right)^2} + \sqrt{m_t^2 + \left(p_T^{\bar{t}}\right)^2}\,, \tag{3.1}$$

where $m_t = 173.3\,\text{GeV}$ is the PDG world average for the top-quark pole mass [67], and p_T^t ($p_T^{\bar{t}}$) is the top (anti-top) transverse momentum. For the top-quark transverse momentum distribution, constructed from the average of the distributions for the top and the anti-top quarks, it has been found that the optimal choice of dynamical scales for the former case is

$$\mu'_R = \mu'_F = \mu' = \sqrt{m_t^2 + \left(p_T^t\right)^2}\Big/2\,, \tag{3.2}$$

with an analogous expression for anti-top quarks (replacing p_T^t by $p_T^{\bar{t}}$). This scale choice leads to an improvement in the convergence of the perturbative series.

The resulting NLO calculations are then supplemented by bin-by-bin $\mathcal{C}$-factors [68], defined as the ratio of the NNLO to NLO calculations,

$$\mathcal{C} = \frac{\widetilde{\sigma}^{\text{nnlo}} \otimes \mathcal{L}^{\text{nnlo}}}{\widetilde{\sigma}^{\text{nlo}} \otimes \mathcal{L}^{\text{nnlo}}}\,, \tag{3.3}$$

JHEP04(2017)044

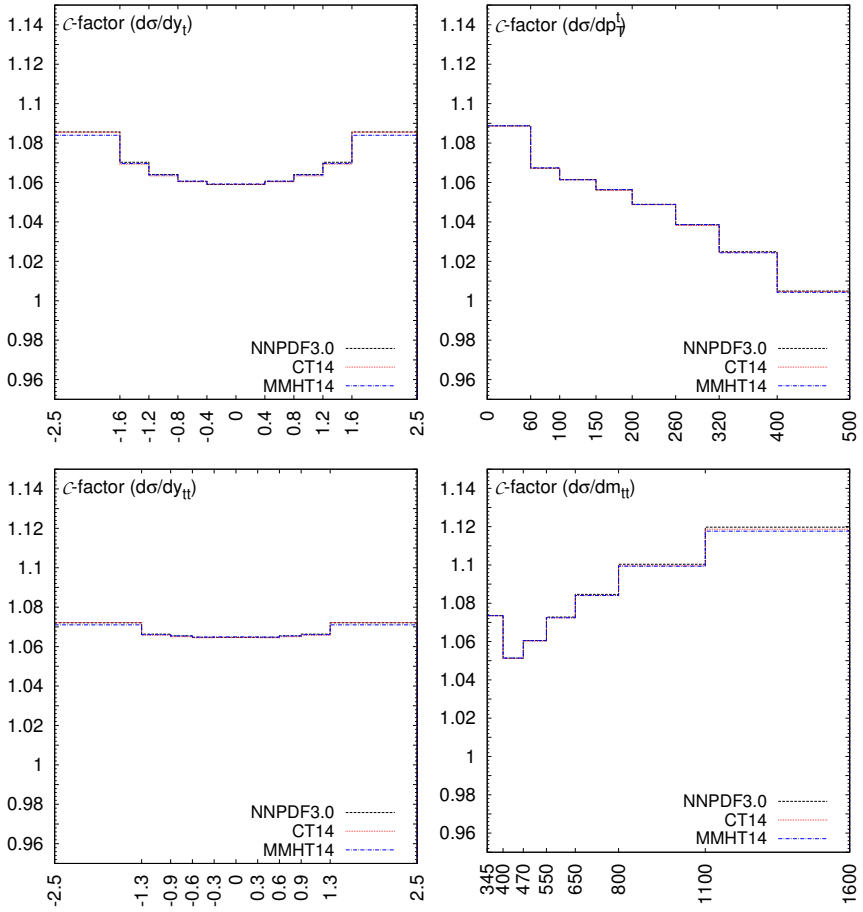

Figure 1. The $\mathcal{C}$-factors eq. (3.3) for the four absolute differential distributions of table 1.

where $\widetilde{\sigma}^{\rm nnlo}$ ($\widetilde{\sigma}^{\rm nlo}$) is the partonic cross-section computed with NNLO (NLO) matrix elements and $\mathcal{L}^{\rm nnlo}$ is the corresponding parton luminosity evaluated with a reference set of NNLO PDFs. The numerator and the denominator in eq. (3.3) were computed with the code of [24].

In figure 1 we compare the $\mathcal{C}$-factors computed with NNPDF3.0 [6], CT14 [8] and MMHT2014 [9] for the absolute differential distributions in the following four variables: the top quark rapidity y_t and transverse momentum p_T^t, and the top-pair rapidity $y_{t\bar{t}}$ and invariant mass $m_{t\bar{t}}$. The binning here is the same as that of the ATLAS and CMS 8 TeV measurements listed in table 1. We find that the dependence of the $\mathcal{C}$-factors on the input PDF set is very small and can be safely neglected. In the case of the y_t and $y_{t\bar{t}}$ distributions, we find NNLO corrections of between 6% and 9%, reasonably flat in the data region. For the p_T^t distribution, the $\mathcal{C}$-factor decreases from 1.09 at low transverse momentum to close to unity for $p_T^t \simeq 500\,\text{GeV}$. For the invariant mass $m_{t\bar{t}}$, the $\mathcal{C}$-factor increases from 5% at low masses to around 12% above 1 TeV.

We note that, exactly as for the corresponding experimental measurements, all NNLO distributions have been normalized with respect to the cross-section integrated over the considered kinematic range. In other words, by construction, the integral of any normalized distribution over its kinematic range is unity.

JHEP04(2017)044

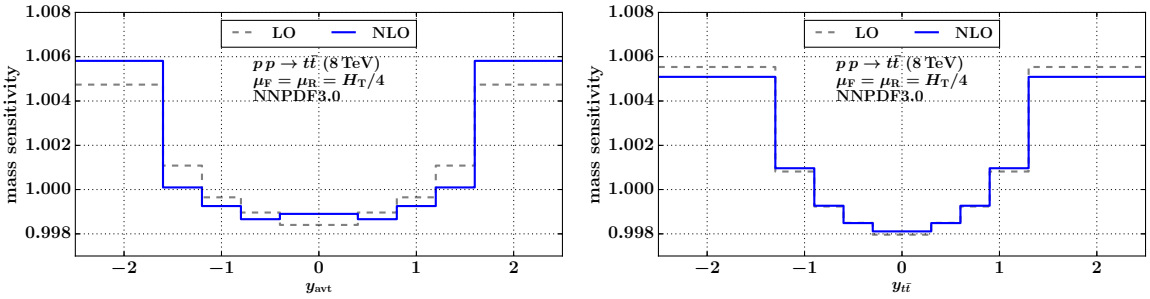

Figure 2. The ratio between $m_t = 172.3\,\text{GeV}$ and $m_t = 173.3\,\text{GeV}$ ("mass sensitivity") at LO and NLO for the normalized y_t (left) and $y_{t\bar{t}}$ (right) distributions at 8 TeV, computed using NNPDF3.0.

As shown in ref. [24], the integration of the differential distributions computed with the optimal dynamical scales eqs. (3.1)–(3.2) returns a total cross-section which is about 2% higher than the NNLO one from `top++` [69], and in close agreement with the NNLO+NNLL `top++` result (recall that the total cross-section in `top++` is computed with fixed scales $\mu_R = \mu_F = m_t$). For this reason, when adding the inclusive cross-section data into PDF fits, it is more appropriate to compute the theory prediction with `top++` at NNLO+NNLL. Nonetheless, in the present work the total inclusive top-pair cross-section and corresponding $\mathcal{C}$-factors are computed using `top++` at NNLO. As explained in section 4.3, and given the exploratory nature of the present work, this choice is adequate since the overall impact of the total cross-sections on the global fits turns out to be small and this 2% difference is thus inconsequential for our study.

The theoretical uncertainties due to the value of m_t deserve special attention. As mentioned above, in this work we use the PDG average of $m_t = 173.3\,\text{GeV}$. The significant spread among the individual measurements contributing to this average, however, suggests that in the future a shift in m_t of up to $\Delta m_t \simeq 1\,\text{GeV}$, or even more, may be possible. The sensitivity upon variations of m_t of the four top-quark differential distributions considered here has been studied in [70]. Shape modifications are pronounced in the $m_{t\bar{t}}$ and p_T^t distributions, especially close to the threshold. On the other hand, the y_t and $y_{t\bar{t}}$ distributions exhibit a much reduced m_t dependence.

To quantify this mass sensitivity, in figure 2 we show the ratio between $m_t = 172.3\,\text{GeV}$ and the PDG average $m_t = 173.3\,\text{GeV}$ for the LO and NLO normalized y_t and $y_{t\bar{t}}$ distributions at 8 TeV. We find that these two distributions are very stable upon a shift of m_t by 1 GeV, varying at most by 0.6%, which is much less than the experimental uncertainties or other sources of theory uncertainty such as PDFs and missing higher orders. This robustness of the normalized y_t and $y_{t\bar{t}}$ distributions with respect to m_t variations is, therefore, an important motivation in favour of using them as input to the PDF fits (see section 4.4).

The region of x for which the LHC differential top data are sensitive to the various PDF flavours can be quantified by computing the correlation coefficient ρ between them and each of the bins of a given differential distribution [71, 72]. Large values of $|\rho|$ indicate regions in x where the top-quark data provide direct sensitivity to each PDF flavour. These correlations are shown in figure 3, for the gluon $g(x, Q^2)$, and in figure 4, for quarks $q(x, Q)$, $q = u, \bar{u}, d, \bar{d}, s, \bar{s}, c, b$. PDFs are evaluated at $Q = 100\,\text{GeV}$ from the NNPDF3.0 NNLO set.

JHEP04(2017)044

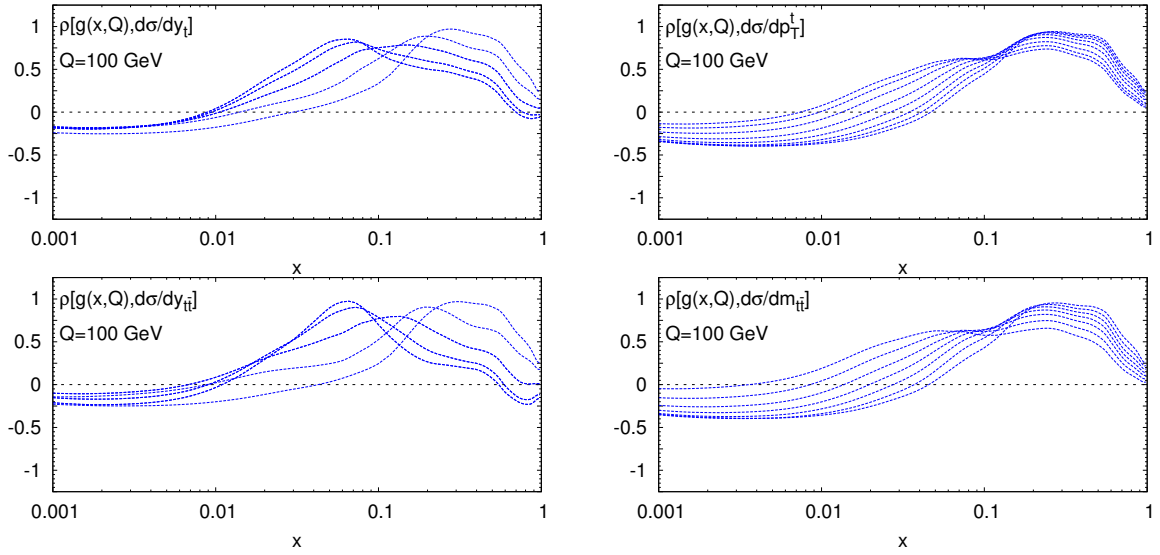

Figure 3. The correlation coefficient ρ between the gluon $g(x, Q^2)$, evaluated at $Q = 100\,\text{GeV}$, and each of the bins of the y_t, p_T^t, $y_{t\bar{t}}$ and $m_{t\bar{t}}$ top-quark differential distributions at the LHC 8 TeV.

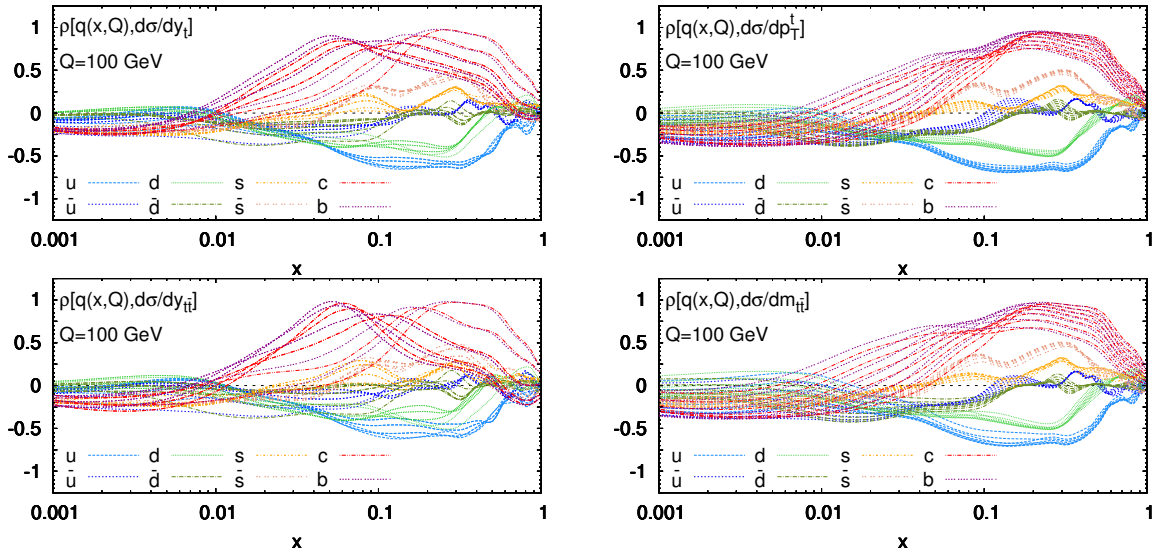

Figure 4. Same as figure 3, but for quarks and antiquarks, $q(x, Q^2)$, $q = u, \bar{u}, d, \bar{d}, s, \bar{s}, c, b$.

In the case of the gluon, we find that already for $x \gtrsim 0.05$ the correlation coefficient can be larger than 0.5, while it peaks in the region between $x \simeq 0.08$ and $x \simeq 0.5$, depending on the specific bin and kinematical distribution. A similar trend is observed for the charm and bottom quarks, as a consequence of the fact that they are generated radiatively through the gluon splitting in a quark-antiquark pair. In the case of light quarks and antiquarks, moderate correlations are observed for u and d, while correlations are almost negligible for $\bar{u}$, $\bar{d}$, s and $\bar{s}$. As we will show in section 4, top-quark data will mostly constrain the gluon, and, as a consequence, the radiatively generated charm and bottom quarks, in the x region where the correlation coefficient $|\rho|$ is larger, roughly $0.08 \lesssim x \lesssim 0.5$.

JHEP04(2017)044

3.2 Comparison with the ATLAS and CMS differential distributions

In order to assess the agreement between the data and the NNLO theoretical predictions based on our current knowledge of PDFs, we perform now a systematic comparison of the calculations described in the previous section and the ATLAS and CMS measurements. This comparison is performed at the level of both absolute and normalized distributions, allowing for an improved understanding of the differences and similarities between PDF sets. This way, one can separate differences induced by the shape of the gluon from those induced by its normalisation.

The NNLO differential distributions with the binning of the ATLAS and CMS measurements have been computed using five different PDFs sets: NNPDF3.0, CT14, MMHT2014, HERAPDF2.0 [38] and ABM12 [7], in the last case with the $n_f = 5$ version. For all these PDF sets, we consistently use the same value of the strong coupling constant as in the NNLO matrix elements. This corresponds to $\alpha_s(m_Z) = 0.118$ for all sets except for ABM12, for which PDFs are only available for their best-fit value of $\alpha_s(m_Z) = 0.113$.

In figure 5 we show the NNLO predictions for the absolute (left) and normalized (right) p_T^t differential distributions compared to the corresponding CMS and ATLAS measurements. The theory calculations are provided for NNPDF3.0, CT14, and MMHT14 and include only PDF uncertainties. The data uncertainties correspond to the square root of the diagonal elements of the experimental covariance matrix. At a qualitative level, we find that the theory calculations based on the three PDF sets used in this comparison are in good agreement both among themselves and with the data. We also see that while at the level of normalized cross-sections the experimental uncertainties are similar between ATLAS and CMS, there are larger differences for absolute distributions. Moreover, we note that the ATLAS and CMS measurements exhibit some degree of tension.

Next, in figure 6 we show the same comparison but now among NNPDF3.0, HERAPDF2.0 and ABM12. In the case of HERAPDF2.0, the PDF error band is the sum in quadrature of the statistical, model and parametrization uncertainties. We note that while HERAPDF2.0 and NNPDF3.0 agree well, in particular for the normalized distribution, this is not the case for ABM12, whose predictions are substantially lower than those of the other PDF sets. This effect is more pronounced for the absolute distributions, and reflects intrinsic differences both in the gluon-gluon luminosity and in the value of $\alpha_s(m_Z)$. We will show that this trend reappears for other kinematical distributions. These differences between ABM12 and the other PDF sets cannot be accommodated by a shift in the value of m_t used. As noted in ref. [70], the sensitivity of the p_T^t absolute differential distribution on the value of m_t is very non-uniform across the whole p_T^t data range. In order for ABM12 to fit the data at the lowest p_T^t, one should require an unreasonably small value of m_t, roughly around $m_t = 169\,\text{GeV}$. However, even with such a shift of m_t, the large p_T^t tail of the distribution will hardly move at all. Therefore, the shape of the ABM12 theoretical prediction will become even more different than that of the measured p_T^t absolute differential distribution. This should remain true also for the normalised p_T^t distribution, since its shape will shift similarly to the absolute one.

JHEP04(2017)044

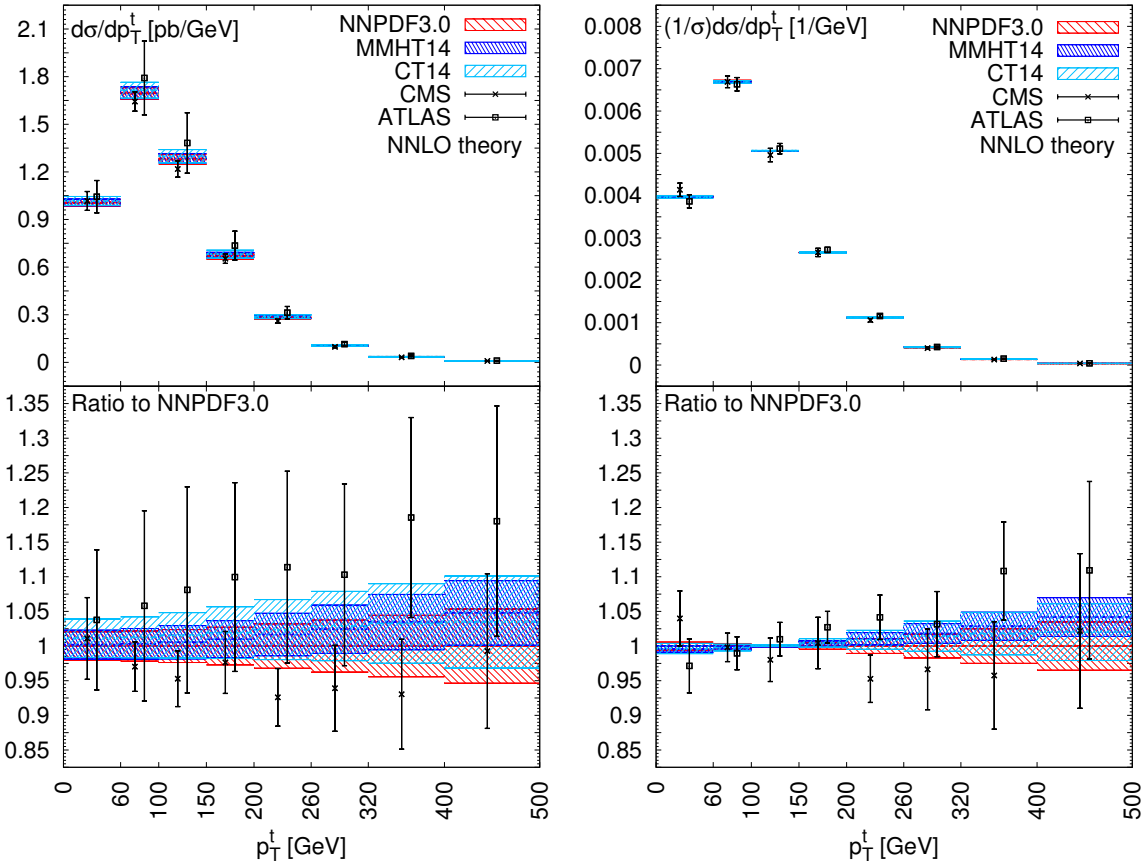

Figure 5. Comparison between the NNLO predictions for the absolute (left) and normalized (right) p_T^t differential distributions in top-quark pair production and the corresponding CMS and ATLAS measurements. The theoretical predictions have been computed with the NNPDF3.0, CT14 and MMHT14 sets and include only the 1–σ PDF uncertainties, while scale uncertainties are not shown. In the lower panels, we display the same results now as the ratio to the central NNPDF3.0 prediction.

In figure 7 we consider now the top quark rapidity distribution, y_t. Here too we find a good agreement among NNPDF3.0, CT14 and MMHT14, both for the absolute and for the normalized distributions. For forward rapidities, the PDF uncertainty in NNPDF3.0 is larger than that of the other two PDF sets. For this distribution, while CMS and ATLAS are consistent in the absolute case, in the normalized case we again observe some discrepancies between the two experiments in the central region. As we will show, this results in some difficulty in being able to achieve a satisfactory fit of the distributions from both experiments simultaneously.

The corresponding comparisons between theory predictions and data for y_t, now among NNPDF3.0, ABM12 and HERAPDF2.0, are shown in figure 8. For the absolute distribution, HERAPDF2.0 is between 5% and 10% lower than NNPDF3.0, with ABM12 lower by a larger amount, between 20% and 30%. These differences are reduced (but then the experimental uncertainties are smaller as well) in the normalized case, where now ABM12 is above NNPDF3.0 and HERAPDF2.0 in the central region and undershoots them in the forward rapidity bins. As we show below, these differences translate into a poor χ^2 when the ABM12 predictions are compared with the experimental data.

JHEP04(2017)044

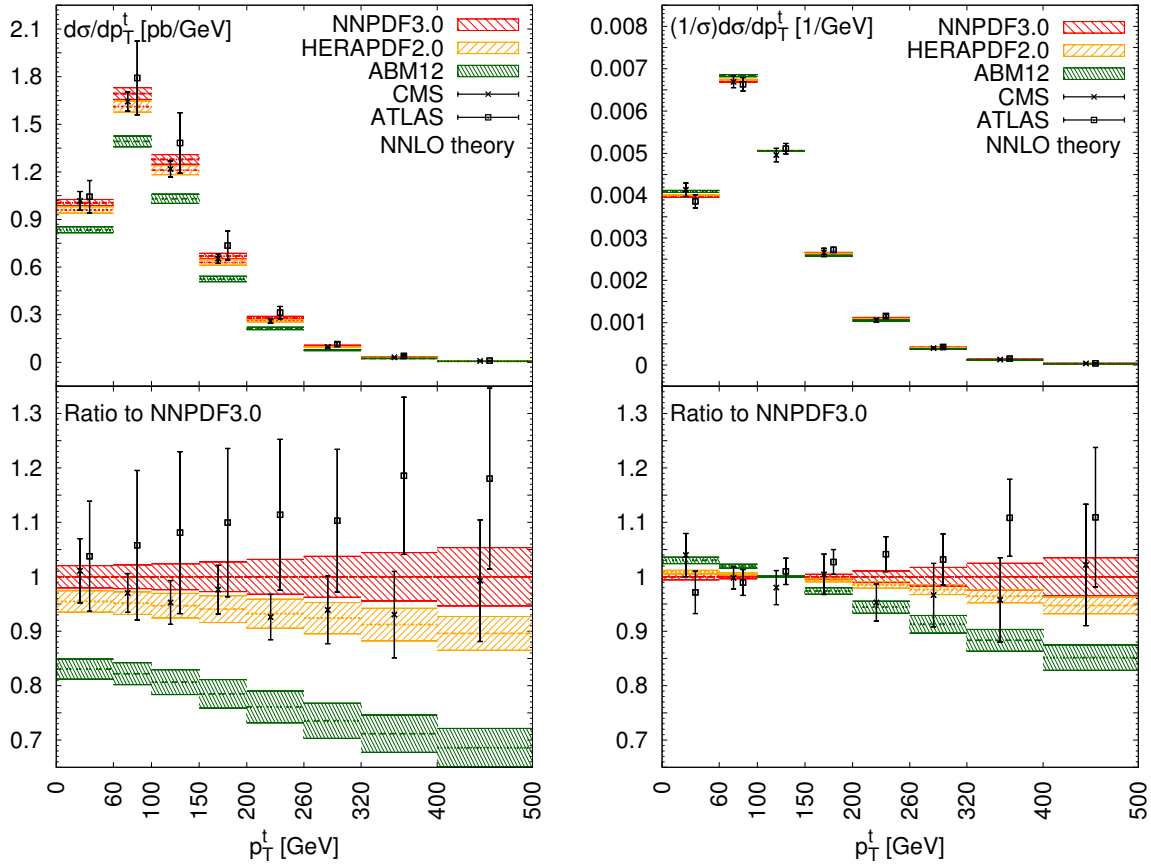

Figure 6. Same as figure 5 for the HERAPDF2.0 and ABM12 PDF sets.

We now move to consider the comparison between data and theory for the kinematical distributions constructed from the top-quark pair kinematics, in particular the rapidity $y_{t\bar{t}}$ and the invariant mass $m_{t\bar{t}}$ of the pair. First of all, in figure 9 we compare the ATLAS and CMS $y_{t\bar{t}}$ measurements with the corresponding NNLO predictions obtained using NNPDF3.0, CT14 and MMHT14. Interestingly, unlike the cases of the p_T^t and y_t distributions, the ATLAS and CMS $y_{t\bar{t}}$ measurements are now in good agreement, both at the level of absolute and normalized distributions, both in the central and forward regions. As we will show, this has the important consequence that $y_{t\bar{t}}$ is the only distribution that can be satisfactorily described when ATLAS and CMS datasets are included together in the same fit. Both for the absolute and the normalized distributions, the theory predictions for $y_{t\bar{t}}$ with the three PDF sets in figure 9 are consistent at the one-sigma level (in units of the PDF uncertainty), and are also in reasonable agreement with the experimental data. As in the case of the y_t distribution, for forward rapidities the PDF uncertainties from NNPDF3.0 are larger than those of the other two sets.

In figure 10 we show the corresponding comparison for the $y_{t\bar{t}}$ distributions, finding a similar trend as in the y_t case in figure 8. For the absolute distribution, HERAPDF2.0 is somewhat lower than NNPDF3.0, with almost touching error bands (this translates into a $\sqrt{2}$ sigma discrepancy between the two sets); ABM12 is lower by an amount between 15% and 30% depending on the specific bin. In the normalized distribution, ABM12

JHEP04(2017)044

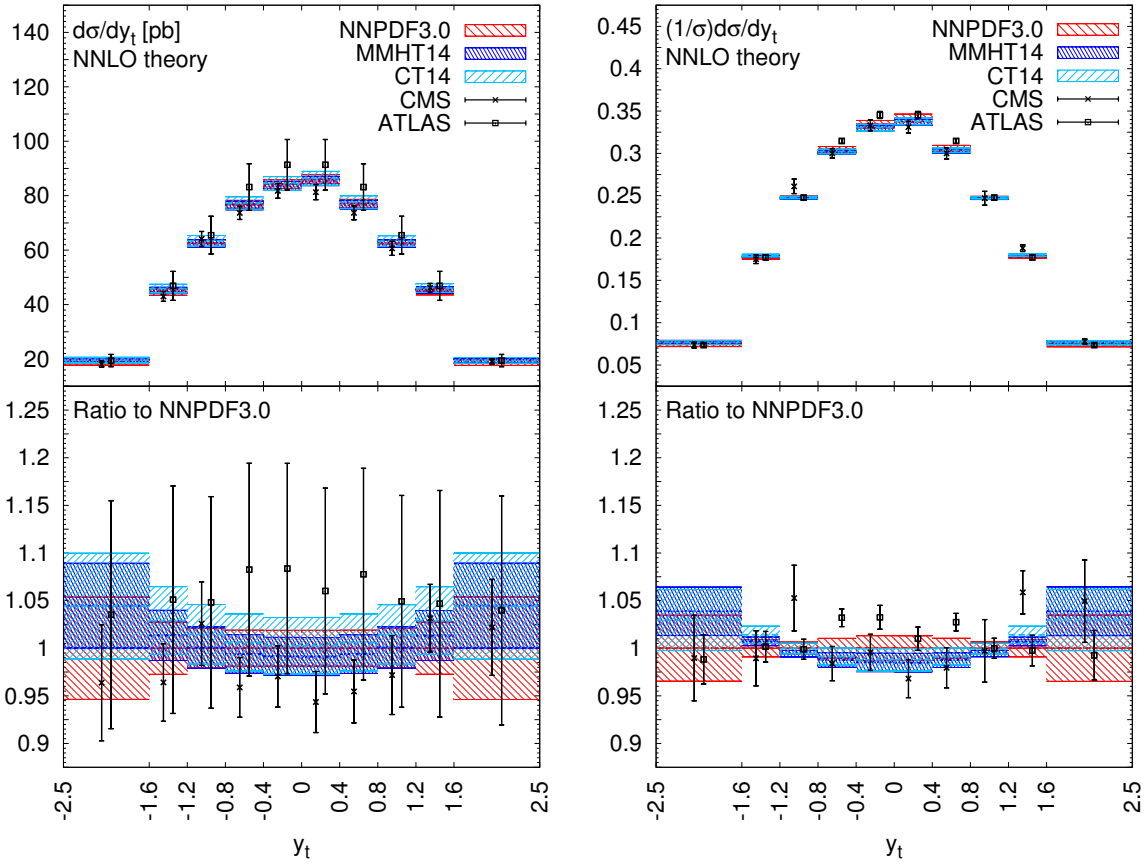

Figure 7. Same as figure 5 for the top quark rapidity distribution y_t.

overshoots the predictions of the other two PDF sets and the data for central rapidities and undershoots them in the forward region.

Finally we consider the differential distribution in the invariant mass of the top-antitop pair, $m_{t\bar{t}}$. An accurate theoretical and experimental understanding of this distribution is crucial in many searches for BSM physics, where new states couple to top quarks. A prime example would be the case of heavy resonances that decay into a $t\bar{t}$ pair. Such decays would appear in the data as an excess in the invariant mass distribution [70, 73–75].

In figure 11 we show the NNLO predictions for the invariant mass distribution of the top-antitop pair, $m_{t\bar{t}}$, using NNPDF3.0, CT14 and MMHT14. The first thing to note is the difference between the ATLAS and CMS measurements, especially in the absolute distribution and for intermediate values of $m_{t\bar{t}}$. The difference in the size of the overall experimental uncertainties is also significant. For instance, despite being based on the same integrated luminosity, the ATLAS uncertainty in the highest $m_{t\bar{t}}$ bin is about four times larger than that of CMS. We also find that the three PDF sets are in good agreement within uncertainties, with NNPDF3.0 exhibiting a somewhat lower central value and larger uncertainties at high $m_{t\bar{t}}$ as compared to the other two sets. While the three PDF sets agree qualitatively with the ATLAS measurements, there seems to be some tension with the CMS data, which exhibits lower central values in the intermediate and high $m_{t\bar{t}}$ regions and has smaller experimental uncertainties.

JHEP04(2017)044

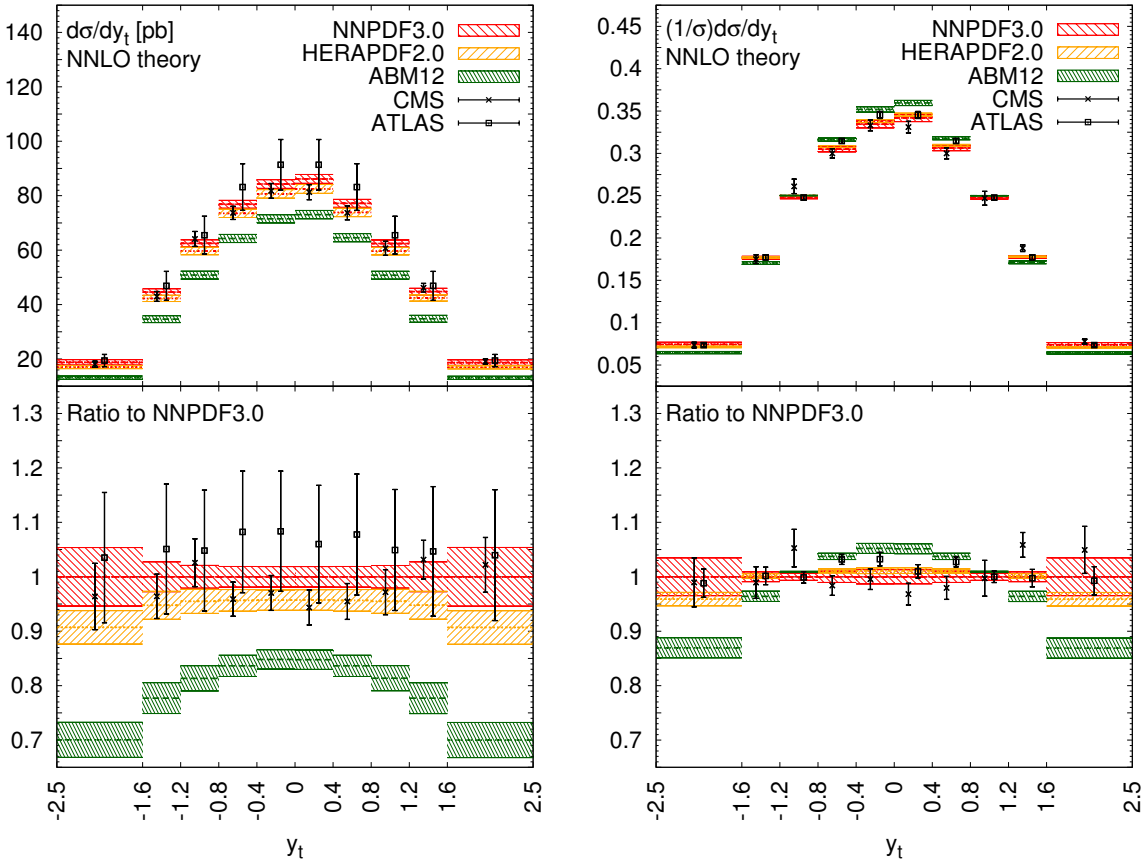

Figure 8. Same as figure 7 now for NNPDF3.0, ABM12 and HERAPDF2.0.

The corresponding comparison between NNPDF3.0, ABM12 and HERAPDF2.0 is shown in figure 12, from which we observe common trends in the absolute and normalized distributions. The HERAPDF2.0 prediction are lower than the NNPDF3.0 ones, with ABM12 being even lower, by up to 40% (25%) in the highest $m_{t\bar{t}}$ bin of the absolute (normalized) distribution. Given that the ATLAS and CMS measurements seem to be pulling in opposite directions, the latter is favored by the ABM12 prediction, while the former is in better agreement with NNPDF3.0 and HERAPDF2.0.

Before moving to a more quantitative assessment of the agreement between data and theory, we would like to compare the NNLO calculations with the experimental measurements of the total cross-section listed in table 2. This comparison is useful because inclusive data provide information on the overall normalisation of the gluon for the cases where normalized distributions are fitted. In figure 13 we show the inclusive cross-sections from ATLAS and CMS at different center-of-mass energies, compared to NNLO theory computed with `top++` for the five PDF sets. Results are shown as ratios to the central NNPDF3.0 predictions. The comparison follows the trend observed at the level of absolute differential distributions, with NNPDF3.0, MMHT14 and CT14 in good agreement both among themselves and with the LHC measurements. On the other hand, HERAPDF2.0 and ABM12 predict cross-sections that are lower by about 6% (4%) and 20% (15%), respectively, at 7 and 8 TeV (13 TeV) as compared to NNPDF3.0.

JHEP04(2017)044

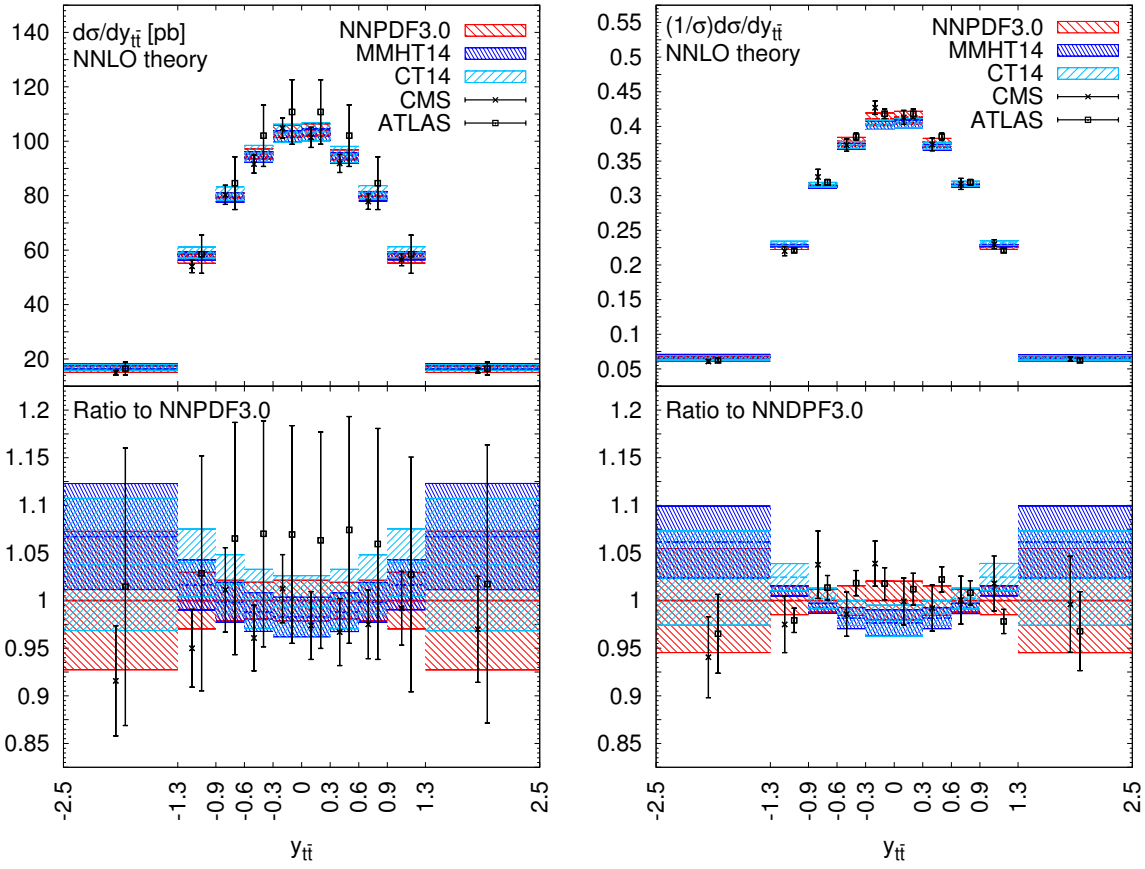

Figure 9. Same as figure 5 for the rapidity distribution of the top-quark pair, $y_{t\bar{t}}$.

3.3 Quantitative assessment of the agreement between theory and data

Due to the presence of large correlated experimental uncertainties (both of statistical and systematic origin), it is not possible to accurately assess the agreement between data and theory solely from the figures above. An adequate measure of this agreement should fully take these correlations into account. To this end we introduce a χ^2 estimator, which depends on the dataset, $\mathcal{D}$, and on the theoretical predictions based on the PDFs f, $\mathcal{T}[f]$. In this work, we use the following definition:

$$\chi^2\left\{\mathcal{T}[f],\mathcal{D}\right\} = \frac{1}{N_{\rm dat}}\sum_{i,j}^{N_{\rm dat}}\left(T_i[f]-D_i\right)C_{ij}^{-1}\left(T_j[f]-D_j\right)\,. \tag{3.4}$$

In this expression, i and j run over the experimental datapoints, D_i are the measured central values, T_i are the corresponding NNLO theoretical predictions computed with a given set of PDFs and C_{ij} is the covariance matrix, constructed from the available information on experimental statistical and systematic uncertainties.

The definition of the covariance matrix in eq. (3.4) is not unique (see, for example, the discussion in refs. [2, 76]). In this section we will use the so-called *experimental* definition:

$$C_{ij}^{\rm exp} \equiv \delta_{ij}\left(\sigma_i^{\rm stat}\right)^2 + \left(\sum_{\alpha=1}^{N_{\rm sysA}}\sigma_{i,\alpha}^{\rm sysA}\sigma_{j,\alpha}^{\rm sysA} + \sum_{\beta=1}^{N_{\rm sysM}}\sigma_{i,\beta}^{\rm sysM}\sigma_{j,\beta}^{\rm sysM}\right)D_iD_j\,, \tag{3.5}$$

JHEP04(2017)044

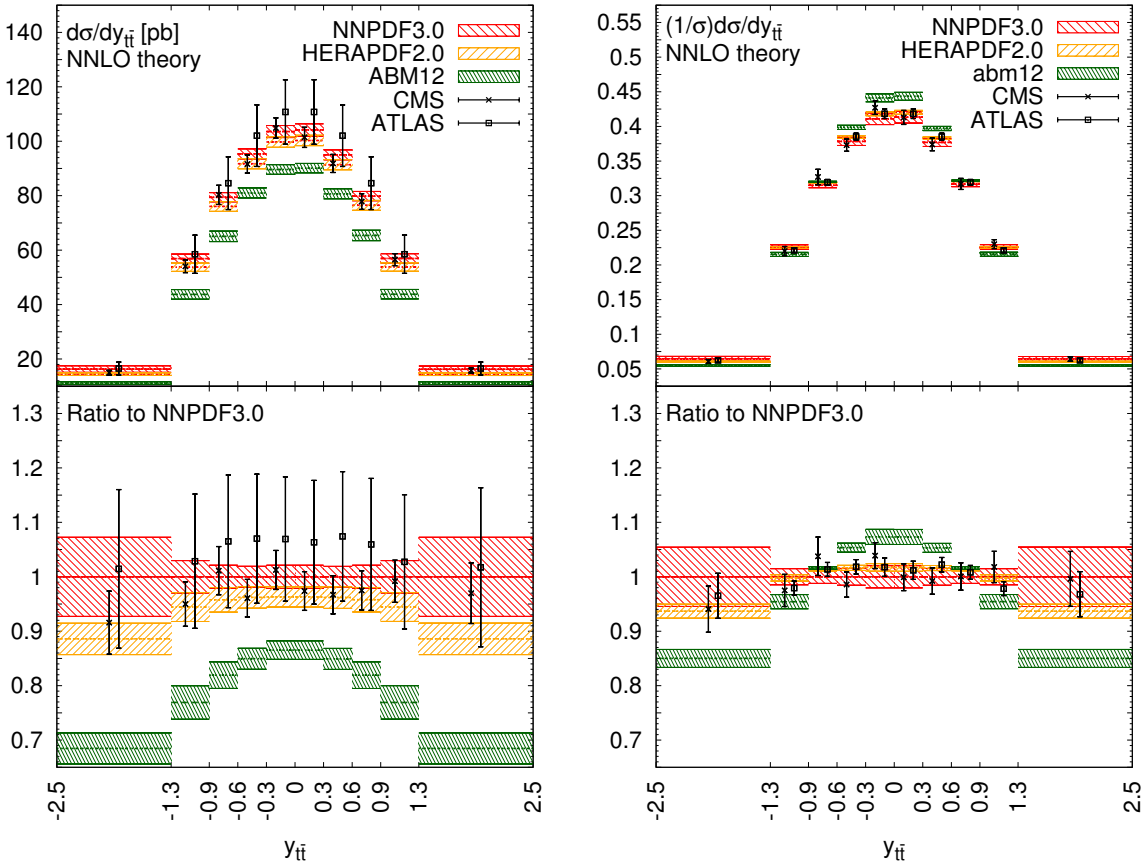

Figure 10. Same as figure 9 now for NNPDF3.0, ABM12 and HERAPDF2.0.

where $\sigma_i^{\rm stat}$ is the uncorrelated uncertainty of the data point i (obtained by adding in quadrature statistical and uncorrelated systematic uncertainties), and $\sigma_{i,\alpha}^{\rm sysA}$ ($\sigma_{i,\alpha}^{\rm sysM}$) are the $N_{\rm sysA}$ ($N_{\rm sysM}$) correlated additive (multiplicative) systematic uncertainties. The total number of correlated uncertainties in this case is $N_{\rm sysM} + N_{\rm sysA}$.

The values of the χ^2 computed using eq. (3.5) for each of the absolute and normalized differential distributions considered in this work, and using the five NNLO PDF sets, are summarized in tables 3 and 4. In order to facilitate the comparison with figures 5–12, we also indicate in parenthesis the χ^2 values computed neglecting bin-by-bin correlations. As expected, the χ^2 reduces substantially if experimental correlations are not accounted for.

At the level of absolute distributions, table 3, we find that for NNPDF3.0 there is good agreement ($\chi^2 \simeq 1$) for all ATLAS distributions, while the agreement is poorer for the CMS distributions except for $y_{t\bar{t}}$ and p_T^t. A similar agreement between data and NNLO theory is found for HERAPDF2.0. The same trend is also partly shared by CT14 and MMHT14, though these two sets lead to a somewhat worse description of the ATLAS and CMS $y_{t\bar{t}}$ distributions as compared to NNPDF3.0 and HERAPDF2.0. On the other hand, for ABM12 one finds a significantly worse χ^2, which reflects the fact that their predictions tend to undershoot the LHC data, as observed in figures 5–11. Concerning the top-quark transverse momentum p_T^t absolute distributions, NNLO theory provides good description of both ATLAS and CMS data for all PDF sets except for ABM12.

JHEP04(2017)044

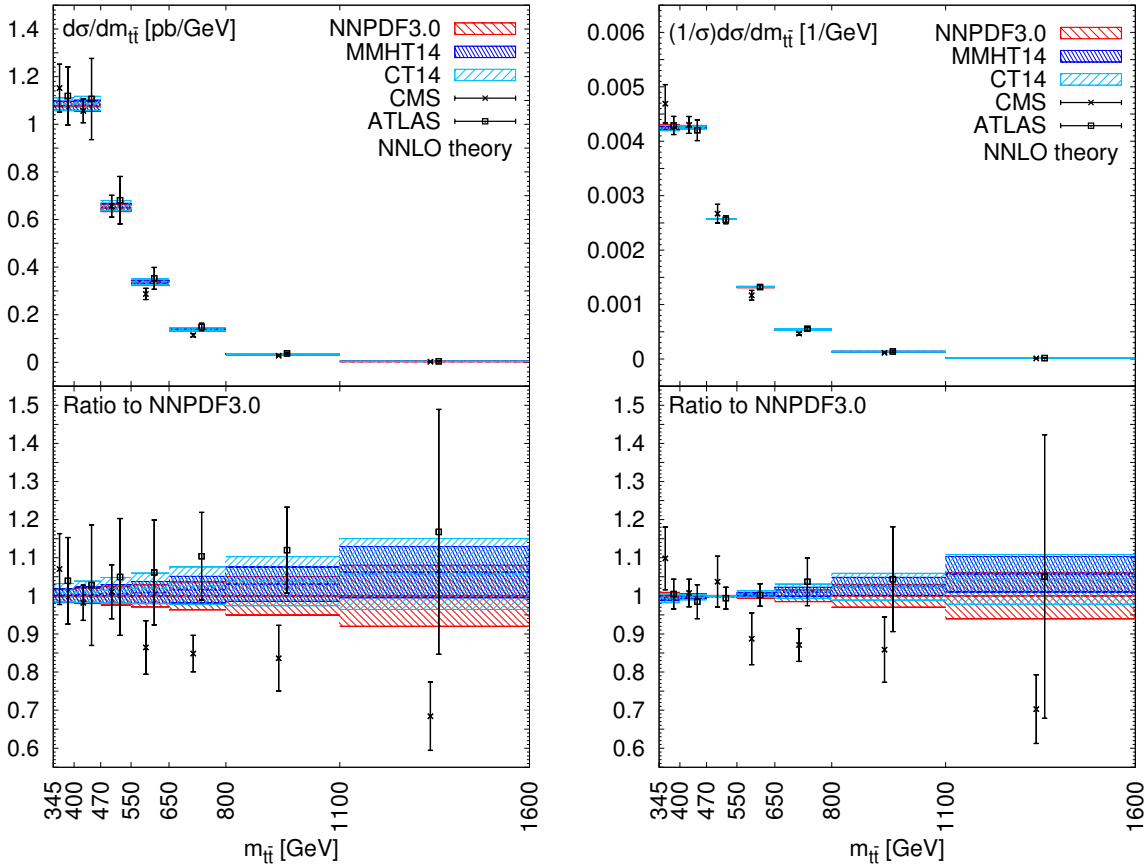

Figure 11. Same as figure 5 for the invariant mass distribution of the top-antitop pair, $m_{t\bar{t}}$.

Moving to normalized distributions, table 4, one finds χ^2 values that are in general higher than those from the absolute case. In the case of the p_T^t distribution, the agreement between normalized data and theory is generally poor for all PDF sets and for both ATLAS and CMS, except for HERAPDF2.0 in the former case. For the normalized y_t and $y_{t\bar{t}}$ distributions, HERAPDF2.0 provides a reasonable description except for the CMS y_t distribution, where one finds $\chi^2 \simeq 5$. None of the other NNLO sets achieves a satisfactory description of these two distributions.

Concerning the normalized invariant mass $m_{t\bar{t}}$ distribution, there is a stark difference between the comparisons of the ATLAS and the CMS measurements with theory. In the former case, NNPDF3.0, CT14 and MMHT14 lead to a good χ^2, while for the latter the same PDF sets lead to a much worse χ^2. For this distribution, HERAPDF2.0 provides a poor description of both ATLAS and CMS data, while ABM12 can successfully describe the CMS data at the price of a very poor χ^2 to the ATLAS measurements. Therefore, it seems not possible to achieve a simultaneous satisfactory description of both the ATLAS and CMS normalized $m_{t\bar{t}}$ distributions. As we will show in the next section, the same conclusions hold after the PDF fit.

A pattern arises from both figures 5–12 and from the χ^2 comparisons in tables 3–4: a certain degree of tension is present between the ATLAS and CMS measurements. This

JHEP04(2017)044

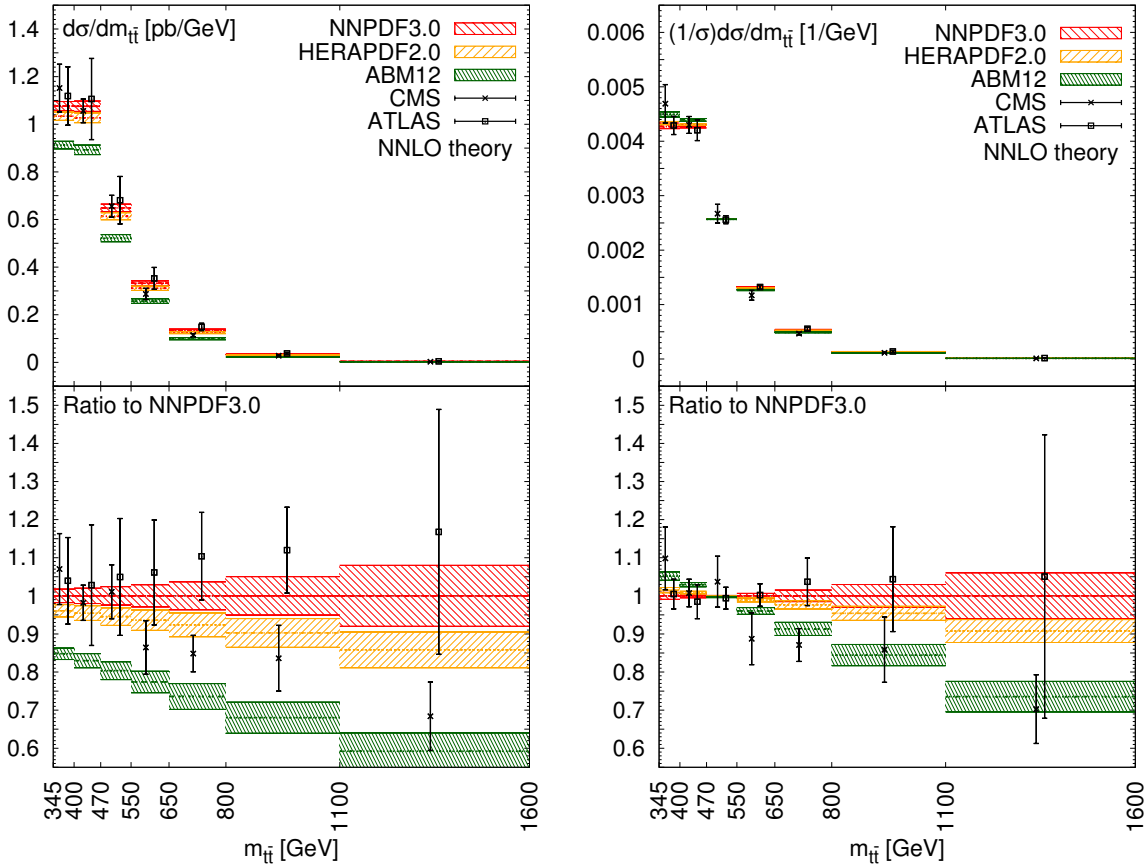

Figure 12. Same as figure 11, now for NNPDF3.0, ABM12 and HERAPDF2.0.

tension is more marked in the normalized distributions, which are characterized by smaller experimental uncertainties. As we will demonstrate next, such tension does not disappear when the top-quark distributions are included in the global PDF fit, though it is significantly alleviated when ATLAS and CMS data are fitted separately. Moreover, as we will show, it is possible to select a combination of ATLAS and CMS data leading to significant constraining power on the large-x gluon while at the same time providing a good χ^2 description of the two experiments.

Before ending this discussion, let us mention that the experimental covariance matrix defined in eq. (3.5), and used in this section, is not suitable to be used in PDF fits, since these would be affected by the D'Agostini bias [77]. A more appropriate definition of the covariance matrix for PDF fits is provided by the t_0-*prescription* [76],

$$C_{ij}^{t_0} \equiv \delta_{ij} \left(\sigma_i^{\rm stat}\right)^2 + \left(\sum_{\alpha=1}^{N_{\rm sysA}} \sigma_{i,\alpha}^{\rm sysA} \sigma_{j,\alpha}^{\rm sysA}\right) D_i D_j + \left(\sum_{\alpha=1}^{N_{\rm sysM}} \sigma_{i,\beta}^{\rm sysM} \sigma_{j,\beta}^{\rm sysM}\right) T_i^{(0)} T_j^{(0)} \,, \qquad (3.6)$$

in which a fixed theory prediction $\{T_i^{(0)}\}$ is used to define the contribution to the χ^2 from the multiplicative systematic uncertainties. The D'Agostini bias, which would otherwise be introduced if the fit were performed using the experimental definition eq. (3.5), is then avoided. Therefore, all the PDF fits presented in the next section will be based on eq. (3.6).

JHEP04(2017)044

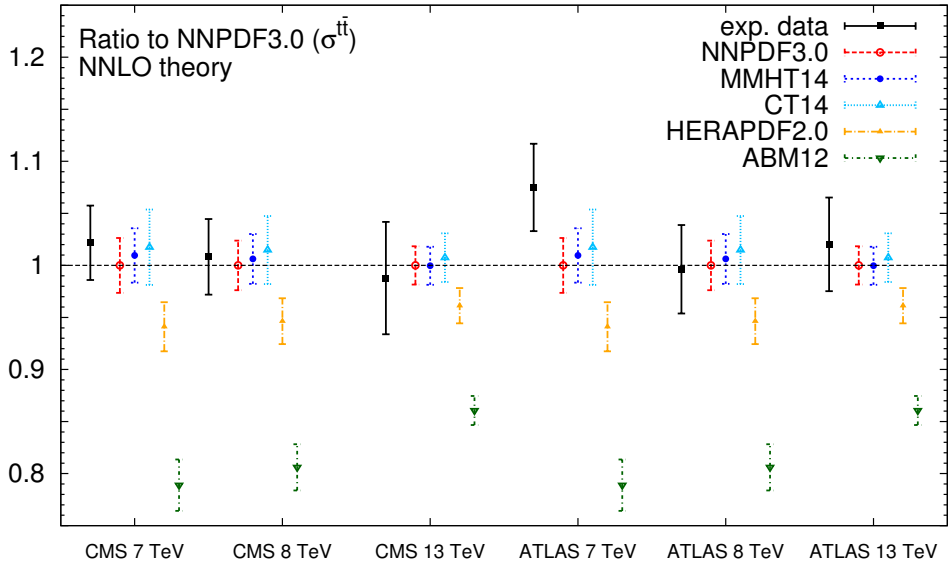

Figure 13. The inclusive cross-sections in top-quark pair production from ATLAS and CMS at different center-of-mass energies $\sqrt{s}$ (see table 2), compared to NNLO theory computed with the program `top++` for the five PDF sets. Results are shown as ratios to the central NNPDF3.0 predictions.

4 PDF fits with top-quark pair differential distributions

We now present the main results of this work, namely, NNLO PDF fits including top-quark pair differential distributions from ATLAS and CMS at $\sqrt{s} = 8\,\text{TeV}$. We begin by describing the settings of the PDF fits, based on the NNPDF framework, and then present the results for various choices of the baseline dataset (HERA-only or global) and of the top-quark differential data (absolute or normalized, and for different kinematic distributions). With this procedure we determine a suitable combination of top-quark measurements to be used in PDF fits. We then quantify the impact of the differential top data on the large-x gluon and on the kinematical distributions not directly included in the fit. Finally, we compare our results with the constraints on the large-x gluon provided by collider inclusive jet measurements.

4.1 Fit settings

The PDF fits presented in this work are based on a variant of the NNPDF3.0 global analysis [6, 37]. PDF evolution and deep-inelastic structure functions are evaluated with the public code `APFEL` [78, 79], with heavy quark structure functions computed in the FONLL-C general-mass scheme [80] with pole masses and with up to $n_f = 5$ active quark flavours. The charm PDF is generated perturbatively from light quarks and gluons. The value of the strong coupling is set to $\alpha_s(m_Z) = 0.118$, consistently with the PDG average [67]. For charm and bottom pole masses we use the values recommended by the Higgs Cross-Section Working Group [81], namely $m_c = 1.51\,\text{GeV}$ and $m_b = 4.92\,\text{GeV}$. DGLAP evolution equations are solved up to NNLO using the truncated solution, and the input PDF parametrization scale is taken to be $Q_0 = 1\,\text{GeV}$.

JHEP04(2017)044

Dataset	PDF set	χ^2		Dataset	PDF set	χ^2	
ATLAS $d\sigma/dp_T^t$				CMS $d\sigma/dp_T^t$			
	NNPDF3.0	0.84	(0.66)		NNPDF3.0	1.24	(0.91)
	CT14	0.76	(0.42)		CT14	1.67	(1.77)
	MMHT14	0.63	(0.44)		MMHT14	1.54	(1.47)
	HERA2.0	1.13	(1.69)		HERA2.0	0.69	(0.34)
	ABM12	6.23	(1.94)		ABM12	12.5	(3.00)
ATLAS $d\sigma/dy_t$				CMS $d\sigma/dy_t$			
	NNPDF3.0	0.73	(0.28)		NNPDF3.0	3.04	(1.05)
	CT14	1.28	(0.20)		CT14	2.23	(1.47)
	MMHT14	1.36	(0.29)		MMHT14	2.12	(0.98)
	HERA2.0	0.72	(0.99)		HERA2.0	3.65	(1.49)
	ABM12	5.32	(1.45)		ABM12	22.1	(9.78)
ATLAS $d\sigma/dy_{t\bar{t}}$				CMS $d\sigma/dy_{t\bar{t}}$			
	NNPDF3.0	0.84	(0.21)		NNPDF3.0	0.99	(0.74)
	CT14	2.69	(0.19)		CT14	1.88	(1.67)
	MMHT14	2.36	(0.29)		MMHT14	2.27	(1.52)
	HERA2.0	0.53	(0.74)		HERA2.0	1.02	(0.78)
	ABM12	4.04	(1.05)		ABM12	18.0	(5.48)
ATLAS $d\sigma/dm_{t\bar{t}}$				CMS $d\sigma/dm_{t\bar{t}}$			
	NNPDF3.0	0.77	(0.38)		NNPDF3.0	5.73	(4.36)
	CT14	0.61	(0.19)		CT14	7.28	(6.06)
	MMHT14	0.58	(0.24)		MMHT14	7.32	(5.74)
	HERA2.0	1.40	(1.30)		HERA2.0	3.32	(1.49)
	ABM12	5.72	(3.81)		ABM12	5.23	(3.22)

Table 3. The χ^2 values for absolute distributions in top-quark pair production from ATLAS and CMS for different NNLO PDF sets. The first number is the χ^2 from the full covariance matrix, eq. (3.5), while the value in parenthesis is obtained by adding in quadrature statistical and systematic errors.

For the fits presented here, we have defined a baseline global dataset which includes: fixed-target neutral-current DIS structure functions from NMC [82, 83], BCDMS [84, 85], and SLAC [86]; the legacy HERA combinations for inclusive [38] and charm [87] reduced cross-sections; charged-current structure functions from CHORUS inclusive neutrino DIS [88] and from NuTeV dimuon production data [89, 90]; fixed-target E605 [91] and E866 [92–94] DY production data; Tevatron collider data including the CDF [95] and D0 [96] Z rapidity distributions; and LHC collider data including ATLAS [97–99], CMS [11, 100–102] and LHCb [103, 104] vector boson production measurements, adding up to a total of $N_{\rm dat} = 3567$ data points.

This baseline global dataset is similar to that of NNPDF3.0 with three important differences. The first is in the HERA inclusive structure functions, where the separate HERA-II measurements from H1 and ZEUS [105–108] have been replaced by the HERA legacy combination [38]. Secondly, inclusive top-quark production cross-sections are excluded from

JHEP04(2017)044

Dataset	PDF set	χ^2		Dataset	PDF set	χ^2	
ATLAS $(1/\sigma)d\sigma/dp_T^t$				CMS $(1/\sigma)d\sigma/dp_T^t$			
	NNPDF3.0	3.13	(0.94)		NNPDF3.0	2.03	(0.51)
	CT14	2.33	(0.62)		CT14	2.88	(0.70)
	MMHT14	2.23	(0.54)		MMHT14	3.15	(0.77)
	HERA2.0	5.19	(1.73)		HERA2.0	1.12	(0.33)
	ABM12	14.0	(4.90)		ABM12	2.80	(0.80)
ATLAS $(1/\sigma)d\sigma/dy_t$				CMS $(1/\sigma)d\sigma/dy_t$			
	NNPDF3.0	4.06	(2.85)		NNPDF3.0	3.29	(1.49)
	CT14	10.3	(5.71)		CT14	2.33	(0.96)
	MMHT14	12.1	(6.82)		MMHT14	2.40	(1.09)
	HERA2.0	1.76	(1.62)		HERA2.0	4.99	(2.29)
	ABM12	15.5	(7.09)		ABM12	17.7	(8.72)
ATLAS $(1/\sigma)d\sigma/dy_{t\bar{t}}$				CMS $(1/\sigma)d\sigma/dy_{t\bar{t}}$			
	NNPDF3.0	3.59	(1.48)		NNPDF3.0	1.17	(0.75)
	CT14	12.7	(5.26)		CT14	2.53	(1.51)
	MMHT14	15.6	(5.49)		MMHT14	3.33	(2.10)
	HERA2.0	1.20	(0.60)		HERA2.0	1.23	(0.73)
	ABM12	20.2	(6.06)		ABM12	8.26	(4.52)
ATLAS $(1/\sigma)d\sigma/dm_{t\bar{t}}$				CMS $(1/\sigma)d\sigma/dm_{t\bar{t}}$			
	NNPDF3.0	1.57	(0.10)		NNPDF3.0	10.6	(3.87)
	CT14	1.09	(0.05)		CT14	13.5	(4.82)
	MMHT14	1.01	(0.05)		MMHT14	13.5	(4.93)
	HERA2.0	4.36	(0.30)		HERA2.0	5.96	(2.28)
	ABM12	21.1	(1.61)		ABM12	1.24	(0.47)

Table 4. Same as table 3 for the normalized differential distributions.

the baseline, as we want to study the impact of top data separately. Finally, in order to ensure a consistent NNLO determination without approximations for the NNLO matrix elements, we exclude jet production measurements [109–112]. The impact of jet data as compared to top data on the large-x gluon is discussed in section 4.5.

The influence of the differential top data on the gluon is assessed in two different scenarios. In the first case, we start from a baseline PDF fit which includes only HERA deep-inelastic structure functions. In the second case, we start from the NNPDF3.0-like baseline PDF fit described above. Subsequently, for each fit, we include either the absolute or normalized top-quark pair differential distributions, in the latter case supplemented with the inclusive total cross-section data. For completeness, we also perform a PDF fit where only total cross-sections are included. An overview of the datasets included in each fit is presented in table 5. We emphasize again that including different distributions from the same experiment would be double counting, since the statistical correlations among them are not available.

4.2 Results from the HERA-only fits

We begin by discussing the results from the HERA-only fits where the baseline dataset is composed exclusively of HERA deep-inelastic structure function measurements. The

JHEP04(2017)044

dataset	Fit ID									
	1	2	3	4	5	6	7	8	9	10
Baseline	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**
ATLAS $d\sigma/dp_T^t$	n	n	**y**	n	n	n	n	n	n	n
ATLAS $d\sigma/dy_t$	n	n	n	**y**	n	n	n	n	n	n
ATLAS $d\sigma/dy_{t\bar{t}}$	n	n	n	n	**y**	n	n	n	n	n
ATLAS $d\sigma/dm_{t\bar{t}}$	n	n	n	n	n	**y**	n	n	n	n
ATLAS $(1/\sigma)d\sigma/dp_T^t$	n	n	n	n	n	n	**y**	n	n	n
ATLAS $(1/\sigma)d\sigma/dy_t$	n	n	n	n	n	n	n	**y**	n	n
ATLAS $(1/\sigma)d\sigma/dy_{t\bar{t}}$	n	n	n	n	n	n	n	n	**y**	n
ATLAS $(1/\sigma)d\sigma/dm_{t\bar{t}}$	n	n	n	n	n	n	n	n	n	**y**
ATLAS $\sigma_{t\bar{t}}$	n	**y**	n	n	n	n	**y**	**y**	**y**	**y**
CMS $d\sigma/dp_T^t$	n	n	**y**	n	n	n	n	n	n	n
CMS $d\sigma/dy_t$	n	n	n	**y**	n	n	n	n	n	n
CMS $d\sigma/dy_{t\bar{t}}$	n	n	n	n	**y**	n	n	n	n	n
CMS $d\sigma/dm_{t\bar{t}}$	n	n	n	n	n	**y**	n	n	n	n
CMS $(1/\sigma)d\sigma/dp_T^t$	n	n	n	n	n	n	**y**	n	n	n
CMS $(1/\sigma)d\sigma/dy_t$	n	n	n	n	n	n	n	**y**	n	n
CMS $(1/\sigma)d\sigma/dy_{t\bar{t}}$	n	n	n	n	n	n	n	n	**y**	n
CMS $(1/\sigma)d\sigma/dm_{t\bar{t}}$	n	n	n	n	n	n	n	n	n	**y**
CMS $\sigma_{t\bar{t}}$	n	**y**	n	n	n	n	**y**	**y**	**y**	**y**

Table 5. Overview of the fits presented in this work. The baseline dataset is composed by either the HERA structure functions or by the NNPDF3.0-like dataset (see text). For each fit, we indicate in boldface which top-quark pair measurement from ATLAS and CMS have been included.

fit quality is assessed by means of the χ^2 computed using the experimental definition of the covariance matrix eq. (3.6). These values are collected in table 6 for each of the ten fits of table 5. The numbers in boldface refer to the fits with the corresponding datasets included, whilst the rest of the entries have been obtained from the predictions of the resultant PDF fit in each column. Note that fits of the normalized differential distributions are supplemented by total cross-sections, and that the first column is the result of the HERA-only baseline fit.

From table 6, we observe that in general it is possible to provide a satisfactory description of most of the fitted differential distributions. However in some cases the fit quality is somewhat poor: in the case of the CMS absolute (normalized) top rapidity distributions y_t, we find that the value of the χ^2 is 1.75 (1.94), while the corresponding ATLAS values are 1.06 (1.48). The worst agreement between NNLO theory and data can be seen in the top-quark pair invariant mass $m_{t\bar{t}}$ normalized distributions. Here we find values of the χ^2 as large as 6.26 and 3.03 for the ATLAS and CMS measurements, respectively. From table 6, we also note that the quality of the description of the HERA data does not deteriorates once top-quark pair differential distributions are added on top of it in the fit. The value of the χ^2 per data point is remarkably stable among all fits and shows limited statistical fluctuations.

JHEP04(2017)044

Dataset	Fit ID									
	1	2	3	4	5	6	7	8	9	10
HERA inclusive	**1.18**	**1.18**	**1.19**	**1.19**	**1.19**	**1.19**	**1.18**	**1.18**	**1.19**	**1.19**
HERA F_2^c	**1.07**	**1.05**	**1.05**	**1.05**	**1.06**	**1.05**	**1.08**	**1.07**	**1.06**	**1.06**
ATLAS $d\sigma/dp_T^t$	2.30	2.48	**0.73**	3.16	3.46	2.04	1.34	3.28	4.88	2.89
ATLAS $d\sigma/dy_t$	0.82	1.14	1.21	**1.06**	0.75	1.04	1.31	0.59	0.75	0.74
ATLAS $d\sigma/dy_{t\bar{t}}$	1.12	1.90	2.40	2.83	**0.45**	4.43	1.96	1.88	0.40	1.49
ATLAS $d\sigma/dm_{t\bar{t}}$	4.27	2.93	2.41	2.81	4.33	**1.53**	2.70	2.88	4.37	5.09
ATLAS $(1/\sigma)d\sigma/dp_T^t$	3.47	2.60	3.80	2.92	3.15	3.91	**1.46**	3.31	3.98	4.01
ATLAS $(1/\sigma)d\sigma/dy_t$	1.21	6.07	3.32	5.95	1.34	2.24	4.27	**1.48**	1.58	1.61
ATLAS $(1/\sigma)d\sigma/dy_{t\bar{t}}$	3.11	12.8	5.09	8.34	0.72	7.04	4.95	3.60	**0.53**	2.60
ATLAS $(1/\sigma)d\sigma/dm_{t\bar{t}}$	8.14	3.07	6.53	4.94	5.42	20.5	6.44	5.61	4.40	**3.03**
ATLAS $\sigma_{t\bar{t}}$	3.88	**0.35**	3.38	0.63	1.58	1.29	**0.87**	**0.37**	**0.42**	**0.66**
CMS $d\sigma/dp_T^t$	2.04	2.29	**0.82**	3.29	2.99	1.52	1.44	2.81	4.16	2.32
CMS $d\sigma/dy_t$	3.38	2.48	2.91	**1.75**	3.51	3.47	2.32	3.03	3.48	4.81
CMS $d\sigma/dy_{t\bar{t}}$	1.00	1.58	2.29	1.68	**1.08**	3.05	1.51	1.34	1.07	1.85
CMS $d\sigma/dm_{t\bar{t}}$	3.96	5.85	4.81	4.70	4.23	**1.73**	4.46	4.23	4.71	3.74
CMS $(1/\sigma)d\sigma/dp_T^t$	2.78	4.86	1.78	5.23	4.05	2.84	**1.57**	4.69	5.29	3.40
CMS $(1/\sigma)d\sigma/dy_t$	5.73	3.15	4.10	2.35	5.04	4.88	3.13	**1.94**	4.60	6.71
CMS $(1/\sigma)d\sigma/dy_{t\bar{t}}$	1.68	2.27	2.62	2.11	1.40	3.42	1.78	1.49	**1.20**	1.98
CMS $(1/\sigma)d\sigma/dm_{t\bar{t}}$	5.30	10.3	7.83	8.24	7.06	2.71	7.45	7.41	8.06	**6.26**
CMS $\sigma_{t\bar{t}}$	6.95	**1.04**	6.17	1.59	3.24	2.75	**1.02**	**1.09**	**1.17**	**1.64**
TOTAL	1.18	1.18	1.17	1.19	1.18	1.19	1.18	1.20	1.18	1.22

Table 6. The χ^2 from the HERA-only PDF fits for various combinations of top-quark data. The numbers in boldface indicate the datasets included in the fit, while the other entries describe the quality of the predictions of the resultant PDF fit for the other distributions.

Because the gluon PDF has little sensitivity to the HERA data in the region where instead it is sensitive to the LHC top differential data (roughly $0.08 \lesssim x \lesssim 0.5$, see figure 3), the poor agreement between data and theory for some distributions cannot be attributed to a tension with one of the other input datasets in the fit. The disagreement therefore appears to be the result of a genuine tension between the ATLAS and CMS measurements. As we will show below, this effect is only exacerbated in the global fits, where there are additional constraints on the gluon from other experiments. Further evidence for an inconsistency is provided by examining fits where ATLAS and CMS data are included separately. In such a case the description of the data by NNLO theory is substantially improved (see appendix A). It is also interesting to note that the inclusion of the total cross-section data in the fit does not necessarily imply a good description of the differential distributions. This highlights the fact that constraints on the large-x gluon stemming from inclusive cross-sections are only a subset of those obtained when fitting the fully differential distributions.

In figure 14 we compare the gluon from the HERA-only baseline fit with those obtained through fits to the various combinations of ATLAS and CMS top quark differential cross-sections. The comparison is performed at the scale $Q = 100\,\text{GeV}$, and the results are shown

JHEP04(2017)044

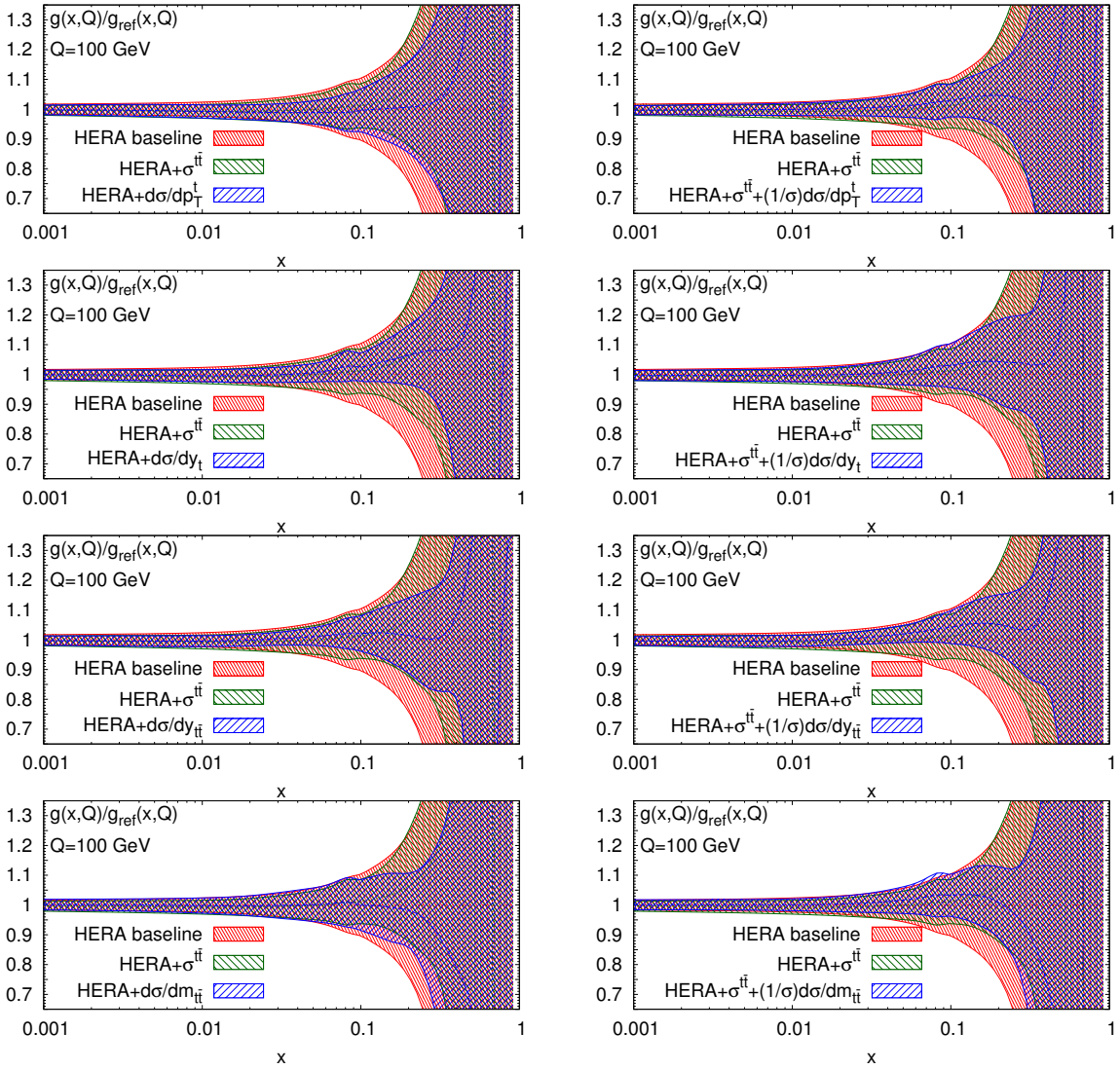

Figure 14. The gluon PDF from the HERA-only fits corresponding to the same input combinations of LHC top-quark data, both at the level of absolute (left) and of normalized distributions (right).

normalized to the central value of the HERA-only baseline. For completeness, we also show the results of the fit where only the total cross-sections $\sigma^{t\bar{t}}$ is included. In figure 15, we also display the central value and the one-sigma uncertainty of the gluon PDF at $Q = 100\,\text{GeV}$ for all the HERA-only fits collected in table 6. We observe that the various distributions demonstrate a fair degree of consistency in their impact on the gluon. For most of the considered kinematic distributions, both normalized and absolute, we find that top-quark data prefers a harder gluon at large x as compared to the HERA-only fit. This trend can also be observed for the fits including only total cross-sections. An exception arises in the fits with the differential $m_{t\bar{t}}$ distributions, which are however those for which the fit quality is worst.

We also observe that the three types of fits (HERA-only, HERA with total cross-sections, and HERA with differential distributions) turn out to be fully consistent within the respective PDF uncertainties. Moreover, the reduction of the PDF uncertainty in

JHEP04(2017)044

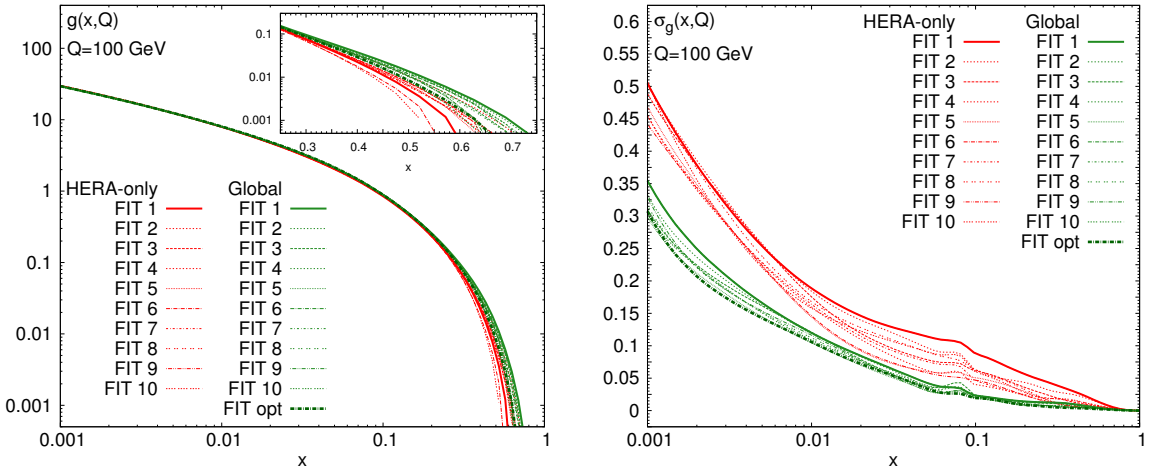

Figure 15. The central value (left) and the one-sigma uncertainty (right) of the gluon PDF at $Q = 100\,\text{GeV}$ for all the HERA-only and the global fits, including our optimal fit to the optimal combination of top-quark pair differential distributions (FIT opt), collected in tables 6–7. The inset in the left plot focuses on the gluon central values in the large-x region.

the large-x gluon appears to be similar for both absolute and normalized distributions. The PDF uncertainty is reduced for $x \gtrsim 0.05$, which is the kinematic range accessed by differential top-quark pair production data (see figure 3). In these HERA-only fits, the considered four kinematic distributions exhibit comparable constraining power.

While the HERA-only fits provide a clean testing ground to validate the implementation of top-quark differential distributions in a PDF fit, it is important to investigate the impact of these datasets in a global analysis, which then could be used for LHC phenomenology. This is done in the next section.

4.3 Results from the global fits

We now present the results of the NNLO fits in which the top-quark data has been added to a baseline fit based on the global dataset. As in the case of the HERA-only fits in table 6, first of all we collect the values of the χ^2 in table 7. Like in the HERA-only case, the description of the global baseline dataset does not deteriorate when any top-quark pair differential distribution is added on top of it in the fit. A slight worsening of the χ^2 per data point is observed for fixed-target DY and ATLAS and CMS vector boson production datasets, though it does not seem to be statistically significant. For some top-quark pair distributions, a good agreement between NNLO theory and data is found after the fit. These include the ATLAS absolute and normalized y_t and absolute $y_{t\bar{t}}$ distributions and the CMS absolute and normalized $y_{t\bar{t}}$ ones.

On the other hand, we also find that for some distributions the values of the χ^2 worsen in the global fits as compared to the HERA-only fits. In the case of the p_T^t distribution, the χ^2 values for the ATLAS and CMS absolute (normalized) distributions are 1.99 and 2.60 (2.96 and 3.56) respectively, to be compared with 0.73 and 0.82 (1.46 and 1.57) in the HERA-only fits. This increase in the χ^2 is also pronounced for the $m_{t\bar{t}}$ distribution, where in the global fits the values of the χ^2 for the ATLAS and CMS data are 4.02 and 5.11 (2.98

JHEP04(2017)044

Dataset	Fit ID										
	1	2	3	4	5	6	7	8	9	10	opt
NMC	**1.39**	**1.37**	**1.37**	**1.38**	**1.37**	**1.38**	**1.38**	**1.40**	**1.39**	**1.39**	**1.39**
SLAC	**0.67**	**0.70**	**0.66**	**0.67**	**0.66**	**0.70**	**0.67**	**0.65**	**0.65**	**0.65**	**0.65**
BCDMS	**1.25**	**1.25**	**1.26**	**1.24**	**1.25**	**1.24**	**1.25**	**1.25**	**1.25**	**1.25**	**1.25**
CHORUS	**1.10**	**1.09**	**1.11**	**1.11**	**1.10**	**1.09**	**1.09**	**1.09**	**1.10**	**1.10**	**1.10**
NuTeV	**0.76**	**0.71**	**0.69**	**0.68**	**0.67**	**0.74**	**0.71**	**0.70**	**0.66**	**0.67**	**0.67**
HERA inlusive	**1.21**	**1.20**	**1.21**	**1.21**	**1.21**	**1.21**	**1.21**	**1.22**	**1.21**	**1.21**	**1.22**
HERA F_2^c	**1.18**	**1.20**	**1.19**	**1.18**	**1.21**	**1.20**	**1.20**	**1.18**	**1.19**	**1.20**	**1.18**
E866	**1.29**	**1.28**	**1.31**	**1.32**	**1.31**	**1.31**	**1.29**	**1.31**	**1.30**	**1.31**	**1.33**
E605	**0.80**	**0.82**	**0.82**	**0.83**	**0.83**	**0.82**	**0.82**	**0.82**	**0.83**	**0.83**	**0.83**
CDF Z rapidity	**1.42**	**1.44**	**1.43**	**1.42**	**1.43**	**1.45**	**1.45**	**1.42**	**1.43**	**1.46**	**1.40**
D0 Z rapidity	**0.59**	**0.59**	**0.60**	**0.59**	**0.59**	**0.60**	**0.59**	**0.59**	**0.59**	**0.60**	**0.60**
LHCb W, Z rapidity	**1.09**	**1.13**	**1.12**	**1.10**	**1.12**	**1.13**	**1.12**	**1.13**	**1.12**	**1.12**	**1.11**
ATLAS W, Z 2010	**1.10**	**1.10**	**1.18**	**1.14**	**1.17**	**1.16**	**1.17**	**1.15**	**1.18**	**1.20**	**1.15**
ATLAS high-mass DY	**1.36**	**1.34**	**1.32**	**1.33**	**1.30**	**1.33**	**1.30**	**1.32**	**1.29**	**1.30**	**1.31**
ATLAS $d\sigma/dp_T^t$	2.37	2.30	**1.99**	2.36	2.24	2.23	2.09	2.18	2.34	2.24	2.19
ATLAS $d\sigma/dy_t$	0.93	0.80	0.74	**1.09**	0.76	0.76	0.86	0.69	0.76	0.66	0.64
ATLAS $d\sigma/dy_{t\bar{t}}$	2.44	2.03	1.96	2.59	**1.32**	2.32	2.11	1.74	1.26	1.80	1.84
ATLAS $d\sigma/dm_{t\bar{t}}$	4.27	4.47	4.68	4.14	4.92	**4.02**	4.34	4.79	4.98	4.99	5.01
ATLAS $(1/\sigma)d\sigma/dp_T^t$	2.93	3.97	3.29	4.36	5.22	4.35	**2.96**	4.26	4.92	5.68	2.49
ATLAS $(1/\sigma)d\sigma/dy_t$	5.00	3.17	2.47	6.36	1.55	2.93	3.94	**1.68**	1.45	1.10	**1.16**
ATLAS $(1/\sigma)d\sigma/dy_{t\bar{t}}$	9.69	5.59	5.89	8.95	2.68	5.73	6.73	3.57	**2.17**	3.73	3.81
ATLAS $(1/\sigma)d\sigma/dm_{t\bar{t}}$	2.30	2.80	3.31	2.67	3.96	4.21	3.09	3.68	3.77	**2.98**	4.55
ATLAS $\sigma_{t\bar{t}}$	0.12	**0.10**	0.21	0.10	0.10	0.12	**0.36**	**0.29**	**0.26**	**0.10**	**0.78**
CMS W electron asy	**0.58**	**0.57**	**0.55**	**0.59**	**0.57**	**0.58**	**0.56**	**0.58**	**0.57**	**0.56**	**0.58**
CMS W muon asy	**1.66**	**1.69**	**1.67**	**1.75**	**1.70**	**1.65**	**1.65**	**1.64**	**1.70**	**1.65**	**1.68**
CMS 2D DY 2011	**1.62**	**1.61**	**1.63**	**1.62**	**1.63**	**1.64**	**1.63**	**1.64**	**1.63**	**1.64**	**1.63**
CMS $d\sigma/dp_T^t$	3.50	3.46	**2.60**	3.50	3.03	3.00	2.85	3.11	3.24	2.92	2.91
CMS $d\sigma/dy_t$	3.48	3.71	4.05	**2.66**	4.18	3.49	3.38	4.23	4.43	4.99	4.98
CMS $d\sigma/dy_{t\bar{t}}$	1.36	1.13	1.00	1.32	**0.89**	0.86	1.00	1.01	1.04	1.24	1.07
CMS $d\sigma/dm_{t\bar{t}}$	7.07	6.27	5.79	6.33	5.09	**5.11**	6.00	5.37	5.21	4.31	4.77
CMS $(1/\sigma)d\sigma/dp_T^t$	4.31	4.00	3.39	4.28	3.65	3.59	**3.56**	3.57	3.73	3.48	3.33
CMS $(1/\sigma)d\sigma/dy_t$	3.66	4.10	4.45	3.10	4.98	4.06	3.65	**4.76**	5.13	6.09	5.78
CMS $(1/\sigma)d\sigma/dy_{t\bar{t}}$	1.59	1.20	1.06	1.73	0.94	1.01	1.20	0.99	**1.05**	1.32	**1.05**
CMS $(1/\sigma)d\sigma/dm_{t\bar{t}}$	12.0	10.8	9.81	11.1	8.72	8.72	10.3	9.15	8.97	**7.27**	8.05
CMS $\sigma_{t\bar{t}}$	0.10	**0.05**	0.26	0.19	0.32	0.21	**0.11**	**0.10**	**0.15**	**0.35**	**0.50**
TOTAL	1.20	1.19	1.20	1.20	1.19	1.21	1.20	1.21	1.20	1.21	1.20

Table 7. Same as table 6 for the global fits, including (last column) our optimal fit to the optimal combination of top-quark pair differential distributions.

JHEP04(2017)044

and 7.27) for the absolute (normalized) distributions, while we found 1.53 and 1.73 (3.03 and 6.26) in the HERA-only fits. For the normalized $m_{t\bar{t}}$ distributions instead, the χ^2 is equally poor in the global and in the HERA-only fits. When the ATLAS and CMS data are included separately in the fit, the χ^2 values exhibit a significant reduction, though they do not turn out to be as good as in the corresponding HERA-only case (see appendix A). This behaviour might be related to a residual tension between some top-quark pair distributions and other experiments included in the global fit, as we will discuss further in appendix B.

The results for the impact on the large-x gluon of adding top-quark pair differential data in the global fits are shown in figures 15–16. Similarly to the case of the HERA-only fits, we find that the four differential distributions, as well as the total cross-section data, have a similar pull on the central value of large-x gluon. Reassuringly, this trend is shared in both absolute and normalized distributions: for $x \gtrsim 0.2$, the LHC top data prefers a softer gluon as compared to the baseline fit. In all cases, the fits with top data are contained within the one-sigma PDF uncertainty band of the baseline fit. A comparison between figure 14 and figure 16 suggests that the relative effect on the gluon PDF is more pronounced in the case of the global fits than in the case of the HERA-only fits. This is explained by observing, first, that the central value of the gluon PDF is smaller for the baseline HERA-only fit than for the baseline global fit, and, second, that one-sigma uncertainties are always larger for the pool of HERA-only fits than for the pool of global fits (see figure 15). As a consequence, the error bands displayed in figures 14–16, i.e. the ratio of the one-sigma uncertainty to the central value (of either the HERA-only or the global baseline fit), is larger in the HERA-only case than in the global case.

From figure 16 we also observe that, when added to a global dataset, normalized distributions tend to exhibit a higher constraining power than the corresponding absolute data. This is especially marked for the y_t and $m_{t\bar{t}}$ distributions, while in the case of the p_T^t and $y_{t\bar{t}}$ distributions the differences between the impact of the absolute and normalized data turns out to be small. A significant reduction of the large-x gluon PDF uncertainty is observed for the normalized distributions, which can be more than a factor of two for $x \gtrsim 0.3$, thus demonstrating the constraining power of top-quark differential measurements for global PDF fits. The exception is the top-quark p_T^t measurement, which leads instead to a smaller impact on the gluon.

Concerning the impact of the inclusive cross-section data (in the fits that do not include differential measurements), we find that their pull on the central value of the gluon is the same as that of normalized distributions. On the other hand, figure 16 also shows that, unlike the case of HERA-only fits, the resulting PDF uncertainties are almost unchanged. We note, however, that a direct comparison with the results of [22] is not straightforward. Firstly, because here we use a smaller number of cross-section data points (only two at 8 TeV, and ignore the 7 TeV and 13 TeV data). Secondly, the dataset that constitutes the present baseline fit is different from that used in [22], NNPDF2.3 [113]. In addition, the results of [22] were based on the Bayesian reweighting method [114, 115], while in the present work top quark measurements are included by means of direct refitting.

JHEP04(2017)044

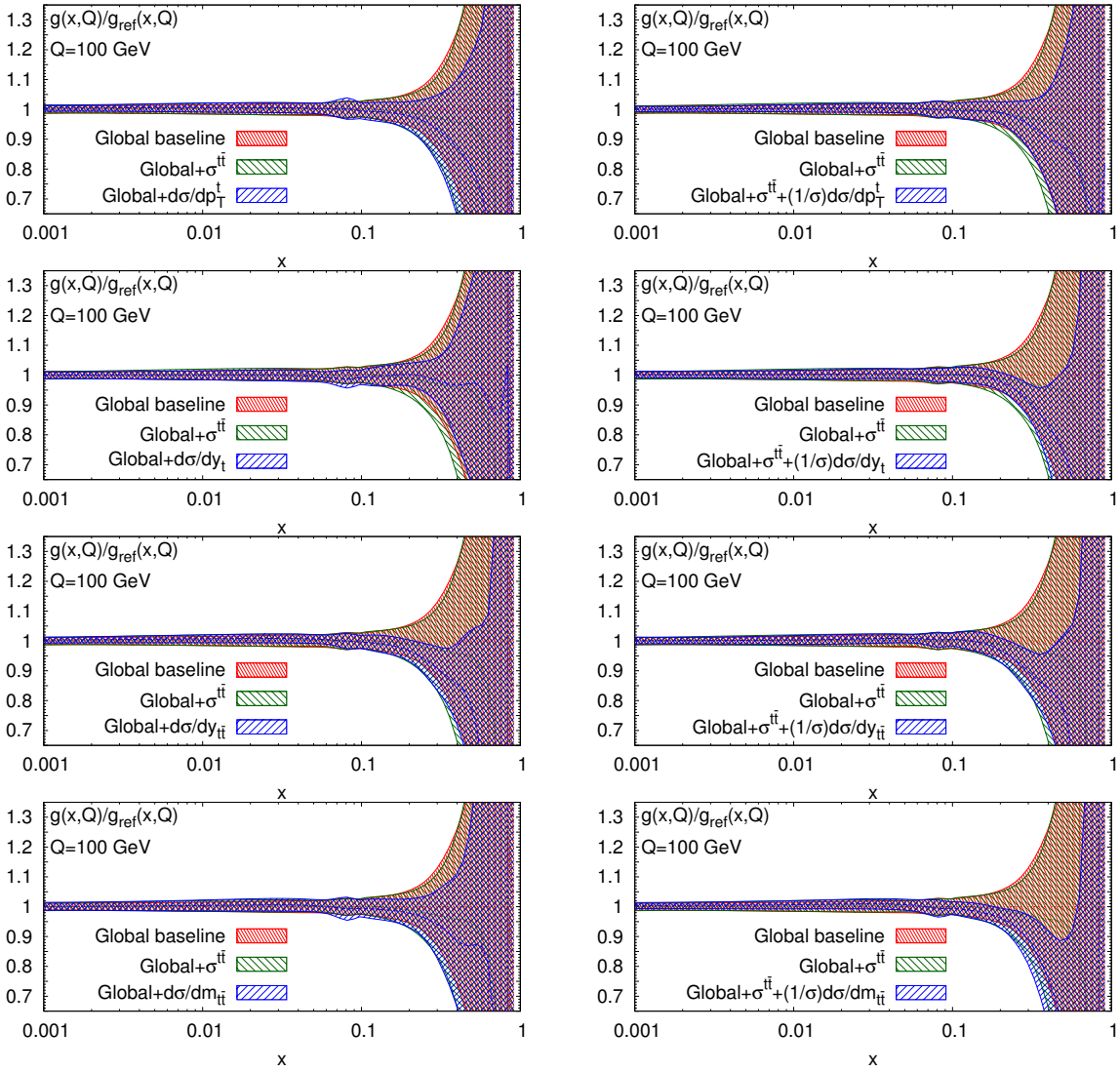

Figure 16. Same as figure 14 for the global fits.

4.4 Impact on the large-x gluon

With these studies at hand, we may now determine a suitable combination of the ATLAS and CMS top-quark pair differential measurements that maximizes the constraints on the large-x gluon while, at the same time, leads to a good agreement between data and theory. First of all, an inspection of figure 16 highlights the fact that, in the global fit, normalized distributions supplemented with the total cross-section have superior constraining power than the corresponding absolute distributions. This is especially the case for the y_t and $m_{t\bar{t}}$ distributions. Secondly, since each distribution provides different kinematic coverage of the gluon, one would like to include in the fit a given distribution from ATLAS and a different one from CMS. Moreover, in order to avoid distortions in the fit due to potential inconsistencies between ATLAS and CMS, it is advisable to include only distributions that can be satisfactory described ($\chi^2 \simeq 1$) when both ATLAS and CMS data are simultaneously

JHEP04(2017)044

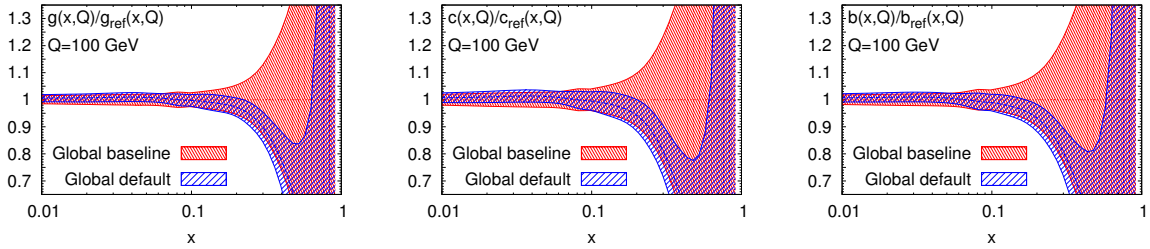

Figure 17. The gluon, charm and bottom PDFs from the global baseline fit compared to the optimal fit including our optimal combination of LHC top-quark data.

included. Finally, the selected distributions should be among the ones leading to the largest reduction of the PDF uncertainty of the large-x gluon.

Taking these guidelines into account, we suggest the following optimal combination:

- the normalized top-quark rapidity distribution $(1/\sigma)d\sigma/dy_t$ from ATLAS;
- the normalized top-quark pair rapidity distribution $(1/\sigma)d\sigma/dy_{t\bar{t}}$ from CMS;
- and the total inclusive cross-section $\sigma_{t\bar{t}}$ from ATLAS and CMS at $\sqrt{s} = 8\,\text{TeV}$.

From the results of figure 16 it also follows that other possible choices, consistent with the above guidelines, would not lead to significantly different results, as the pull of the ATLAS and CMS measurements on the large-x gluon is consistent among all distributions.

We have therefore performed a final global PDF fit using this optimal combination of LHC top data, and checked explicitly its features. The values of the χ^2 per data point for each dataset included in the fit are collected in the last column of table 7. The central value and one-sigma uncertainty of the corresponding gluon PDF are displayed in figure 15 (thick dashed line). In figure 17, we show the gluon, the charm and bottom quark PDFs from our global baseline fit and from our optimal fit including our optimal choice of top-quark data. Results are computed at $Q = 100\,\text{GeV}$ and are normalized to the global baseline fit. Other quark and antiquark PDFs are marginally affected by top data, as expected, and hence are not shown in figure 17. We now explore the impact of the new fit both on luminosities and on kinematic distributions not included in the fit.

First of all, we compute the PDF luminosities at $\sqrt{s} = 13\,\text{TeV}$ for this fit as a function of the invariant mass M_X of the produced final state. The factorization scale is set to $\mu_F = M_X$. In figure 18 we show the gg and the $q\bar{q}$ luminosities comparing the global baseline fit with the fit including LHC top data, together with the corresponding one-sigma PDF uncertainties. For the gg luminosity, the results of figure 18 confirm the substantial PDF uncertainty reduction reported in figure 17, which now translates into a reduction of the uncertainty for large invariant masses $M_X \gtrsim 600\,\text{GeV}$. For example, in the production of a final state with invariant mass $M_X \simeq 2\,\text{TeV}$ (3 TeV), PDF uncertainties are reduced from 12% (20%) down to around 5% (8%). Such a reduction has clear implications for BSM searches involving top quarks. The quark PDF uncertainties are also reduced, essentially as a consequence of the improved determination of heavy quarks, which follows in turn from a better determination of the gluon PDF. For the $q\bar{q}$ luminosity, for example, we

JHEP04(2017)044

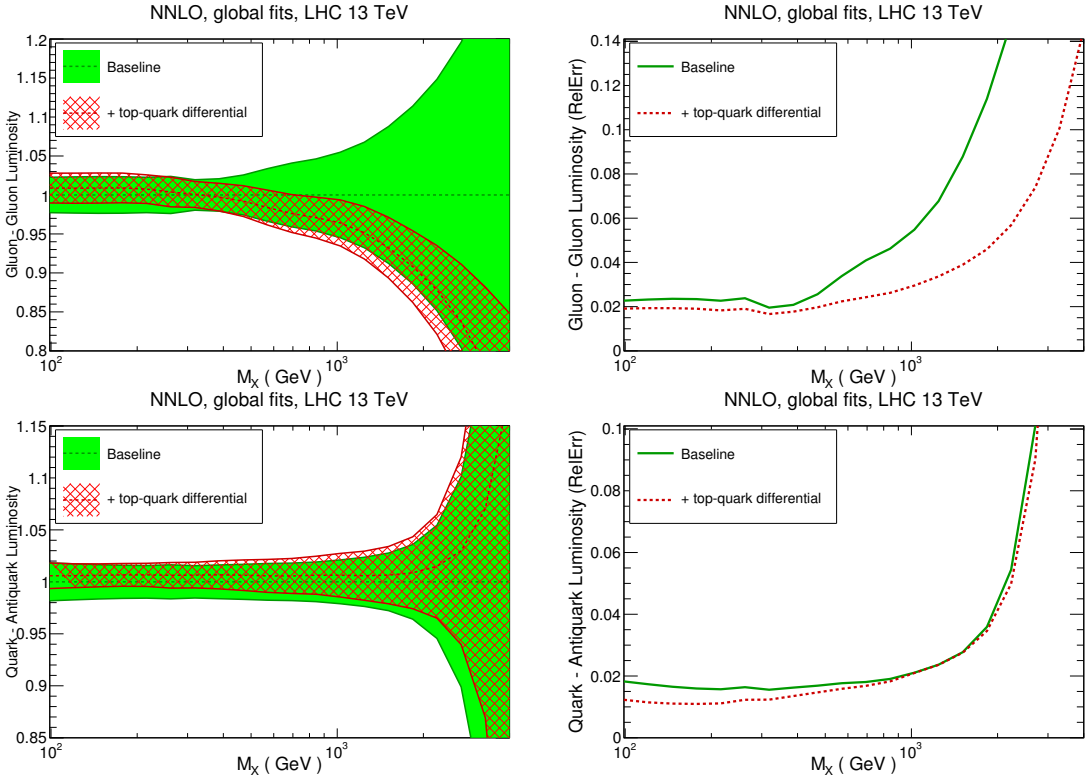

Figure 18. The gluon-gluon (upper) and quark-antiquark (lower) NNLO luminosities (left) and their relative 1-σ PDF uncertainties (right) at the LHC with $\sqrt{s} = 13\,\text{TeV}$. We compare the global baseline fit with the fit including the optimal combination of LHC top-quark pair differential data.

observe only a moderate uncertainty reduction in the region with $M_X \gtrsim 1\,\text{TeV}$, while PDF uncertainties are reduced from 2% to 1% around $M_X \sim 100\,\text{GeV}$.

Next, we study how the theoretical predictions are modified for those top-quark pair differential distributions not included in the fit. In figures 19 and 20 we show the NNLO calculations for the absolute and normalized $m_{t\bar{t}}$ and p_T^t distributions, respectively, obtained from the global PDF fit before and after the LHC top-quark data has been included. In the lower panels, we show the results normalized to the baseline fit. Note that none of the ATLAS and CMS data shown in figures 19 and 20 has actually been used as input in the fit.

The quality of the description of the p_T^t and $m_{t\bar{t}}$ data improves in most cases, both for absolute and normalized distributions, as quantified by the decrease in the values of the χ^2 per data point collected in table 7: for ATLAS absolute (normalized) p_T^T distribution, the χ^2 drops down from 2.37 (2.93) to 2.19 (2.49); for CMS absolute (normalized) p_T^T distribution from 3.50 (4.31) to 2.91 (3.33); for CMS absolute (normalized) $m_{t\bar{t}}$ distribution from 7.07 (12.0) to 4.77 (8.05). An exception is represented by ATLAS absolute (normalized) $m_{t\bar{t}}$ distribution, where instead the χ^2 increases from 4.27 (2.30) to 5.01 (4.55). Indeed, the fit tends to move towards the CMS data, which is more precise than the ATLAS data, but in clear tension with the latter.

JHEP04(2017)044

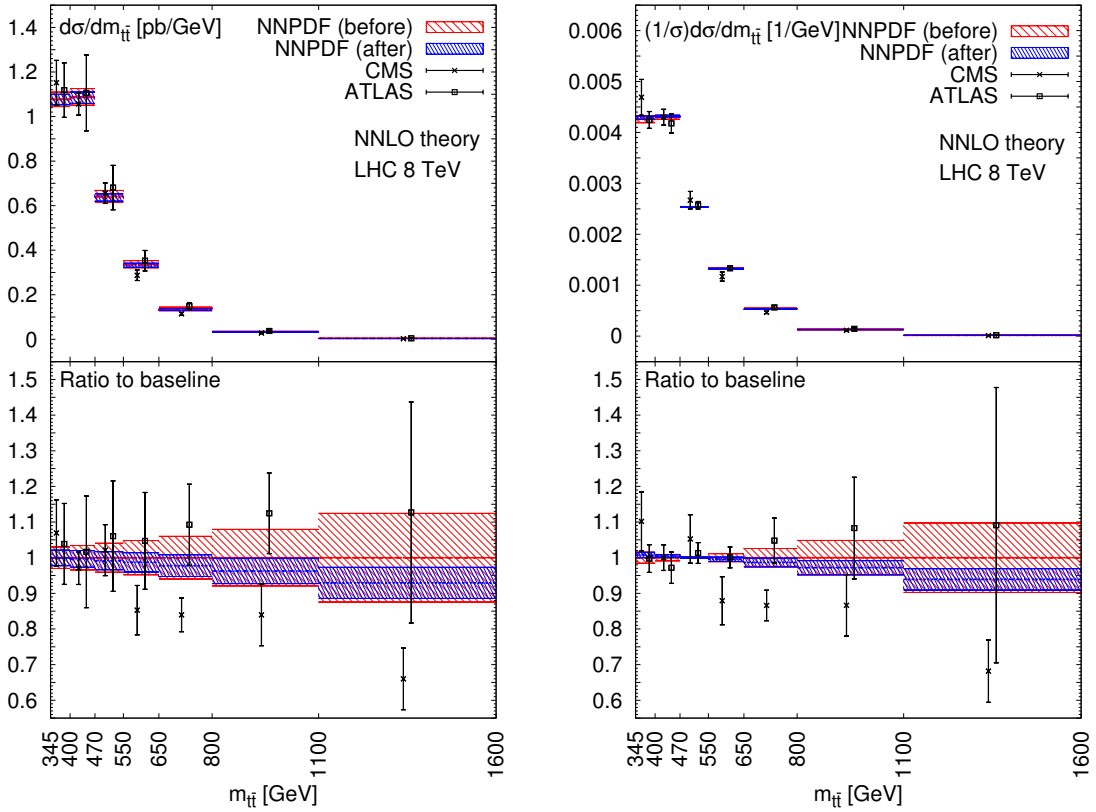

Figure 19. The NNLO theoretical predictions for the absolute (left) and normalized (right) $m_{t\bar{t}}$ distributions at the LHC 8 TeV, obtained from the global PDF fit before and after the optimal combination of top data has been included. The theory predictions include only the 1–σ PDF uncertainty band, while scale uncertainties are not shown. The lower panels show the results as a ratio to the baseline fit.

In comparison to the global baseline fit, theoretical predictions for the $m_{t\bar{t}}$ and p_T^t distributions are more precise in the optimal fit with our optimal choice of top-quark data included. This is a direct consequence of the large-x gluon constraints derived from fitting the y_t and $y_{t\bar{t}}$ distributions. For the top-quark pair invariant mass distributions, the PDF uncertainties in the rightmost bin, a region which is crucial for BSM searches, are reduced by more than a factor of two. This reduction would be even more pronounced for larger $m_{t\bar{t}}$, as can be inferred from the gg luminosity in figure 18. For the case of the top quark p_T^t distribution, we also observe a sizable PDF uncertainty reduction in the entire range probed by the LHC measurements, which can be again as large as a factor of two for $p_T^t \simeq 500\,\text{GeV}$.

Figures 19 and 20 highlight the potential of a comprehensive program of measurements of top-quark pair production to achieve a self-consistent reduction of theoretical uncertainties with the subsequent improvement of the prospects for BSM searches. In the specific case studied in this work, we have shown how the inclusion in the global fit of the normalized y_t and $y_{t\bar{t}}$ distributions leads to improved theory predictions for ATLAS and CMS p_T^t distributions and for CMS $m_{t\bar{t}}$ distributions. A corresponding improvement in the ATLAS $m_{t\bar{t}}$ distributions is not observed, though it might become evident once the apparent ten-

JHEP04(2017)044

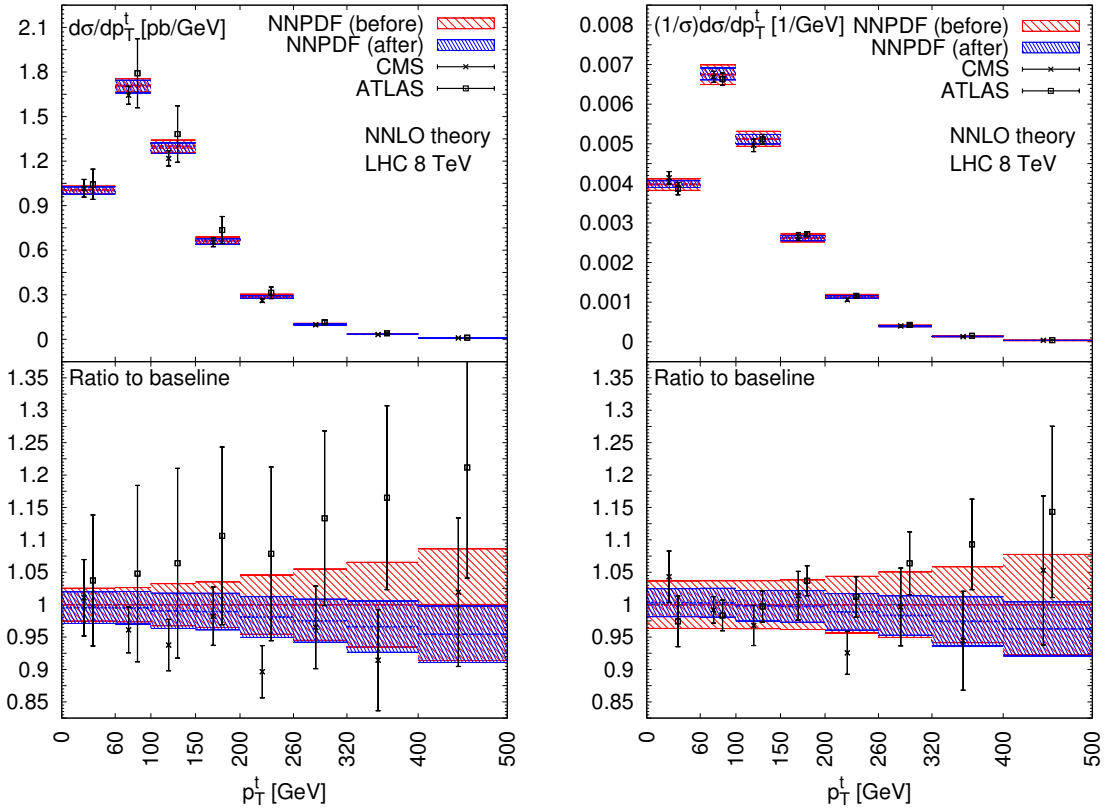

Figure 20. Same as figure 19 for the top quark pair p_T^t distribution.

sion between ATLAS and CMS data will be understood. Similar improvements will apply for other LHC processes, either in the SM or beyond, that are driven by the gg luminosity at large invariant masses.

It is important to emphasize that, with our choice of top-quark distributions to be used in the PDF fit, the possibility for contamination in the resulting gluon from BSM effects is reduced. The reason for this is that heavy new resonances are likely to be kinematically suppressed in the rapidity distributions but not in the tails of the $m_{t\bar{t}}$ and p_T^t distributions. Therefore, constraining the gluon from the y_t and $y_{t\bar{t}}$ measurements and using the resulting PDF to predict the $m_{t\bar{t}}$ and p_T^t distributions represents a robust strategy in the context of BSM searches.

4.5 Comparison with the constraints from jet data

As discussed in section 4.1, the global dataset used for the baseline fits excludes the jet production measurements from the Tevatron and the LHC that were part of NNPDF3.0. The rationale for this choice is that the NNLO calculation for jet production has become available only very recently [34], and we aim at providing a fully consistent determination of the large-x gluon at NNLO.

It is anyway instructive to assess how the PDF uncertainty reduction on the large-x gluon driven by top-quark data in the global fits (figure 16) compares with that from inclusive jet measurements. This way, it is possible to ascertain whether available differential top

JHEP04(2017)044

measurements provide competitive constraints as compared to those from jet production. To address this question, we have performed a NNLO fit where now the global baseline dataset is supplemented with collider inclusive jet production measurements, without any top-quark data. For these fits, theoretical calculations of the inclusive jet cross-section have been performed with NNLO DGLAP evolution and α_s running, but NLO matrix elements. This approximation is justified here since we are not interested in the shift in the central value of the large-x gluon as a result of the inclusion of the jet data, but only in the relative reduction of the PDF uncertainties.

In particular, we have added the inclusive jet production cross-sections from CDF Run II (k_t algorithm) [109]; from ATLAS at $\sqrt{s} = 2.76\,\text{TeV}$ [112] and 7 TeV [110], in the latter case from the 2010 run, together with their cross-correlations; and from CMS at 7 TeV [111] from the 2011 data-taking period. These four datasets were already part of the NNPDF3.0 fits. Moreover, we have added two additional inclusive jet measurements, from CMS at $\sqrt{s} = 2.76\,\text{TeV}$ and from ATLAS at 7 TeV from the 2011 run [116].[2] The resulting inclusive jet cross-sections add up to 76 points for CDF, 180 for ATLAS, and 214 from CMS, for a total of $N_{\text{dat}} = 470$ points.

In figure 21 we show the relative PDF uncertainty on the large-x gluon (left) and on the gg luminosity at large values of M_X (right) in the global baseline fit, compared to the corresponding fits including either top-quark pair differential measurements or jet production cross-sections. Interestingly, we find that the constraints on the large-x gluon from collider jet measurements turn out to be similar to those from the LHC top differential data. This result is particularly remarkable since, as indicated in table 1, the LHC data included in these fits amounts to $N_{\text{dat}} = 17$ data points (including the total cross-section measurements), while the collider jet dataset is substantially larger, $N_{\text{dat}} = 470$ points. On the other hand, while jet production is sensitive to the qg luminosity, and can have a large contribution for qq luminosity at high p_T, top quark production is driven instead by the gg one, which partly explains the comparable impact on the large-x gluon despite the different number of points. Note that PDF uncertainties in the gg luminosity at high masses are slightly reduced in the fits with top data than in the fits with jet data, despite the fact that for the gluon PDF itself the situation is opposite. This indicates that the top data induces a somewhat more stringent correlation between different x regions of the gluon as compared to jet data, thereby leading to smaller fluctuations in the gg luminosity as compared to those observed in $g(x, Q^2)$

The results in figure 21 indicate that the constraining power of top-quark pair differential distributions at 8 TeV on the large-x gluon is already similar to that of collider jet production measurements. Moreover, accounting for additional measurements at 8 TeV in other final states and with boosted kinematics, as well as available and upcoming 13 TeV measurements, will further strengthen the conclusions and make top-quark data even more competitive. On the other hand, figure 21 also indicates that ultimate accuracy on the large-x gluon can only be achieved by means of the simultaneous inclusion in the global analysis of both top and jet data.

[2]Details on the implementation of this two new datasets will be discussed in a forthcoming publication [117].

JHEP04(2017)044

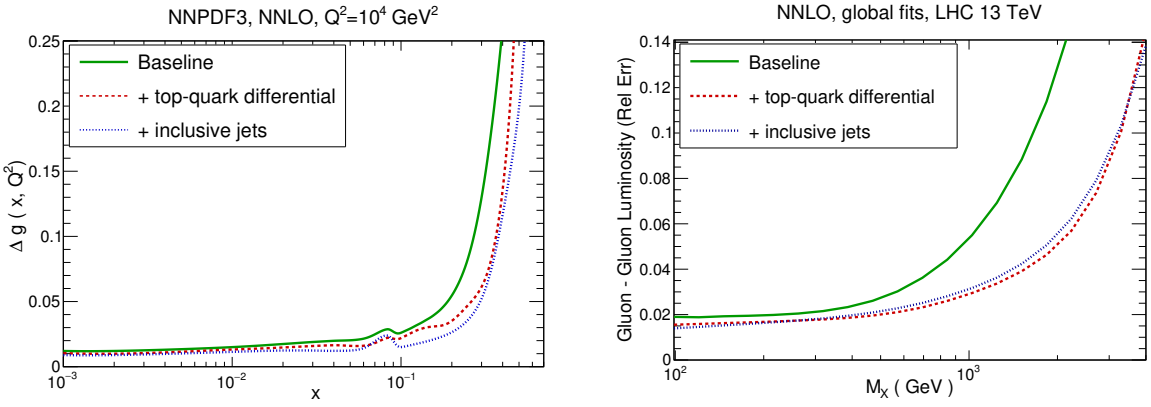

Figure 21. The relative PDF uncertainty on the large-x gluon (left) and on the gluon-gluon luminosity at large values of M_X (right plot) in the global baseline fit, compared with the corresponding fits including either top-quark pair differential measurements or jet production cross-sections.

5 Conclusions and outlook

Recent developments in higher-order QCD calculations of LHC processes require parton distributions with matching accuracy. PDFs in general, and the limited knowledge of the gluon at large x in particular, are often the dominant source of theory uncertainty for top-quark pair differential distributions [24]. This motivates a self-consistent two-step program where top-quark pair data from the LHC is first used to constrain the gluon within a NNLO global analysis, and then the improved gluon PDF is used to provide updated predictions for other top-quark observables or gluon-driven processes. This way one achieves a significant reduction of theory uncertainties, improving the prospects of both precision SM measurements and of BSM searches.

In this work we have quantified the impact on the large-x gluon of ATLAS and CMS $\sqrt{s} = 8\,\text{TeV}$ measurements of top-quark pair differential distributions using the NNPDF framework. We have shown how differential measurements can improve PDFs by extending the constraints on the gluon obtained from total-cross-section data. We have also studied the consistency between the ATLAS and CMS measurements, identifying certain tension among them. While the origin of this tension is still not understood, when the ATLAS and CMS data are included separately in the fits we find an improved agreement with NNLO theory for most kinematical distributions.

Our analysis indicates that normalized distributions, supplemented with the total inclusive cross-sections, have in general better constraining power than absolute ones. We have determined a suitable combination of ATLAS and CMS data to use as input to NNLO fits. This dataset has both high constraining power and leads to a good agreement between theory and data for the two experiments. Based on this analysis, our recommendation concerning the use of LHC top-quark pair production measurements into PDF fits would be to include:

- the normalized y_t distribution from ATLAS at $\sqrt{s} = 8\,\text{TeV}$ (lepton+jets channel),
- the normalized $y_{t\bar{t}}$ distribution from CMS at $\sqrt{s} = 8\,\text{TeV}$ (lepton+jets channel),
- total inclusive cross-sections at $\sqrt{s} = 7$, 8 and 13 TeV (all available data).

JHEP04(2017)044

Differential distributions should be included using NNLO theory, while inclusive cross-sections should be consistently computed at either NNLO+NNLL if fixed scales are used (as is done in `top++`), or at NNLO if dynamic scales are used. Future studies should be able to indicate which of the other available top-quark differential measurements, in particular in the dilepton channel at $\sqrt{s} = 8\,\text{TeV}$ and in the lepton+jets and dilepton channels at $\sqrt{s} = 13\,\text{TeV}$, can be used to complement the above list.

We have performed a global fit including this optimal combination of LHC top-quark data, and found that the uncertainty of the large-x gluon is substantially reduced in comparison to the baseline fit. As an illustration, the PDF error of the gg luminosity at $\sqrt{s} = 13\,\text{TeV}$ decreases from 6% (11%) down to 3% (5%) at $m_X = 1\,\text{TeV}$ (2 TeV), with an even larger reduction for yet higher values of m_X. We have then shown that the constraints on the large-x gluon provided by top-quark differential data are comparable to those derived from inclusive jet production, despite that the top data is based on a much smaller number of data points. Our results, therefore, provide a strong motivation for the inclusion of present and future LHC top-quark pair differential distributions into the next generation of PDF analyses.

In this work we have assumed the current world average of the top mass, $m_t = 173.3\,\text{GeV}$. However, the spread among individual m_t measurements leaves open the possibility of a future $\mathcal{O}(1\,\text{GeV})$ shift in the m_t central value. Such a shift would impact on the shape of normalized distributions, potentially affecting the resulting PDF fits. The optimal combination of LHC top-quark measurements used in our PDF fits is based on the y_t and $y_{t\bar{t}}$ distributions, which turn out to be those with the smallest shape sensitivity to m_t variations. Therefore, our results should be robust against future $\mathcal{O}(1\,\text{GeV})$ shifts in the central value of m_t.

Another important property of the top-quark distributions that we have used as input to the PDF fits is that, in general, they reduce the risk of a possible contamination in the gluon from BSM effects in top-quark pair production. For example, heavy resonances would be kinematically suppressed in the rapidity distributions, but not in the tails of the $m_{t\bar{t}}$ and p_T^t ones, where most searches are instead performed. Therefore, the gluon fitted from data on y_t and $y_{t\bar{t}}$ is safer to be used in BSM searches employing $m_{t\bar{t}}$ and p_T^t distributions.

The studies presented in this work could be extended in several directions. First of all, the inclusion of LHC measurements at 13 TeV with increased statistics and reduced systematic uncertainties will improve both the kinematic reach and the constraining power of top-quark pair data in PDF fits. Another avenue worth exploring is to quantify the impact on the gluon PDF of boosted top quark production, with invariant masses $m_{t\bar{t}}$ in the multi-TeV region. This program requires the inclusion of higher-order QCD and electroweak effects [51] as well as the photon PDF. The latter has been recently calculated in terms of DIS structure functions [118], improving on previous model-independent estimates [119] and reducing the impact of photon-initiated contributions in top-quark production.

Another important direction for future work would be the exploitation of particle-level distributions in top-quark pair production for PDF fits, which however requires NNLO calculations with top quark decays. This would be particularly useful in view of the reported tension between the ATLAS and CMS measurements of top-quark level distributions, and

JHEP04(2017)044

would remove the need to resort to theory-driven extrapolations in top-quark measurements, which introduce model dependence with associated uncertainties and biases that are difficult to quantify.

Ultimately, the best constraints on the large-x gluon will be obtained from the consistent combination of inclusive jet and dijet data with top-quark pair production measurements. The recent NNLO calculation of inclusive jet production [34] will make it possible in the near future. This way, it should be possible to achieve an even greater reduction in the gluon PDF uncertainty, providing a milestone contribution to the precision LHC phenomenology program.

Acknowledgments

We are grateful to Maria Aldaya, Frederic Deliot, Andrea Gianmanco, Alison Lister, Andreas Meyer, Mark Owen, Pedro Silva and Francesco Spano for continous assistance with the ATLAS and CMS inclusive and differential top-quark measurements as well as for many illuminating discussions and comments on this project. We thank David Heymes for related collaboration. We acknowledge extensive discussions on top quark production and PDF fits with the members of the NNPDF Collaboration.

M.C. is supported in part by grants of the DFG and BMBF. N.H. and J.R. are supported by an European Research Council Starting Grant "PDF4BSM". The work of A.M. is supported by the UK STFC grants ST/L002760/1 and ST/K004883/1 and by the European Research Council Consolidator Grant "NNLOforLHC2". E.R.N. is supported by the UK STFC grant ST/M003787/1.

A On the compatibility between the ATLAS and CMS data

One of the most puzzling aspects of the comparison between the NNLO theoretical calculations and the ATLAS and CMS top-quark pair differential cross-sections reported in this work is the apparent tension between some of the distributions from the two experiments. This tension was first observed in the comparisons between data and theory of section 3, and then further quantified by the χ^2 analysis from the HERA-only and global fits in section 4. There we found that achieving a good simultaneous description of several of the ATLAS and CMS distributions was not possible.

In this appendix we study further the issue of the compatibility between the ATLAS and CMS data by performing additional PDF fits where the two experiments are included separately. Our aim is to disentangle a genuine tension between the ATLAS and CMS measurements from alternative explanations of the poor χ^2 reported in section 4, for instance, the inadequacy of NNLO theory to describe the LHC data, or tension between the top-quark data and other experiments included in the global fit. To find out which is the correct explanation, we have repeated the HERA-only fits, as well as a selection of the global fits, but now adding the ATLAS and CMS distributions separately. These fits should lead to improved χ^2 values as compared to tables 6 and 7, provided that NNLO

JHEP04(2017)044

Dataset	Fit ID							
HERA only +	3*	4*	5*	6*	7*	8*	9*	10*
ATLAS $d\sigma/dp_T^t$	**0.44**	3.01	2.98	2.29	1.06	2.66	3.76	2.60
ATLAS $d\sigma/dy_t$	1.27	**0.47**	0.75	3.01	2.17	0.50	0.75	1.56
ATLAS $d\sigma/dy_{t\bar{t}}$	1.75	2.12	**0.43**	7.06	3.95	1.77	0.44	2.32
ATLAS $d\sigma/dm_{t\bar{t}}$	2.37	4.03	3.85	**0.39**	1.96	4.26	4.07	3.01
ATLAS $(1/\sigma)d\sigma/dp_T^t$	3.06	3.86	4.10	3.16	**0.60**	3.40	3.55	2.15
ATLAS $(1/\sigma)d\sigma/dy_t$	3.55	1.09	1.60	2.65	16.6	**0.75**	1.25	11.3
ATLAS $(1/\sigma)d\sigma/dy_{t\bar{t}}$	3.32	5.00	2.49	4.82	2.48	3.94	**0.45**	14.5
ATLAS $(1/\sigma)d\sigma/dm_{t\bar{t}}$	5.00	7.46	10.1	2.65	2.61	8.29	7.13	**0.55**
ATLAS $\sigma_{t\bar{t}}$	2.76	2.60	3.96	0.10	**0.99**	**0.88**	**1.02**	**0.71**
Dataset	Fit ID							
HERA only +	3**	4**	5**	6**	7**	8**	9**	10**
CMS $d\sigma/dp_T^t$	**0.82**	2.96	2.36	1.83	0.60	2.82	3.09	1.93
CMS $d\sigma/dy_t$	3.80	**1.30**	3.05	5.17	6.20	1.25	3.19	6.36
CMS $d\sigma/dy_{t\bar{t}}$	1.29	3.88	**0.74**	2.51	3.16	3.46	0.66	3.69
CMS $d\sigma/dm_{t\bar{t}}$	3.69	5.47	3.81	**1.28**	2.67	5.50	5.13	0.78
CMS $(1/\sigma)d\sigma/dp_T^t$	1.46	5.62	3.28	2.13	**0.85**	5.67	4.83	2.29
CMS $(1/\sigma)d\sigma/dy_t$	5.83	1.82	4.46	8.33	8.98	**1.70**	4.05	9.55
CMS $(1/\sigma)d\sigma/dy_{t\bar{t}}$	1.61	5.40	0.94	3.05	3.71	4.95	**0.75**	4.32
CMS $(1/\sigma)d\sigma/dm_{t\bar{t}}$	5.69	9.42	6.15	1.41	4.10	9.41	8.90	**0.92**
CMS $\sigma_{t\bar{t}}$	5.53	1.91	4.41	5.73	**0.57**	**0.79**	**0.70**	**0.80**

Table 8. Same as table 6 for the fits where the ATLAS and CMS data are included separately.

QCD is accurate enough to describe the experimental data, and that, in the case of the global fits, there are no tensions with other experiments.

The results of the fits to HERA data supplemented with ATLAS (CMS) top-quark pair differential distributions, with CMS (ATLAS) data excluded, are summarized in the upper (lower) part of table 8. As in table 6, we indicate the values of the χ^2 obtained from each fit, with numbers in boldface indicating the datasets that have been included in each fit. A comparison with table 6 shows that when the ATLAS or CMS measurements are included in the HERA-only fit separately, a better agreement between data and theory is obtained for all the kinematic distributions, both absolute and normalized. Note that this good agreement is not guaranteed: in several cases, the χ^2 for individual kinematical distributions is poor unless they are used in the fit, even when other top-quark distributions are being fitted. This behaviour reflects the fact that each distribution contains independent information on the gluon PDF.

The inclusion of perturbative corrections beyond NNLO, if they were known, would be unlikely to improve this picture. First, the size of the $\mathcal{C}$-factors in figure 1, which can be taken as a measure of the perturbative convergence, is approximately the same for all distributions. This suggests that they all converge with similar rapidity. Second, we

JHEP04(2017)044

Dataset	Fit ID				Dataset	Fit ID			
Global +	7 *	8*	9 *	10*	Global +	7**	8**	9**	10**
ATLAS $d\sigma/dp_T^t$	2.25	2.25	2.20	2.33	CMS $d\sigma/dp_T^t$	2.58	3.34	2.86	2.36
ATLAS $d\sigma/dy_t$	1.17	0.64	0.77	1.35	CMS $d\sigma/dy_t$	3.91	2.37	3.99	7.83
ATLAS $d\sigma/dy_{t\bar{t}}$	2.82	1.80	1.22	3.19	CMS $d\sigma/dy_{t\bar{t}}$	0.84	2.06	0.89	3.21
ATLAS $d\sigma/dm_{t\bar{t}}$	4.12	5.12	5.06	4.03	CMS $d\sigma/dm_{t\bar{t}}$	4.67	7.06	5.24	2.69
ATLAS $(1/\sigma)d\sigma/dp_T^t$	**2.38**	5.10	4.80	2.71	CMS $(1/\sigma)d\sigma/dp_T^t$	**3.03**	4.10	3.40	3.73
ATLAS $(1/\sigma)d\sigma/dy_t$	8.07	**1.11**	1.32	9.74	CMS $(1/\sigma)d\sigma/dy_t$	4.58	**2.66**	4.54	11.0
ATLAS $(1/\sigma)d\sigma/dy_{t\bar{t}}$	12.2	3.94	**2.12**	14.7	CMS $(1/\sigma)d\sigma/dy_{t\bar{t}}$	0.93	2.85	**0.93**	4.44
ATLAS $(1/\sigma)d\sigma/dm_{t\bar{t}}$	2.11	4.85	4.23	**1.88**	CMS $(1/\sigma)d\sigma/dm_{t\bar{t}}$	7.92	12.1	8.93	**4.12**
ATLAS $\sigma_{t\bar{t}}$	**0.78**	**0.11**	**0.45**	**0.15**	CMS $\sigma_{t\bar{t}}$	**0.23**	**0.17**	**0.51**	**1.52**

Table 9. Same as table 7, but for global fits including normalized distributions only.

explicitly checked that the size of the ratio of NNLO to NLO corrections is smaller than the size of the relative uncertainties of the data. This suggests that the data will be hardly sensitive to beyond-NNLO perturbative corrections within its present precision.

Therefore, when the ATLAS and CMS measurements are included separately in the HERA-only fit, we find no evidence of a tension between data and NNLO theory, indicating that the poor values of χ^2 in table 6 arise from a genuine incompatibility between the two experiments. One particularly illustrative example of this improvement is provided by the invariant mass $m_{t\bar{t}}$ normalized distribution. In this case, from table 6 we find that, for the fits including both experiments, the χ^2 is 3.03 and 6.26 for ATLAS and CMS respectively, while from table 8 we see that the corresponding values are 0.55 and 0.92 when each experiment is included separately.

In the case of the global fits, in table 9 we show the χ^2 values for a selection of global fits with the ATLAS and CMS data included separately. In particular, the fits shown include either the $m_{t\bar{t}}$ or the y_t normalized distributions from one of the two experiments. As before, the numbers in boldface indicate the specific distributions included in each case. By comparing with table 7, we find a picture that is qualitatively similar to the case of HERA-only fits. In general, also in this case improved χ^2 values are found when the ATLAS and CMS distributions are fitted separately.

On the other hand, even for the global fits which include separately the ATLAS and CMS data, the description of some of the top-quark distributions is still not optimal. For instance, when the two experiments are included simultaneously, we find that the χ^2 values of the normalized y_t ($m_{t\bar{t}}$) distributions for ATLAS and CMS are 1.68 (2.98) and 4.76 (7.27), respectively. When each experiment is included separately, the corresponding χ^2 values are instead 1.11 (1.88) and 2.66 (4.12). Therefore, while there is a significant improvement, the χ^2 values tend to be worse than those from the corresponding HERA-only fits in table 8, especially for CMS data. This behaviour might be related to a tension between some top-quark distributions and other experiments (see section 4.1) included in the global fit. Some additional insight on this issue is provided in appendix B.

JHEP04(2017)044

Dataset	Fit ID							
	3*	4*	5*	6*	7*	8*	9*	10*
NMC	9.38	9.39	8.56	8.76	11.1	10.1	9.80	9.95
SLAC	2.04	2.13	2.35	1.91	2.37	2.18	2.14	2.07
BCDMS	6.90	5.58	5.75	7.09	6.61	6.08	6.42	6.47
CHORUS	7.37	23.5	22.3	7.76	6.58	19.3	29.6	13.4
NuTeV	109	22.9	28.0	47.4	52.2	26.4	29.9	20.4
E866	371	440	776	35.6	68.1	612	163	33.4
E605	1.35	3.23	4.77	2.99	1.29	4.02	8.15	3.18
CFD Z rapidity	2.97	3.37	3.76	3.71	3.95	2.90	3.60	4.16
D0 Z rapidity	1.84	1.74	1.79	1.99	2.15	1.62	1.93	2.16
LHCb W, Z rapidity	3.07	1.91	1.82	1.94	2.13	2.19	1.44	2.09
ATLAS W, Z 2010	6.55	4.03	3.84	3.95	5.78	4.83	3.81	3.90
ATLAS high-mass DY	1.41	1.46	1.53	1.13	1.10	1.48	1.30	1.37
CMS W electron asy	17.0	24.7	10.2	29.3	12.7	36.0	4.91	15.2
CMS W muon asy	141	79.1	52.2	74.9	108	90.3	68.3	77.3
CMS 2D DY 2011	2.19	2.22	2.41	2.08	2.18	2.26	2.16	2.12

Table 10. The χ^2 per data point of the experiments not included in the HERA-only fits performed in appendix A (see also table 8), computed with the outcome of the corresponding HERA-only fits.

B Fitting top data with non-global datasets

A second puzzling issue, which we have encountered in this study, is the apparent tension between top-pair differential distributions, both absolute and normalized, and the rest of the dataset included in the global fits. Indeed, the quality of the description of top-pair data is significantly worse in the global fits (presented in section 4.3) than in the HERA-only fits (presented in section 4.2). This is apparent by comparing the values of the χ^2 in table 7 with their counterparts in table 6. Such a discrepancy persists even when ATLAS and CMS distributions, which were demonstrated to show some signs of tension in appendix A, are included separately in the fits.

In order to identify the data which originates the tension with top-pair differential distributions in the global fit, we compute the χ^2 for the experiments included in the global fits, but not in the HERA-only fits, based on the outcome of the HERA-only fits performed in appendix A. The results are collected in tables 10–11, which integrate the information contained in table 8. We have explicitly checked that the χ^2 obtained for HERA inclusive and charm reduced cross-sections does not significantly change from the χ^2 obtained in a simultaneous fit to ATLAS and CMS top-pair data (reported in table 6).

From tables 10–11, it is apparent that the HERA-only fits provide a very poor description of most of the data not included in them, especially of those sets which are expected

JHEP04(2017)044

Dataset	Fit ID							
	3**	4**	5**	6**	7**	8**	9**	10**
NMC	10.3	11.0	10.7	7.66	9.33	10.5	10.3	8.25
SLAC	2.45	2.46	2.25	1.85	2.42	2.47	2.37	1.94
BCDMS	6.90	7.07	7.91	6.26	6.71	6.10	5.97	6.19
CHORUS	6.29	13.5	23.4	17.1	9.28	7.48	18.0	10.1
NuTeV	75.9	39.6	24.5	15.8	38.2	48.3	21.9	33.2
E866	33.2	986	83.7	272	268	296	718	695
E605	1.35	8.53	11.9	10.2	2.85	5.42	4.00	19.0
CFD Z rapidity	2.72	4.12	3.97	3.22	2.12	3.51	3.75	2.68
D0 Z rapidity	1.65	2.17	2.07	1.61	2.05	1.84	1.92	1.48
LHCb W, Z rapidity	2.30	1.41	1.73	1.92	2.59	1.26	1.39	1.43
ATLAS W, Z 2010	4.90	5.99	4.10	3.64	4.64	5.49	4.26	2.94
ATLAS high-mass DY	1.83	0.93	1.28	1.45	3.04	0.95	1.04	1.26
CMS W electron asy	24.4	24.4	24.0	24.1	33.2	28.8	21.7	24.1
CMS W muon asy	82.2	108	71.5	59.1	82.5	104	86.2	45.2
CMS 2D DY 2011	2.17	2.14	2.16	2.18	2.21	2.10	2.05	2.10

Table 11. Same as table 10, but for the HERA-only fits including CMS top-pair differential distributions.

to constrain individual quark flavours at large x. This can be understood since the HERA data provides information only on the total quark singlet, and only very little on quark flavour separation. A rearrangement of the quark flavour separation is then needed in the global fits to obtain a good description of the whole data set. Such an improved description can be achieved, as proven by the values of the χ^2 collected in table 7 (very similar values are obtained in the global fits to ATLAS and CMS top-pair data separately). However, this makes the gluon PDF less flexible in accommodating the top-pair data, which is then described worse in the global fits than in the HERA-only fits. We note that in principle some datasets will require PDFs which are genuinely incompatible with HERA plus top data, while some other dataset will not. For instance, the strange quark can be presumably modified to fit the NuTeV data, which is mostly sensitive to it, without causing much change in the fit to HERA plus top data.

In order to further investigate this issue, we have performed a series of fits to a reduced dataset, which we defined as the global dataset except all fixed-target DIS data; one top-pair normalized distribution, separately from ATLAS or from CMS, and the corresponding total cross-section have been retained on top of the reduced dataset. In all cases, we have found that the quality of the description of the ATLAS and CMS top-pair data significantly improves, with respect to the corresponding global fits, and becomes comparable to that obtained in the HERA-only fits. The relevant values of the χ^2 are similar to those reported in boldface in table 8. For example, in the case of the normalized p_T^t ($m_{t\bar{t}}$) distribution, they

JHEP04(2017)044

are 0.79 (0.61) for ATLAS (in the fits including the corresponding top-pair distributions only from ATLAS, in addition to the reduced data set) and 0.90 (1.01) for CMS (in the fits including the corresponding top-pair distributions only from CMS, in addition to the reduced data set). The corresponding values in the case of the HERA-only fits are, from table 8, 0.60 (0.55) for ATLAS and 0.85 (0.92) for CMS.

These studies indicate that most of the tension between some of the top-pair differential distributions and the rest of the dataset in the global fits can be alleviated by removing the fixed-target DIS data. Of course, our studies do not indicate whether the tension comes from a specific fixed-target DIS experiment, or from the general constraint applied by fixed-target DIS data at a particular x. In principle, the information collected in table 7 would have provided some insight into this issue, if the top data had carried enough weight to result in a significant deterioration in fit quality to the data in tension.

It is however beyond the scope of this paper to draw a definite conclusion from this fact, since in principle all distributions have a similar correlation to the underlying gluons and quarks, as shown in figures 3–4. A comprehensive disclosure of the origin of the tension between top data and fixed-target DIS data can be addressed by performing a series of additional fits in which one fixed-target DIS experiment is removed at a time from the global data set. However, such an exercise will require a non negligible amount of extra computational effort and is therefore left to future study. Future comparisons between theory and LHC data for particle-level observables, as well as with the $\sqrt{s} = 13\,\text{TeV}$ measurements, might also shed more light on this issue.

References

[1] S. Forte, *Parton distributions at the dawn of the LHC*, *Acta Phys. Polon.* **B 41** (2010) 2859 [arXiv:1011.5247] [INSPIRE].

[2] R.D. Ball et al., *Parton Distribution Benchmarking with LHC Data*, *JHEP* **04** (2013) 125 [arXiv:1211.5142] [INSPIRE].

[3] S. Forte and G. Watt, *Progress in the Determination of the Partonic Structure of the Proton*, *Ann. Rev. Nucl. Part. Sci.* **63** (2013) 291 [arXiv:1301.6754] [INSPIRE].

[4] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys.* **G 43** (2016) 023001 [arXiv:1510.03865] [INSPIRE].

[5] J. Rojo et al., *The PDF4LHC report on PDFs and LHC data: Results from Run I and preparation for Run II*, *J. Phys.* **G 42** (2015) 103103 [arXiv:1507.00556] [INSPIRE].

[6] NNPDF collaboration, R.D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040 [arXiv:1410.8849] [INSPIRE].

[7] S. Alekhin, J. Blumlein and S. Moch, *The ABM parton distributions tuned to LHC data*, *Phys. Rev.* **D 89** (2014) 054028 [arXiv:1310.3059] [INSPIRE].

JHEP04(2017)044

[8] S. Dulat et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev.* **D 93** (2016) 033006 [arXiv:1506.07443] [INSPIRE].

[9] L.A. Harland-Lang, A.D. Martin, P. Motylinski and R.S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J.* **C 75** (2015) 204 [arXiv:1412.3989] [INSPIRE].

[10] W.J. Stirling and E. Vryonidou, *Charm production in association with an electroweak gauge boson at the LHC*, *Phys. Rev. Lett.* **109** (2012) 082002 [arXiv:1203.6781] [INSPIRE].

[11] CMS collaboration, *Measurement of associated W + charm production in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **02** (2014) 013 [arXiv:1310.1138] [INSPIRE].

[12] S.A. Malik and G. Watt, *Ratios of W and Z Cross sections at Large Boson p_T as a Constraint on PDFs and Background to New Physics*, *JHEP* **02** (2014) 025 [arXiv:1304.2424] [INSPIRE].

[13] R.M. Chatterjee, M. Guchait and R. Placakyte, *QCD analysis of the CMS inclusive differential Z production data at $\sqrt{s} = 8$ TeV*, *Phys. Rev.* **D 94** (2016) 034035 [arXiv:1603.09619] [INSPIRE].

[14] D. d'Enterria and J. Rojo, *Quantitative constraints on the gluon distribution function in the proton from collider isolated-photon data*, *Nucl. Phys.* **B 860** (2012) 311 [arXiv:1202.1762] [INSPIRE].

[15] R. Gauld, J. Rojo, L. Rottoli and J. Talbert, *Charm production in the forward region: constraints on the small-x gluon and backgrounds for neutrino astronomy*, *JHEP* **11** (2015) 009 [arXiv:1506.08025] [INSPIRE].

[16] PROSA collaboration, O. Zenaiev et al., *Impact of heavy-flavour production cross sections measured by the LHCb experiment on parton distribution functions at low x*, *Eur. Phys. J.* **C 75** (2015) 396 [arXiv:1503.04581] [INSPIRE].

[17] M. Cacciari, M.L. Mangano and P. Nason, *Gluon PDF constraints from the ratio of forward heavy-quark production at the LHC at $\sqrt{S} = 7$ and 13 TeV*, *Eur. Phys. J.* **C 75** (2015) 610 [arXiv:1507.06197] [INSPIRE].

[18] R. Gauld and J. Rojo, *Precision determination of the small-x gluon from charm production at LHCb*, *Phys. Rev. Lett.* **118** (2017) 072001 [arXiv:1610.09373] [INSPIRE].

[19] M. Czakon, P. Fiedler and A. Mitov, *Total Top-Quark Pair-Production Cross section at Hadron Colliders Through $O(\alpha_S^4)$*, *Phys. Rev. Lett.* **110** (2013) 252004 [arXiv:1303.6254] [INSPIRE].

[20] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, *JHEP* **01** (2013) 080 [arXiv:1210.6832] [INSPIRE].

[21] P. Bärnreuther, M. Czakon and A. Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, *Phys. Rev. Lett.* **109** (2012) 132001 [arXiv:1204.5201] [INSPIRE].

[22] M. Czakon, M.L. Mangano, A. Mitov and J. Rojo, *Constraints on the gluon PDF from top quark pair production at hadron colliders*, *JHEP* **07** (2013) 167 [arXiv:1303.7215] [INSPIRE].

[23] M. Czakon, D. Heymes and A. Mitov, *High-precision differential predictions for top-quark pairs at the LHC*, *Phys. Rev. Lett.* **116** (2016) 082003 [arXiv:1511.00549] [INSPIRE].

JHEP04(2017)044

[24] M. Czakon, D. Heymes and A. Mitov, *Dynamical scales for multi-TeV top-pair production at the LHC*, arXiv:1606.03350 [INSPIRE].

[25] M. Czakon, P. Fiedler and A. Mitov, *Resolving the Tevatron Top Quark Forward-Backward Asymmetry Puzzle: Fully Differential Next-to-Next-to-Leading-Order Calculation*, *Phys. Rev. Lett.* **115** (2015) 052001 [arXiv:1411.3007] [INSPIRE].

[26] M. Czakon, P. Fiedler, D. Heymes and A. Mitov, *NNLO QCD predictions for fully-differential top-quark pair production at the Tevatron*, *JHEP* **05** (2016) 034 [arXiv:1601.05375] [INSPIRE].

[27] M. Guzzi, K. Lipka and S.-O. Moch, *Top-quark pair production at hadron colliders: differential cross section and phenomenological applications with DiffTop*, *JHEP* **01** (2015) 082 [arXiv:1406.0386] [INSPIRE].

[28] ATLAS collaboration, *Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, *Eur. Phys. J.* **C 76** (2016) 538 [arXiv:1511.04716] [INSPIRE].

[29] CMS collaboration, *Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J.* **C 75** (2015) 542 [arXiv:1505.04480] [INSPIRE].

[30] J. Rojo, *Constraints on parton distributions and the strong coupling from LHC jet data*, *Int. J. Mod. Phys.* **A 30** (2015) 1546005 [arXiv:1410.7728] [INSPIRE].

[31] S.D. Ellis, Z. Kunszt and D.E. Soper, *The one jet inclusive cross-section at order α_s^3 quarks and gluons*, *Phys. Rev. Lett.* **64** (1990) 2121 [INSPIRE].

[32] S.D. Ellis, Z. Kunszt and D.E. Soper, *Two jet production in hadron collisions at order α_s^3 in QCD*, *Phys. Rev. Lett.* **69** (1992) 1496 [INSPIRE].

[33] Z. Nagy, *Three jet cross-sections in hadron hadron collisions at next-to-leading order*, *Phys. Rev. Lett.* **88** (2002) 122003 [hep-ph/0110315] [INSPIRE].

[34] J. Currie, E.W.N. Glover and J. Pires, *NNLO QCD predictions for single jet inclusive production at the LHC*, *Phys. Rev. Lett.* **118** (2017) 072002 [arXiv:1611.01460] [INSPIRE].

[35] A. Gehrmann-De Ridder, T. Gehrmann, E.W.N. Glover and J. Pires, *Second order QCD corrections to jet production at hadron colliders: the all-gluon contribution*, *Phys. Rev. Lett.* **110** (2013) 162003 [arXiv:1301.7310] [INSPIRE].

[36] J. Currie, A. Gehrmann-De Ridder, E.W.N. Glover and J. Pires, *NNLO QCD corrections to jet production at hadron colliders from gluon scattering*, *JHEP* **01** (2014) 110 [arXiv:1310.3993] [INSPIRE].

[37] NNPDF collaboration, R.D. Ball et al., *A Determination of the Charm Content of the Proton*, *Eur. Phys. J.* **C 76** (2016) 647 [arXiv:1605.06515] [INSPIRE].

[38] ZEUS, H1 collaborations, H. Abramowicz et al., *Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data*, *Eur. Phys. J.* **C 75** (2015) 580 [arXiv:1506.06042] [INSPIRE].

[39] NNPDF collaboration, J. Rojo, *Progress in the NNPDF global analysis and the impact of the legacy HERA combination*, PoS(EPS-HEP2015)506 [arXiv:1508.07731] [INSPIRE].

[40] T. Gleisberg et al., *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) 007 [arXiv:0811.4622] [INSPIRE].

JHEP04(2017)044

[41] L. Del Debbio, N.P. Hartland and S. Schumann, *MCgrid: projecting cross section calculations on grids*, *Comput. Phys. Commun.* **185** (2014) 2115 [arXiv:1312.4460] [INSPIRE].

[42] ATLAS collaboration, *Measurements of top quark pair relative differential cross-sections with ATLAS in pp collisions at* $\sqrt{s} = 7$ *TeV*, *Eur. Phys. J.* **C 73** (2013) 2261 [arXiv:1207.5644] [INSPIRE].

[43] ATLAS collaboration, *Measurements of normalized differential cross sections for* $t\bar{t}$ *production in pp collisions at* $\sqrt{s} = 7$ *TeV using the ATLAS detector*, *Phys. Rev.* **D 90** (2014) 072004 [arXiv:1407.0371] [INSPIRE].

[44] ATLAS collaboration, *Differential top-antitop cross-section measurements as a function of observables constructed from final-state particles using pp collisions at* $\sqrt{s} = 7$ *TeV in the ATLAS detector*, *JHEP* **06** (2015) 100 [arXiv:1502.05923] [INSPIRE].

[45] CMS collaboration, *Measurement of differential top-quark pair production cross sections in pp colisions at* $\sqrt{s} = 7$ *TeV*, *Eur. Phys. J.* **C 73** (2013) 2339 [arXiv:1211.2220] [INSPIRE].

[46] NNPDF collaboration, L. Del Debbio et al., *Unbiased determination of the proton structure function F(2)**p with faithful uncertainty estimation*, *JHEP* **03** (2005) 080 [hep-ph/0501067] [INSPIRE].

[47] ATLAS collaboration, *Measurement of top quark pair differential cross-sections in the dilepton channel in pp collisions at* $\sqrt{s} = 7$ *and* 8 *TeV with ATLAS*, *Phys. Rev.* **D 94** (2016) 092003 [arXiv:1607.07281] [INSPIRE].

[48] CMS collaboration, *Measurement of particle level differential ttbar cross sections in the dilepton channel at* $\sqrt{s} = 13$ *TeV*, CMS-PAS-TOP-16-007.

[49] ATLAS collaboration, *Measurement of the differential cross-section of highly boosted top quarks as a function of their transverse momentum in* $\sqrt{s} = 8$ *TeV proton-proton collisions using the ATLAS detector*, *Phys. Rev.* **D 93** (2016) 032009 [arXiv:1510.03818] [INSPIRE].

[50] CMS collaboration, *Measurement of the integrated and differential* $t\bar{t}$ *production cross sections for high-*p_t *top quarks in pp collisions at* $\sqrt{s} = 8$ *TeV*, *Phys. Rev.* **D 94** (2016) 072002 [arXiv:1605.00116] [INSPIRE].

[51] D. Pagani, I. Tsinikos and M. Zaro, *The impact of the photon PDF and electroweak corrections on* $t\bar{t}$ *distributions*, *Eur. Phys. J.* **C 76** (2016) 479 [arXiv:1606.01915] [INSPIRE].

[52] ATLAS collaboration, *Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at* $\sqrt{s} = 13$ *TeV using the ATLAS detector*, ATLAS-CONF-2016-040.

[53] CMS collaboration, *Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV*, submitted to *Phys. Rev. D* (2016) [arXiv:1610.04191] [INSPIRE].

[54] CMS collaboration, *Measurement of the ttbar production cross section in the emu channel in pp collisions at* 7 *and* 8 *TeV*, CMS-PAS-TOP-13-004.

[55] ATLAS collaboration, *Measurement of the* $t\bar{t}$ *production cross-section using* $e\mu$ *events with b-tagged jets in pp collisions at* $\sqrt{s} = 7$ *and* 8 *TeV with the ATLAS detector*, *Eur. Phys. J.* **C 74** (2014) 3109 [arXiv:1406.5375] [INSPIRE].

JHEP04(2017)044

[56] ATLAS collaboration, *Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector*, *Phys. Lett.* **B 761** (2016) 136 [arXiv:1606.02699] [INSPIRE].

[57] CMS collaboration, *Measurement of the t-tbar production cross section in the e-mu channel in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV*, *JHEP* **08** (2016) 029 [arXiv:1603.02303] [INSPIRE].

[58] CMS collaboration, *Measurement of the $t\bar{t}$ production cross section using events in the $e\mu$ final state in pp collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J.* **C 77** (2017) 172 [arXiv:1611.04040] [INSPIRE].

[59] T. Carli et al., *A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project*, *Eur. Phys. J.* **C 66** (2010) 503 [arXiv:0911.2985] [INSPIRE].

[60] FASTNLO collaboration, M. Wobisch et al., *Theory-Data Comparisons for Jet Measurements in Hadron-Induced Processes*, arXiv:1109.1310 [INSPIRE].

[61] Z. Nagy, *Next-to-leading order calculation of three jet observables in hadron hadron collision*, *Phys. Rev.* **D 68** (2003) 094002 [hep-ph/0307268] [INSPIRE].

[62] J.M. Campbell and R.K. Ellis, *Next-to-leading order corrections to $W + 2$ jet and $Z + 2$ jet production at hadron colliders*, *Phys. Rev.* **D 65** (2002) 113007 [hep-ph/0202176] [INSPIRE].

[63] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [arXiv:1405.0301] [INSPIRE].

[64] V. Bertone, R. Frederix, S. Frixione, J. Rojo and M. Sutton, *aMCfast: automation of fast NLO computations for PDF fits*, *JHEP* **08** (2014) 166 [arXiv:1406.7693] [INSPIRE].

[65] A. Buckley et al., *Rivet user manual*, *Comput. Phys. Commun.* **184** (2013) 2803 [arXiv:1003.0694] [INSPIRE].

[66] F. Cascioli, P. Maierhofer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, *Phys. Rev. Lett.* **108** (2012) 111601 [arXiv:1111.5206] [INSPIRE].

[67] PARTICLE DATA GROUP collaboration, K.A. Olive et al., *Review of Particle Physics*, *Chin. Phys.* **C 38** (2014) 090001 [INSPIRE].

[68] NNPDF collaboration, R.D. Ball et al., *Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO*, *Nucl. Phys.* **B 855** (2012) 153 [arXiv:1107.2652] [INSPIRE].

[69] M. Czakon and A. Mitov, *Top++: A program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, *Comput. Phys. Commun.* **185** (2014) 2930 [arXiv:1112.5675] [INSPIRE].

[70] M. Czakon, D. Heymes and A. Mitov, *Bump hunting in LHC $t\bar{t}$ events*, *Phys. Rev.* **D 94** (2016) 114033 [arXiv:1608.00765] [INSPIRE].

[71] F. Demartin, S. Forte, E. Mariani, J. Rojo and A. Vicini, *The impact of PDF and α_s uncertainties on Higgs Production in gluon fusion at hadron colliders*, *Phys. Rev.* **D 82** (2010) 014002 [arXiv:1004.0962] [INSPIRE].

[72] R.D. Ball et al., *Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology*, *Nucl. Phys.* **B 849** (2011) 296 [arXiv:1101.1300] [INSPIRE].

JHEP04(2017)044

[73] R. Frederix and F. Maltoni, *Top pair invariant mass distribution: A window on new physics*, *JHEP* **01** (2009) 047 [arXiv:0712.2355] [INSPIRE].

[74] B. Hespel, F. Maltoni and E. Vryonidou, *Signal background interference effects in heavy scalar production and decay to a top-anti-top pair*, *JHEP* **10** (2016) 016 [arXiv:1606.04149] [INSPIRE].

[75] M. Carena and Z. Liu, *Challenges and opportunities for heavy scalar searches in the $t\bar{t}$ channel at the LHC*, *JHEP* **11** (2016) 159 [arXiv:1608.07282] [INSPIRE].

[76] NNPDF collaboration, R.D. Ball et al., *Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties*, *JHEP* **05** (2010) 075 [arXiv:0912.2276] [INSPIRE].

[77] G. D'Agostini, *On the use of the covariance matrix to fit correlated data*, *Nucl. Instrum. Meth.* **A 346** (1994) 306 [INSPIRE].

[78] V. Bertone, S. Carrazza and J. Rojo, *APFEL: A PDF Evolution Library with QED corrections*, *Comput. Phys. Commun.* **185** (2014) 1647 [arXiv:1310.1394] [INSPIRE].

[79] V. Bertone, S. Carrazza and N.P. Hartland, *APFELgrid: a high performance tool for parton density determinations*, *Comput. Phys. Commun.* **212** (2017) 205 [arXiv:1605.02070] [INSPIRE].

[80] S. Forte, E. Laenen, P. Nason and J. Rojo, *Heavy quarks in deep-inelastic scattering*, *Nucl. Phys.* **B 834** (2010) 116 [arXiv:1001.2312] [INSPIRE].

[81] LHC Higgs Cross section Working Group collaboration, D. de Florian et al., *Handbook of LHC Higgs Cross sections: 4. Deciphering the Nature of the Higgs Sector*, arXiv:1610.07922 [INSPIRE].

[82] New Muon collaboration, M. Arneodo et al., *Accurate measurement of F_2^d/F_2^p and $R_d - R_p$*, *Nucl. Phys.* **B 487** (1997) 3 [hep-ex/9611022] [INSPIRE].

[83] New Muon collaboration, M. Arneodo et al., *Measurement of the proton and deuteron structure functions, F_2^p and F_2^d, and of the ratio σ_L/σ_T*, *Nucl. Phys.* **B 483** (1997) 3 [hep-ph/9610231] [INSPIRE].

[84] BCDMS collaboration, A.C. Benvenuti et al., *A High Statistics Measurement of the Proton Structure Functions $F_2(x, Q^2)$ and R from Deep Inelastic Muon Scattering at High Q^2*, *Phys. Lett.* **B 223** (1989) 485 [INSPIRE].

[85] BCDMS collaboration, A.C. Benvenuti et al., *A High Statistics Measurement of the Deuteron Structure Functions $F_2(x, Q^2)$ and R From Deep Inelastic Muon Scattering at High Q^2*, *Phys. Lett.* **B 237** (1990) 592 [INSPIRE].

[86] L.W. Whitlow, E.M. Riordan, S. Dasu, S. Rock and A. Bodek, *Precise measurements of the proton and deuteron structure functions from a global analysis of the SLAC deep inelastic electron scattering cross-sections*, *Phys. Lett.* **B 282** (1992) 475 [INSPIRE].

[87] ZEUS, H1 collaborations, H. Abramowicz et al., *Combination and QCD Analysis of Charm Production Cross section Measurements in Deep-Inelastic ep Scattering at HERA*, *Eur. Phys. J.* **C 73** (2013) 2311 [arXiv:1211.1182] [INSPIRE].

[88] CHORUS collaboration, G. Onengut et al., *Measurement of nucleon structure functions in neutrino scattering*, *Phys. Lett.* **B 632** (2006) 65 [INSPIRE].

JHEP04(2017)044

[89] NuTeV collaboration, M. Goncharov et al., *Precise measurement of dimuon production cross-sections in ν_μ Fe and $\bar{\nu}_\mu$ Fe deep inelastic scattering at the Tevatron*, *Phys. Rev.* **D 64** (2001) 112006 [hep-ex/0102049] [INSPIRE].

[90] D.A. Mason, *Measurement of the strange-antistrange asymmetry at NLO in QCD from NuTeV dimuon data*, FERMILAB-THESIS-2006-01 [INSPIRE].

[91] G. Moreno et al., *Dimuon production in proton-copper collisions at $\sqrt{s} = 38.8$ GeV*, *Phys. Rev.* **D 43** (1991) 2815 [INSPIRE].

[92] NuSea collaboration, J.C. Webb et al., *Absolute Drell-Yan dimuon cross-sections in 800 GeV/c pp and pd collisions*, hep-ex/0302019 [INSPIRE].

[93] J.C. Webb, *Measurement of continuum dimuon production in 800-GeV/C proton nucleon collisions*, hep-ex/0301031 [INSPIRE].

[94] NuSea collaboration, R.S. Towell et al., *Improved measurement of the $\bar{d}/\bar{u}$ asymmetry in the nucleon sea*, *Phys. Rev.* **D 64** (2001) 052002 [hep-ex/0103030] [INSPIRE].

[95] CDF collaboration, T.A. Aaltonen et al., *Measurement of $d\sigma/dy$ of Drell-Yan e^+e^- pairs in the Z Mass Region from $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Lett.* **B 692** (2010) 232 [arXiv:0908.3914] [INSPIRE].

[96] D0 collaboration, V.M. Abazov et al., *Measurement of the shape of the boson rapidity distribution for $p\bar{p} \to Z/gamma^* \to e^+e^- + X$ events produced at $\sqrt{s}$ of 1.96 TeV*, *Phys. Rev.* **D 76** (2007) 012003 [hep-ex/0702025] [INSPIRE].

[97] ATLAS collaboration, *Measurement of the inclusive $W^\pm$ and Z/γ^* cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys. Rev.* **D 85** (2012) 072004 [arXiv:1109.5141] [INSPIRE].

[98] ATLAS collaboration, *Measurement of the high-mass Drell-Yan differential cross-section in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys. Lett.* **B 725** (2013) 223 [arXiv:1305.4192] [INSPIRE].

[99] ATLAS collaboration, *Measurement of the Transverse Momentum Distribution of W Bosons in pp Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector*, *Phys. Rev.* **D 85** (2012) 012005 [arXiv:1108.6308] [INSPIRE].

[100] CMS collaboration, *Measurement of the electron charge asymmetry in inclusive W production in pp collisions at $\sqrt{s} = 7$ TeV*, *Phys. Rev. Lett.* **109** (2012) 111806 [arXiv:1206.2598] [INSPIRE].

[101] CMS collaboration, *Measurement of the muon charge asymmetry in inclusive $pp \to W + X$ production at $\sqrt{s} = 7$ TeV and an improved determination of light parton distribution functions*, *Phys. Rev.* **D 90** (2014) 032004 [arXiv:1312.6283] [INSPIRE].

[102] CMS collaboration, *Measurement of the differential and double-differential Drell-Yan cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **12** (2013) 030 [arXiv:1310.7291] [INSPIRE].

[103] LHCb collaboration, *Inclusive W and Z production in the forward region at $\sqrt{s} = 7$ TeV*, *JHEP* **06** (2012) 058 [arXiv:1204.1620] [INSPIRE].

[104] LHCb collaboration, *Measurement of the cross-section for $Z \to e^+e^-$ production in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **02** (2013) 106 [arXiv:1212.4620] [INSPIRE].

JHEP04(2017)044

[105] H1 collaboration, F.D. Aaron et al., *Inclusive Deep Inelastic Scattering at High Q^2 with Longitudinally Polarised Lepton Beams at HERA*, *JHEP* **09** (2012) 061 [arXiv:1206.7007] [INSPIRE].

[106] H1 collaboration, F.D. Aaron et al., *Measurement of the Inclusive $e^{\pm}p$ Scattering Cross section at High Inelasticity y and of the Structure Function F_L*, *Eur. Phys. J.* **C 71** (2011) 1579 [arXiv:1012.4355] [INSPIRE].

[107] ZEUS collaboration, H. Abramowicz et al., *Measurement of high-Q2 neutral current deep inelastic e^+p scattering cross sections with a longitudinally polarized positron beam at HERA*, *Phys. Rev.* **D 87** (2013) 052014 [arXiv:1208.6138] [INSPIRE].

[108] ZEUS collaboration, H. Abramowicz et al., *Measurement of high-Q^2 charged current deep inelastic scattering cross sections with a longitudinally polarised positron beam at HERA*, *Eur. Phys. J.* **C 70** (2010) 945 [arXiv:1008.3493] [INSPIRE].

[109] CDF collaboration, A. Abulencia et al., *Measurement of the Inclusive Jet Cross section using the $k_{\rm T}$ algorithmin $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV with the CDF II Detector*, *Phys. Rev.* **D 75** (2007) 092006 [*Erratum ibid.* **D 75** (2007) 119901] [hep-ex/0701051] [INSPIRE].

[110] ATLAS collaboration, *Measurement of inclusive jet and dijet production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *Phys. Rev.* **D 86** (2012) 014022 [arXiv:1112.6297] [INSPIRE].

[111] CMS collaboration, *Measurements of differential jet cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV with the CMS detector*, *Phys. Rev.* **D 87** (2013) 112002 [arXiv:1212.6660] [INSPIRE].

[112] ATLAS collaboration, *Measurement of the inclusive jet cross section in pp collisions at $\sqrt{s} = 2.76$ TeV and comparison to the inclusive jet cross section at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *Eur. Phys. J.* **C 73** (2013) 2509 [arXiv:1304.4739] [INSPIRE].

[113] R.D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys.* **B 867** (2013) 244 [arXiv:1207.1303] [INSPIRE].

[114] NNPDF collaboration, R.D. Ball et al., *Reweighting NNPDFs: the W lepton asymmetry*, *Nucl. Phys.* **B 849** (2011) 112 [*Erratum ibid.* **B 854** (2012) 926] [arXiv:1012.0836] [INSPIRE].

[115] R.D. Ball et al., *Reweighting and Unweighting of Parton Distributions and the LHC W lepton asymmetry data*, *Nucl. Phys.* **B 855** (2012) 608 [arXiv:1108.1758] [INSPIRE].

[116] ATLAS collaboration, *Measurement of the inclusive jet cross-section in proton-proton collisions at $\sqrt{s} = 7$ TeV using 4.5 fb^{-1} of data with the ATLAS detector*, *JHEP* **02** (2015) 153 [*Erratum ibid.* **09** (2015) 141] [arXiv:1410.8857] [INSPIRE].

[117] NNPDF collaboration, R.D. Ball et al., *NNPDF3.1: parton distributions for precision LHC physics*, in preparation.

[118] A. Manohar, P. Nason, G.P. Salam and G. Zanderighi, *How bright is the proton? A precise determination of the photon parton distribution function*, *Phys. Rev. Lett.* **117** (2016) 242002 [arXiv:1607.04266] [INSPIRE].

[119] NNPDF collaboration, R.D. Ball et al., *Parton distributions with QED corrections*, *Nucl. Phys.* **B 877** (2013) 290 [arXiv:1308.0598] [INSPIRE].

JHEP04(2017)044