THE EUROPEAN
PHYSICAL JOURNAL C

CrossMark

Regular Article - Theoretical Physics

Towards ultimate parton distributions at the high-luminosity LHC

Rabah Abdul Khalek[1,4]**, Shaun Bailey**[2]**, Jun Gao**[3]**, Lucian Harland-Lang**[2]**, Juan Rojo**[1,4,a]

[1] Department of Physics and Astronomy, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands
[2] Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK
[3] Institute of Nuclear and Particle Physics, Shanghai Key Laboratory for Particle Physics and Cosmology, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, China
[4] Nikhef Theory Group, Science Park 105, 1098 XG Amsterdam, The Netherlands

Abstract Since its start of data taking, the LHC has provided an impressive wealth of information on the quark and gluon structure of the proton. Indeed, modern global analyses of parton distribution functions (PDFs) include a wide range of LHC measurements of processes such as the production of jets, electroweak gauge bosons, and top quark pairs. In this work, we assess the ultimate constraining power of LHC data on the PDFs that can be expected from the complete dataset, in particular after the High-Luminosity (HL) phase, starting in around 2025. The huge statistics of the HL-LHC, delivering $\mathcal{L} = 3$ ab^{-1} to ATLAS and CMS and $\mathcal{L} = 0.3$ ab^{-1} to LHCb, will lead to an extension of the kinematic coverage of PDF-sensitive measurements as well as to an improvement in their statistical and systematic uncertainties. Here we generate HL-LHC pseudo-data for different projections of the experimental uncertainties, and then quantify the resulting constraints on the PDF4LHC15 set by means of the Hessian profiling method. We find that HL-LHC measurements can reduce PDF uncertainties by up to a factor of 2 to 4 in comparison to state-of-the-art fits, leading to few-percent uncertainties for important observables such as the Higgs boson transverse momentum distribution via gluon-fusion. Our results illustrate the significant improvement in the precision of PDF fits achievable from hadron collider data alone, and motivate the continuation of the ongoing successful program of PDF-sensitive measurements by the LHC collaborations.

Contents

1 Introduction

A detailed understanding of the quark and gluon structure of the proton [1–3] is an essential ingredient of theoretical predictions for hadron colliders such as the LHC. This is quantified by the parton distribution functions (PDFs), which determine how the proton's momentum is shared among its constituents in a hard–scattering collision. PDF uncertainties represent one of the dominant theoretical systematic errors in many important LHC processes, including the profiling of the Higgs boson sector [4]; direct searches for new heavy beyond the Standard Model (BSM) particles [5]; indirect BSM searches by means of the SM Effective Field Theory (SMEFT) [6]; as well as in the measurement of fundamental SM parameters such as the W boson mass [7], the Weinberg mixing angle [8], and the strong coupling constant [9] and its running [10].

Since the start of data taking in 2009, the LHC has provided an impressive wealth of information on the proton's PDFs. Indeed, modern global PDF fits [11–14] include a wide

a e-mail: juan.rojo@cern.ch

Springer

range of LHC measurements in processes such as the production of jets, weak gauge bosons, and top quark pairs. Crucially, the recent breakthroughs in the calculation of NNLO QCD and NLO QED and electroweak corrections (including photon–induced ones) to most PDF–sensitive processes have been instrumental in allowing for the full exploitation of the information provided by the LHC measurements. The impact of high precision LHC data combined with state–of–the art perturbative calculations has been quantified for many of the processes of interest, such as top quark pair production [15,16], the transverse momentum spectrum of Z bosons [17], direct photon production [18,19], D meson production in the forward region [20,21], W production in association with charm quarks [22–24], and inclusive jet production [25,26]. See the reviews [1,2] for a more extensive list of references.

With experimentalists warming up to analyse the complete Run II dataset, the high energy physics community is already busy looking ahead to the future. Following Run III, around 2023, a major upgrade of the LHC accelerator and detector systems will make the start of its High Luminosity (HL) operation phase possible. The ten–fold increase in its instantaneous luminosity will lead to the collection of huge datasets, with the HL–LHC expected to deliver around $\mathcal{L} = 3$ ab^{-1} to ATLAS and CMS and around $\mathcal{L} = 0.3$ ab^{-1} to LHCb. This unprecedented dataset will open new exciting physics opportunities, such as the measurement of the Higgs boson couplings to second generation fermions as well as of its self–interactions. These opportunities will be summarised in a CERN Yellow Report [27] to be presented before the end of 2018, in order to contribute to the update of the European Strategy for Particle Physics (ESPP).[1]

From the point of view of PDF determinations, the availability of these immense data samples will permit a significant extension of the kinematic coverage of PDF–sensitive measurements as well as a marked improvement in their statistical and systematic uncertainties. With this motivation, the goal of this work is to quantify the impact of the future HL–LHC measurements on the proton PDFs. In other words, we aim to assess the ultimate constraining power of hadron collider data on the PDFs. In turn, the resulting projections for the expected PDF uncertainties will feed into other related projections for HL–LHC processes, which will benefit from the associated reduction of theoretical errors.

It is important to emphasise here that while this type of study has previously been carried out in the context of lepton–hadron colliders such as the Large Hadron electron Collider (LHeC) and the Electron Ion Collider (EIC) [29–37], this is the first time that such a systematic effort has been devoted to determine the PDF–constraining potential of a future hadron collider. Clearly, being able to compare the information on PDFs that will be provided by the HL–LHC with that from proposed electron–proton colliders such as the LHeC represents an important input to inform the upcoming ESPP update.

Our analysis has been carried out as follows. First, we have generated HL–LHC pseudo–data for a number of PDF–sensitive processes: Drell–Yan production (both at high dilepton invariant mass and in the forward rapidity regions); W production in association with charm quarks (central and forward regions); inclusive jet and prompt photon production; the transverse momentum of Z bosons; and differential distributions in top quark pair production. We have selected those processes that should benefit more directly from the increased statistics available at the HL–LHC. We consider measurements such as inclusive W, Z production in the central region, which are already completely limited by systematic uncertainties [38,39], with no significant improvement anticipated from increased statistics alone. For each process, the binning and kinematic cuts applied to the pseudo–data is constructed from a suitable extension of reference measurements at $\sqrt{s} = 8$ and 13 TeV. We consider different scenarios for the expected systematic uncertainties, from a conservative one with approximately the same systematics as the corresponding baseline measurements from Run I and a factor of 2 reduction for those from Run II, to an optimistic one with a reduction by a factor 2.5 (5) as compared to Run I (II).

Subsequently, we quantify the constraints of the HL–LHC pseudo–data on the PDF4LHC15_100 set [40–43] by means of the Hessian Profiling method [44] (see also [45]). We have chosen the PDF4LHC15 set since it broadly represents the state–of–the–art understanding of the proton structure. While it is beyond the scope of this work to construct forecasts of the experimental correlation models, we account for their effective impact by using available Run I and II measurements as benchmarks. The resulting profiled sets are then implemented in the LHAPDF6 interface [46], thus being readily available for phenomenological applications.

By performing this analysis, we find that the legacy HL–LHC measurements can reduce the uncertainties in the PDF luminosities by a factor between 2 and 5 in comparison to state–of–the–art fits, depending on the specific flavour combination of the initial state and the invariant mass of the produced final state. We also show that our projections for the PDF error reduction, which are predominantly driven by the increased statistics of the HL–LHC data sample, depend only moderately on the specific scenario adopted for the reduction of the experimental systematic errors.

We then explore the implications of the profiled PDFs for representative LHC cross sections at $\sqrt{s} = 14$ TeV, both within the Standard Model (SM) and beyond it. Our analysis highlights how $\mathcal{O}(1\%)$ PDF uncertainties are within

---

[1] See also [28] for a recent review of the potential of the LHCb experiment in the HL–LHC era.
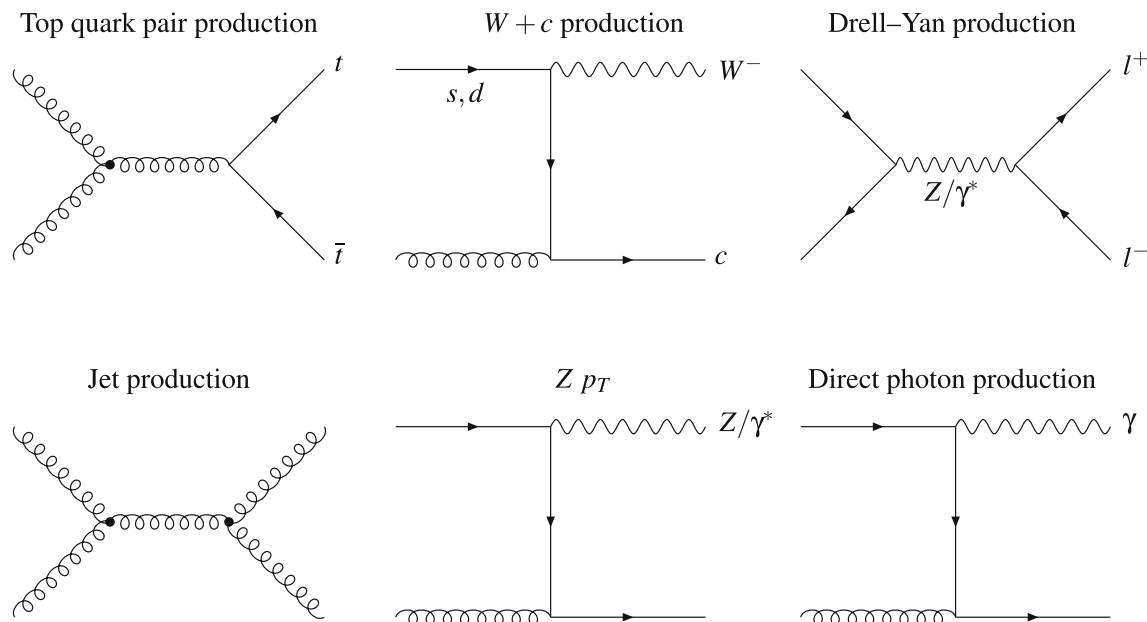
Fig. 1 Representative Feynman diagrams at the Born level of the six types of collider processes for which HL–LHC pseudo–data has been generated in this analysis: the production of top quark pairs, W bosons in association with charm quarks, and the neutral and charged current Drell–Yan processes; the production of inclusive jets, Z bosons at finite transverse momentum, and direct photons

the reach of the HL–LHC for key observables such as the transverse momentum distribution in Higgs production from gluon fusion. Therefore, our study illustrates the significant improvement in the precision of PDF determinations achievable from hadron collider data alone, and motivates the continuation of the ongoing successful program of PDF–sensitive measurements at the LHC.

The outline of the paper is the following. First, in Sect. 2 we describe the features of the PDF–sensitive processes used to generate the HL–LHC pseudo–data. Then in Sect. 3 we quantify the constraints on the PDFs of individual processes using the Hessian profiling method. The full set of HL–LHC pseudo–data is combined in Sect. 4 to construct the ultimate HL–LHC parton distributions, which is then used to assess their phenomenological implications for different processes both in the SM and beyond it. Finally, in Sect. 5 we summarise our results and indicate how they are made publicly available.

2 Pseudo–data generation

In this section we present the PDF–sensitive processes for which HL–LHC pseudo–data have been generated, provide details about the binning and kinematic cuts, and also describe the baseline Run I and II measurements that are used to model the experimental systematic uncertainties expected in the HL–LHC era.

2.1 PDF–sensitive processes

We start by describing the PDF–sensitive processes that will be considered in this study to generate HL–LHC pseudo–data. Our analysis is based on six different types of processes: the production of top quark pairs, jets, direct photons, and W bosons in association with charm quarks, the transverse momentum of Z bosons, and the neutral and charged current Drell–Yan processes. In Fig. 1 we show representative Feynman diagrams at the Born level for all of these processes, in order to illustrate their sensitivity to the different partonic initial states. For instance, we see that jets, photon, and top quark pair production are directly dependent on the gluon content of the proton, while W+charm is sensitive to strangeness, and the Drell–Yan process to the quark–antiquark luminosity.

This choice of input processes is driven by the fact that some types of hard–scattering reactions should benefit more directly from the increased statistics offered by the HL–LHC than others. Indeed, some of the existing LHC measurements, such as inclusive W, Z production in the central region [38,39], are already limited by systematic uncertainties, and therefore are unlikely to improve significantly at higher luminosities. On the other hand, our selection of processes will greatly benefit from the huge HL–LHC dataset either because they are relatively rare, such as W+charm, or because their kinematic coverage can be extended to regions of large invariant masses and transverse momentum or forward rapidities where event rates exhibit a steep fall–

off. While these pseudo–data sets do include some regions which are currently systematics dominated, i.e. towards central rapidity and lower mass/transverse momentum, as we will see the dominant PDF impact comes from the regions which are not, where the existing data are less constraining and the contributing PDFs are currently less well determined.

In more detail, the specific processes for which HL–LHC pseudo–data have been generated are the following:

- High–mass Drell–Yan, specifically the dilepton invariant mass differential distribution $d\sigma(pp \rightarrow ll)/dm_{ll}$ for $m_{ll} \gtrsim 110$ GeV for a central rapidity acceptance, $|\eta_l| \leq 2.4$.

  This process is particularly useful for quark flavour separation, specifically to constrain the poorly known large–x sea quarks.

  Here the ATLAS 8 TeV measurement of differential Drell–Yan cross sections [47] is taken as reference, with additional bins in the high m_{ll} region included to benefit from the enhanced kinematic coverage.

- The differential distributions for on–peak W and Z boson production in the forward region, $2.0 \leq \eta_l \leq 4.5$, covered by the LHCb experiment.

  These measurements constrain quark flavour separation, including the strange and charm content of the proton, in the large and small x region [48], complementary to the data from the central region.

  The reference analysis is the LHCb measurement of the rapidity distributions of W and Z bosons in the muon final state at 8 TeV [49].

  In comparison to the reference measurement, a finer binning by a factor of 2 to 5 has been adopted as allowed by the increased event rates.

  Events are selected if $p_T^l \geq 20$ GeV, the lepton rapidities fall in the LHCb acceptance, and, in the case of Z production, there is the additional requirement that 60 GeV $\leq m_{ll} \leq 120$ GeV.

- Differential distributions in top quark pair production, providing direct information on the large x gluon [15].

  Specifically, we consider the top quark transverse momentum p_T^t and rapidity y_t, and the top quark pair rapidity $y_{t\bar{t}}$ and invariant mass $m_{t\bar{t}}$.

  The reference measurements here are the ATLAS 8 TeV differential distributions in the lepton+jets final state [50]. We assume that the statistical correlations between different distributions will be available, as is the case for the 8 TeV data [51], and therefore include the four distributions simultaneously in the fit. To account for the increased statistics of the HL–LHC, the number of bins in the rapidity distributions is doubled, while the p_T^t and $m_{t\bar{t}}$ distributions are extended to higher values in the TeV region.

- The transverse momentum distribution of the Z bosons in the dilepton final state, 20 GeV $< p_T^{ll} < 3.5$ TeV region for central rapidities $|\eta_Z| \leq 2.4$ and different bins of the dilepton invariant mass m_{ll}.

  This process is relevant to constrain the gluon and the antiquarks at intermediate values of x [17].

  For the reference analysis, we take the ATLAS measurements of the transverse momentum of lepton pairs at 8 TeV [52].

  The pseudo–data is generated for six different bins of the dilepton invariant mass m_{ll}, with boundaries 12, 20, 30, 40, 66, 116, and 150 GeV respectively.

  In each of the invariant mass m_{ll} bins, additional bins are added to the p_T^{ll} distribution to exploit the improved coverage of the large transverse momentum region.

- The production of W bosons in association with charm quarks.

  This process provides a sensitive handle on the strangeness content of the proton [23,53], which is the least well known of the light quark PDFs.

  The pseudo–data for this process has been generated as a function of the lepton psuedorapidity η_l from the W boson decay, and is inclusive over the kinematics of the charm quark provided it satisfies the selection cuts.

  For this process, pseudo–data have been generated both for the central rapidity region relevant for ATLAS and CMS as well as for the forward region covered by LHCb. In the central rapidity region, $|\eta^l| \leq 2.4$, the reference measurement is the CMS analysis at 13 TeV [24], where events are selected provided that $p_T^c \geq 5$ GeV and $p_T^l \geq 26$ GeV with l indicating the result of the $W \rightarrow l\nu$ decay. At forward rapidities, $2 \leq \eta^l \leq 4.5$, we use a dedicated selection strategy with $2.2 \leq \eta^c \leq 4.2$, $p_T^\mu \geq 20$ GeV, $p_T^c \geq 20$ GeV, and $p_T^{\mu+c} \geq 20$ GeV [54]. We take the acceptance to be 30% due to c–jet tagging and an overall normalisation error of 5%.

- Prompt isolated photon production represents a complementary probe of the gluon PDF at intermediate values of x [19].

  Here the pseudo–data has been generated as differential distributions in the photon transverse momentum p_T^γ for different bins in the photon rapidity η^γ.

  The reference measurements here is the ATLAS 13 TeV analysis of [55], where additional bins have been added to the p_T^γ distribution in each rapidity bins to benefit from the improved coverage of the large p_T^γ region.

- Finally, we consider the inclusive production of hadronic jets in different bins of their rapidity up to $|y_{\text{jet}}| \leq 3$ as a function of their p_T^{jet}.

  This process provides direct information on the gluon and the valence quarks at large–x.

Here the jets are reconstructed using the anti–k_T algorithm with $R = 0.4$ as radius parameter.

The reference measurement here is the 13 TeV ATLAS analysis of inclusive jet and dijet production based on a luminosity $\mathcal{L} = 3.2$ fb^{-1} from the 2015 data–taking period. The coverage of the high–p_T region has been extended to the $p_T^{\text{jet}} \simeq 2 - 3$ TeV in comparison to these.

It is important to emphasise that the list of processes considered in this work is by no means exhaustive. Clearly, there are other important processes that will provide useful information on the proton PDF in the HL–LHC era. Among these, one could consider dijet production [56] and single top quark production [57], providing information on the gluon and on the quark flavour separation respectively. In both cases, the NNLO corrections for differential distributions are available, as well as reference LHC measurements at 8 and 13 TeV. Our choices of processes are in addition generally geared towards the high and intermediate x region. Other PDF–sensitive processes, such as inclusive D meson production, can play a role at lower x. Although in this case it is unlikely to benefit from the high–luminosity phase, as it already occurs at very high rates, this may not be true for other rarer processes sensitive to this region.

In addition, one should take into account that progress from both the experimental and theoretical sides could lead to novel processes being added to the PDF fitting toolbox, for instance more exclusive processes or processes for which the standard DGLAP description breaks down. With these caveats, the set of processes adopted in this work is representative enough to provide a reasonable snapshot of the PDF–constraining potential of the HL–LHC.

It is also important to mention that the HL–LHC projections presented in this work are based on pseudo–data generated specifically for this study, and that they are not endorsed by the LHC collaborations. However, we have taken into account all the feedback and suggestions received from the ATLAS, CMS, and LHCb contacts involved in the Yellow Report studies.

2.2 Theory calculations and pseudo–data generation

For the various processes described above, we have generated pseudo–data for a centre–of–mass energy of $\sqrt{s} = 14$ TeV assuming a total integrated luminosity of $\mathcal{L} = 3$ ab^{-1} for the CMS and ATLAS experiments, and of $\mathcal{L} = 0.3$ ab^{-1} for the LHCb experiment. Note that in the former case we explicitly include pseudo–data from both experiments. Statistical uncertainties are evaluated from the expected number of events per bin, taking into account branching ratios and acceptance corrections determined from the corresponding reference analysis. Systematic uncertainties are taken to be

those of the 13 or 8 TeV baseline analyses and then rescaled appropriately. We consider various scenarios for the reduction of systematic errors, from a more conservative one to a more optimistic one.

Theoretical predictions are computed at next–to–leading order (NLO) in the QCD expansion using MCFM [58] interfaced to APPLgrid [59] to produce the corresponding fast grids. The only exception is inclusive jet production, for which the NLO calculation is obtained from the NLOJET++ program [60]. The central value of the pseudo–data initially coincides with the corresponding prediction obtained using this NLO calculation with the PDF4LHC15 NNLO set as input. Subsequently, this central value is fluctuated according to the corresponding experimental uncertainties. This implies that, by construction, one should find $\chi^2/N_{\text{dat}} \simeq 1$ from the fit to the pseudo–data.

Specifically, if σ_i^{th} is the theoretical cross section for bin i of a given process, then the central value of the HL–LHC pseudo–data σ_i^{exp} is constructed by means of

$$\sigma_i^{\text{exp}} = \sigma_i^{\text{th}} \times \left(1 + r_i \cdot \delta_{\text{tot},i}^{\text{exp}} + \lambda \cdot \delta_{\mathcal{L}}^{\text{exp}} + s \cdot \delta_{\mathcal{N}}^{\text{exp}} \right), \quad (2.1)$$

where r_i, λ, and s are univariate Gaussian random numbers, $\delta_{\text{tot},i}^{\text{exp}}$ is the total (relative) experimental uncertainty corresponding to this specific bin (excluding the luminosity and normalization uncertainties), and $\delta_{\mathcal{L}}^{\text{exp}}$ is the luminosity uncertainty, which is fully correlated among all the pseudo–data bins of the same experiment (but uncorrelated among different experiments). We take this luminosity uncertainty to be $\delta_{\mathcal{L}}^{\text{exp}} = 1.5\%$ for the three LHC experiments. $\delta_{\mathcal{N}}^{\text{exp}}$ are possible additional normalization uncertainties as in the case of W boson production in association with charm quarks that will be explained later.

In Eq. (2.1), the total experimental uncertainty $\delta_{\text{tot},i}^{\text{exp}}$ is defined as

$$\delta_{\text{tot},i}^{\text{exp}} \equiv \left(\left(\delta_{\text{stat},i}^{\text{exp}} \right)^2 + \left(f_{\text{corr}} \times f_{\text{red}} \times \delta_{\text{sys},i}^{\text{exp}} \right)^2 \right)^{1/2}. \quad (2.2)$$

In this expression, the relative statistical error $\delta_{\text{stat},i}^{\text{exp}}$ is computed as

$$\delta_{\text{stat},i}^{\text{exp}} = \left(f_{\text{acc}} \times N_{\text{ev},i} \right)^{-1/2}, \quad (2.3)$$

where $N_{\text{ev},i} = \sigma_i^{\text{th}} \times \mathcal{L}$ is the expected number of events in bin i at the HL–LHC with $\mathcal{L} = 3 \ (0.3)$ ab^{-1}. In Eq. (2.3), $f_{\text{acc}} \leq 1$ is an acceptance correction which accounts for the fact that, for some of the processes considered, such as top quark pair production, there is a finite experimental acceptance for the final state products and/or one needs to include the effects of branching fractions. The value of f_{acc} is determined by extrapolation using the reference dataset, except for forward W+charm production (where there is no baseline measurement) where the acceptance is set to $f_{\text{acc}} = 0.3$, due dominantly to the c–jet tagging efficiency.

Table 1 Summary of the features of the HL–LHC pseudo–data generated for the present study For each process we indicate the kinematic coverage, the number of pseudo–data points used across all detectors N_{dat}, the values of the correction factors f_{corr} and f_{red}; and finally the reference from the 8 TeV or 13 TeV measurement used as baseline to define the binning and the systematic uncertainties of the HL–LHC pseudo–data, as discussed in the text

Process	Kinematics	N_{dat}	f_{corr}	f_{red}	Baseline
$Z\ p_T$	$20\,\text{GeV} \leq p_T^{ll} \leq 3.5\,\text{TeV}$	338	0.5	(0.4, 1)	[52] (8 TeV)
	$12\,\text{GeV} \leq m_{ll} \leq 150\,\text{GeV}$				
	$|y_{ll}| \leq 2.4$				
High-mass Drell-Yan	$p_T^{l1(2)} \geq 40(30)\,\text{GeV}$	32	0.5	(0.4, 1)	[47] (8 TeV)
	$|\eta^l| \leq 2.5, m_{ll} \geq 116\,\text{GeV}$				
Top quark pair	$m_{t\bar{t}} \simeq 5\,\text{TeV}, |y_t| \leq 2.5$	110	0.5	(0.4, 1)	[50] (8 TeV)
W+charm (central)	$p_T^\mu \geq 26\,\text{GeV}, p_T^c \geq 5\,\text{GeV}$	12	0.5	(0.2, 0.5)	[24] (13 TeV)
	$|\eta^\mu| \leq 2.4$				
W+charm (forward)	$p_T^\mu \geq 20\,\text{GeV}, p_T^c \geq 20\,\text{GeV}$	10	0.5	(0.4, 1)	LHCb projection
	$p_T^{\mu+c} \geq 20\,\text{GeV}$				
	$2 \leq \eta^\mu \leq 4.5, 2.2 \leq \eta^c \leq 4.2$				
Direct photon	$E_T^\gamma \lesssim 3\,\text{TeV}, |\eta_\gamma| \leq 2.5$	118	0.5	(0.2, 0.5)	[55] (13 TeV)
Forward W, Z	$p_T^l \geq 20\,\text{GeV}, 2.0 \leq \eta^l \leq 4.5$	90	0.5	(0.4, 1)	[49] (8 TeV)
	$60\,\text{GeV} \leq m_{ll} \leq 120\,\text{GeV}$				
Inclusive jets	$|y| \leq 3, R = 0.4$	58	0.5	(0.2, 0.5)	[61] (13 TeV)
Total		768			

In Eq. (2.2), $\delta_{\text{sys},i}^{\text{exp}}$ indicates the total systematic error of bin i taken from the reference LHC measurement at either 8 TeV or 13 TeV, while $f_{red} \leq 1$ is a correction factor that accounts for the fact that on average systematic uncertainties will decrease at the HL–LHC in comparison to Run II due to both detector improvements and the enlarged dataset for calibration. Finally, f_{corr} represents an effective correction factor that accounts for the fact that data with correlated systematics may be more constraining than the same data where each source of error is simply added in quadrature, as we do in this analysis. We discuss below in Sect. 2.3 how the value of f_{corr} can be determined by means of available LHC measurements for which the full information on correlated systematics is available.

Concerning the theoretical calculations adopted here, since the present study relies on pseudo–data, it is not necessary to account for higher–order QCD effects or electroweak corrections. Indeed, by far the dominant contribution to the PDF sensitivity of hadron collider processes is contained within the NLO calculation. As in the case of PDF closure tests [62], here we are only interested in the relative reduction of the PDF uncertainties once the HL–LHC data is included in the fit, while the central value itself will be essentially unaffected. Note that this also holds for the contribution of photon–initiated (PI) processes, since the photon PDF is very well know [63–65]. Therefore, PI processes effectively induce an overall rescaling of the cross section which becomes irrelevant when generating pseudo–data.

In Table 1 we present the summary of the main features of the HL–LHC pseudo–data generated for the present study. For each process, we indicate the kinematic coverage, the

number of pseudo–data points used N_{dat}, the values of the correction factors f_{acc}, f_{corr}, and f_{red}; and finally the reference for the 8 TeV or 13 TeV measurement used as baseline to define the binning and the systematic uncertainties of the HL–LHC pseudo–data. A total of around $N_{dat} = 768$ pseudo–data points are then used in the PDF profiling. The values of the reduction factor for the systematic errors f_{red} is varied between 1 (0.5) and 0.4 (0.2) in the conservative and optimistic scenarios for a 8 TeV (13 TeV) baseline measurement. This different treatment is motivated by the fact that available 13 TeV measurements are based on a smaller dataset and therefore tend to have larger systematic errors in comparison to the 8 TeV case. Thus we can expect some improvement here at the HL–LHC even in the most conservative scenario; Run II measurements based on the complete integrated luminosity will certainly benefit from reduced systematics.
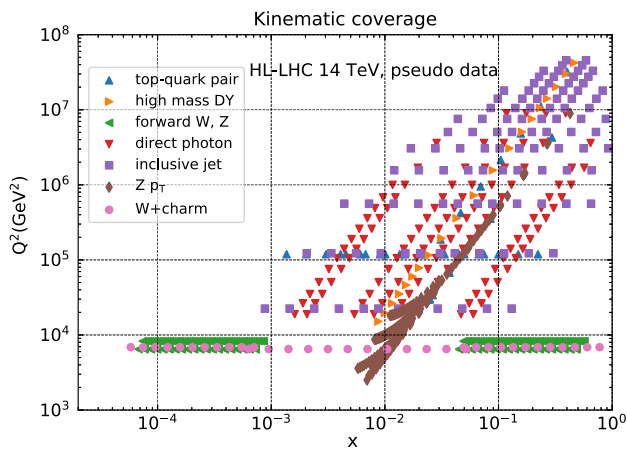
In Fig. 2 we show the kinematical coverage in the (x, Q^2) plane of the HL–LHC pseudo–data included in this analysis. For each data point, the values of (x_1, Q) and (x_2, Q) corresponding to the two colliding partons are determined approximately from leading–order kinematics, which is sufficient for illustration purposes. We assume $x_1 = x_2$ if rapidities are not specified for the final states. We see that the HL–LHC pseudo–data covers a wide kinematic region, including the large momentum transfers up to $Q \simeq 6$ TeV, as well as the large-x region, with several different processes. Specifically, the input pseudo–data spans the range $6 \times 10^{-5} \lesssim x \lesssim 0.7$ and $40\,\text{GeV} \lesssim Q \lesssim 7\,\text{TeV}$ in the (x, Q) kinematic plane. Note that the LHCb measurements are instrumental to con-

Fig. 2 The kinematical coverage in the (x, Q^2) plane of the HL–LHC pseudo–data included in this analysis. For each data point, the values of (x_1, Q^2) and (x_2, Q^2) corresponding to each of the two colliding partons are determined approximately from the corresponding leading–order kinematics. We assume $x_1 = x_2$ if rapidities are not specified for the final states. The HL–LHC pseudo–data therefore spans a wide region in the kinematic plane, namely $6 \times 10^{-5} \lesssim x \lesssim 0.7$ and 40 GeV $\lesssim Q \lesssim 7$ TeV

strain the small–x region, $6 \times 10^{-5} \lesssim x \lesssim 10^{-3}$, beyond the acceptance of ATLAS and CMS.

2.3 Impact of correlating uncertainties

As we will also discuss in Sect. 3, when constructing the χ^2 estimator for the HL–LHC pseudo–data we will not explicitly include the correlations between the systematic errors. Instead, we add statistical and systematic uncertainties in quadrature as indicated in Eq. (2.2). This choice is motivated by the fact that it is already challenging to estimate how specific systematic uncertainties will be reduced at the HL–LHC, let alone how their mutual correlations will be modified. Note that even restricting ourselves to Run I measurements, the determination of the experimental correlation model is a delicate problem, and can in some cases complicate the PDF interpretation of measurements such as inclusive jet production [66].

On the other hand, completely neglecting the effects of the experimental correlations may artificially reduce the impact of the pseudo–data into the fit. Precisely for this reason, we have introduced the correction factor $f_{\rm corr}$ in Eq. (2.2). Its value has been tuned to the LHC measurements of the top quark pair differential distributions [50,67] at $\sqrt{s} = 8$ TeV and of the central W+charm rapidity distribution [24] at $\sqrt{s} = 13$ TeV, for which the full breakdowns of systematic errors are available. Although not shown explicitly here, in the latter case we also check against the published 7 TeV data [68], finding similar results to the preliminary 13 TeV. In the following comparison, we use a value for the tolerance of $T = 1$ (defined in the next section) to exaggerate the effect of

$f_{\rm corr}$ due to the new data having a more dominate role in the χ^2, enabling an easier determination of the optimal value.

In Fig. 3 we compare the baseline PDF4LHC15 set and the sets profiled with these two LHC datasets, with or without the correlations between the experimental systematic uncertainties accounted for. In the latter case, the $f_{\rm corr}$ factor is chosen to reproduce the results of the profiling when the correlations are included. We can see that for the two considered datasets, rather different values of $f_{\rm corr}$ are preferred; for the top data, we require $f_{\rm corr} \sim 0.25$ while for the W+charm data we require instead $f_{\rm corr} \sim 1$. Clearly the precise value of this correction therefore appears to depend quite sensitively on the considered datasets, in terms of the corresponding breakdown of systematic uncertainties and overall PDF impact.

The results of Fig. 3 might suggest that, for projections which are dominantly driven by the potential improvement in systematic uncertainties, our approach could be questionable and require a more complete treatment of experimental correlations. However, here we have explicitly chosen our input dataset to be composed of those processes for which the PDF impact will be driven instead by the improvement in the statistics and extension to unconstrained kinematic regions. Indeed, we will see later on that the specific value of this parameter does not have a large impact on the final results, and we will simply take $f_{\rm corr} = 0.5$ in what follows as an average, somewhat weighted towards the value required by the top quark differential data, as this shows a larger PDF impact and would therefore be more important to account for accurately.

3 HL–LHC constraints from individual processes

In this section, we study the constraints on the PDFs that are expected from individual HL–LHC measurements listed in Table 1. First of all, we review the formulation of the Hessian profiling used in this work to quantify the PDF constraints. Then we present the results for the various HL–LHC processes and study how the description of the pseudo–data is affected. The complete set of processes is combined together into a single profiled PDF set in the next section.

3.1 The Hessian profiling method

Quantifying the impact of new experimental data into a Hessian PDF set such as PDF4LHC15_100 can be efficiently carried out by means of the Hessian Profiling technique [44,45]. This approach is based on the minimization of the following figure of merit:
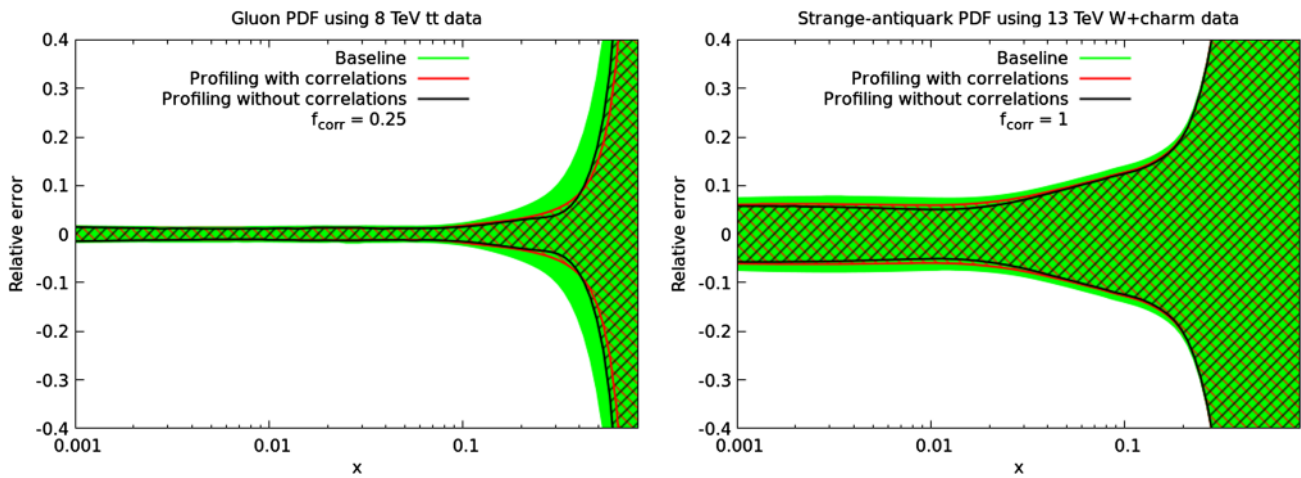
Fig. 3 Comparison between the baseline PDF4LHC15 set and the sets profiled with the LHC data, either with or without the correlations between the experimental systematic uncertainties accounted for. In the latter case, the f_{corr} factor is chosen to reproduce the results of the profiling when the correlations are included, see text. We show here

the results of profiling with the top differential distributions at $\sqrt{s} = 8$ TeV with $f_{corr} = 0.25$ (left) and the W+charm rapidity distribution at $\sqrt{s} = 13$ TeV with $f_{corr} = 1$ (right plot). A tolerance factor of $T = 1$ has been used for this specific comparison

$$
\chi^2 \left(\beta_{\mathrm{exp}}, \beta_{\mathrm{th}} \right)
$$
$$
= \sum_{i=1}^{N_{\mathrm{dat}}} \frac{1}{\left(\delta_{\mathrm{tot},i}^{\mathrm{exp}} \sigma_i^{\mathrm{th}} \right)^2} \left(\sigma_i^{\mathrm{exp}} + \sum_j \Gamma_{ij}^{\mathrm{exp}} \beta_{j,\mathrm{exp}} - \sigma_i^{\mathrm{th}} + \sum_k \Gamma_{ik}^{\mathrm{th}} \beta_{k,\mathrm{th}} \right)^2
$$
$$
+ \sum_j \beta_{j,\mathrm{exp}}^2 + T^2 \sum_k \beta_{k,\mathrm{th}}^2 , \tag{3.1}
$$

where σ_i^{exp} (σ_i^{th}) are the central values of a given experimental measurement (theory prediction), see Eq. (2.1), $\beta_{j,\mathrm{exp}}$ are the nuisance parameters corresponding to the set of fully correlated experimental systematic uncertainties, $\beta_{k,\mathrm{th}}$ are the nuisance parameters corresponding to the PDF Hessian eigenvectors, N_{dat} is the number of data points and T is the tolerance factor. The matrices $\Gamma_{ij}^{\mathrm{exp}}$ and $\Gamma_{ik}^{\mathrm{th}}$ encode the effects of the corresponding nuisance parameters on the experimental data and on the theory predictions, respectively.

The minimisation of Eq. (3.1) produces approximately equivalent results to carrying out the corresponding Hessian fit from scratch, provided settings such as the input PDF parameterisations, the tolerance factor T, and the theoretical calculations are unchanged. An advantage of the Hessian profiling method in comparison to related techniques such as the Bayesian reweighting method [69,70], relevant for Monte Carlo PDF sets, is that there is no information loss even when the added measurements provide significant new information. This property is crucial in the present analysis, since the HL–LHC pseudo–data induces significant constraints on the PDFs.

As mentioned in Sect. 2.2, in this study we add in quadrature statistical and experimental uncertainties (except for the luminosity, which is kept fully correlated), and then account

for the effects of the missing correlations by means of the factor f_{corr}. For this reason, we only consider nuisance parameters for the luminosity errors, as well as for an overall normalization uncertainty of 5% in forward W+charm production, arising from charm–jet tagging. If we then minimise Eq. (3.1) with respect to these experimental nuisance parameters we obtain

$$
\chi^2 \left(\beta_{\mathrm{th}} \right) = \sum_{i,j=1}^{N_{\mathrm{dat}}} \left(\sigma_i^{\mathrm{exp}} - \sigma_i^{\mathrm{th}} + \sum_k \Gamma_{ik}^{\mathrm{th}} \beta_{k,\mathrm{th}} \right) (\mathrm{cov})_{ij}^{-1}
$$
$$
\times \left(\sigma_j^{\mathrm{exp}} - \sigma_j^{\mathrm{th}} + \sum_m \Gamma_{jm}^{\mathrm{th}} \beta_{m,\mathrm{th}} \right) + T^2 \sum_k \beta_{k,\mathrm{th}}^2 , \tag{3.2}
$$

where we have defined the experimental covariance matrix as follows:

$$
(\mathrm{cov})_{ij} = \delta_{ij} \left(\delta_{\mathrm{tot},i}^{\mathrm{exp}} \sigma_i^{\mathrm{th}} \right)^2 + \Gamma_{i,\mathrm{lumi}}^{\mathrm{exp}} \Gamma_{j,\mathrm{lumi}}^{\mathrm{exp}} + \Gamma_{i,\mathrm{norm}}^{\mathrm{exp}} \Gamma_{j,\mathrm{norm}}^{\mathrm{exp}} . \tag{3.3}
$$

Note that since Eq. (3.3) is defined in terms of a fixed theoretical prediction (rather than of the fit output itself), our results are resilient with respect to the D'Agostini bias [71,72].

At this point, the minimisation of Eq. (3.2) with respect to the Hessian PDF nuisance parameters $\beta_{k,\mathrm{th}}$ can be interpreted as leading to PDFs that have been optimized to describe this new specific measurement. The resulting Hessian matrix in the $\beta_{k,\mathrm{th}}$ parameter space at the minimum can be diagonalized to construct the new eigenvector directions, and PDF uncertainties are determined from the $\Delta\chi^2 = T^2$ criteria. In the studies presented here, we use $T = 3$, which roughly
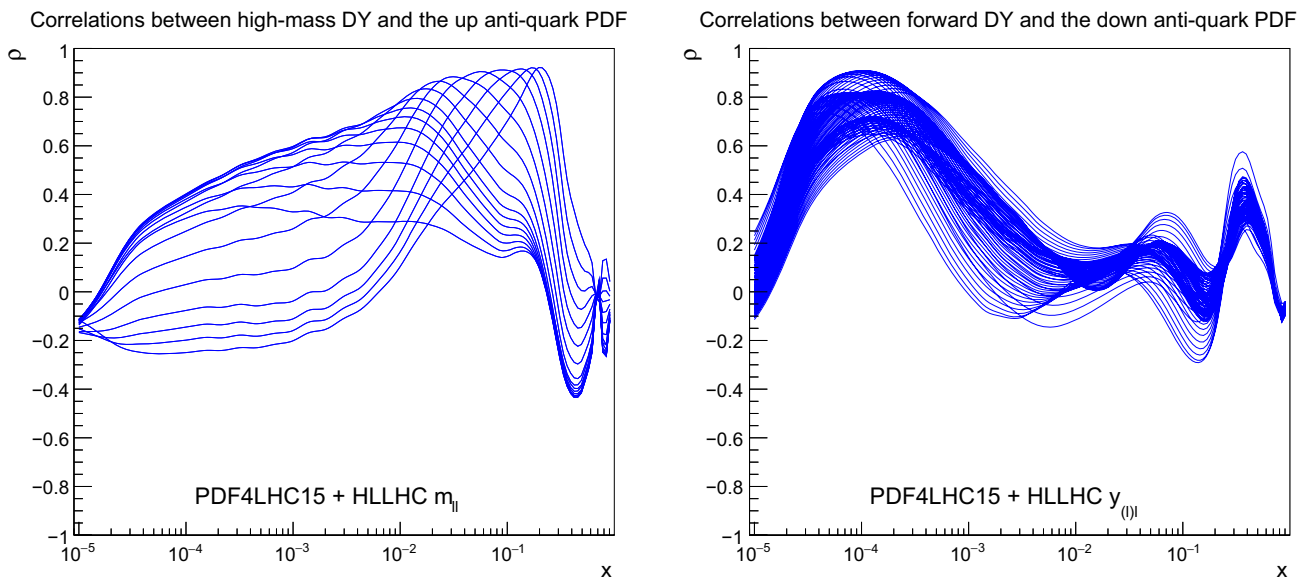
Correlations between high-mass DY and the up anti-quark PDF

Correlations between forward DY and the down anti-quark PDF

Fig. 4 The correlation coefficients ρ between the PDFs and the HL–LHC pseudo–data. Left: the correlation between the anti–up quark and the high–mass Drell–Yan pseudo–data as a function of x for $Q = 100$ GeV. Right: the correlation between the anti–down quark and the inclu-sive W, Z production process in the forward region. In each plot, the different curves correspond to each of the bins of the pseudo–data used in the fit

corresponds to the average tolerance determined dynamically in the CT14 and MMHT14 analyses. The resulting profiled PDF set[2] can be straightforwardly used for phenomenology using the uncertainty prescription of symmetric Hessian sets, and the default output format is compliant with the `LHAPDF` interface.

3.2 Inclusive gauge boson production

We now present results for the Hessian profiling of the PDF4LHC15 set after the inclusion of HL–LHC pseudo–data from individual processes. Then in Sect. 4 we will consider the results of the combination for all the processes together. We begin with the inclusive gauge boson production processes listed in Table 1 and described in Sect. 2.1. We consider two sets of pseudo–data: inclusive γ^*/Z production in the central rapidity $|\eta_{ll}| \leq 2.4$ and high invariant mass $m_{ll} \geq 116$ GeV regions, and inclusive W^+, W^-, γ^*/Z production in the forward region, $2.0 \leq \eta^l \leq 4.5$.

In this section, we will use the same structure to discuss the impact on the PDFs of the individual HL–LHC processes that are being considered. First, we will display representative examples of the correlations between the PDFs and the pseudo–data, to illustrate the sensitivity of the latter. Second, we will show how the description of the HL–LHC pseudo–data is modified once it is included in the PDF4LHC15 set

by means of profiling. Finally, we will assess its impact on the PDFs in a specific scenario for the projections of the experimental systematic errors. In particular, we adopt the 'optimistic' choice of Table 2, i.e. $F \equiv f_{\rm corr} \cdot f_{\rm red} = 0.2$, which corresponds to a value $f_{\rm red} = 0.4$ for the reduction of the systematic uncertainties compared to the 8 TeV baseline measurements. As discussed above, for 13 TeV baselines, in this scenario we take a lower value of $f_{\rm red} = 0.2$, to account for the smaller 13 TeV datasets these are based on.

We start by discussing the correlations. In Fig. 4 we show the correlation coefficients ρ between the PDFs and the HL–LHC pseudo–data on the Drell–Yan process. The left (right) plot displays the correlation between the anti–up (anti–down) quark as a function of x for $Q = 100$ GeV for the high–mass (forward) Drell–Yan pseudo–data. A value of ρ close to 1 (-1) in a given region of x indicates that this process is strongly (anti–) correlated with the input PDFs in this same region, and thus that could potentially be used to reduce PDF uncertainties there.

As we can see from Fig. 4, in the case of high–mass Drell–Yan we have $\rho \geq 0.9$ for $0.05 \lesssim x \lesssim 0.5$, indicating that this process can provide information on the large–x antiquarks. In the case of the forward W, Z production measurements the correlation coefficient for the $\bar{d}$ PDF peaks at $x \simeq 10^{-4}$, highlighting that the forward kinematic coverage of LHCb allows the quark flavour separation to be pinned down to small values of x.

We next assess the impact of inclusive gauge boson production, after profiling. In Fig. 5 we show the comparison between the HL–LHC pseudo–data and the theoretical pre-

---

[2] Note that sometimes in this paper we will for brevity use the shorthand 'fit', but it always understood that a profiling has been performed rather than a full refit.
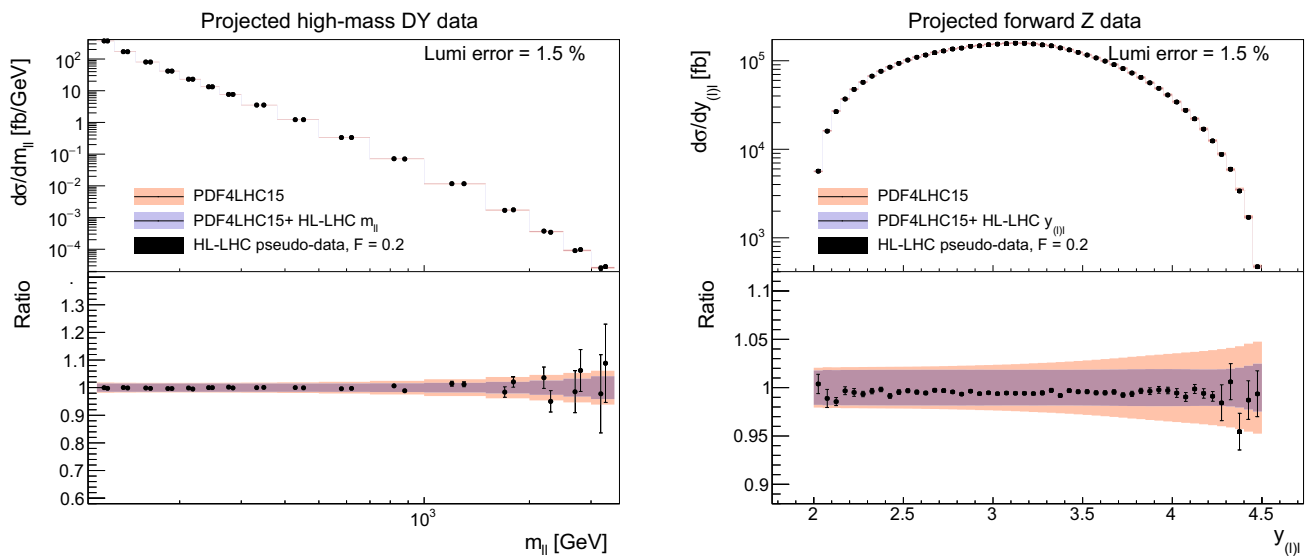
Fig. 5 Comparison between the HL–LHC pseudo–data and the theoretical predictions for high–mass (left) and forward (right) Drell–Yan production. The theory calculations are shown both before (PDF4LHC15) and after profiling. Luminosity uncertainties are not shown in the experimental errors. In the bottom panel, we show the same results normalised to the central value of the original theory calculation. Note in the right plot the comparison are only made for forward Z data though both W and Z data are included in the profiling

dictions for high–mass (left) and forward (right) Drell–Yan production. Note in the right plot the comparison is only made for forward Z data, but both W and Z data are included in the profiling. In addition, in the left plot, in each bin there are two experimental pseudo–data points, corresponding to the ATLAS and CMS projections; this is true for all central rapidity pseudo–datasets which follow. The theory calculations are shown both before (that is, using the PDF4LHC15 set) and after profiling. Luminosity uncertainties are not shown in the experimental errors, but are included in the profiling. In the bottom panel, we show the same results normalised to the central value of the original theory calculation.

From these comparisons, we see that the impact of the high–mass Drell–Yan pseudo–data on the PDFs is rather moderate, presumably because even at the HL–LHC the expected precision of the measurements is comparable or larger than current PDF uncertainties, in particular in the high m_{ll} range. On the other hand, for the W, Z measurements that will be carried out by LHCb we can observe a marked error reduction of up to a factor two, highlighting the usefulness of the forward kinematic coverage. Note that in both cases the central values of the theoretical predictions are relatively unaffected, with the dominant impact being on the uncertainties. This is expected, as by construction we assume the datasets are consistent with the underlying theory and PDFs.

Concerning the corresponding impact of the HL–LHC pseudo–data on the PDFs, in Fig. 6 we show the reduction of the PDF uncertainties found upon the inclusion of the high–mass Drell–Yan (left) and the forward W, Z (right) pseudo–data on the PDF4LHC15 set. We display the same

PDF flavours as those used in the calculation of the correlation coefficients in Fig. 4, namely the up and down antiquarks respectively. What we find is consistent with Fig. 5: a rather moderate effects on the up antiquark from the high–mass Drell–Yan process, while a more marked effect on the down antiquark from the forward W, Z process specially in the small–x region.

3.3 Top quark pair production

Here we will focus on the gluon PDF, given that at the LHC top quark pairs are mostly produced via gluon fusion. As explained in Sect. 2.1, we include four different distributions simultaneously: p_T^t, y_t, $m_{t\bar{t}}$, and $y_{t\bar{t}}$, assuming that the statistical correlations among them will be available. First, in Fig. 7 we show the same correlation coefficients as in Fig. 4 now between the gluon PDF and the various bins of $m_{t\bar{t}}$, the invariant mass distribution of top quark pairs. The fact that ρ peaks in the large–x region indicates that adding the $t\bar{t}$ distributions will directly constrain the gluon PDF here.

Next, in Fig. 8 (left) we show the same comparison as in Fig. 5 now for the $m_{t\bar{t}}$ distribution. We can observe a very marked PDF uncertainty reduction at large values of the invariant mass. As expected, we find in Fig. 8 (right) that the addition of the HL–LHC $t\bar{t}$ pseudo–data leads to a significant reduction in the PDF uncertainties in the gluon PDF at large–x, highlighting the good constraining power of this type of measurements.
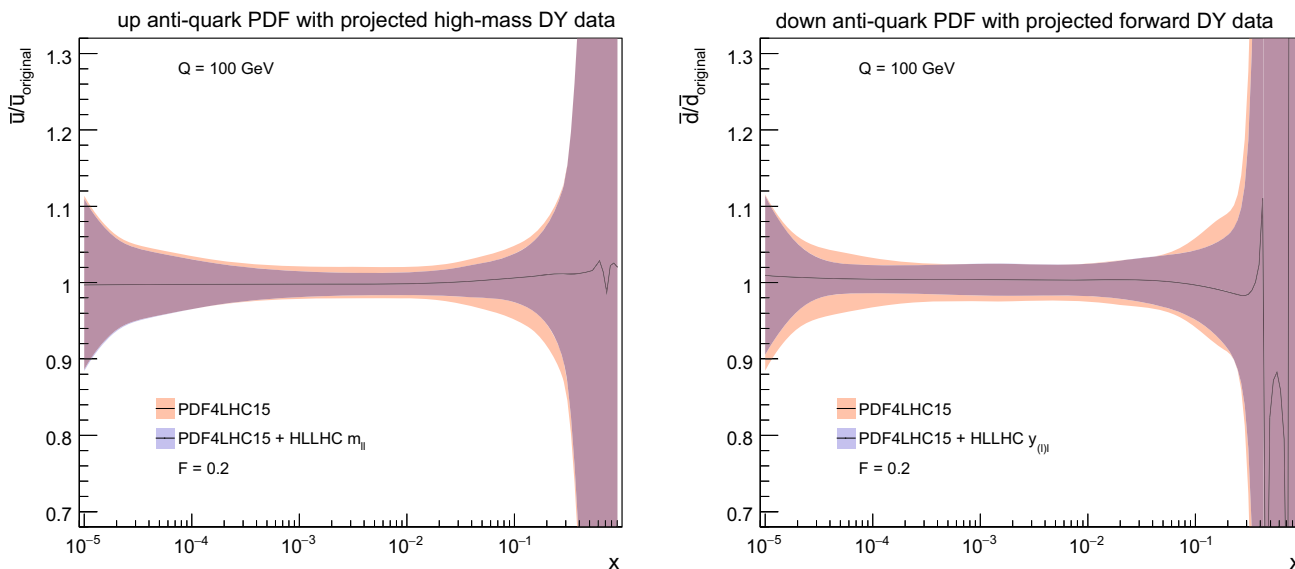
Fig. 6 The impact of the HL–LHC pseudo–data on the PDFs at $Q = 100$ GeV. Left: impact of high–mass Drell–Yan production on the up antiquark. Right: impact of forward W, Z process on the down antiquark
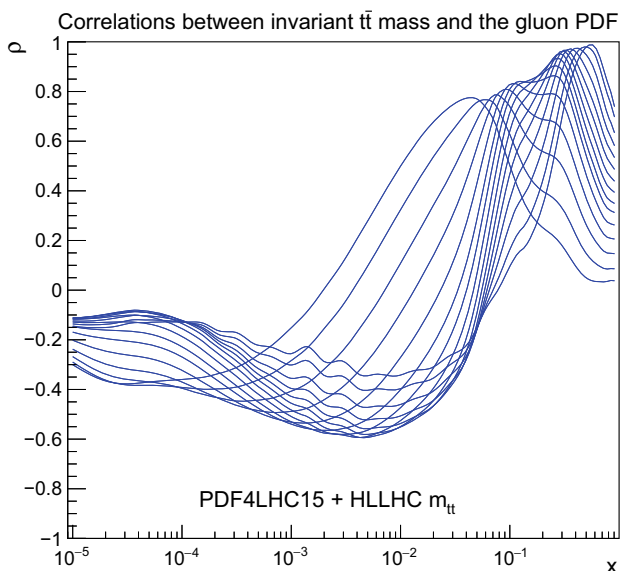

Fig. 7 As in Fig. 4, now for the correlation coefficient between the gluon PDF and the various bins of $m_{t\bar{t}}$, the invariant mass of the top quark pair

3.4 Jet and photon production

We now turn to consider two of the processes that can be used to provide information on the gluon: inclusive jet production and direct photon production. Note that these measurements also provide a handle on the valence quark distributions, due to the significant fraction of events that originate from quark–gluon scattering. In Fig. 9 we display the correlation coefficient between the gluon PDF and the central rapidity bin of the inclusive jet (left) and direct pho-

ton (right) pseudo–data. From this comparison we see that the correlation profiles are similar for the two processes. In both cases the correlation coefficient is significant around $x \simeq 10^{-2}$, is then reduced a bit, and then becomes large again, peaking at $x \simeq 0.5$. One can verify that the value of ρ decreases as we move to more forward rapidities, due to the enhanced contribution from quark–initiated diagrams.

The corresponding comparison between the theory predictions and the HL–LHC pseudo–data, before and after adding the latter in the fit, is shown in Fig. 10. Note that while we show the results only for central rapidity bins, the PDF fits include the constraints from all the available rapidity bins. It is interesting to observe the excellent coverage that the HL–LHC will offer in the TeV region for these two processes. In the case of inclusive jet (direct photon) production, we expect to measure the differential cross sections up to $p_T^{\rm jet} \simeq 2 - 3$ TeV ($E_T^\gamma \simeq 3$ TeV), a marked improvement in comparison to current coverage. Given the back–to–back topology at Born level for these two processes, we see that they can probe scales up to $Q \simeq 6$ TeV. We find that the effect of adding the HL–LHC pseudo–data is to reduce the PDF uncertainties in a range of $p_T^{\rm jet}$ and E_T^γ. In the case of direct photon production, the effects are seen across most of the E_T^γ range, while in the case of inclusive jet production they are more localised around the $p_T^{\rm jet} \sim 1$ TeV region.

Concerning the impact of these two types of HL–LHC pseudo–data on the PDFs, shown in Fig. 11, one sees that in both cases there is a visible reduction on the gluon uncertainties at both intermediate and large values of x,
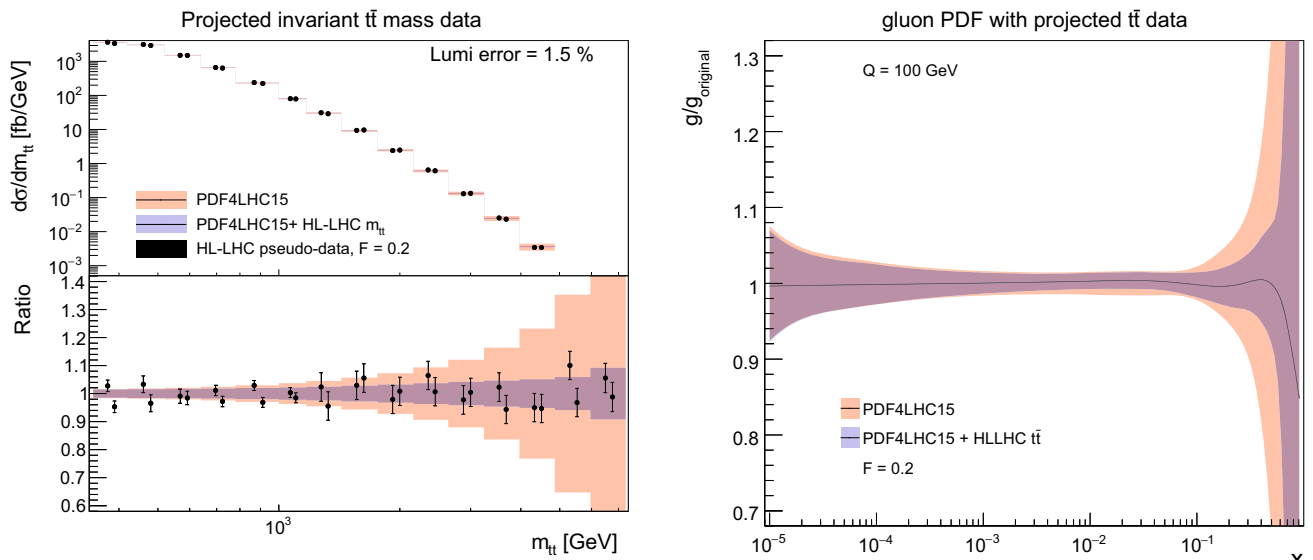
Fig. 8 Left: As in Fig. 5, now for the $m_{t\bar{t}}$ distribution in top quark pair production. Right: As in Fig. 6, now for the gluon PDF after including the HL–LHC $t\bar{t}$ pseudo–data in the fit
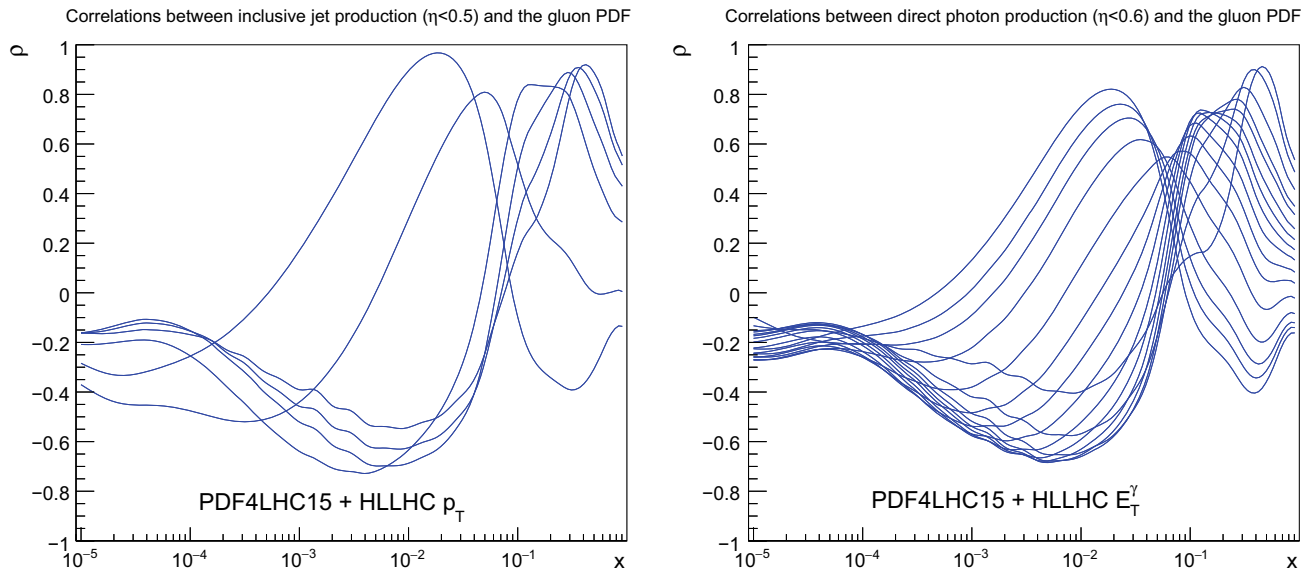

Fig. 9 As in Fig. 4, now for the correlation coefficient between the gluon PDF and the central rapidity bin of the inclusive jet (left) and direct photon (right) pseudo–data

of comparable size at high x, while at intermediate x the isolated photon data is somewhat more constraining. Note again that as expected the shift in the central values of the PDFs after profiling are much smaller than the PDF uncertainties themselves. Taking into account the results found when adding top quark pair production data into the fit, see Fig. 8, a clear picture emerges showing that the HL–LHC measurements will provide particularly stringent constraints on the large–x gluon PDF, in addition to those on the quarks.

3.5 W production in association with charm quarks

We now consider the impact of W production in association with charm quarks, which provides direct information on the strange content of the proton. As explained in Sect. 2.1 we have generated pseudo–data both in the central rapidity region, relevant for ATLAS and CMS, and in the forward rapidity region, relevant for LHCb. In Fig. 12 we show the correlation coefficient between the strange PDF and the lepton rapidity distributions in W+charm production pseudo–data both for the central and the forward rapidity regions.
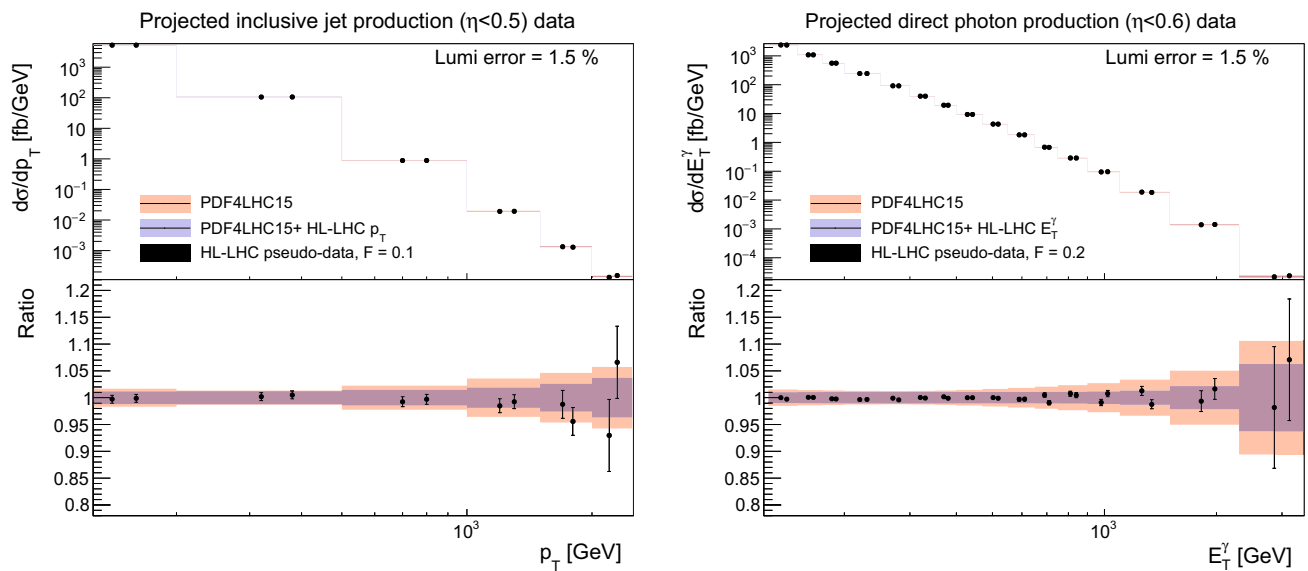
Fig. 10 As in Fig. 5, now for the central rapidity bins of the inclusive jet production (left plot) and the direct photon production (right plot) HL–LHC pseudo–data
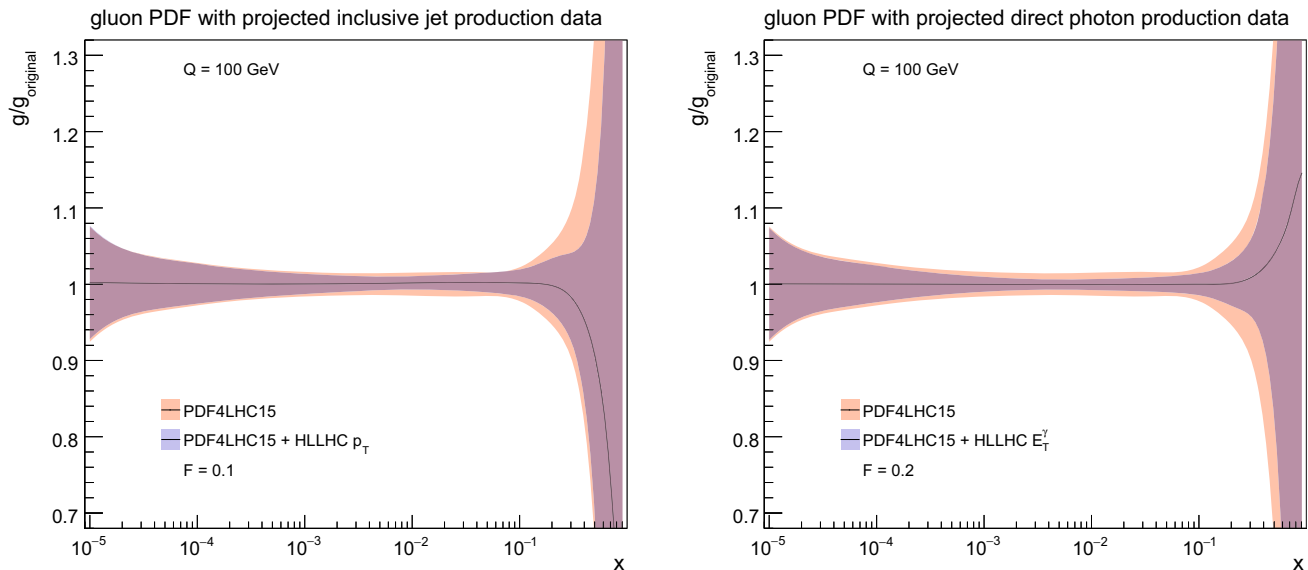

Fig. 11 As in Fig. 6, now for the gluon PDF after including the HL–LHC pseudo–data on inclusive jet production (left plot) and on direct photon production (right plot)

We can see that indeed there is a large correlation between the strange PDF and the W+charm production pseudo–data in a broad range of x values. For the case of central production, we find $\rho \geq 0.9$ in the range of $10^{-3} \leq x \leq 0.1$, while for forward production the correlation coefficient ρ peaks at a somewhat smaller value, and covers a broader range in x, with in particular a coverage of the small and large–x regions that is complementary to the central production pseudo–data.

The comparison between the HL–LHC pseudo–data and the corresponding theoretical predictions for W+charm production both in the central and forward regions are collected

in Fig. 13. In the central region, we see a clear reduction of the PDF uncertainties after including the pseudo–data into the fit, by around a factor two. This reduction of uncertainty is approximately constant as a function of the lepton rapidity. At forward rapidities instead, we find that before adding the pseudo–data the PDF uncertainties grow very fast with rapidity, reaching up to 30% for $\eta_l \simeq 4.5$, while after including it they are markedly reduced and become more or less constant with rapidity as in the central region. Taking into account the correlation coefficients shown in Fig. 12, these results indicates that W+charm production in the forward region pro-
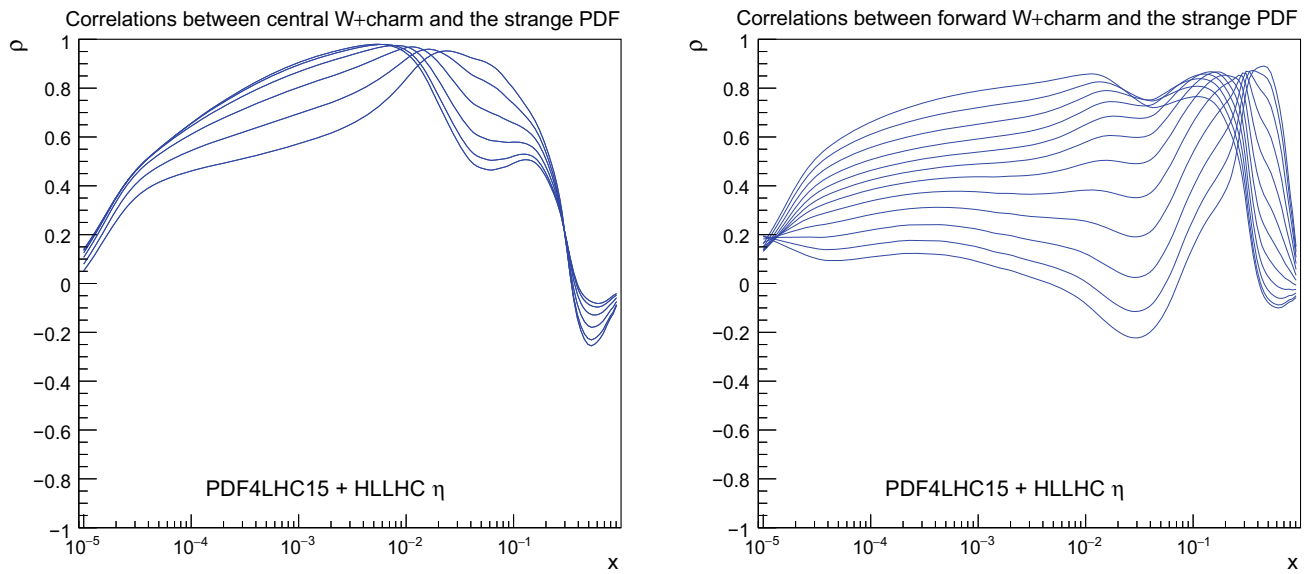
Fig. 12 As in Fig. 4, now for the correlation coefficient between the strange PDF and the lepton rapidity distributions in W+charm production pseudo–data in the central rapidity region (left) and in the forward region (right plot)
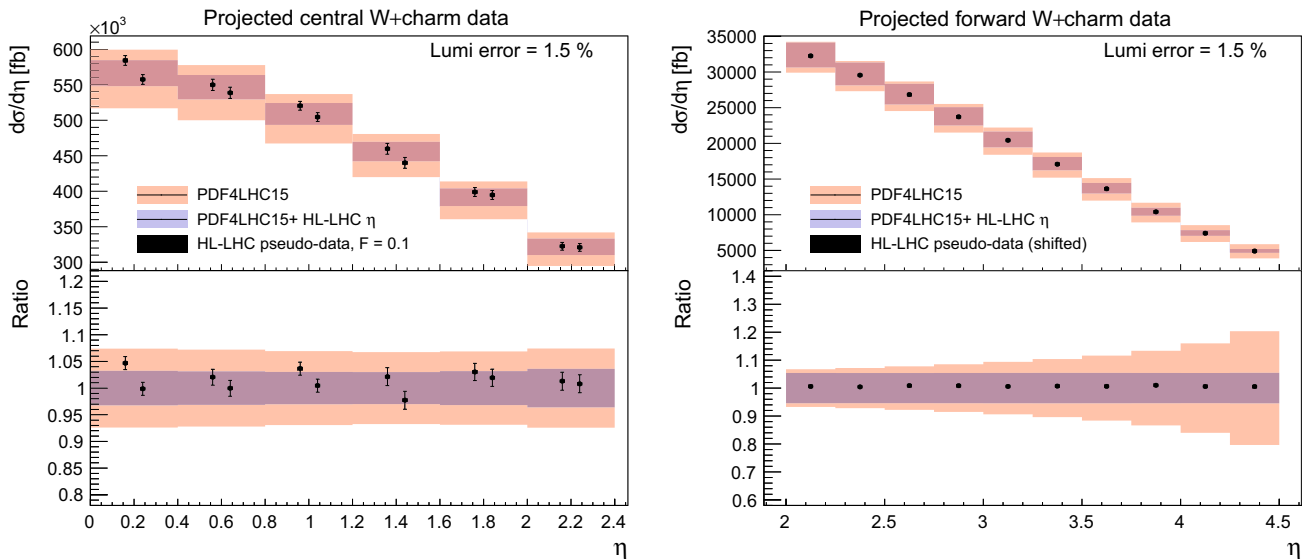

Fig. 13 As in Fig. 5, now for W+charm production in the central (left plot) and forward (right plot) rapidity regions. In the right plot only the statistical errors are shown, while the data have been shifted by the

dominant correlated source of uncertainty, namely the 5% normalization uncertainty

vides valuable constraints on the large–x strangeness, which is currently affected by large uncertainties.

This PDF uncertainty reduction on strangeness upon the addition of the W+charm pseudo–data is quantified in Fig. 14. For central production, we find that indeed most of the PDF uncertainty reduction is concentrated in the region $10^{-3} \lesssim x \lesssim 0.1$, while the large–$x$ region is affected only moderately. For the pseudo–data in the forward region instead, there is a superior reduction of the PDF uncertainties in the large–x region. The nice complementarity seen from

Fig. 14 illustrates the usefulness of combine PDF–sensitive measurements in the central rapidity region with those of the forward region.

3.6 The transverse momentum of Z bosons

We complete the study of the impact of individual HL–LHC pseudo–data on the PDFs with the transverse momentum of Z bosons, a process which is known to influence the gluon and the quark PDFs at intermediate values of x. We will
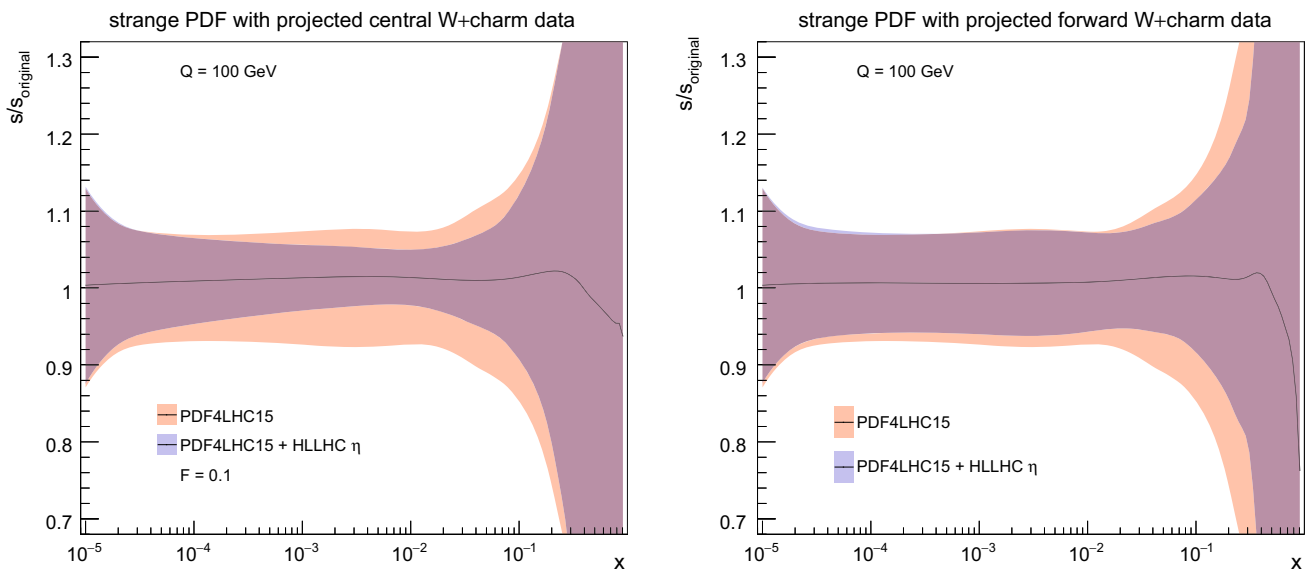
Fig. 14 As in Fig. 6, now for the strange quark PDF including the HL–LHC pseudo–data on W+charm production in the central (left plot) and in the forward region (right plot)
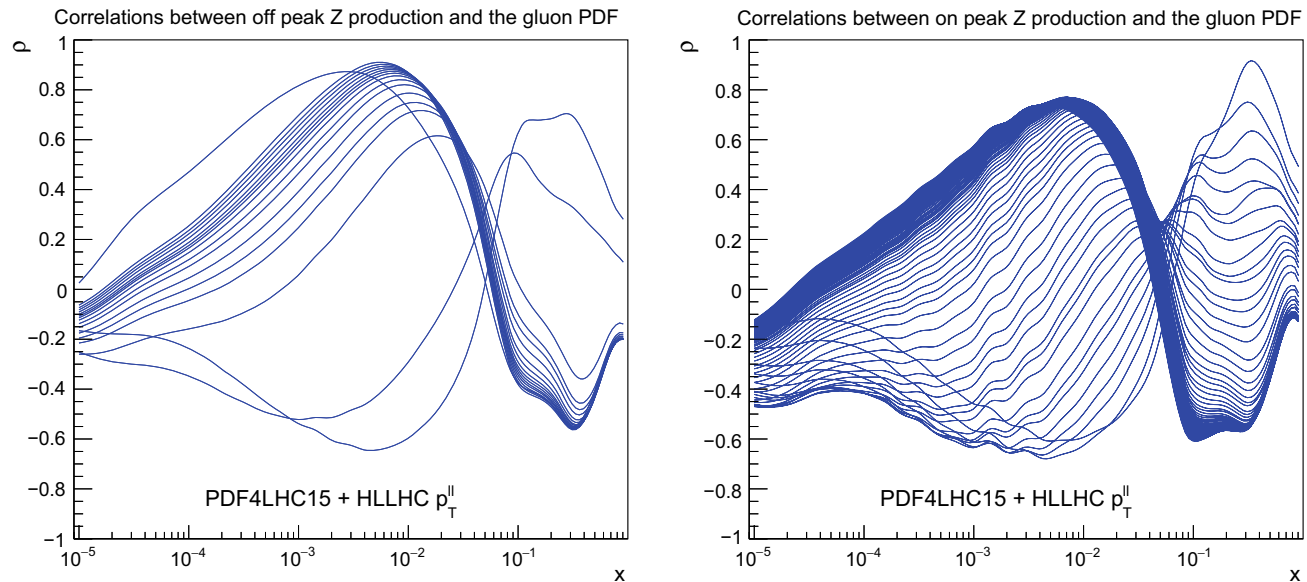

Fig. 15 As in Fig. 4, now for the correlation coefficients between the gluon PDF and the Z transverse momentum distributions in the central rapidity region, for the dilepton invariant mass $10\,\mathrm{GeV} \leq m_{ll} \leq 20\,\mathrm{GeV}$ (left plot) and $66\,\mathrm{GeV} \leq m_{ll} \leq 116\,\mathrm{GeV}$ (right plot)

show the data vs. theory comparison only for the dominant dilepton invariant mass bin, $66\,\mathrm{GeV} \leq m_{ll} \leq 116\,\mathrm{GeV}$, although in the PDF profiling the effects of all the six m_{ll} bins are being taken into account. To begin with, in Fig. 15 we show the correlation coefficients between the gluon PDF and the Z transverse momentum distributions in the central rapidity region, for the dilepton invariant mass $10\,\mathrm{GeV} \leq m_{ll} \leq 20\,\mathrm{GeV}$ and $66\,\mathrm{GeV} \leq m_{ll} \leq 116\,\mathrm{GeV}$. As we can see from this comparison, the pseudo–data on the Z p_T provides information on the gluon PDF, being mostly sensitive at $x \simeq$

10^{-2}, as well as around $x \simeq 0.3$ in the case of the on–peak dilepton invariant mass bin.

The comparison between HL–LHC pseudo–data and theoretical predictions in the on–peak bin defined by $66\,\mathrm{GeV} \leq m_{ll} \leq 116\,\mathrm{GeV}$ is shown in Fig. 16, where we can see that coverage up to $p_T^{ll} \simeq 3\,\mathrm{TeV}$ is expected, similar as in the case of direct photon production. We find a moderate reduction in the PDF uncertainties once the HL–LHC pseudo–data is added to the fit by means of Hessian profiling. Concerning its effects on the gluon, we see that the Z p_T measurements provide valuable information in the intermediate x region
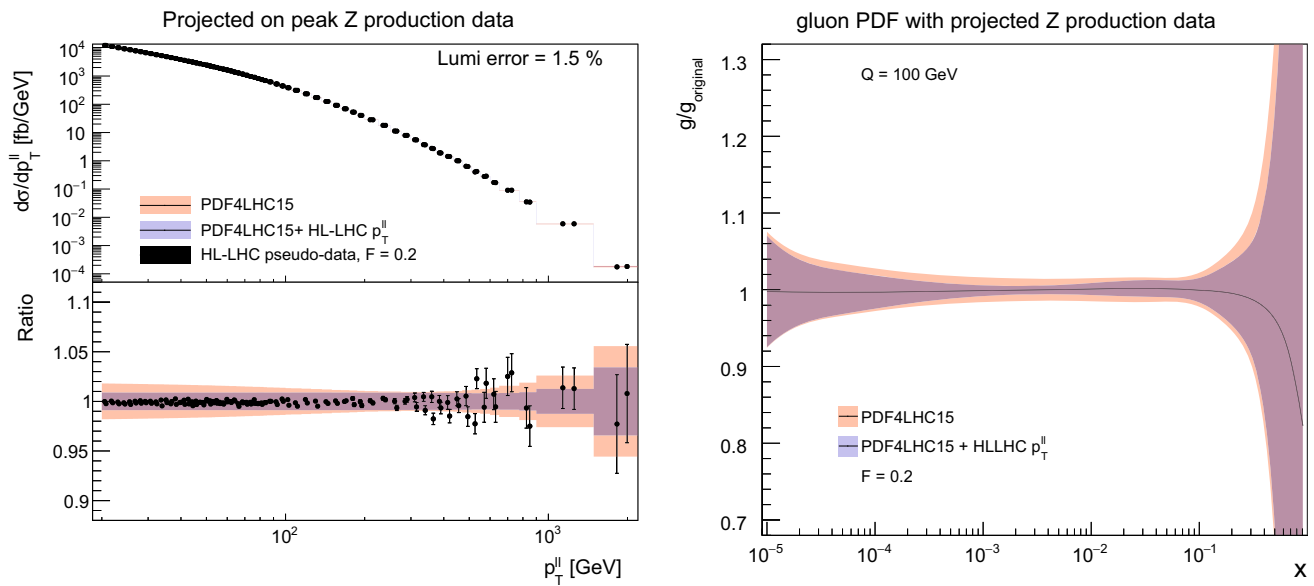
Fig. 16 Left: as in Fig. 5, now for the p_T distribution of Z bosons in the dilepton final state, in the on–peak bin defined by $66\,\text{GeV} \le m_{ll} \le 116\,\text{GeV}$. Right: as in Fig. 6, now for the gluon PDF after including the HL–LHC Z transverse momentum pseudo–data in the fit

Table 2 The three scenarios for the systematic uncertainties of the HL–LHC pseudo–data that we assume in the present study. These scenarios, ranging from conservative to optimistic, differ among them in the reduction factor f_{red}, Eq. (2.2), applied to the systematic errors of the reference 8 TeV or 13 TeV measurements. We also indicate in each case the name of the corresponding LHAPDF grid

Scenario	f_{red} (8 TeV)	f_{red} (13 TeV)	LHAPDF set	Comments
A	1	0.5	PDF4LHC_nnlo_hllhc_scen1	Conservative
B	0.7	0.36	PDF4LHC_nnlo_hllhc_scen2	Intermediate
C	0.4	0.2	PDF4LHC_nnlo_hllhc_scen3	Optimistic

between 10^{-3} and 10^{-2} with a clear reduction of PDF uncertainties even if in this region these were quite small to begin with.

4 Ultimate PDFs with HL–LHC pseudo–data

In this section we combine the complete set of HL–LHC pseudo–data listed in Table 1 to produce the final profiled PDF sets, which quantify the impact of future HL–LHC measurements on our knowledge of the quark and gluon structure of the proton.

In Table 2 we list the three scenarios for the systematic uncertainties of the HL–LHC pseudo–data that we assume in the present analysis. These scenarios, ranging from more conservative to more optimistic, differ among them in the reduction factor f_{red}, Eq. (2.2), applied to the systematic errors of the reference 8 TeV or 13 TeV measurements. In particular, in the optimistic scenario we assume a reduction of the systematic errors by a factor 2.5 (5) as compared to the reference 8 TeV (13 TeV) measurements, while for the conservative scenario we assume no reduction in systematic errors with

respect to 8 TeV reference. We also indicate in each case the name of the corresponding LHAPDF grid. Reassuringly, as we show below, the qualitative results of our study depend only mildly in the specific assumption for the values of f_{red}.

In this section, we study how the HL–LHC pseudo–data constraints the parton distributions and the PDF luminosities for proton–proton collisions at $\sqrt{s} = 14$ TeV. Then we present an initial study with some representative implications of the ultimate PDFs for LHC phenomenology.

4.1 Parton distributions

In Fig. 17 we present a comparison of the baseline PDF4LHC15 set with the profiled sets based on HL–LHC pseudo–data from scenarios A (conservative) and C (optimistic) as defined in Table 2. Specifically, we show the gluon, down quark, up anti–quark, and total strangeness at $Q = 10$ GeV, normalized to the central value of the PDF4LHC15 baseline. In this comparison, the bands correspond to the one–sigma PDF uncertainties.

First of all, we observe that the impact of the HL–LHC pseudo–data is reasonably similar in the conserva-
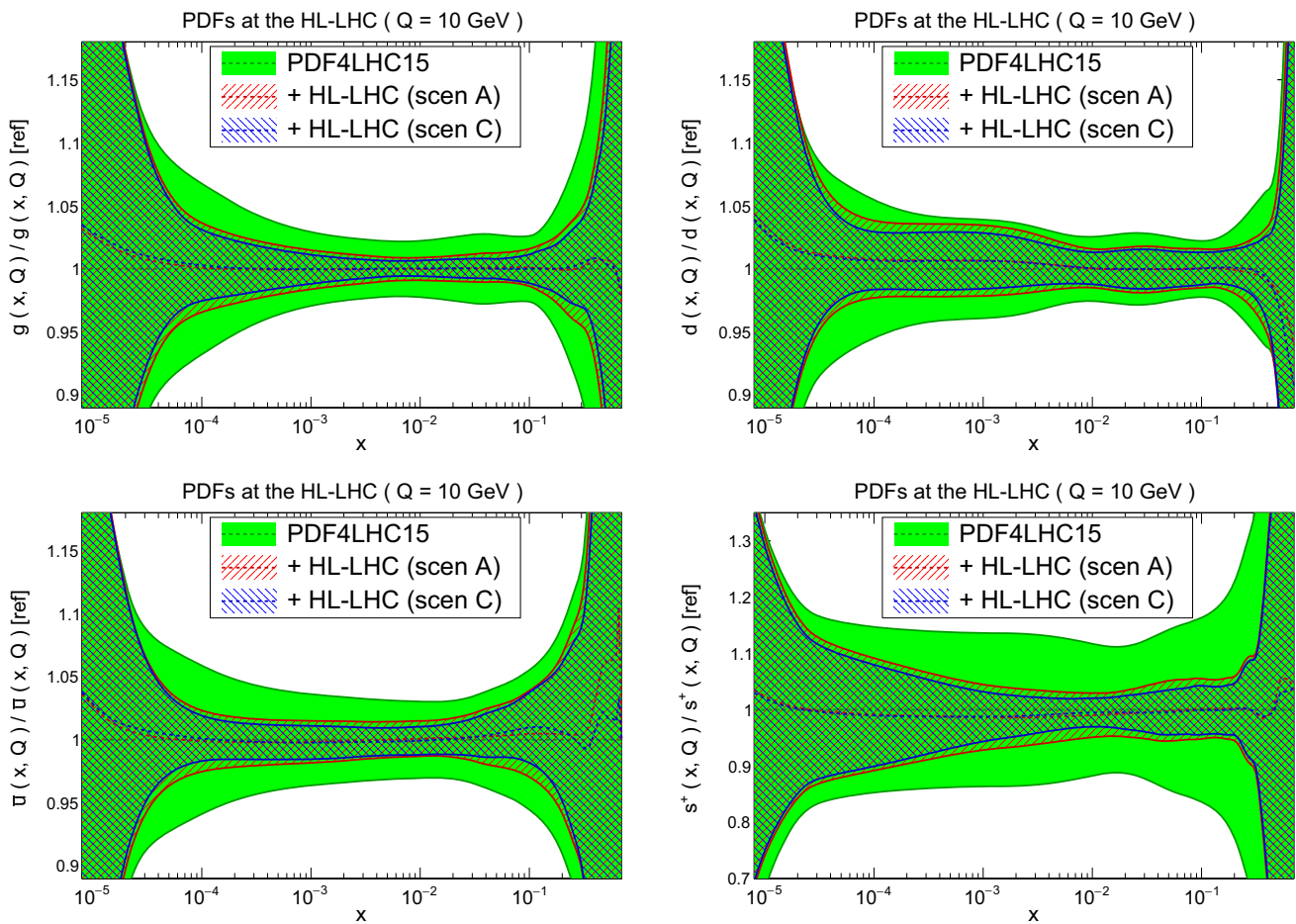
Fig. 17 Comparison of the PDF4LHC15 set with the HL–LHC profiled set in scenarios A and C, defined in Table 2. We show the gluon, down quark, up anti–quark, and total strangeness at $Q = 10$ GeV, normalized to the central value of the PDF4LHC15 baseline. The bands correspond to the one–sigma PDF uncertainties

tive and optimistic scenarios. This is not so surprising, as we have explicitly chosen those datasets which will benefit from a significant improvement in statistics, and these tend to lie in kinematic regions where the PDFs themselves are generally less well determined, see the discussion in Sect. 2. Therefore, the dominant reason for the observed reduction of PDF uncertainties is the increased statistics and the corresponding extended kinematic reach that becomes available at the HL–LHC, rather than the specific assumptions about the systematic uncertainties. This demonstrates that our results are robust against the details of the projections of how the experimental systematic uncertainties will be reduced in the HL–LHC era.

From Fig. 17 we observe a marked reduction of the PDF uncertainties in all cases. This is particularly significant for the gluon and the sea quarks, for the reason that these are currently affected by larger uncertainties than in the case of the valence quarks. In the case

of the gluon PDF, there is an improvement of uncertainties across a very broad range of x. This is a direct consequence of the fact that we have included several HL–LHC processes that have direct sensitivity to the gluon content of the proton, namely jet, direct photon, and top quark pair production, as well as the transverse momentum of Z bosons.

Another striking feature of Fig. 17 concerns the strange PDF. In this case, the PDF uncertainties are reduced by almost a factor 4, from around 15% to a few percent, in a wide region of x. This result highlights the importance of the W+charm measurements at the HL–LHC, specially those in the forward region by LHCb, see Fig. 12, which represent a unique handle on the poorly known strange content of the proton. In turn, such an improved understanding of the strange PDF will feed into a reduction of theory uncertainties in crucial HL–LHC measurements such as those of M_W or $\sin^2 \theta_W$.

4.2 Partonic luminosities

Next we take a look at the partonic luminosities, to quantify the improvement in the PDF uncertainties in different initial–state partonic combinations from the HL–LHC pseudo–data. In Fig. 18 we show the reduction of PDF uncertainties in the gg, qg, $q\bar{q}$, and qq, $s\bar{s}$, and $s\bar{u}$ luminosities at $\sqrt{s} = 14$ TeV that can be expected as a consequence of adding the HL–LHC pseudo–data on top of the PDF4LHC15 baseline. Note that a value of 1 in these plots corresponds to no uncertainty reduction. As in the case of the PDF comparisons, results are shown both for the conservative (A) and optimistic (C) scenarios for our projections of the experimental systematic uncertainties.

In addition, in Table 3 we also report the average values of these PDF uncertainty reductions for three different invariant mass bins. In particular, we consider low invariant masses, 10 GeV $\leq M_X \leq$ 40 GeV, relevant for instance for Monte Carlo tuning and QCD studies; intermediate masses, 40 GeV $\leq M_X \leq$ 1 TeV, relevant for electroweak, top, and Higgs measurements; and large invariance masses, 1 TeV $\leq M_X \leq$ 6 TeV, relevant for searches of new heavy particles. These averages are computed from 10 points per mass bin, logarithmically spaced in M_X. In Table 3, the values shown outside (inside) the brackets correspond to the optimistic (conservative) scenario.

From the comparisons in Fig. 18 and Table 3, we observe again that the reduction in the uncertainties of the PDF luminosities is rather robust with respect to the assumed projections for the experimental systematic uncertainties. For instance, for intermediate values of the final–state invariant mass, 40 GeV $\leq M_X \leq$ 1 TeV, we find that for all the partonic initial states the reduction factor varies between 0.28 and 0.40 (0.42 and 0.45, 0.31 and 0.40) in the optimistic and conservative scenario for the gluon–gluon (gluon–quark, quark–antiquark) luminosities. These results again reinforce our conclusion that the results of this study are only mildly sensitive to the details of the projected pseudo–data.

We find that in the intermediate M_X bin the reduction of PDF uncertainties ranges approximately between a factor 2 and a factor 5, depending on the specific partonic channel and the scenario for the systematic errors. For example, for the gluon–gluon luminosity in the range relevant for Higgs production in gluon fusion, one finds a reduction by almost a factor 4 in the optimistic scenario. The improvement in the strange–initiated processes is also remarkable, for example the PDF uncertainties in the $s\bar{s}$ luminosity are expected to be reduced by a factor 5 (3) in the optimistic (conservative) scenario. Recall that strange–initiated processes are important for a variety of LHC analysis, from measurements of M_W and $\sin^2 \theta_W$ to searches for BSM W' bosons. We also find that the uncertainties in quark–antiquark luminosities, relevant for example for precision electroweak measurements,

are expected to be reduced by up to a factor 3 in this invariant mass range.

Similar improvements in the PDF luminosities are found in the high mass region, $M_X \geq 1$ TeV, directly relevant for BSM searches. For instance, in the optimistic scenario, the PDF error reduction at higher masses is expected to be as large as a factor 5 for the gluon–gluon luminosity. Again this is a consequence of the inclusion in the profiling of gluon–dominated processes such as $t\bar{t}$ and inclusive jets that at the HL–LHC, which cover the region up to 6 TeV, see Fig. 2. The impact of the HL–LHC pseudo–data is less marked for the quark–quark and quark–antiquark luminosities in this high–mass region, due to the fact that of the data points included in the profiling only a fraction of them are both quark–initiated and cover the large–x region.

It is worth emphasizing again here that the list of processes studied in this work and summarised in Table 3 are just a subset of those HL–LHC measurements with PDF–constraining potential. Therefore, it is conceivable that the actual reduction of PDF errors presented in Table 3 would actually be more significant than our estimates here.

4.3 Implications for HL–LHC phenomenology

We now turn to present some representative results of the phenomenological implications that these "ultimate" PDFs will have at the HL–LHC, both for processes within the SM and beyond it. It is beyond the scope of this work to carry out a comprehensive phenomenological study, and we refer the reader to the upcoming Yellow Report [27] describing the physics case of the HL–LHC, where more detailed projections and analyses will be presented.

Let us begin by assessing the PDF impact of HL–LHC measurements on representative Standard Model processes. In particular, we consider diphoton production, dijet production, and Higgs production in gluon fusion, both inclusive and in association with a hard jet. In the following all cross sections have been computed at $\sqrt{s} = 14$ TeV using leading order (LO) matrix elements with MCFMv8.2 [58] and applying the standard ATLAS/CMS central acceptance cuts. Since the comparison is restricted to ratios of cross sections, the LO calculation is sufficient to illustrate the impact of the improvement in the PDF uncertainties in each of these processes. Indeed, we are only interested here in illustrating the relative impact of the PDF error reduction, rather than providing state–of–the–art predictions for the rates, which will be presented elsewhere in the Yellow Report [27].

First of all, we show the production cross sections of pairs of photons (left) and of jets (right) in the upper panels of Fig. 19. We compare the PDF4LHC15 baseline with the HL–LHC profiled PDF sets in the conservative (A) and optimistic (C) scenarios of Table 2, normalised to the central value of PDF4LHC15. In the considered kinematic regions, these two
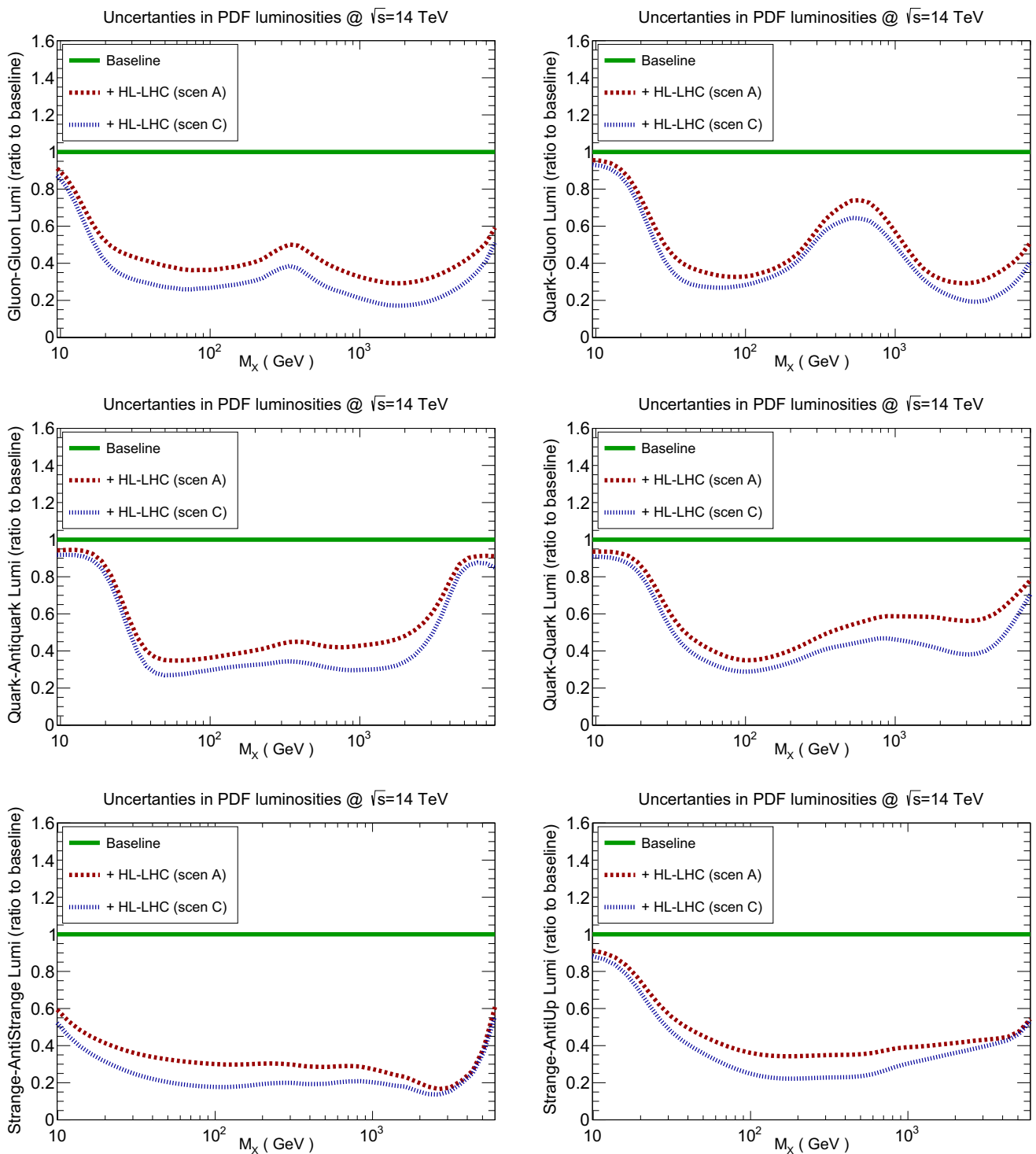
Fig. 18 The reduction of the uncertainties in the PDF luminosities at $\sqrt{s} = 14$ TeV once the HL–LHC pseudo–data has been included, compared to the PDF4LHC15 baseline. We show the gg, qg, $q\bar{q}$, qq, $s\bar{s}$, and $s\bar{u}$ luminosities for the conservative (A) and optimistic (C) scenarios. The average values for the PDF uncertainty reduction in different bins of M_X is also reported in this figure

Table 3 The reduction of the PDF uncertainties compared to the PDF4LHC15 baseline for different initial partonic combinations (that is, a value of 1 corresponds to no reduction at all). Results are presented for three different bins of the invariant mass M_X of the produced system, averaging over 10 points logarithmically spaced within each bin. The values shown outside (inside) the brackets correspond to the optimistic (conservative) scenario. The corresponding results differential in M_X are presented in Fig. 18

Ratio to baseline	10 GeV $\leq M_X \leq$ 40 GeV	40 GeV $\leq M_X \leq$ 1 TeV	1 TeV $\leq M_X \leq$ 6 TeV
gluon–gluon	0.50 (0.60)	0.28 (0.40)	0.22 (0.34)
gluon–quark	0.66 (0.72)	0.42 (0.45)	0.28 (0.37)
quark–quark	0.74 (0.79)	0.37 (0.46)	0.43 (0.59)
quark–antiquark	0.71 (0.76)	0.31 (0.40)	0.50 (0.60)
strange–antistrange	0.34 (0.44)	0.19 (0.30)	0.23 (0.27)
strange–antiup	0.67 (0.73)	0.27 (0.38)	0.38 (0.43)

processes are mostly sensitive to the quark–antiquark initial state, and to the quark–gluon and quark–(anti)quark initial states, respectively. The cross sections are presented as a function of the minimum invariant mass of the final state, $M_{\gamma\gamma}^{\min}$ and $M_{jj}^{\min}$ respectively, in order to facilitate their comparison with the corresponding PDF luminosities shown in Fig. 18.

From this comparison, we find again that both the optimistic and conservative scenarios lead to similar results in terms of the expected reduction of the PDF uncertainties in the entire kinematical range accessible at the HL–LHC. In the case of dijet production, we find that PDF uncertainties could be reduced down to $\simeq 2\%$ even for invariant masses as large as $M_{jj} = 6$ TeV. The resulting improved theory predictions for dijet production should also become relevant at the HL–LHC for BSM searches, *e.g.* for quark compositeness [73,74]. Note that since the initial partonic states are the same, similar improvements are expected for inclusive jet production, see also Fig. 10, as well as for multijet production. Similar considerations apply for diphoton production, where the expected PDF error reduction is a bit less marked since it is driven by the quark–antiquark luminosity, see also the comparisons in Table 3.

In the two lower plots of Fig. 19, we present the corresponding comparisons for the case of Higgs boson production via gluon fusion, using heavy top quark effective theory. In the case of inclusive production with decay into bottom quarks (left plot), we find that the constraints from HL–LHC measurements are expected to reduce PDF uncertainties down to the 1% level. Needless to say, this will directly benefit the characterisation of the Higgs sector at the HL–LHC, where a few percent is the typical uncertainty target for the determination of its couplings. In the case of Higgs boson production in association with a hard jet (right plot), also there we find a marked error reduction, indicating that PDF uncertainties in the Higgs transverse momentum distribution could be reduced down to the $\simeq 2\%$ level in the entire kinematical range relevant at the HL–LHC. We recall that the large Higgs transverse momentum region is sensitive to

new heavy particles running in the loops as well as to BSM effects such as partial Higgs compositeness [75].

As we have discussed above in Sects. 4.1 and 4.2, the impact of the HL–LHC pseudo–data is also significant in the large–x region, which in turn corresponds to large invariant masses for the PDF luminosities. This is of course an important region for the searches of BSM heavy particles, where PDF uncertainties often represent the dominant source of theoretical uncertainty. With this motivation, to illustrate the benefits that HL–LHC measurements will provide for BSM searches we consider here high–mass supersymmetric (SUSY) particle production at $\sqrt{s} = 14$ TeV, where the HL–LHC reach extends to sparticles masses up to around $M \simeq 3$ TeV. While we use SUSY production as a benchmark process, our results also apply to the production of other heavy particles predicted in different BSM scenarios.

In Fig. 20 we show the comparison between the PDF4LHC15 predictions with the corresponding results from the profiled PDF sets with HL–LHC pseudo–data, normalised to the central value of the PDF4LHC15 baseline. As in Fig. 19, we provide results for scenarios A and C, the conservative and optimistic ones respectively. Specifically, we show the cross sections for gluino–gluino and squark–gluino production at $\sqrt{s} = 14$ TeV – similar conclusions are derived from squark–squark and squark–antisquark production. The theoretical calculations have been obtained using leading order (LO) matrix elements with `Pythia8.235` [76] and assuming the SLHA2 benchmark point [77], for a range of sparticle masses within the HL–LHC reach. For simplicity, underlying event and multiple interactions have been switched off in the calculation. Again, we are not interested here in providing state–of–the–art predictions for the event rates, which can be found elsewhere [78].

From the comparisons in Fig. 20, we can see that the constraints on the PDFs expected from the HL–LHC data permit a significant reduction of the uncertainties in the high–mass SUSY cross sections. The size of this reduction is consistent with the corresponding results at the level of luminosities, reported in Fig. 18 and Table 3, recalling that gluino–gluino
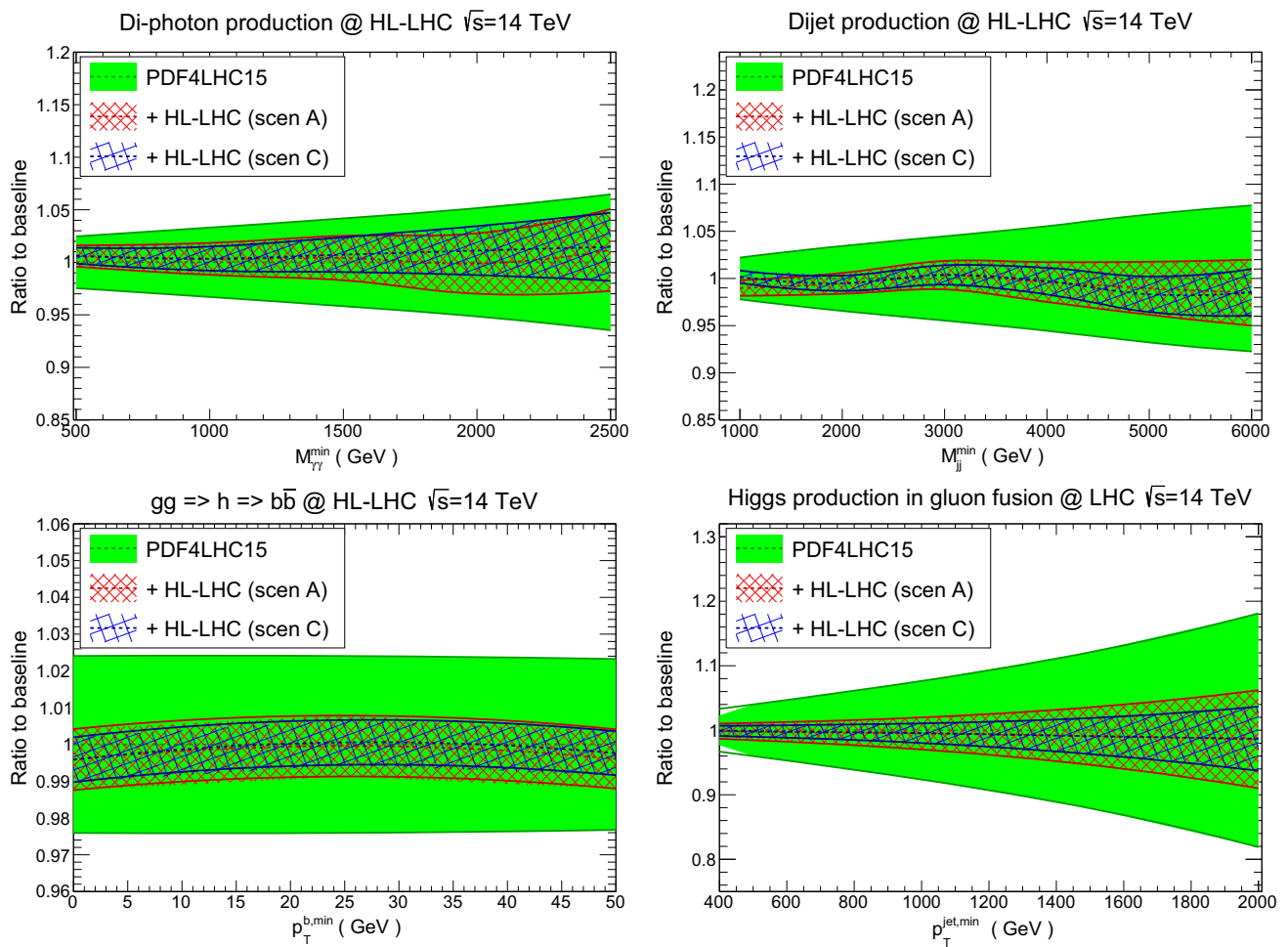
Fig. 19 Comparison of the predictions for representative SM cross sections at $\sqrt{s} = 14$ TeV between the PDF4LHC15 baseline and the HL–LHC profiled sets in the conservative (A) and optimistic (C) scenarios. Results are shown normalised to the central value of PDF4LHC15. The upper plots show the diphoton (left) and dijet (right plot) production cross sections as a function of the minimum invariant masses of the final state, $M_{\gamma\gamma}^{\min}$ and $M_{jj}^{\min}$ respectively. The bottom plots show Higgs boson production in gluon fusion with heavy top quark effective theory, both inclusive and decaying into $b\bar{b}$ as a function of $p_T^{b,\min}$ (left), and then in association with a hard jet as a function its transverse momentum $p_T^{jet,\min}$ (right plot). The calculations have been performed using MCFM8.2 with leading–order matrix elements

and gluino–squark production are driven by the gluon–gluon and gluon–quark initial states respectively [5]. For instance, for gluino pair–production with $M_{\tilde{g}} = 3$ TeV, the PDF uncertainties are reduced from $\simeq 60\%$ to $\simeq 20\%$ in the optimistic scenario. A somewhat milder reduction is found for the squark–gluino cross sections. For squark–squark and squark–antisquark production, driven by the quark–quark and quark–antiquark initial states respectively, a PDF uncertainty reduction by around a factor two at high masses is found, consistently with Table 3.

To summarise, the initial phenomenological study presented in this section nicely illustrates the internal coherence of the HL–LHC physics program: high precision SM measurements will lead to a much improved understanding of the quark and gluon structure of protons, which in turn will ben-

efit many other important analyses, from the characterisation of the Higgs sector to the searches of new heavy particles.

5 Summary

In this study, we have quantified the expected constraints that precision HL–LHC measurements will impose on the quark and gluon structure of the proton. To achieve this goal, we have assessed the impact of a range of relevant PDF–sensitive processes, from weak gauge boson and jet production to top quark and photon production. Moreover, we have studied the robustness of our results with respect to different projections for the experimental systematic uncertainties, from a more conservative one, where systematics are assumed to have the
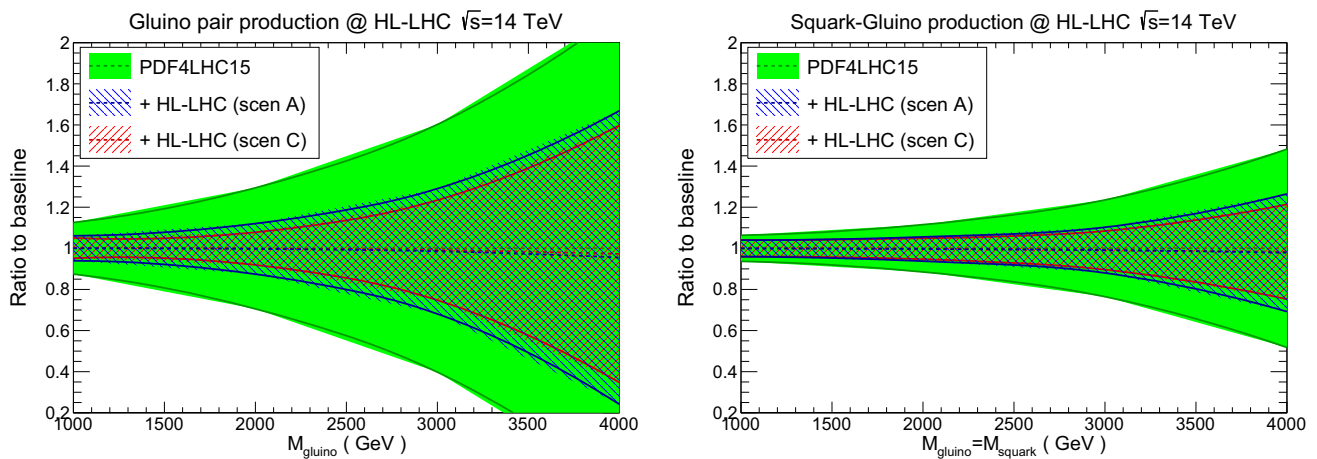
Fig. 20 The cross sections for high–mass supersymmetric particle production at $\sqrt{s} = 14$ TeV, comparing the predictions of the PDF4LHC15 baseline with those of the HL–LHC PDF sets in the conservative (A) and optimistic (C) scenarios, normalised to the central value of PDF4LHC15. We show the results corresponding to gluino pair production (left) and squark–gluino production (right). The cross sections have been evaluated with `Pythia8.235` using leading–order matrix elements and the SLHA2 benchmark point as model input

same size as in current measurements, to a more optimistic one, where they are markedly reduced.

Our main finding is that HL–LHC data has the potential to significantly reduce the PDF uncertainties in a wide kinematic range and for all relevant partonic final states. This is true both for the region of intermediate invariant masses, relevant for precision Higgs, electroweak, and top quark measurements, as well as in the TeV region relevant for searches of new heavy particles. Even in the most conservative scenario, in the region $M_X \gtrsim 40$ GeV we find that HL–LHC measurements can reduce PDF uncertainties by at least a factor between 2 and 3 as compared to the current PDF4LHC15 baseline. The PDF constraining information from the HL–LHC is expected to be specially significant for gluon– and for strange–initiated processes. We also find that the quark–antiquark luminosity at the electroweak scale, a central input for legacy LHC measurements such as M_W and $\sin^2 \theta_W$, could be improved by more than a factor 3 in the optimistic scenario.

This improved knowledge of the quark and gluon structure of the proton which will become possible at the HL–LHC will directly benefit a number of phenomenologically important process, due to the reduction of the associated theoretical errors. For instance, the PDF uncertainties in Higgs production in gluon fusion can be reduced down to $\lesssim 2\%$ for the entire range of Higgs transverse momenta accessible at the HL–LHC. Likewise, PDF uncertainties in high–mass supersymmetric particle production can be decreased by up to a factor 3, with a similar impact expected for other BSM scenarios. This improvement should strengthen the bounds derived in the case of null searches, or facilitate their characterisation in the case of an eventual discovery. Similar improvements are found for Standard Model process, for

example dijet production, which provides a unique opportunity to measurement the running of the strong coupling constant at the TeV scale. More detailed studies of the phenomenological implications of our study will be presented in the upcoming HL–LHC Yellow Report. Two caveats are relevant at this point. First, it should be emphasised again that in this study we have only considered a subset of all possible measurements of relevance for PDF fits. There are certainly processes for which data is and will be available, such as multijet production and single top production, that we have not considered here. Moreover, we can also reasonably expect that various new processes may be added to the PDF toolbox on the rather long timescales we consider here. Thus, we may certainly expect further constraints to become available for PDF studies by the end of HL–LHC running.

Second, in this study we have ignored any possible issues such as data incompatibilities, limitations of the theoretical calculations, or issues affecting the data correlation models. These are common in PDF fits, and indeed have already been found when comparing theory calculations against existing LHC data from Runs I and II. Such potential problems may eventually limit the PDF constraining power, in comparison to the estimates presented in this work, when the actual global fit with real HL–LHC data is performed. Clearly, such questions can only be tackled once the HL–LHC measurements are carried out, and indeed doing so will present an important programme of experimental and theoretical PDF–related work on its own. We cannot anticipate such work in our present study, which instead represents our best quantitative projections using our current knowledge. The results of this study are made publicly available in the `LHAPDF6` format [46], with the grid names listed in Table 2 for the three

scenarios that have been considered. These three grid files can be downloaded from:

https://data.nnpdf.science/HLLHC_YR/PDF4LHC15_nnlo_hllhc_scen1.tgz

https://data.nnpdf.science/HLLHC_YR/PDF4LHC15_nnlo_hllhc_scen2.tgz

https://data.nnpdf.science/HLLHC_YR/PDF4LHC15_nnlo_hllhc_scen3.tgz

The "ultimate" PDFs produced in this exercise can then be straightforwardly applied to other physics projections of HL–LHC processes, taking into account our improved knowledge of the partonic structure of the proton which is expected by then. We believe that the results of this work represent an important ingredient towards sharpening as much as possible the physics reach of the LHC in its upcoming high–luminosity era.

References

1. J. Gao, L. Harland-Lang, J. Rojo, The structure of the proton in the LHC precision era. Phys. Rept. **742**, 1–121 (2018). arXiv:1709.04922
2. 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**, 103103 (2015). arXiv:1507.00556
3. S. Forte, G. Watt, Progress in the determination of the partonic structure of the proton. Ann. Rev. Nucl. Part. Sci. **63**, 291 (2013). arXiv:1301.6754
4. 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
5. W. Beenakker, C. Borschensky, M. Kramer, A. Kulesza, E. Laenen, S. Marzani, J. Rojo, NLO+NLL, squark and gluino production cross-sections with threshold-improved parton distributions. Eur. Phys. J. C **76**(2), 53 (2016). arXiv:1510.00375
6. S. Alioli, M. Farina, D. Pappadopulo, J.T. Ruderman, Precision probes of QCD at high energies. JHEP **07**, 097 (2017). arXiv:1706.03068
7. ATLAS Collaboration, M., Aaboud et al., Measurement of the W-boson mass in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. Eur. Phys. J. C **78**(2), 110 (2018). arXiv:1701.07240
8. CDF, D0 Collaboration, T. A. Aaltonen et al., Tevatron Run II combination of the effective leptonic electroweak mixing angle. Phys. Rev. D **97**(11), 112007 (2018). arXiv:1801.06283
9. NNPDF Collaboration, R.D. Ball, S. Carrazza, L. Del Debbio, S. Forte, Z. Kassabov, J. Rojo, E. Slade, M. Ubiali, Precision determination of the strong coupling constant within a global PDF analysis. Eur. Phys. J. C **78**(5), 408 (2018). arXiv:1802.03398
10. D. Becciolini, M. Gillioz, M. Nardecchia, F. Sannino, M. Spannowsky, Constraining new colored matter from the ratio of 3 to 2 jets cross sections at the LHC. Phys. Rev. D **91** (1), 015010 (2015). arXiv:1403.7411 [Addendum: Phys. Rev.D92,no.7,079905(2015)
11. NNPDF Collaboration, R.D. Ball et al., Parton distributions from high-precision collider data. Eur. Phys. J. C **77**(10), 663 (2017). arXiv:1706.00428
12. S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, C .P. Yuan, New parton distribution functions from a global analysis of quantum chromodynamics. Phys. Rev. D **93**(3), 033006 (2016). arXiv:1506.07443
13. L.A. Harland-Lang, A.D. Martin, P. Motylinski, R.S. Thorne, Parton distributions in the LHC era: MMHT 2014 PDFs. Eur. Phys. J. C **75**, 204 (2015). arXiv:1412.3989
14. S. Alekhin, J. Blümlein, S. Moch, R. Placakyte, Parton distribution functions, $alpha_s$, and heavy-quark masses for LHC Run II., Phys. Rev. D **96**(1), 014011 (2017). arXiv:1701.05838
15. M. Czakon, N.P. Hartland, A. Mitov, E.R. Nocera, J. Rojo, Pinning down the large-x gluon with NNLO top-quark pair differential distributions. JHEP **04**, 044 (2017). arXiv:1611.08609
16. M. Guzzi, K. Lipka, S.-O. Moch, Top-quark pair production at hadron colliders: differential cross section and phenomenological applications with DiffTop. JHEP **01**, 082 (2015). [arXiv:1406.0386]
17. R. Boughezal, A. Guffanti, F. Petriello, M. Ubiali, The impact of the LHC Z-boson transverse momentum data on PDF determinations. JHEP **07**, 130 (2017). arXiv:1705.00343
18. D. d'Enterria, J. Rojo, Quantitative constraints on the gluon distribution function in the proton from collider isolated-photon data. Nucl. Phys. B **860**, 311–338 (2012). arXiv:1202.1762
19. J.M. Campbell, J. Rojo, E. Slade, C. Williams, Direct photon production and PDF fits reloaded. Eur. Phys. J. C **78**(6), 470 (2018). arXiv:1802.03021
20. 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**(8), 396 (2015). arXiv:1503.04581
21. R. Gauld, J. Rojo, Precision determination of the small-x gluon from charm production at LHCb, Phys. Rev. Lett. **118**(7), 072001 (2017). arXiv:1610.09373
22. ATLAS Collaboration, G. Aad, et al., Measurement of the production of a W boson in association with a charm quark in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, JHEP **1405**, 068 (2014). arXiv:1402.6263
23. CMS Collaboration, S. Chatrchyan et al., Measurement of the muon charge asymmetry in inclusive pp to WX production at $\sqrt{s} = 7$ TeV and an improved determination of light parton distribution functions. Phys. Rev. D **90**, 032004 (2014). arXiv:1312.6283
24. CMS Collaboration, Measurement of associated production of W bosons with charm quarks in proton-proton collisions at $\sqrt{s} = 13$ TeV with the CMS experiment at the LHC, Tech. Rep. CMS-PAS-SMP-17-014, CERN, Geneva (2018)
25. J. Currie, E.W.N. Glover, J. Pires, Next-to-next-to leading order QCD predictions for single jet inclusive production at the LHC. Phys. Rev. Lett. **118**(7), 072002 (2017). arXiv:1611.01460

26. J. Rojo, Constraints on parton distributions and the strong coupling from LHC jet data. Int. J. Mod. Phys. A **30**, 1546005 (2015). arXiv:1410.7728
27. P. Azzi et al., The Physics at the HL/HE-LHC Yellow Report
28. LHCb Collaboration, I. Bediaga et al., Physics case for an LHCb Upgrade II—opportunities in flavour physics, and beyond, in the HL-LHC era. arXiv:1808.08865
29. LHeC Study Group Collaboration, J. Abelleira Fernandez et al., A large hadron electron collider at cern: report on the physics and design concepts for machine and detector. J. Phys. G **39** (2012) 075001. arXiv:1206.2913
30. LHeC Study Group Collaboration, J. L. Abelleira Fernandez et al., On the Relation of the LHeC and the LHC. arXiv:1211.5102
31. LHeC study Group Collaboration, H. Paukkunen, An update on nuclear PDFs at the LHeC. PoS DIS **2017**, 109 (2018). arXiv:1709.08342
32. NNPDF Collaboration, R.D. Ball, S. Forte, A. Guffanti, E.R. Nocera, G. Ridolfi, J. Rojo, Polarized Parton Distributions at an Electron-Ion. Collider. Phys. Lett. B **728**, 524–531 (2014). arXiv:1310.0461
33. C. Marquet, M.R. Moldes, P. Zurita, Unveiling saturation effects from nuclear structure function measurements at the EIC. Phys. Lett. B **772**, 607–614 (2017). arXiv:1702.00839
34. E.C. Aschenauer, S. Fazio, M.A.C. Lamont, H. Paukkunen, P. Zurita, Nuclear structure functions at a future electron-ion collider. Phys. Rev. D **96**(11), 114005 (2017). arXiv:1708.05654
35. E.C. Aschenauer, R. Sassot, M. Stratmann, Unveiling the proton spin decomposition at a future electron-ion collider. Phys. Rev. D **92**(9), 094030 (2015). arXiv:1509.06489
36. D. Boer, M. Diehl, R. Milner, R. Venugopalan, W. Vogelsang, et al., Gluons and the quark sea at high energies: distributions, polarization, tomography. arXiv:1108.1713
37. LHeC study Group Collaboration, A. M. Cooper-Sarkar, Improved measurement of parton distribution functions and $alpha_s(M_Z)$ with the LHeC. PoS DIS2016 **274** (2016). arXiv:1605.08579
38. ATLAS Collaboration, M. Aaboud et al., Precision measurement and interpretation of inclusive W^+, W^- and $Z/gamma^*$ production cross sections with the ATLAS detector. Eur. Phys. J. C **77**(6), 367 (2017). arXiv:1612.03016
39. CMS Collaboration, V. Khachatryan et al., Measurement of the differential cross section and charge asymmetry for inclusive pp → W$^\pm$ + X production at $\sqrt{s}$ = 8 TeV. Eur. Phys. J. C **76**(8), 469 (2016). arXiv:1603.01803
40. J. Butterworth et al., PDF4LHC recommendations for LHC Run II. J. Phys. G **43**, 023001 (2016). arXiv:1510.03865
41. J. Gao, P. Nadolsky, A meta-analysis of parton distribution functions. JHEP **1407**, 035 (2014). arXiv:1401.0013
42. S. Carrazza, J.I. Latorre, J. Rojo, G. Watt, A compression algorithm for the combination of PDF sets. Eur. Phys. J. C **75**, 474 (2015). arXiv:1504.06469
43. S. Carrazza, S. Forte, Z. Kassabov, J.I. Latorre, J. Rojo, An unbiased hessian representation for Monte Carlo PDFs. Eur. Phys. J. C **75**(8), 369 (2015). arXiv:1505.06736
44. H. Paukkunen, P. Zurita, PDF reweighting in the Hessian matrix approach. JHEP **12**, 100 (2014). arXiv:1402.6623
45. C. Schmidt, J. Pumplin, C. P. Yuan, P. Yuan, Updating and Optimizing Error PDFs in the Hessian Approach. arXiv:1806.07950
46. A. Buckley, J. Ferrando, S. Lloyd, K. Nordstrm, B. Page et al., LHAPDF6: parton density access in the LHC precision era. Eur. Phys. J. C **75**, 132 (2015). arXiv:1412.7420
47. ATLAS Collaboration, G. Aad et al., Measurement of the double-differential high-mass Drell-Yan cross section in pp collisions at $sqrts$ = 8 TeV with the ATLAS detector. JHEP **08**, 009 (2016). arXiv:1606.01736

48. NNPDF Collaboration, J. Rojo, Improving quark flavor separation with forward W and Z production at LHCb. PoS DIS **2017**, 198 (2018). arXiv:1705.04468
49. LHCb Collaboration, R. Aaij, et al., Measurement of forward W and Z boson production in pp collisions at $sqrts$ = 8. TeV, JHEP **01**, 155 (2016). arXiv:1511.08039
50. ATLAS Collaboration, G., Aad et al., Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $sqrts$ = 8 TeV using the ATLAS detector. Eur. Phys. J. C **76**(10), 538 (2016). arXiv:1511.04716
51. ATLAS Collaboration Collaboration, Determination of the parton distribution functions of the proton from ATLAS measurements of differential W and Z/γ^* and $t\bar{t}$ cross sections, Tech. Rep. ATL-PHYS-PUB-2018-017, CERN, Geneva (2018)
52. ATLAS Collaboration, G., Aad et al., Measurement of the transverse momentum and phi^*_{eta} distributions of Drell-Yan lepton pairs in proton-proton collisions at $sqrts$ = 8 TeV with the ATLAS detector. Eur. Phys. J. C **76**(5), 291 (2016). arXiv:1512.02192
53. W. Stirling, E. Vryonidou, Charm production in association with an electroweak gauge boson at the LHC. Phys. Rev. Lett. **109**, 082002 (2012). arXiv:1203.6781
54. Will Barter, Stephen Farry, private communication
55. ATLAS Collaboration, M., Aaboud, et al., Measurement of the cross section for inclusive isolated-photon production in pp collisions at $sqrts$ = 13 TeV using the ATLAS detector. Phys. Lett. B **770**, 473–493 (2017). arXiv:1701.06882
56. J. Currie, A. Gehrmann-De Ridder, T. Gehrmann, E.W.N. Glover, A. Huss, J. Pires, Precise predictions for dijet production at the LHC. Phys. Rev. Lett. **119**(15), 152001 (2017). arXiv:1705.10271
57. E.L. Berger, J. Gao, H.X. Zhu, Differential distributions for t-channel single top-quark production and decay at next-to-next-to-leading order in QCD. JHEP **11**, 158 (2017). arXiv:1708.09405
58. R. Boughezal, J.M. Campbell, R.K. Ellis, C. Focke, W. Giele, X. Liu, F. Petriello, C. Williams, Color singlet production at NNLO in MCFM. Eur. Phys. J. C **77**(1), 7 (2017). arXiv:1605.08011
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**, 503 (2010). arXiv:0911.2985
60. Z. Nagy, Next-to-leading order calculation of three-jet observables in hadron hadron collision. Phys. Rev. D **68**, 094002 (2003). hep-ph/0307268
61. CMS Collaboration, V. Khachatryan et al., Measurement of the double-differential inclusive jet cross section in proton–proton collisions at $\sqrt{s}$ = 13 TeV. Eur. Phys. J. C **76**(8), 451 (2016). https://doi.org/10.1140/epjc/s10052-016-4286-3
62. NNPDF Collaboration, R.D. Ball et al., Parton distributions for the LHC Run II. JHEP **04** 040, (2015). arXiv:1410.8849
63. A.V. Manohar, P. Nason, G.P. Salam, G. Zanderighi, The photon content of the proton. JHEP **12**, 046 (2017). arXiv:1708.01256
64. NNPDF Collaboration, V. Bertone, S. Carrazza, N.P. Hartland, J. Rojo, Illuminating the photon content of the proton within a global PDF analysis. SciPost Phys. **5**, 008 (2018). arXiv:1712.07053
65. R. Nathvani, L. Harland-Lang, R. Thorne, A. Martin, Ad Lucem: The Photon in the MMHT PDFs, in 26th International Workshop on Deep Inelastic Scattering and Related Subjects (DIS 2018) Port Island, Kobe, Japan, April 16–20, 2018, (2018). arXiv:1807.07846
66. L.A. Harland-Lang, A.D. Martin, R.S. Thorne, The impact of LHC jet data on the MMHT PDF Fit at NNLO. Eur. Phys. J. C **78**(3), 248 (2018). arXiv:1711.05757
67. CMS Collaboration, V. Khachatryan et al., Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s}$ = 8 TeV. Eur. Phys. J. C **75**(11), 542 (2015). arXiv:1505.04480

68. CMS Collaboration, S. Chatrchyan et al., Measurement of associated W + charm production in pp collisions at $\sqrt{s}$ = 7 TeV. JHEP **02**, 013 (2014). arXiv:1310.1138
69. R.D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte et al., Reweighting and unweighting of parton distributions and the LHC W lepton asymmetry data. Nucl. Phys. B **855**, 608–638 (2012). arXiv:1108.1758
70. The NNPDF Collaboration, R.D. Ball et al., Reweighting NNPDFs: the W lepton asymmetry. Nucl. Phys. B **849**, 112–143 (2011). arXiv:1012.0836
71. The NNPDF Collaboration, R.D. Ball et al., Fitting parton distribution data with multiplicative normalization uncertainties. JHEP **05**, 075 (2010). arXiv:0912.2276
72. R.D. Ball, S. Carrazza, L. Del Debbio, S. Forte, J. Gao et al., Parton distribution benchmarking with LHC data. JHEP **1304**, 125 (2013). arXiv:1211.5142
73. J. Gao, C.S. Li, C.P. Yuan, NLO QCD corrections to dijet production via quark contact interactions. JHEP **07**, 037 (2012). arXiv:1204.4773
74. CMS Collaboration, V. Khachatryan et al., Search for quark contact interactions and extra spatial dimensions using dijet angular distributions in protonproton collisions at $\sqrt{s}$ = 8 TeV. Phys. Lett. B **746**, 79–99 (2015). arXiv:1411.2646
75. C. Grojean, E. Salvioni, M. Schlaffer, A. Weiler, Very boosted Higgs in gluon fusion. JHEP **05**, 022 (2014). arXiv:1312.3317
76. T. Sjostrand, S. Mrenna, P.Z. Skands, A brief introduction to PYTHIA 8.1. Comput. Phys. Commun. **178**, 852–867 (2008). arXiv:0710.3820
77. B.C. Allanach, SUSY Les Houches Accord 2. Comput. Phys. Commun. **180**, 8–25 (2009). arXiv:0801.0045
78. W. Beenakker, C. Borschensky, M. Krmer, A. Kulesza, E. Laenen, NNLL-fast: predictions for coloured supersymmetric particle production at the LHC with threshold and Coulomb resummation. JHEP **12**, 133 (2016). arXiv:1607.07741