



# Proton-PDF uncertainties in extracting nuclear PDFs from $W^\pm$ production in p+Pb collisions

Kari J. Eskola<sup>1,2,a</sup>, Petja Paakkinen<sup>1,2,3,b</sup>, Hannu Paukkunen<sup>1,2,c</sup>, Carlos A. Salgado<sup>3,d</sup>

<sup>1</sup> Department of Physics, University of Jyväskylä, P.O. Box 35, 40014 University of Jyväskylä, Finland

<sup>2</sup> Helsinki Institute of Physics, University of Helsinki, P.O. Box 64, 00014 University of Helsinki, Finland

<sup>3</sup> Instituto Galego de Física de Altas Enerxías (IGFAE), Universidade de Santiago de Compostela, 15782 Galicia, Spain

Received: 11 February 2022 / Accepted: 28 February 2022 / Published online: 29 March 2022  
© The Author(s) 2022

**Abstract** We discuss the recent CMS Collaboration measurement of  $W^\pm$  boson production in p+Pb collisions at 8.16 TeV in terms of the constraining power on nuclear parton distribution functions (PDFs). The impact of the free-proton PDF uncertainties on the nuclear PDF extraction is quantified by using a theoretical covariance-matrix method and Hessian PDF reweighting. We discuss different ways to mitigate these theoretical uncertainties, including self-normalization, forward-to-backward ratios and nuclear-modification ratios. It is found that none of these methods offer perfect cancellation of the free-proton PDFs but, with the present data uncertainties, the residual free-proton-PDF dependence has, conveniently for the global analyses, little effect on the extraction of the nuclear modifications. Based on a simple estimate of obtainable statistics at the LHC Run 3, we argue that this will change in the near future and it becomes more important to propagate the proton-PDF uncertainties accordingly. Using the obtained information on the correlations of the free-proton uncertainties, we also identify a new charge asymmetry ratio, where the cancellation of the proton-PDF uncertainties is found to be extremely good.

## 1 Introduction

The parton distribution functions (PDFs) of heavy nuclei, like their free-proton counterparts, are currently obtained most reliably from global analyses of experimental data. The bulk of these data comes from deep inelastic scattering (DIS) measurements which probe the nuclear structure directly and uniquely, with the precision limited at high enough scales

only by experimental uncertainties and perturbative accuracy. However, to constrain the full flavour dependence of the nuclear PDFs, it is necessary to additionally use proton–nucleus (p+A) processes, including fixed-target Drell–Yan (DY) dilepton as well as collider electroweak (EW) boson and (di)jet production data. For these processes, the collinearly factorized cross sections contain a convolution of the nuclear and free-proton PDFs, and as a consequence, the nuclear PDFs extracted from such data become inherently dependent on the assumed free-proton PDFs.

One could then envisage two systematic approaches to treat the proton-PDF uncertainties in the nuclear-PDF analyses: First, one can try to reduce the proton-PDF uncertainties by using observables where one probes instead the *nuclear modifications* of the PDFs, which are then parametrized and fitted, and the “baseline” free-proton PDF dependence effectively drops out, as has been done systematically in the EKS–EPPS line of analyses [1–7], and by others [8–12]. This approach has been particularly attractive since much of the older DIS and DY data are in any case available only in terms of nuclear ratios. Or, second, one could allow using also absolute cross sections, taking into account all possible correlations with the free-proton PDFs, a program which has more recently been undertaken by the nNNPDF collaboration [13–15].

In some other instances the treatment of the baseline proton-PDF dependence have been less explicit [16–25], with the inherent assumption being that the free-proton uncertainties are in any case smaller than the nuclear-PDF ones and thus do not cause a significant bias in the fit. Until very recently, this has been a justifiable approximation. However, as the precision of data from the LHC p+A program improves, it can become necessary to either propagate or mitigate the proton-PDF uncertainties in extracting the nuclear PDFs.

<sup>a</sup> e-mail: kari.eskola@jyu.fi

<sup>b</sup> e-mail: petja.k.m.paakkinen@jyu.fi (corresponding author)

<sup>c</sup> e-mail: hannu.paukkunen@jyu.fi

<sup>d</sup> e-mail: carlos.salgado@usc.es

One of the latest additions to the nuclear-PDF constraints is the CMS Collaboration measurement of  $W^\pm$  boson production in LHC Run 2 p+Pb collisions at 8.16 TeV [26], with an eight-fold increase in the statistics compared to the Run 1 data taking at 5.02 TeV [27]. These data have been already included in nuclear-PDF analyses, where they have been seen to give constraints either specifically on the gluon and sea-quark PDFs [25], strangeness [22], or on the flavour separation in more general [14]. The level at which the proton-PDF uncertainties are treated in these analyses varies, with Refs. [22, 25] taking the proton-PDFs as fixed, ignoring their uncertainties, and Ref. [14] propagating the proton-PDF uncertainties in the analysis, but not discussing their importance in the fit.

The role of the proton-PDF uncertainties in  $W^\pm$  production in p+Pb collisions has been considered previously in Ref. [28]. In this paper, we elaborate their significance in the context of the aforementioned CMS Collaboration measurement [26] (Sect. 2) and different ratios constructed from the data (Sect. 4). We also extend the analysis of the importance of the proton-PDF uncertainties in nuclear-modification fitting by the tools of theoretical covariance matrix (Sect. 3) and Hessian PDF reweighting (Sect. 5).

## 2 Proton-PDF uncertainties in $W^\pm$ production in p+Pb collisions

It is conventional to write the PDFs  $f_i^A$  of a nucleus with  $Z$  protons and  $N$  neutrons in terms of bound-nucleon PDFs at momentum fraction  $x$  and scale  $Q^2$  as

$$f_i^A(x, Q^2) = Zf_i^{p/A}(x, Q^2) + Nf_i^{n/A}(x, Q^2), \quad (1)$$

taking the bound-neutron PDFs  $f_i^{n/A}$  to be related to the bound-proton ones  $f_i^{p/A}$  by the isospin symmetry:

$$\begin{aligned} u^{n/A}(x, Q^2) &= d^{p/A}(x, Q^2), \quad d^{n/A}(x, Q^2) = u^{p/A}(x, Q^2), \\ \bar{u}^{n/A}(x, Q^2) &= \bar{d}^{p/A}(x, Q^2), \quad \bar{d}^{n/A}(x, Q^2) = \bar{u}^{p/A}(x, Q^2), \end{aligned} \quad (2)$$

and  $f_i^{n/A} = f_i^{p/A}$  for other flavours. This can be seen as an effective prescription, where the bound-nucleon PDFs should be understood as carrying information on the parton content of the ‘‘average’’ nucleon, used only to simplify the treatment of isospin dependence. In the region where  $x \leq 1$ , one can further write

$$f_i^{p/A}(x, Q^2) = R_i^{p/A}(x, Q^2)f_i^p(x, Q^2), \quad (3)$$

where the nuclear modification factors  $R_i^{p/A}$ , given the free-proton PDFs  $f_i^p$ , now encode all the information on the partonic structure of nuclei.

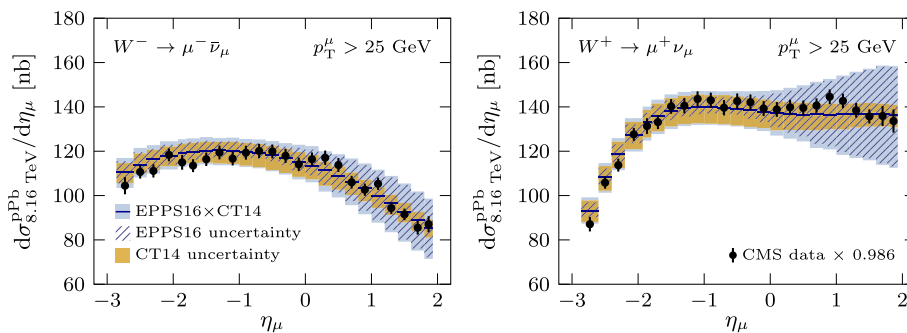
It should be noted that the above steps can be taken without any loss of generality. In the end, if *all* correlations between the proton and nuclear PDFs are correctly taken into account, it should not matter whether one parametrizes the absolute nuclear PDFs or the nuclear modifications. One is simply mapping a set of unknown functions  $f_i^A$  to the same number of functions  $R_i^{p/A}$ . In practice, however, simplifying assumptions are used in the nuclear-PDF analyses. We will use here the nuclear modifications from the EPPS16 analysis [6], and for full consistency, we use the CT14 NLO free-proton PDFs [29] but also validate the robustness of the results by comparing to the CT18 NLO PDFs [30]. This assumes that  $R_i^{p/A}$  depend only on the nuclear mass number  $A = Z + N$  (i.e. that there is no non-trivial isospin dependence in them) and that they are uncorrelated to  $f_i^p$ .<sup>1</sup> As mentioned in Sect. 1, the latter is true as a first approximation due to the use of appropriate ratio observables in EPPS16, but we will discuss later in this article the validity of this assumption in the presence of increasingly precise electroweak data.

The advantage of this framework is that we can study the relative importance of the nuclear-modification and free-proton-PDF uncertainties in any observable of interest. Here, we study these in the context of  $W^\pm$  production in p+Pb collisions at 8.16 TeV, as measured by the CMS Collaboration in the muon decay channel [26]. The lepton-rapidity differential cross sections, with a cut on lepton transverse momentum  $p_T^\mu > 25$  GeV, is presented in Fig. 1. The theoretical next-to-leading order (NLO) perturbative QCD predictions are obtained with MCFM [31], and the PDF uncertainties from EPPS16 and CT14 are calculated with the conventional asymmetric prescription at the 90% confidence level. As can be seen from the figure, the baseline CT14 free-proton PDF errors (shown as yellow boxes) contribute significantly to the total theoretical uncertainty budget (light blue boxes) and can even exceed those from the EPPS16 nuclear modifications (blue hatching) in some bins. The smallness of nuclear-modification uncertainties in the negative (backward) rapidities originates from the good neutral and charged-current DIS constraints at the probed values of  $x$ . Going to positive (forward) rapidities, we enter the less-constrained small- $x$  region and the nuclear-modification uncertainties begin to grow.

As was shown already in Ref. [26], the agreement between the CMS measurement and the NLO predictions from EPPS16  $\times$  CT14 is excellent. The goodness of fit for this data set is given by

$$\begin{aligned} \chi_C^2 &= (D - f_{\text{norm.}T})^T C^{-1} (D - f_{\text{norm.}T}) \\ &\quad + \left( \frac{f_{\text{norm.}} - 1}{\sigma_{\text{norm.}}} \right)^2, \end{aligned} \quad (4)$$

<sup>1</sup> Note that  $f_i^{p/A}$  and  $f_i^A$  are still correlated to  $f_i^p$  through Eqs. (3) and (1).



**Fig. 1** Lepton-rapidity differential  $W^\pm$  production cross sections in p+Pb collisions at 8.16 TeV with a breakdown of the theoretical uncertainties (EPPS16  $\times$  CT14, light-blue boxes) into those from free-proton PDFs (CT14 NLO, yellow boxes) and from the nuclear modifications

(EPPS16, blue hatching). The data from the CMS measurement [26] are presented with black markers, scaled with the optimal normalization factor explained in the text

where  $D$  and  $T$  are vectors of dimension  $N_{\text{data}}$  containing the data and theory values, and we have extracted the normalization uncertainty  $\sigma_{\text{norm.}} = 3.5\%$  from the data covariance matrix  $C$ , thus avoiding the D’Agostini bias [32]. By doing so, the optimal normalization factor  $f_{\text{norm.}}$  can be solved analytically, and in the figures we multiply the data with a factor

$$1/f_{\text{norm.}} = \frac{1 + \sigma_{\text{norm.}}^2 T^T C^{-1} T}{1 + \sigma_{\text{norm.}}^2 D^T C^{-1} D}. \tag{5}$$

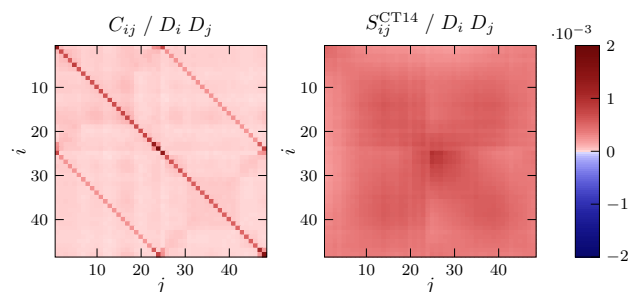
Taking  $T$  as the central prediction from EPPS16  $\times$  CT14, we then have  $1/f_{\text{norm.}} = 0.986$  and  $\chi^2_C/N_{\text{data}} = 1.12$ , confirming the visibly good data-to-theory agreement. When fitting to these data, we therefore do not expect the central nuclear PDFs to change much from the EPPS16 results, but as the experimental uncertainties are much smaller than the nuclear-modification uncertainties especially at forward rapidities, we can expect a significant reduction in the latter. The large baseline free-proton uncertainties can however affect the obtainable constraints and we need to find a way to either quantify or mitigate the impact.

### 3 Theoretical covariance matrix

One way to quantify the impact of a certain theoretical source of uncertainty is to use the method of theoretical covariance matrix [33]. For the free-proton uncertainties, taken from the CT14 PDFs, this matrix is given by

$$S_{ij}^{\text{CT14}} = \sum_k \left( \frac{T_i[S_{\text{CT14}}^{k,+}] - T_i[S_{\text{CT14}}^{k,-}]}{2 \times 1.645} \times \frac{T_j[S_{\text{CT14}}^{k,+}] - T_j[S_{\text{CT14}}^{k,-}]}{2 \times 1.645} \right), \tag{6}$$

where the sum goes over the CT14 parameter eigendirections  $k$  and  $T_i[S_{\text{CT14}}^{k,\pm}]$  are the corresponding predictions for the  $i$ th data point with positive and negative parameter variations in



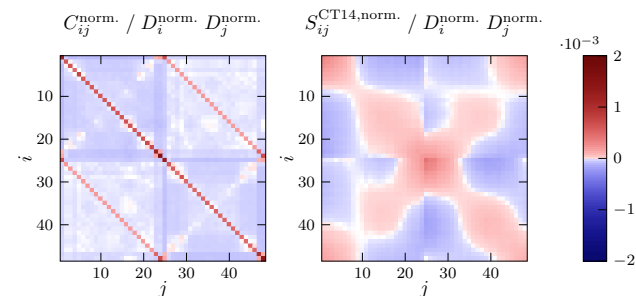
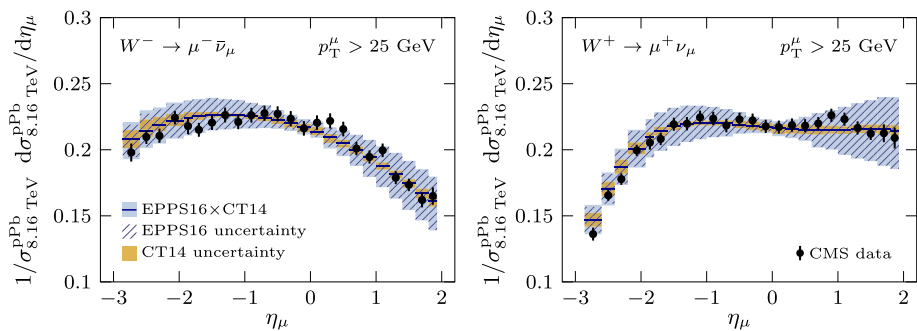
**Fig. 2** The experimental (excluding overall normalization uncertainty) and theoretical free-proton-PDF covariance matrices for the p+Pb  $W^\pm$  measurement at 8.16 TeV. Indices  $i, j$  follow the same ordering as the data points in Fig. 1, with the indices 1 through 24 corresponding to the  $W^-$  production and 25 through 48 to  $W^+$

that eigendirection (all calculated with the central EPPS16 nuclear modifications). The factors 1.645 in the denominators are used to scale the nominally 90% confidence-level uncertainties of CT14 to a 68% (one standard deviation) level in order not to overestimate their impact with respect to the experimental uncertainties.

The CT14 theoretical covariance matrix is presented in Fig. 2 with a comparison to the experimental covariance matrix. To ease the interpretation, we have excluded the large fully correlated luminosity component from the experimental covariance matrix, as in Eq. (4), and divided each matrix element with the product of the corresponding data values. We see that the proton-PDF uncertainties are comparable or larger than the correlated non-luminosity experimental uncertainties but still mostly smaller than the combined statistical and non-luminosity systematical uncertainties in the diagonal elements. Clearly, the free-proton PDFs contribute a non-negligible uncertainty component to the nuclear-modification fitting.

With the theoretical CT14 proton-PDF uncertainties taken into account, the figure of merit for the nuclear-modification d.o.f.s takes the form [33]

**Fig. 3** As Fig. 1, but now for the self-normalized cross sections



**Fig. 4** As Fig. 2, but now for the self-normalized cross sections. Indices  $i, j$  follow the same ordering as the data points in Fig. 3, with the indices 1 through 24 corresponding to the  $W^-$  production and 25 through 48 to  $W^+$

$$\chi^2_{C+S^{CT14}} = (D - f_{\text{norm.}} T)^T (C + S^{CT14})^{-1} (D - f_{\text{norm.}} T) + \left( \frac{f_{\text{norm.}} - 1}{\sigma_{\text{norm.}}} \right)^2, \tag{7}$$

and we find  $\chi^2_{C+S^{CT14}}/N_{\text{data}} = 0.85$  for EPPS16. Comparing this to the value  $\chi^2_C/N_{\text{data}} = 1.12$  for EPPS16  $\times$  CT14, we see that the chosen proton PDFs can indeed have a significant impact on the level of agreement with the data.

Interestingly, the proton-PDF uncertainties are strongly positively correlated, behaving almost like an additional normalization uncertainty. It is exactly this positive correlation (and the positive correlation with the corresponding proton-proton cross section) which makes the uncertainty reduction with the ratios discussed in Sect. 4 possible. We note also that the optimal data normalization that we find for EPPS16  $\times$  CT14 from Eq. (5) is  $1/f_{\text{norm.}} = 0.986$ , well within the 3.5% normalization uncertainty. This should be compared to the value of 0.960 in the nCTEQ15WZ fit for these data [22]. Since the proton PDFs contribute significantly to the normalization of the predictions, we can speculate whether the larger than  $1 \times \sigma_{\text{norm.}}$  normalization shift in the nCTEQ15WZ analysis originates from the used free-proton baseline. This possibility is also corroborated by the fact that when testing the robustness of the results presented here by changing the free-proton PDFs to CT18 NLO, the main effect was a change in the normalization, with the optimal data-scaling factor  $1/f_{\text{norm.}}$  changing to a value 0.997. The CT18 uncertainties were also observed to be slightly

less correlated across different rapidities, but the uncertainties were found to be almost the same, and this had no impact on our conclusions.

### 4 Reducing proton-PDF uncertainties

As we have shown that the free-proton uncertainties are important in describing the p+Pb  $W^\pm$  data, it makes sense to explore ways to reduce these uncertainties. Since the covariance matrix of the CMS measurement is available to us, we can propagate the data uncertainties to any desired observable keeping also track of the correlations by using

$$C^{\text{new}} = J C J^T, \tag{8}$$

where  $J$  is the Jacobian of the transformation. We note that also perturbative higher-order corrections can (partially) cancel in many of the considered ratios, which supports their use in nuclear-PDF analyses, but the importance of missing higher orders is left outside the scope of this article.

#### 4.1 Self-normalized cross sections

Since the free-proton PDF uncertainties were found to be strongly correlated, almost normalization-like, a viable option to reduce them is by self-normalizing the cross sections

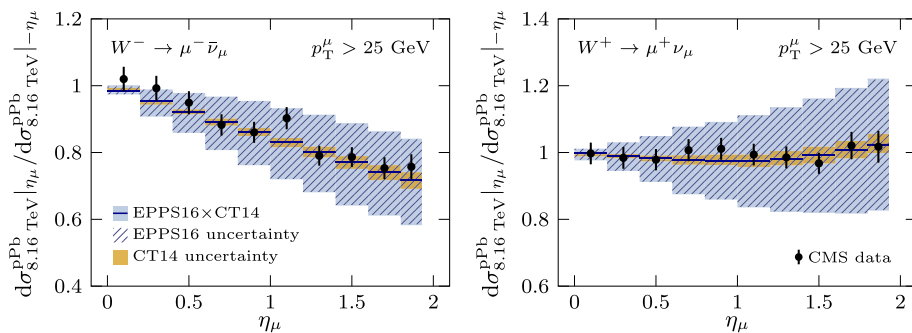
$$d\sigma_{\text{pPb}}^{W^\pm, \text{norm.}}/d\eta_\mu = \frac{1}{\sigma_{\text{pPb}}^{W^\pm}} d\sigma_{\text{pPb}}^{W^\pm}/d\eta_\mu, \tag{9}$$

where

$$\sigma_{\text{pPb}}^{W^\pm} = \int_{-2.86}^{1.93} d\eta_\mu d\sigma_{\text{pPb}}^{W^\pm}/d\eta_\mu \tag{10}$$

are the fiducial integrated cross sections of each  $W^\pm$  charge. We perform here the normalization for each charge separately, but it would be also possible to do this by dividing with the charge-summed integrated cross section. The obtained normalized cross sections are presented in Fig. 3 and the experimental  $C^{\text{norm.}}$  (from Eq. (8)) and theoretical  $S^{\text{CT14, norm.}}$  (calculated directly from the CT14 error sets) covariance matrices in Fig. 4. We observe a very good, but not perfect cancellation of the free-proton uncertainties, with

**Fig. 5** As Fig. 1, but now for the forward-to-backward ratios



the remaining CT14 uncertainties being largest at the large negative rapidities,  $\eta_\mu < -1.93$ .

In addition to cancelling the normalization uncertainty, the self-normalization changes the correlation pattern of the remaining statistical and systematic experimental uncertainties, which become mostly anticorrelated across different rapidity bins. Importantly, even the originally uncorrelated (statistical) uncertainties become correlated in the self-normalized cross sections. Another important thing to notice here is that neither the experimental nor the theoretical covariance matrix is invertible, with  $\det C^{\text{norm.}} = \det S^{\text{CT14, norm.}} = 0$ . This simply follows from the fact that a self-normalized set of data forms an overdetermined system: given all but one data point, the last one can be solved from the requirement that the data integrate to one. Another way to see this is to notice that the self-normalization is not a bijection, with an immediate consequence that  $\det J = 0$ . For this property, one should fit to the self-normalized data by leaving one point out. Due to the fully correlated nature of the normalized data, it does not matter which data point is left out, manifesting the loss of information in the normalization.<sup>2</sup>

### 4.2 Forward-to-backward ratios

The forward-to-backward ratios

$$R_{\text{FB}}^{W^\pm} = \frac{d\sigma_{\text{pPb}}^{W^\pm} / d\eta_\mu |_{\eta_\mu}}{d\sigma_{\text{pPb}}^{W^\pm} / d\eta_\mu |_{-\eta_\mu}} \tag{11}$$

have been considered earlier in Ref. [28], where it was realised that they do not yield as good a cancellation of the free-proton uncertainties as e.g. the same ratios for Z-boson or dijet [36] production. Since the experimental acceptance is not symmetric with respect to the p+Pb center-of-mass

frame, one has to drop part of the data points, in this case those for  $\eta_\mu < -1.93$ . Furthermore, by taking the ratio, the number of data points is still halved, leading to a significant loss of information. The remaining ten data points for each charge are shown in Fig. 5, where we see that the proton-PDF cancellation is good, but starts to worsen towards larger rapidities.

This can be understood by taking the large-rapidity limit  $\eta_\mu \gg 0$ , where we can take the large momentum-fraction  $x_1$  to probe only valence quarks and the small momentum-fraction  $x_2$  then probes the sea quarks. At this limit, neglecting the Cabibbo suppressed quark-mixing effects and denoting  $x_{1,2} := x_{1,2} |_{\eta_\mu} = x_{2,1} |_{-\eta_\mu}$ , we can approximate at leading order

$$R_{\text{FB}}^{W^-} \underset{x_1 \text{ large}, x_2 \text{ small}}{\approx} \underset{\eta_\mu \gg 0}{\approx} \frac{Z R_{\bar{u}}^{p/A}(x_2) + N \frac{\bar{d}^p(x_2)}{\bar{u}^p(x_2)} R_d^{p/A}(x_2)}{Z R_{d_V}^{p/A}(x_1) + N \frac{u_V^p(x_1)}{d_V^p(x_1)} R_{u_V}^{p/A}(x_1)} \tag{12}$$

and

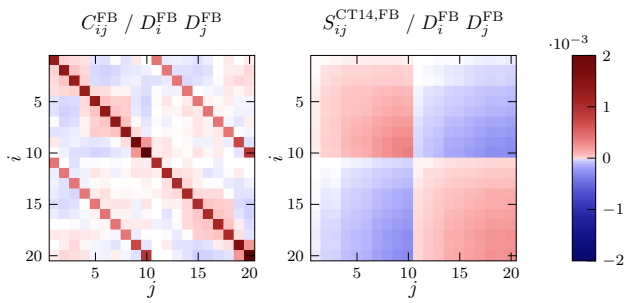
$$R_{\text{FB}}^{W^+} \underset{x_1 \text{ large}, x_2 \text{ small}}{\approx} \underset{\eta_\mu \gg 0}{\approx} \frac{Z R_d^{p/A}(x_2) + N \frac{\bar{u}^p(x_2)}{\bar{d}^p(x_2)} R_{\bar{u}}^{p/A}(x_2)}{Z R_{u_V}^{p/A}(x_1) + N \frac{d_V^p(x_1)}{u_V^p(x_1)} R_{d_V}^{p/A}(x_1)} \tag{13}$$

where we have suppressed for simplicity the relevant phase-space integrations and the scale-dependence of the PDFs. We note that this approximation is not exact in the data region as there can still be sizeable (but subleading) contributions also from the  $\bar{c} + s$  and  $c + \bar{s}$  channels [38]. In any case, we see that the forward-to-backward ratios depend on the free-proton PDFs through  $u_V/d_V$  and  $\bar{u}/\bar{d}$  ratios, which determine the relative size of the contributions from the different nuclear modifications and give a non-cancelling contribution to the theoretical uncertainty.

The resulting covariance matrices for the forward-to-backward ratios are presented in Fig. 6, where we observe a positive correlation of the proton-PDF uncertainties between same-charge bins, but an anticorrelation between different charges. Indeed, one could reduce the proton-PDF uncertainties further by taking the forward-to-backward ratio of

<sup>2</sup> We note, however, that e.g. in the case of the CMS measurement of self-normalized dijet cross sections in p+p and p+Pb collisions at 5.02 TeV [34], which we have studied in Ref. [35], the data correlations were not published and it is less clear how to treat the data statistically accurately in a fit. Without knowing the correlations, it would matter which data point was left out.





**Fig. 6** As Fig. 2, but now for the forward-to-backward ratio. Indices  $i, j$  follow the same ordering as the data points in Fig. 5, with the indices 1 through 10 corresponding to the  $W^-$  production and 11 through 20 to  $W^+$

the differential cross section summed over the two charges, as considered in Ref. [26], but this leads to a further loss of information compared to taking the ratio separately for different charges, and the constraints for nuclear-PDF analyses are rather limited.

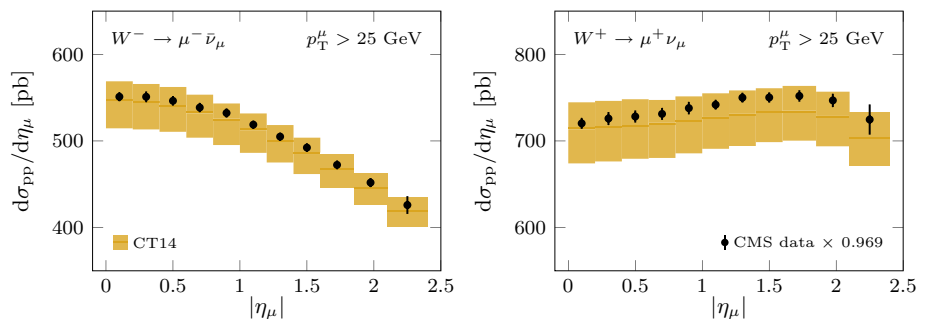
### 4.3 Nuclear-modification ratios

We now study the possibility of using nuclear-modification ratios to cancel free-proton uncertainties. For the 8.16 TeV p+Pb data, no same-energy p+p reference is available, but one could construct “mixed-energy” ratios

$$R_{pPb}^{W^\pm} = \frac{d\sigma_{pPb, 8.16 \text{ TeV}}/d\eta_\mu}{A d\sigma_{pp, 8.0 \text{ TeV}}/d\eta_\mu} \quad (14)$$

with the p+p measurements at 8.0 TeV, where the probed  $x$ -ranges are almost the same between the two energies. We use here the measurements from the CMS Collaboration [37], shown in Fig. 7 along with the predictions from the CT14 PDFs. Again, the 2.6% normalization uncertainty is taken into account in presenting the data. The optimal shift, 0.969, is slightly larger than what we found for p+Pb. The agreement in normalization could be again improved by using CT18

**Fig. 7** Lepton-rapidity differential  $W^\pm$  production cross sections in p+p collisions at 8.0 TeV with theoretical uncertainties from the free-proton PDFs (CT14 NLO, yellow boxes). The data from the CMS Collaboration measurement [37] are presented with black markers, scaled with the optimal normalization factor



**Fig. 8** The Jacobian matrix  $J_{ij}^{pp}$  for propagating the p+p uncertainties into  $R_{pPb}$  uncertainties (cf. Eq. (15)) and the experimental (excluding overall normalization uncertainty) covariance matrix  $C_{ij}^{pp}$  for the p+p  $W^\pm$  measurement at 8.0 TeV. Indices  $i, j$  in  $C_{ij}^{pp}$  and the index  $j$  in  $J_{ij}^{pp}$  follow the ordering of the data points in Fig. 7, with the indices 1 through 11 corresponding to the  $W^-$  production and 12 through 22 to  $W^+$ , and the index  $i$  in  $J_{ij}^{pp}$  follows the ordering in Fig. 9, with the indices 1 through 22 corresponding to the  $W^-$  production and 23 through 44 to  $W^+$

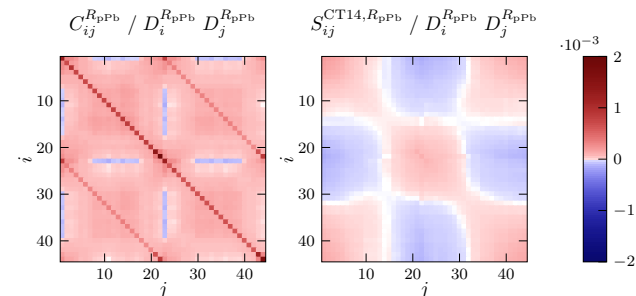
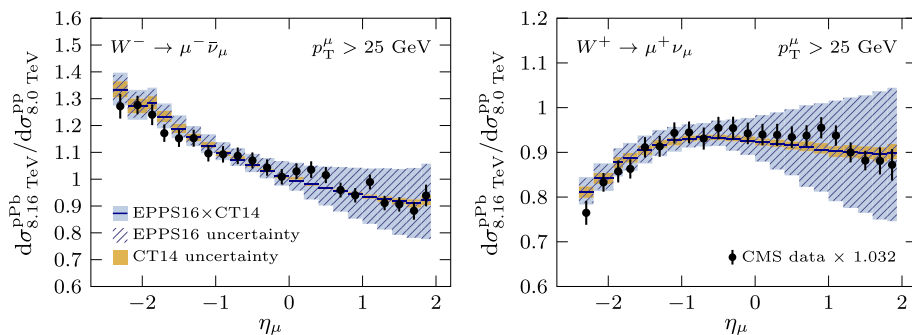
PDFs, to 0.996, but the size of the PDF uncertainties stays almost the same.

The rapidity binning is the same in p+p and p+Pb measurements up to  $|\eta_\mu| < 1.6$ . For larger rapidities, we associate the p+Pb bins with the most-overlapping one in p+p. Therefore, for the p+Pb bins with  $1.6 < |\eta_\mu| < 1.8$  we take the ratio with p+p bin  $1.6 < |\eta_\mu| < 1.85$  and the p+p bin  $1.85 < |\eta_\mu| < 2.1$  is used in obtaining three of the  $R_{pPb}$  bins:  $1.8 < \eta_\mu < 1.93$ ,  $-1.93 < \eta_\mu < -1.8$  and  $-2.2 < \eta_\mu < -1.93$ . Finally, for the p+Pb bin  $-2.4 < \eta_\mu < -2.2$ , the p+p bin  $2.1 < |\eta_\mu| < 2.4$  is used. For  $\eta_\mu < -2.4$  we run out of p+p bins and we discard the remaining two p+Pb data points for both  $W^\pm$  charges. The loss of information is therefore slightly larger than in the self-normalized cross sections, but significantly smaller than in the forward-to-backward ratios.

Since the correlations between the p+Pb and p+p measurements are not known, the covariance matrix for the ratio is calculated with

$$C^{R_{pPb}} = J^{pPb} C^{pPb} (J^{pPb})^T + J^{pp} C^{pp} (J^{pp})^T. \quad (15)$$

**Fig. 9** As Fig. 1, but now for the nuclear-modification ratio with the p+p reference taken from Ref. [37]



**Fig. 10** As Fig. 2, but now for the nuclear-modification ratio with the p+p reference taken from Ref. [37]. Indices  $i, j$  follow the same ordering as the data points in Fig. 9, with the indices 1 through 22 corresponding to the  $W^-$  production and 23 through 44 to  $W^+$

This is a conservative estimate: in a direct experimental analysis some of the systematic uncertainties could be cancelled in the ratio. The Jacobian  $J^{PP}$  and the covariance matrix  $C^{PP}$  for the p+p data are presented in Fig. 8, visualising also how each p+p point contributes to multiple  $R_{pPb}$  bins. The correlations arising from this are then taken correctly into account in Eq. (15).

The obtained mixed-energy nuclear-modification ratios and the corresponding experimental and theoretical covariance matrices are presented in Figs. 9 and 10, respectively. The data are again well described by the EPPS16 x CT14 predictions, but due to the larger optimal downward normalization shift in p+p compared to p+Pb, we find the optimal shift for the nuclear-modification ratio to be  $1/f_{\text{norm.}} = 1.032$ , still within the combined normalization uncertainty of 4.36%. Compared to the previously discussed ratios, we can expect a more “local” cancellation of the proton-PDF dependence. However, since we use different collision energies for p+p and p+Pb and the rapidity binning does not exactly match outside mid-rapidity, the probed  $x$  regions in the ratio can be slightly different, which can make the proton-PDF cancellation less than perfect.<sup>3</sup> We observe still a very good cancella-

tion, comparable or better than in the previous ratios, and the CT14 uncertainties in Fig. 10 are now clearly smaller than the diagonal elements of the experimental covariance matrix in all bins (note that we have also here omitted the overall experimental normalization uncertainty from the presentation of the matrix, in accordance with Eq. (4)). Consequently, the free-proton PDFs have smaller impact on the agreement with data, as we find  $\chi_C^2/N_{\text{data}} = 0.77$  with the pure experimental uncertainties and  $\chi_{C+SCT14}^2/N_{\text{data}} = 0.75$  after taking the CT14 uncertainties into account.

The cancellation somewhat deteriorates towards larger rapidities. In the far-backward region, the momentum-fraction from the nuclear side  $x_2$  is large and we can approximate, at leading order and neglecting the small shifts in the momentum fractions due to the different energies,

$$R_{pPb}^{W^-} \underset{x_2 \text{ large}}{\underset{\eta_\mu \ll 0}{\approx}} \frac{Z}{A} R_{dV}^{p/Pb}(x_2) + \frac{N}{A} \frac{u_V^p(x_2)}{d_V^p(x_2)} R_{uV}^{p/Pb}(x_2) \quad (16)$$

and

$$R_{pPb}^{W^+} \underset{x_2 \text{ large}}{\underset{\eta_\mu \ll 0}{\approx}} \frac{Z}{A} R_{uV}^{p/Pb}(x_2) + \frac{N}{A} \frac{d_V^p(x_2)}{u_V^p(x_2)} R_{dV}^{p/Pb}(x_2), \quad (17)$$

where we see that the proton  $u_V/d_V$  ratio again sets the limit to how well the proton-PDF uncertainties are cancelled. Note that in Eq. (16) we have the ratio  $u_V/d_V$ , leading to an enhancement at the probed backward rapidities, whereas in Eq. (17) we have its reciprocal, leading to a suppression, even in absence of nuclear modifications.

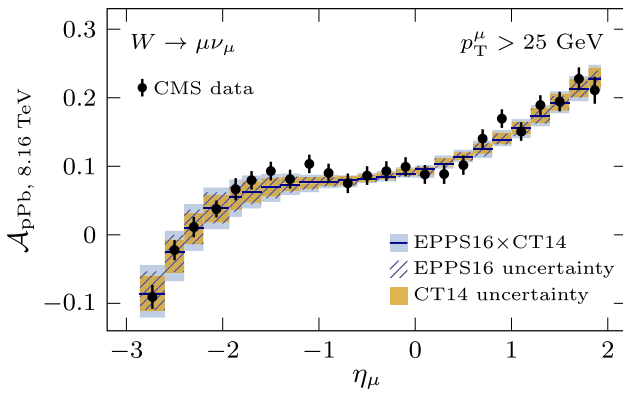
In the far-forward region the nuclear momentum-fraction  $x_2$  is small and we have

$$R_{pPb}^{W^-} \underset{x_2 \text{ small}}{\underset{\eta_\mu \gg 0}{\approx}} \frac{Z}{A} R_{\bar{u}}^{p/Pb}(x_2) + \frac{N}{A} \frac{\bar{d}^p(x_2)}{\bar{u}^p(x_2)} R_{\bar{d}}^{p/Pb}(x_2) \quad (18)$$

and

$$R_{pPb}^{W^+} \underset{x_2 \text{ small}}{\underset{\eta_\mu \gg 0}{\approx}} \frac{Z}{A} R_{\bar{d}}^{p/Pb}(x_2) + \frac{N}{A} \frac{\bar{u}^p(x_2)}{\bar{d}^p(x_2)} R_{\bar{u}}^{p/Pb}(x_2). \quad (19)$$

<sup>3</sup> In a direct experimental measurement of the ratio, one could consider binning the data in a shifted rapidity variable  $y_{\text{ref}}$  as in Ref. [39] to minimize the effect of using different energies.



**Fig. 11** The  $W^\pm$  production charge asymmetry, with a breakdown of theory uncertainties as in Fig. 1

At this limit, we see from Fig. 9 that the ratios of both charges approach the value 0.9, reflecting the fact that at these large scales, one is probing an almost flavour symmetric quark sea generated through  $g \rightarrow q\bar{q}$  splittings, and therefore also the probed nuclear modifications are almost the same.

#### 4.4 Charge asymmetries

As discussed already in Ref. [28], the traditional charge asymmetry

$$\mathcal{A}_{pPb} = \frac{d\sigma_{pPb}^{W^+}/d\eta_\mu - d\sigma_{pPb}^{W^-}/d\eta_\mu}{d\sigma_{pPb}^{W^+}/d\eta_\mu + d\sigma_{pPb}^{W^-}/d\eta_\mu} \quad (20)$$

is very sensitive to the free-proton uncertainties. As shown in Fig. 11, the excellent data-to-theory agreement continues to be valid also in this observable, but as now the free-proton and nuclear-modification uncertainties are of the same order, one is probing a non-trivial combination of the two, and the usefulness for nuclear-PDF fits is rather limited. In particular, at large positive rapidities this observable probes mostly the  $u_V - d_V$  asymmetry in proton [40], with the nuclear uncertainties having a strong cancellation.

It is, however, possible to construct asymmetries with more direct sensitivity to the nuclear modifications. In Ref. [28], a charge ratio of forward–backward differences

$$\tilde{\mathcal{A}}_{pPb} = \frac{d\sigma_{pPb}^{W^+}/d\eta_\mu|_{\eta_\mu} - d\sigma_{pPb}^{W^+}/d\eta_\mu|_{-\eta_\mu}}{d\sigma_{pPb}^{W^-}/d\eta_\mu|_{\eta_\mu} - d\sigma_{pPb}^{W^-}/d\eta_\mu|_{-\eta_\mu}}, \quad (21)$$

$$\mathcal{A}_{pPb}^* \underset{\substack{\eta_\mu \gg 0 \\ x_1 \text{ large} \\ x_2 \text{ small}}}{\approx} \frac{Z \left[ R_d^{p/A}(x_2) - \frac{\bar{u}^p(x_2)}{d^p(x_2)} \frac{d_V^p(x_1)}{u_V^p(x_1)} R_{d_V}^{p/A}(x_1) \right] + N \frac{\bar{u}^p(x_2)}{d^p(x_2)} \left[ R_u^{p/A}(x_2) - R_{u_V}^{p/A}(x_1) \right]}{Z \left[ R_{u_V}^{p/A}(x_1) - \frac{\bar{u}^p(x_2)}{d^p(x_2)} \frac{d_V^p(x_1)}{u_V^p(x_1)} R_u^{p/A}(x_2) \right] + N \frac{d_V^p(x_1)}{u_V^p(x_1)} \left[ R_{d_V}^{p/A}(x_1) - R_d^{p/A}(x_2) \right]}. \quad (23)$$

shown in Fig. 12 (left), was proposed, motivated by the finding that for cross sections differential in the  $W^\pm$  boson rapidity, the proton-PDF uncertainties cancel extremely well in this quantity. Here, with the experimentally measurable lepton rapidity, we find the cancellation to be slightly worse, but it still gives far better access to the nuclear modifications than the traditional charge asymmetry. In particular, the measured data differ significantly from the predictions with free-proton PDFs taking into account the isospin effects only, i.e. neglecting the nuclear modifications in the bound-nucleon PDFs. Close to midrapidity this observable is experimentally problematic since the denominator approaches zero, and we find with the linear error propagation the statistics to be insufficient for any constraints at  $|\eta_\mu| < 0.4$ .

We can now use our knowledge of the proton-PDF correlations to our advantage. As can be seen from Figs. 4 and 9, after taking away the overall normalization-like contribution, there is an anticorrelation in the proton-PDF uncertainties between the same-charge forward and backward cross sections. This anticorrelation is reflected also in the imperfect free-proton-PDF cancellation in the forward-to-backward ratios. Conversely, and quite unexpectedly, there appears to be a *positive* correlation between the forward production of one charge and the backward production of the other and an anticorrelation between the two charges at the same rapidity. Based on the approximation in Eqs. (16) through (19), this appears to be possible only if the ratio  $\bar{u}^p(x_2)/\bar{d}^p(x_2)$  at small  $x_2$  and  $u_V^p(x_1)/d_V^p(x_1)$  at large  $x_1$  are positively correlated. We find this to be true at the EW scale for CT14 (and also for CT18, MSHT20 [41] and NNPDF4.0 [42]) as long as  $x_2 < 0.03$ , independently of  $x_1$ .

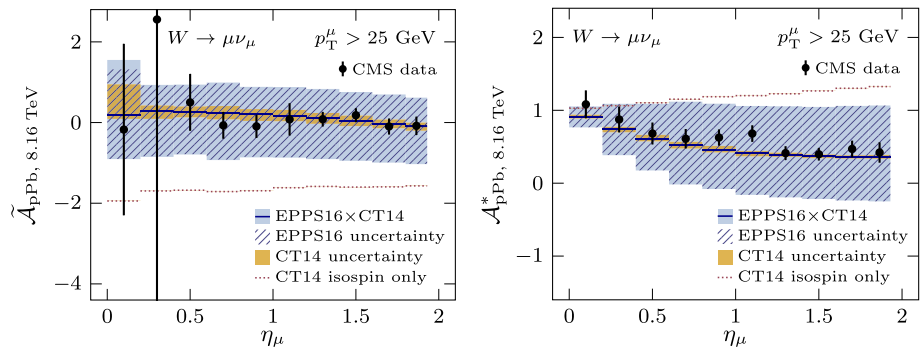
With this information in mind, we construct here a new forward-to-backward ratio of rapidity-mirrored charge difference

$$\mathcal{A}_{pPb}^* = \frac{d\sigma_{pPb}^{W^+}/d\eta_\mu|_{\eta_\mu} - d\sigma_{pPb}^{W^-}/d\eta_\mu|_{-\eta_\mu}}{d\sigma_{pPb}^{W^+}/d\eta_\mu|_{-\eta_\mu} - d\sigma_{pPb}^{W^-}/d\eta_\mu|_{\eta_\mu}}, \quad (22)$$

shown in Fig. 12 (right). The free-proton-PDF uncertainties in this observable are negligible and it avoids the problem of vanishing denominator that appeared in Eq. (21) as the  $W^+$  cross section is always larger than the  $W^-$  one. Within the approximation used in Eqs. (12) and (13), it can be written in the large-rapidity limit as



**Fig. 12** As Fig. 1, but now for the alternative charge-asymmetry ratios



The proton-PDF cancellation is still not exact, but we see that unlike in the forward-to-backward ratios where the proton-PDF correlations lead to a strictly additive behaviour of the uncertainties, in this asymmetry the correlations contribute in a destructive way in the terms proportional to  $Z$ . There is also large cancellation of the  $N$ -proportional terms in the differences, and at the “isospin only” limit these terms vanish completely. Indeed, one notices that in a (hypothetical) free-proton–free-neutron collision at leading order the dominant  $u + \bar{d}$  and  $\bar{u} + d$  contributions cancel in the difference  $d\sigma_{pn}^{W^+}/d\eta_\mu|_{\eta_\mu} - d\sigma_{pn}^{W^-}/d\eta_\mu|_{-\eta_\mu}$ . As  $d\sigma_{pp}^{W^+}/d\eta_\mu|_{\eta_\mu} - d\sigma_{pp}^{W^-}/d\eta_\mu|_{-\eta_\mu}$  is instead forward-to-backward symmetric, the asymmetry in the “isospin only” limit is then approximately one.

When the nuclear modifications are taken into account, this asymmetry is seen to probe a fairly non-trivial combination of different nuclear-modification terms, and the achievable constraints will depend on their relative uncertainties. Similarly to the forward-to-backward ratios, both a nuclear suppression in the sea quarks at small  $x$  and an enhancement in the valence quarks at large  $x_1$  cause a suppression in the ratio, explaining the strong deviation from the “isospin only” prediction.

For the other alternative asymmetry one finds that

where the proton-PDF uncertainty reduction is again apparent in the  $Z$ -proportional terms. It is now these terms that have the large cancellation and vanish at the “isospin only” limit, owing to the forward-to-backward symmetry of the  $p+p$  system, and the asymmetry reduces to minus one in this approximation.

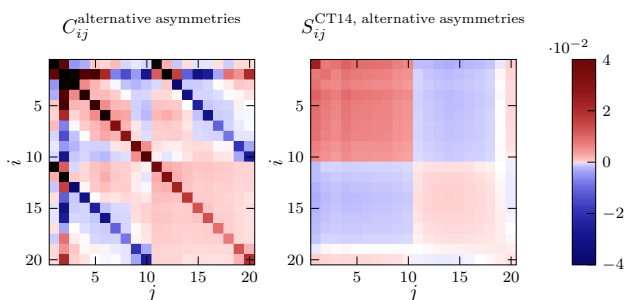
Since the same cross sections are used in each of them, the experimental uncertainties of the different asymmetries are strongly correlated, as can be seen from Fig. 13 for the alternative asymmetries, and must be taken into account when performing a simultaneous fit. We also note that in constructing these ratios, like in the case of forward-to-backward ratios, one again had to discard part of the  $p+Pb$  data, which might lead to reduced constraints.

### 5 Hessian reweighting

To further test the importance of the free-proton uncertainties in a nuclear-PDF fit, we have employed here the Hessian PDF-reweighting method [35, 43–46]. This proceeds by supplementing Eqs. (4) and (7) with a penalty term  $P_{EPPS16}$ , such that

$$\Delta\chi_{\text{total}, c}^2(\mathbf{z}) = \chi_c^2(\mathbf{z}) + P_{EPPS16}(\mathbf{z}) \tag{25}$$

$$\tilde{\mathcal{A}}_{pPb} \underset{\substack{\eta_\mu \gg 0 \\ x_1 \text{ large} \\ x_2 \text{ small}}}{\approx} \frac{Z \left[ R_{\bar{d}}^{p/A}(x_2) - R_{u\nu}^{p/A}(x_1) \right] + N \left[ \frac{\bar{u}^p(x_2)}{d^p(x_2)} R_{\bar{u}}^{p/A}(x_2) - \frac{d_V^p(x_1)}{u_V^p(x_1)} R_{d\nu}^{p/A}(x_1) \right]}{Z \frac{\bar{u}^p(x_2)}{d^p(x_2)} \frac{d_V^p(x_1)}{u_V^p(x_1)} \left[ R_{\bar{u}}^{p/A}(x_2) - R_{d\nu}^{p/A}(x_1) \right] + N \left[ \frac{d_V^p(x_1)}{u_V^p(x_1)} R_{\bar{d}}^{p/A}(x_2) - \frac{\bar{u}^p(x_2)}{d^p(x_2)} R_{\mu\nu}^{p/A}(x_1) \right]}, \tag{24}$$



**Fig. 13** As Fig. 2, but now for the alternative asymmetries shown in Fig. 12. Note that here we do not normalize by the data values as they can be smaller than the uncertainties in  $\tilde{A}_{p\text{pb}}$ . The scale of the heat map is thus different from the other figures. Indices  $i, j$  follow the same ordering as the data points in Fig. 12, with the indices 1 through 10 corresponding to  $\tilde{A}_{p\text{pb}}$  and 11 through 20 to  $\tilde{A}_{p\text{pb}}^*$

and

$$\Delta\chi_{\text{total}, C+S^{\text{CT14}}}^2(\mathbf{z}) = \chi_{C+S^{\text{CT14}}}^2(\mathbf{z}) + P_{\text{EPPS16}}(\mathbf{z}) \quad (26)$$

approximate the change in the total figure of merit of a global analysis, where the CMS data would be included on top of those in the EPPS16 analysis, as a function of the eigenparameters  $\mathbf{z}$  of the EPPS16 analysis, either without or with the CT14 theoretical covariance matrix included. As in Ref. [35], we take into account the leading non-quadratic terms by taking

$$P_{\text{EPPS16}}(\mathbf{z}) = \sum_k (a_k z_k^2 + b_k z_k^3) \quad (27)$$

and in  $\chi_C^2$  and  $\chi_{C+S^{\text{CT14}}}^2$  by including the first non-linear terms in

$$T(\mathbf{z}) = T_0 + \sum_k (d_k z_k + e_k z_k^2), \quad (28)$$

where the central prediction  $T_0$  as well as the coefficients  $a_k, b_k \in \mathbb{R}$  and  $d_k, e_k \in \mathbb{R}^{N_{\text{data}}}$  are calculated with the information provided in the EPPS16 analysis [6]. By minimizing Eqs. (25) or (26), one finds the new, updated best estimate of the nuclear PDFs, and, by diagonalizing the Hessian matrix at the found minimum, one can define the new error sets, as explained in Ref. [35].

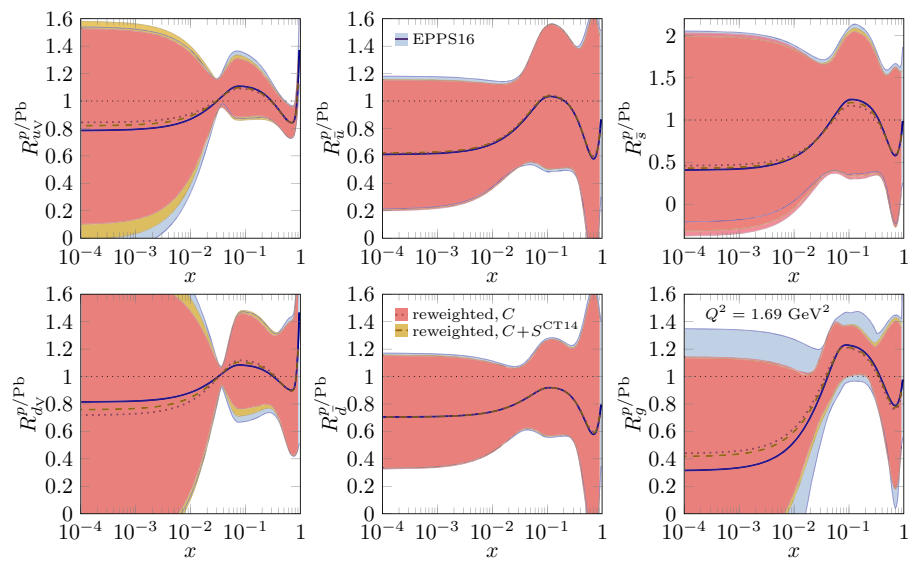
Figure 14 shows the nuclear modifications after reweighting with the absolute cross sections from Fig. 1, presented at the EPPS16 parametrization scale  $Q^2 = 1.69 \text{ GeV}$ , both without (“C”) and with (“C +  $S^{\text{CT14}}$ ”) the CT14 uncertainties included. The results might appear surprising, in the sense that the  $W^\pm$  production data which are directly sensitive to the  $u_V$  and  $d_V$  distributions of the nucleus do not put stronger constraints on the valence-quark nuclear modification factors. This is due to the existing constraints from CHORUS neutrino–nucleus DIS data [47] included in the EPPS16 analysis, on top of which these additional constraints from  $W^\pm$  bosons are rather modest. For sea quarks,

we find the constraints at the parametrization scale to be negligible. This originates from the EW-scale quark sea being dictated by the parametrization-scale gluons. For this reason, at the parametrization scale, we obtain the largest constraints for the gluon nuclear modifications, consistent with those [35, 48, 49] from dijet and  $D^0$ -meson production measurements [34, 50]. We note that these hadronic observables favour a larger small- $x$  suppression compared to EPPS16, whereas the opposite is true for the  $W^\pm$  bosons. Everything is still consistent within uncertainties and the universality of nuclear PDFs appears to hold.

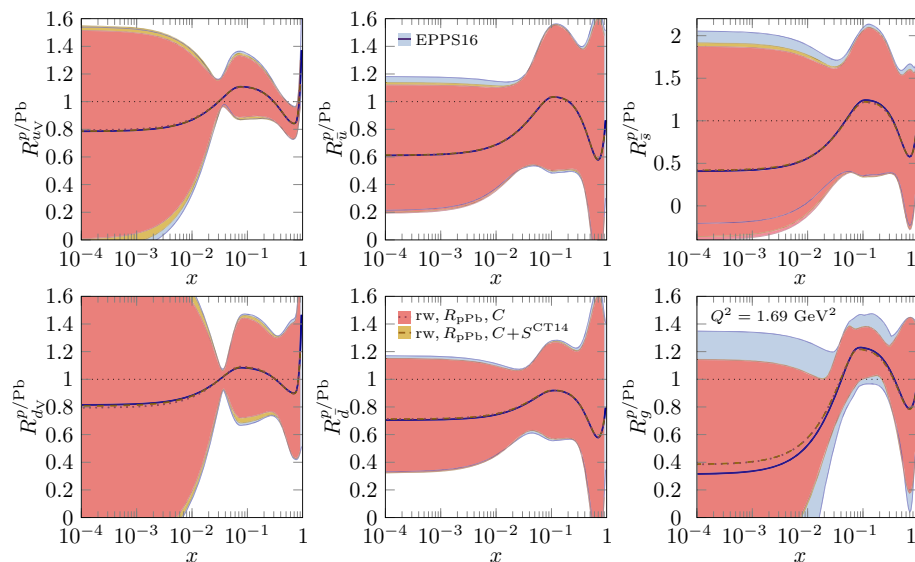
While the impact of the data on the valence modifications is not very large, there is still some dependence in the results on whether the proton-PDF uncertainties are taken into account or not. Based on the very large free-proton-PDF uncertainties we saw in Fig. 1, one could have expected these to have even larger impact in the reweighting. The smallness of the impact can be understood based on our observation that the dominant proton-PDF uncertainty in the absolute cross sections was normalization-like. Combined with the significant experimental luminosity uncertainty, it therefore has a minor impact on the nuclear modifications. However, while the difference in the “C” versus “C +  $S^{\text{CT14}}$ ” might not appear large, it should be noticed that including the proton-PDF uncertainties washes away a large part of the valence-quark flavour-separation constraints that we would have otherwise had from these data.

We have tested also the constraints obtainable from the different ratios discussed in Sect. 4. The results are shown for the nuclear-modification ratios in Fig. 15, and we discuss the impact with the other ratios below. Overall, the constraints from the nuclear-modification ratios are even surprisingly similar to those with absolute cross sections. What we have gained by reducing the free-proton uncertainties, we have lost by introducing additional experimental uncertainties from the p+p reference. Importantly though, we do not lose any constraining power in the process. This could be improved further if one would be able to cancel some of the systematical uncertainties in the ratio. Still, using the nuclear-modification ratios reduces the difference in the reweighted uncertainties between the “C” and “C +  $S^{\text{CT14}}$ ” extractions, the results being now almost identical. Also, while there is some pull on the central values of valence-quark nuclear modifications when using absolute cross sections, the size of which depends on whether the CT14 uncertainties are included, this effect vanishes with the nuclear modification ratio. Thus, with this observable and present data precision, the residual proton-PDF dependence in the nuclear-modification extraction is found to be small, conveniently for the global analyses. We observe also a slightly larger impact on the sea quarks compared to the absolute cross sections, but within the large uncertainties this is hardly significant. We emphasise that the nuclear-modification ratios were constructed here from

**Fig. 14** The results of reweighting the EPPS16 (blue bands and solid lines) nuclear modifications with the  $W^\pm$  absolute cross sections, both without (“C”, red bands and dotted lines) and with (“C +  $S^{CT14}$ ”, yellow bands and dashed lines) the CT14 uncertainties included



**Fig. 15** As Fig. 14, but now for reweighting with the nuclear-modification ratios from Fig. 9

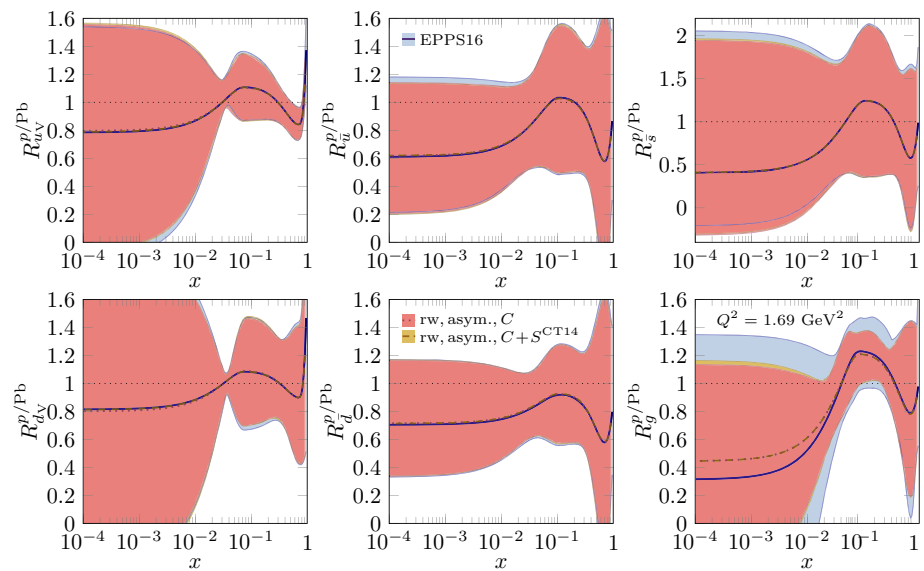


separate measurements of p+Pb and p+p, and we therefore cannot cancel systematical uncertainties in the ratio, which could have improved the impact further. A direct measurement of the nuclear-modification ratios with the future LHC Run 3 data would thus be most welcome.

The relative smallness of free-proton-PDF impact is found to extend also to nuclear-modification extraction with the other ratios. For the self-normalized cross sections, where we fit by leaving one point out as explained in Sect. 4.1, the results are almost identical to those in Fig. 15, with only marginally larger proton-PDF dependence due to the less direct cancellation in the ratio. For the forward-to-backward ratios and charge asymmetries the loss of information restricts the achievable constraints from individual observables. As a result, with the forward-to-backward ratios essentially all constraints on the valence flavour separation are lost, and the resulting effect is on the more poorly con-

strained sea-quarks and gluons, exactly as in the middle and right panels of Fig. 15. As we have discussed above, the traditional charge asymmetry  $\mathcal{A}_{pPb}$  does not provide strong free-proton-PDF cancellation, and as a result the constraints on the valence quarks are the same as with the absolute cross sections in Fig. 14, but due to the cancellation of nuclear effects in the forward region, practically all constraints on the gluons are lost. When used separately, the alternative asymmetries  $\tilde{\mathcal{A}}_{pPb}$  and  $\mathcal{A}_{pPb}^*$  have essentially the same effect as the forward-to-backward ratios, with the  $\mathcal{A}_{pPb}^*$  being slightly more constraining than  $\tilde{\mathcal{A}}_{pPb}$ , and constrain only the small- $x$  uncertainties which dominate in these observables. When used together, as shown in Fig. 16, these asymmetries begin to have some constraining power also on the valence quarks. The impact is small, but appears to be completely free from the proton-PDF uncertainties, as should be expected.

**Fig. 16** As Fig. 14, but now for reweighting with the alternative symmetries from Fig. 12



The impact on the valence-quark flavour separation appears to be limited mainly by the p+Pb statistical uncertainties. This will change after the LHC Run 3 data taking, where we expect a factor of 3–4 increase in the attainable statistics [51]. For a simple test, we have performed a reweighting with mock data with an altered covariance matrix  $\tilde{C}$ , where we take the current 8.16 TeV measurement and reduce the statistical uncertainties by a factor of two. As a result, the valence-quark nuclear-modification uncertainties become smaller when neglecting the free-proton uncertainties, but *not* when the latter are taken into account. This holds even for relatively free-proton-PDF insensitive observables such as the nuclear-modification ratios, the results for which are shown in Fig. 17. As the  $W^\pm$  boson measurements in p+p collisions are already dominated by systematical uncertainties (see Ref. [52] for discussion about the expected constraints from high-luminosity LHC) whereas in p+Pb there is still a significant statistical contribution to the uncertainty, it is plausible that the constraints on nuclear PDFs evolve faster than those on the proton PDFs (we found that changing CT14 to CT18 NLO PDFs did not have a significant effect on the free-proton uncertainties, but it should be noted that CT18 has relatively conservative uncertainty estimates compared to other contemporary analyses). The proton-PDF uncertainties can therefore in the worst case become even a limiting factor for extracting the bound-nucleon modifications in the future, and must then be propagated accordingly or suppressed with special ratios such as the alternative asymmetries discussed above.

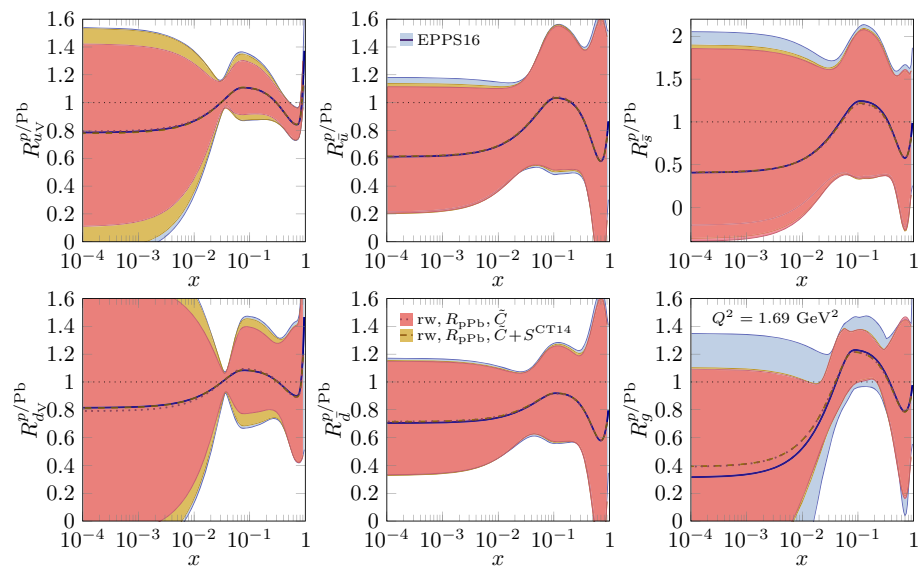
## 6 Summary and discussion

In this paper, we have systematically studied the importance of the free-proton-PDF uncertainties in extracting the nuclear

modifications of bound-nucleon PDFs from  $W^\pm$  production data. We have done this in the context of the most recent lepton-rapidity-differential  $W^\pm$  cross sections at 8.16 TeV [26], and various ratios constructed from the data. We have demonstrated that none of the considered ratios yield a perfect cancellation of the proton-PDFs uncertainties, with the exception of the alternative charge-asymmetry ratios discussed in Sect. 4.4. By using the methods of theoretical covariance matrices and Hessian PDF reweighting, we however find that the residual free-proton uncertainties in the ratios are small enough compared to the present data uncertainties that they do not pose significant bias in the nuclear-modification extraction.

The Hessian reweighting performed in this study also gives us information on what the impact of the 8.16 TeV  $W^\pm$  production data would have been, had they been included in the EPPS16 analysis. We find that, after a consistent propagation or a reduction of the proton-PDF uncertainties, the impact on the flavour separation of the nuclear modifications is rather mild, owing to the existing constraints from CHORUS neutrino-DIS data [47] included in the EPPS16 analysis, and the main new information is on the nuclear gluons, which were poorly constrained in EPPS16. Importantly, the new gluon-PDF constraints are consistent with those [35, 48, 49] from dijet and  $D^0$ -meson production measurements [34, 50], and the  $W^\pm$  production data also fully agrees with the flavour-separation constraints from the CHORUS data. This supports the universality of nuclear PDFs. The simple reweighting study performed here falls short in few aspects which we address in a concurrent global analysis [7] (see also the work in Refs. [13, 14]): First, to reliably quantify the full free-proton-PDF dependence of the nuclear modifications, it is necessary to include them in all the relevant fitted observables. And second, once the free-proton uncertainties are included in the nuclear-modification fitting,

**Fig. 17** As Fig. 15, but using mock data with reduced statistical uncertainties



one should find a way to consistently propagate these uncertainties into any desired observable (this problem is analogous to the one in evaluating scale uncertainties both in the PDF fits and predictions for observables, see Refs. [53,54]).

Even though the impact of the free-proton PDFs was found here to be small in the nuclear-PDF extraction, especially when using the experimental nuclear-modification ratios, this can radically change in the future with yet another 3–4 fold increase in the p+Pb statistics expected from the LHC Run 3. As argued in Sect. 5, the proton-PDF uncertainties can then become a significant source of uncertainty in the extraction of the nuclear modifications, which highlights the importance of understanding the large- $x$  nucleon structure [55,56] and accurate determination of the free-proton PDFs and their uncertainties also from the nuclear PDF point of view. If we no longer can take the (effective) bound-nucleon nuclear modifications to be independent of the free-nucleon PDFs themselves, this can have also significant consequences e.g. for the attempts to understand the physical cause behind the EMC effect [57]. In this case, special observables like the alternative asymmetries can prove to be useful in testing different models.

We stress that this study (in its present extent) would not have been possible without having access to the experimental data correlations. When we strive for an increased precision in the nuclear-PDF extraction with the upcoming LHC runs and future experiments, it becomes increasingly important to publish the full covariance matrices of the measurements or, even better, a full breakdown of the systematic-uncertainty contributions, source by source, optimally with the full information on the used statistical model published as well [58]. This would enable nuclear-PDF fits to use the maximal amount of information, with the possibility to even “cross calibrate” different measurements [59].

**Acknowledgements** We thank Francois Arleo for discussions concerning the asymmetry observables in  $W^\pm$  production. We acknowledge financial support from Xunta de Galicia (Centro singular de investigación de Galicia accreditation 2019-2022), European Union ERDF, the “María de Maeztu” Units of Excellence program MDM-2016-0692 and the Spanish Research State Agency, European Research Council project ERC-2018-ADG-835105 YoctoLHC (C.A.S. and P.P.), and from the Academy of Finland projects 308301 (H.P.) and 330448 (K.J.E. and P.P.). This research was funded as a part of the Center of Excellence in Quark Matter of the Academy of Finland (project 346325).

**Data Availability Statement** This manuscript has no associated data or the data will not be deposited. [Authors’ comment: The analysis output data are available from the authors upon request.]

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. Funded by SCOAP<sup>3</sup>.

## References

1. K.J. Eskola, V.J. Kolhinen, P.V. Ruuskanen, Nucl. Phys. B **535**, 351 (1998). [https://doi.org/10.1016/S0550-3213\(98\)00589-6](https://doi.org/10.1016/S0550-3213(98)00589-6)
2. K.J. Eskola, V.J. Kolhinen, C.A. Salgado, Eur. Phys. J. C **9**, 61 (1999). <https://doi.org/10.1007/s100520050513>
3. K.J. Eskola, V.J. Kolhinen, H. Paukkunen, C.A. Salgado, JHEP **05**, 002 (2007). <https://doi.org/10.1088/1126-6708/2007/05/002>
4. K.J. Eskola, H. Paukkunen, C.A. Salgado, JHEP **07**, 102 (2008). <https://doi.org/10.1088/1126-6708/2008/07/102>
5. K.J. Eskola, H. Paukkunen, C.A. Salgado, JHEP **04**, 065 (2009). <https://doi.org/10.1088/1126-6708/2009/04/065>



6. K.J. Eskola, P. Paakkinen, H. Paukkunen, C.A. Salgado, *Eur. Phys. J. C* **77**(3), 163 (2017). <https://doi.org/10.1140/epjc/s10052-017-4725-9>
7. K.J. Eskola, P. Paakkinen, H. Paukkunen, C.A. Salgado, EPPS21: a global QCD analysis of nuclear PDFs (2021). [arXiv:2112.12462](https://arxiv.org/abs/2112.12462) [hep-ph]
8. M. Hirai, S. Kumano, M. Miyama, *Phys. Rev. D* **64**, 034003 (2001). <https://doi.org/10.1103/PhysRevD.64.034003>
9. M. Hirai, S. Kumano, T.H. Nagai, *Phys. Rev. C* **70**, 044905 (2004). <https://doi.org/10.1103/PhysRevC.70.044905>
10. M. Hirai, S. Kumano, T.H. Nagai, *Phys. Rev. C* **76**, 065207 (2007). <https://doi.org/10.1103/PhysRevC.76.065207>
11. S. Atashbar Tehrani, *Phys. Rev. C* **86**, 064301 (2012). <https://doi.org/10.1103/PhysRevC.86.064301>
12. H. Khanpour, S. Atashbar Tehrani, *Phys. Rev. D* **93**(1), 014026 (2016). <https://doi.org/10.1103/PhysRevD.93.014026>
13. R. Abdul Khalek, J.J. Ethier, J. Rojo, *Eur. Phys. J. C* **79**(6), 471 (2019). <https://doi.org/10.1140/epjc/s10052-019-6983-1>
14. R. Abdul Khalek, J.J. Ethier, J. Rojo, G. van Weelden, *JHEP* **09**, 183 (2020). [https://doi.org/10.1007/JHEP09\(2020\)183](https://doi.org/10.1007/JHEP09(2020)183)
15. R.A. Khalek, R. Gauld, T. Giani, E.R. Nocera, T.R. Rabemananjara, J. Rojo, nNNPDF3.0: evidence for a modified partonic structure in heavy nuclei (2022). [arXiv:2201.12363](https://arxiv.org/abs/2201.12363) [hep-ph]
16. D. de Florian, R. Sassot, *Phys. Rev. D* **69**, 074028 (2004). <https://doi.org/10.1103/PhysRevD.69.074028>
17. D. de Florian, R. Sassot, P. Zurita, M. Stratmann, *Phys. Rev. D* **85**, 074028 (2012). <https://doi.org/10.1103/PhysRevD.85.074028>
18. M. Walt, I. Helenius, W. Vogelsang, *Phys. Rev. D* **100**(9), 096015 (2019). <https://doi.org/10.1103/PhysRevD.100.096015>
19. H. Khanpour, M. Soleymaninia, S. Atashbar Tehrani, H. Spiesberger, V. Guzey, *Phys. Rev. D* **104**(3), 034010 (2021). <https://doi.org/10.1103/PhysRevD.104.034010>
20. I. Schienbein, J.Y. Yu, K. Kovarik, C. Keppel, J.G. Morfin, F. Olness, J.F. Owens, *Phys. Rev. D* **80**, 094004 (2009). <https://doi.org/10.1103/PhysRevD.80.094004>
21. K. Kovarik et al., *Phys. Rev. D* **93**(8), 085037 (2016). <https://doi.org/10.1103/PhysRevD.93.085037>
22. A. Kusina et al., *Eur. Phys. J. C* **80**(10), 968 (2020). <https://doi.org/10.1140/epjc/s10052-020-08532-4>
23. E.P. Segarra et al., *Phys. Rev. D* **103**(11), 114015 (2021). <https://doi.org/10.1103/PhysRevD.103.114015>
24. P. Duventäster, L.A. Husová, T. Ježo, M. Klasen, K. Kovarik, A. Kusina, K.F. Muzakka, F.I. Olness, I. Schienbein, J.Y. Yu, *Phys. Rev. D* **104**(9), 094005 (2021). <https://doi.org/10.1103/PhysRevD.104.094005>
25. I. Helenius, M. Walt, W. Vogelsang, TUJU21: NNLO nuclear parton distribution functions with electroweak-boson production data from the LHC (2021). [arXiv:2112.11904](https://arxiv.org/abs/2112.11904) [hep-ph]
26. A.M. Sirunyan et al., *Phys. Lett. B* **800**, 135048 (2020). <https://doi.org/10.1016/j.physletb.2019.135048>
27. V. Khachatryan et al., *Phys. Lett. B* **750**, 565 (2015). <https://doi.org/10.1016/j.physletb.2015.09.057>
28. H. Paukkunen, C.A. Salgado, *JHEP* **03**, 071 (2011). [https://doi.org/10.1007/JHEP03\(2011\)071](https://doi.org/10.1007/JHEP03(2011)071)
29. S. Dulat, T.J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, C.P. Yuan, *Phys. Rev. D* **93**(3), 033006 (2016). <https://doi.org/10.1103/PhysRevD.93.033006>
30. T.J. Hou et al., *Phys. Rev. D* **103**(1), 014013 (2021). <https://doi.org/10.1103/PhysRevD.103.014013>
31. J.M. Campbell, R.K. Ellis, W.T. Giele, *Eur. Phys. J. C* **75**(6), 246 (2015). <https://doi.org/10.1140/epjc/s10052-015-3461-2>
32. G. D'Agostini, *Nucl. Instrum. Methods A* **346**, 306 (1994). [https://doi.org/10.1016/0168-9002\(94\)90719-6](https://doi.org/10.1016/0168-9002(94)90719-6)
33. R. Abdul Khalek, et al., *Eur. Phys. J. C* **79**(11), 931 (2019). <https://doi.org/10.1140/epjc/s10052-019-7401-4>
34. A.M. Sirunyan et al., *Phys. Rev. Lett.* **121**(6), 062002 (2018). <https://doi.org/10.1103/PhysRevLett.121.062002>
35. K.J. Eskola, P. Paakkinen, H. Paukkunen, *Eur. Phys. J. C* **79**(6), 511 (2019). <https://doi.org/10.1140/epjc/s10052-019-6982-2>
36. K.J. Eskola, H. Paukkunen, C.A. Salgado, *JHEP* **10**, 213 (2013). [https://doi.org/10.1007/JHEP10\(2013\)213](https://doi.org/10.1007/JHEP10(2013)213)
37. V. Khachatryan et al., *Eur. Phys. J. C* **76**(8), 469 (2016). <https://doi.org/10.1140/epjc/s10052-016-4293-4>
38. A. Kusina, T. Stavreva, S. Berge, F.I. Olness, I. Schienbein, K. Kovarik, T. Jezo, J.Y. Yu, K. Park, *Phys. Rev. D* **85**, 094028 (2012). <https://doi.org/10.1103/PhysRevD.85.094028>
39. F. Arleo, E. Chapon, H. Paukkunen, *Eur. Phys. J. C* **76**(4), 214 (2016). <https://doi.org/10.1140/epjc/s10052-016-4049-1>
40. A. Kusina, F. Lyonnet, D.B. Clark, E. Godat, T. Jezo, K. Kovarik, F.I. Olness, I. Schienbein, J.Y. Yu, *Eur. Phys. J. C* **77**(7), 488 (2017). <https://doi.org/10.1140/epjc/s10052-017-5036-x>
41. S. Bailey, T. Cridge, L.A. Harland-Lang, A.D. Martin, R.S. Thorne, *Eur. Phys. J. C* **81**(4), 341 (2021). <https://doi.org/10.1140/epjc/s10052-021-09057-0>
42. R.D. Ball, et al., The path to proton structure at one-percent accuracy (2021). [arXiv:2109.02653](https://arxiv.org/abs/2109.02653) [hep-ph]
43. H. Paukkunen, C.A. Salgado, *Phys. Rev. Lett.* **110**(21), 212301 (2013). <https://doi.org/10.1103/PhysRevLett.110.212301>
44. H. Paukkunen, P. Zurita, *JHEP* **12**, 100 (2014). [https://doi.org/10.1007/JHEP12\(2014\)100](https://doi.org/10.1007/JHEP12(2014)100)
45. C. Schmidt, J. Pumplin, C.P. Yuan, P. Yuan, *Phys. Rev. D* **98**(9), 094005 (2018). <https://doi.org/10.1103/PhysRevD.98.094005>
46. T.J. Hou, Z. Yu, S. Dulat, C. Schmidt, C.P. Yuan, *Phys. Rev. D* **100**(11), 114024 (2019). <https://doi.org/10.1103/PhysRevD.100.114024>
47. G. Onengut et al., *Phys. Lett. B* **632**, 65 (2006). <https://doi.org/10.1016/j.physletb.2005.10.062>
48. K.J. Eskola, I. Helenius, P. Paakkinen, H. Paukkunen, *JHEP* **05**, 037 (2020). [https://doi.org/10.1007/JHEP05\(2020\)037](https://doi.org/10.1007/JHEP05(2020)037)
49. A. Kusina, J.P. Lansberg, I. Schienbein, H.S. Shao, *Phys. Rev. Lett.* **121**(5), 052004 (2018). <https://doi.org/10.1103/PhysRevLett.121.052004>
50. R. Aaij et al., *JHEP* **10**, 090 (2017). [https://doi.org/10.1007/JHEP10\(2017\)090](https://doi.org/10.1007/JHEP10(2017)090)
51. Z. Citron, et al., CERN Yellow Rep. Monogr. **7**, 1159 (2019). <https://doi.org/10.23731/CYRM-2019-007.1159>
52. R. Abdul Khalek, S. Bailey, J. Gao, L. Harland-Lang, J. Rojo, *Eur. Phys. J. C* **78**(11), 962 (2018). <https://doi.org/10.1140/epjc/s10052-018-6448-y>
53. L.A. Harland-Lang, R.S. Thorne, *Eur. Phys. J. C* **79**(3), 225 (2019). <https://doi.org/10.1140/epjc/s10052-019-6731-6>
54. R.D. Ball, R.L. Pearson, *Eur. Phys. J. C* **81**, 830 (2021). <https://doi.org/10.1140/epjc/s10052-021-09602-x>
55. A. Accardi, W. Melnitchouk, J.F. Owens, M.E. Christy, C.E. Keppel, L. Zhu, J.G. Morfin, *Phys. Rev. D* **84**, 014008 (2011). <https://doi.org/10.1103/PhysRevD.84.014008>
56. J. Dudek et al., *Eur. Phys. J. A* **48**, 187 (2012). <https://doi.org/10.1140/epja/i2012-12187-1>
57. J. Arrington, N. Fomin, *Phys. Rev. Lett.* **123**(4), 042501 (2019). <https://doi.org/10.1103/PhysRevLett.123.042501>
58. K. Cranmer, et al., *SciPost Phys.* **12**, 037 (2022). <https://doi.org/10.21468/SciPostPhys.12.1.037>
59. I. Helenius, J. Lajoie, J.D. Osborn, P. Paakkinen, H. Paukkunen, *Phys. Rev. D* **100**(1), 014004 (2019). <https://doi.org/10.1103/PhysRevD.100.014004>