Impact of heavy quark and quarkonium data on nuclear gluon PDFs

P. Duwentäster,[1,*] T. Ježo,[1,†] M. Klasen,[1] K. Kovařík,[1] A. Kusina,[2] K. F. Muzakka,[1] F. I. Olness,[3] R. Ruiz,[2] I. Schienbein,[4] and J. Y. Yu[4]

[1]*Institut für Theoretische Physik, Westfälische Wilhelms-Universität Münster, Wilhelm-Klemm-Straße 9, D-48149 Münster, Germany*
[2]*Institute of Nuclear Physics Polish Academy of Sciences, PL-31342 Krakow, Poland*
[3]*Southern Methodist University, Dallas, Texas 75275, USA*
[4]*Laboratoire de Physique Subatomique et de Cosmologie, Université Grenoble-Alpes, CNRS/IN2P3, 53 avenue des Martyrs, 38026 Grenoble, France*

(Received 28 April 2022; accepted 13 June 2022; published 27 June 2022)

A clear understanding of nuclear parton distribution functions (nPDFs) plays a crucial role in the interpretation of collider data taken at the Relativistic Heavy Ion Collider, the Large Hadron Collider (LHC), and in the near future at the Electron-Ion Collider. Even with the recent inclusions of vector boson and light meson production data, the uncertainty of the gluon PDF remains substantial and limits the interpretation of heavy ion collision data. To obtain new constraints on the nuclear gluon PDF, we extend our recent nCTEQ15WZ + SIH analysis to inclusive quarkonium and open heavy-flavor meson production data from the LHC. This vast new data set covers a wide kinematic range and puts strong constraints on the nuclear gluon PDF down to $x \lesssim 10^{-5}$. The theoretical predictions for these data sets are obtained from a data-driven approach, where proton-proton data are used to determine effective scattering matrix elements. This approach is validated with detailed comparisons to existing next-to-leading order calculations in nonrelativistic QCD for quarkonia and in the general-mass variable-flavor-number scheme for the open heavy-flavored mesons. In addition, the uncertainties from the data-driven approach are determined using the Hessian method and accounted for in the PDF fits. This extension of our previous analyses represents an important step toward the next generation of PDFs not only by including new data sets, but also by exploring new methods for future analyses.

DOI: 10.1103/PhysRevD.105.114043

I. INTRODUCTION

Parton distribution functions (PDFs) are fundamental quantities required to calculate predictions for processes involving hadronic initial states. The underlying theoretical framework is based on factorization theorems, which have been proven from first principles of QCD for a number of collider processes in ep and pp collisions [1]. This formalism provides both a field theoretical definition of the PDFs, and the definition of the short distance hard scattering cross sections at the partonic level. Additionally, it includes a statement about the error of the collinear factorization formula, which is inversely proportional to some power of the hard scale of the process. The predictive power of this formalism lies in the fact that the PDFs are universal, i.e., process independent, whereas the process dependent short distance cross sections can be systematically calculated in perturbation theory. This approach has been widely used in analyses of proton PDFs, which have been constrained with great precision [2–14]. Assuming that the twist-2 collinear factorization remains valid also in the case of eA and pA collisions, nuclear parton distribution functions (nPDFs) have been determined [6,15–30] as well, albeit with significantly larger uncertainties compared to the case of proton PDFs. In particular, the poorly constrained nuclear gluon PDF has been the focus of recent nCTEQ studies [15,31,32] and studies by other groups [16,18,24,26–28], but unfortunately the uncertainties have remained substantial. Through DGLAP evolution, the gluons also produce significant uncertainties on other flavors. Even the vector boson production and single inclusive hadron production data sets, that were included in our recent nCTEQ15WZ(+SIH) studies [31,32], do not constrain the gluon below $x \approx 10^{-3}$.

In this paper, we perform a global analysis of nuclear PDFs in the nCTEQ framework, including open heavy flavor meson and heavy quarkonium data. Heavy quarkonium and open heavy-flavor meson production data (in the

*pit.duw@uni-muenster.de
†tomas.jezo@uni-muenster.de

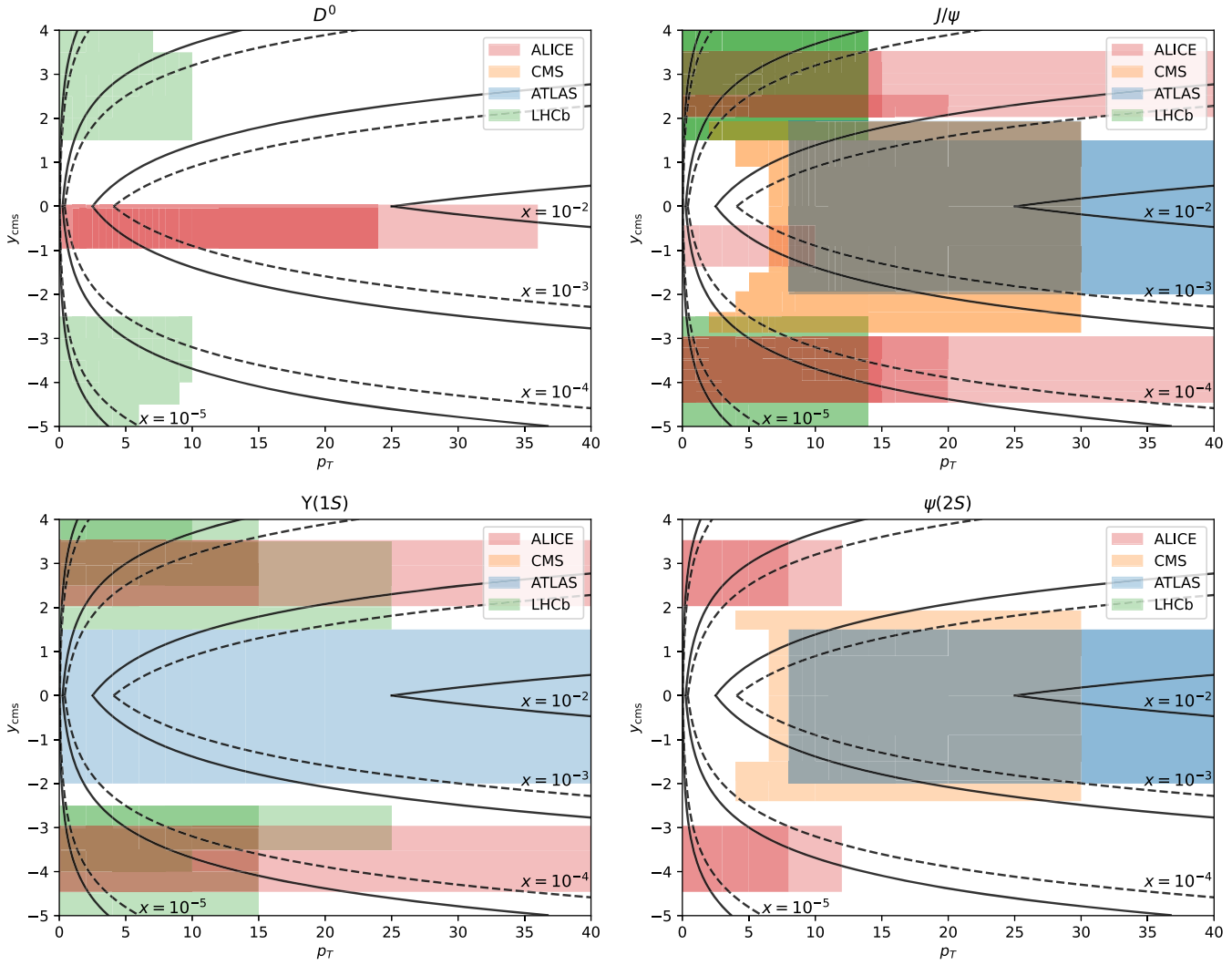

FIG. 1. Coverage of the kinematic (p_T, y_{cms})-plane of the quarkonium and open heavy quark production data sets from proton-lead collisions. ALICE data is shown in red, ATLAS in blue, CMS in orange, and LHCb in green. The dashed and solid contours show the estimated x dependence for $\sqrt{s} = 5$ and 8 TeV, respectively.

following collectively called heavy quark production, HQ) has the potential to yield new constraints on the gluon PDF, because the gluon-gluon channel contributes the dominant part to the overall cross section of these processes. This was first shown in the case of proton PDFs in Ref. [33], before the reweighting study presented in Ref. [34] was the first to demonstrate that the LHC heavy-quark data is also useful to constrain the nuclear gluon PDF at small x. The x dependence of the data can be estimated at leading order as

$$x \approx \frac{2p_T}{\sqrt{s}} \exp(-|y|), \tag{1}$$

where y is the rapidity in the center-of-mass frame.[1] Figure 1 shows the kinematic coverage of the available data with contours for the corresponding x dependence according to the estimate above. This shows the data to be sensitive to x values below 10^{-5} for the most forward or backward rapidity at low p_T.

D-meson production data in particular has been used in recent nuclear PDF analyses to reduce the uncertainty on the gluon PDF [24,28,35]. In Ref. [35], a reweighting study was performed by comparing pQCD predictions in the GMVFNS [36–38] to the ratio of double differential D-meson production cross sections between proton-proton and proton-lead data. Reference [24] then included the same data directly in the global analysis. In contrast, Ref. [28] compares the same data to predictions in a fixed-flavor number scheme with POWHEG and Pythia8 and uses this to perform a Bayesian reweighting of the PDFs. The aforementioned studies also include dijet data taken by CMS [39], which should provide additional constraints on the gluon PDF.

[1]All rapidities mentioned in the following are in the center-of-mass frame and are therefore denoted simply by y.

In this investigation we focus on the study of single inclusive[2] production of open heavy-flavor mesons and heavy quarkonia in proton-proton and proton-lead collisions. Including these processes in a global PDF fit is not as straight-forward as others, since there is no universally accepted theoretical model for quarkonia. Therefore we introduce a data-driven approach, that relies only on the following assumptions:

(i) the gluon-gluon channel contributes the dominant fraction of the total cross section,
(ii) it is sufficient to focus on subprocesses with $2 \to 2$ kinematics and subprocesses with more than two hard final state particles can be neglected.

If these assumptions are valid, the data-driven approach has considerable advantages over the other available calculations in perturbative QCD. Most importantly, the studied processes can be accurately described across a large kinematic region with well-controlled uncertainties. A secondary advantage is the speed of the calculation, which is significantly faster than the available full pQCD calculations [40]. While perturbative calculations can be used in global analysis with efficient gridding, as has been done in Ref. [24], the data driven approach still provides a faster calculation, which is very convenient in the lengthy process of a global fit.

The focus of this study will be on incorporating this process into a new global analysis, while accounting for the uncertainty of the data-driven theory, to determine the nuclear gluon PDF with greater accuracy than previously possible. A precise knowledge of the nuclear PDFs is important for the following reasons: Firstly, they provide a description of the hadronic or nuclear structure in terms of quark and gluon degrees of freedom. In the context of the standard pQCD formalism the PDFs are universal and therefore required to make predictions for a wide range of collider observables. Finally, they provide a starting point for comparisons with microscopic models predicting the nuclear modifications (at the twist-2 level) in different x regions. This includes microscopic models for nuclear effects on PDFs in the shadowing region [41–44], the antishadowing region [44–46], or the EMC effect [44,47–51]. At small x and moderately hard scales, the density of gluons becomes very large such that the assumptions underlying collinear factorization are expected to break down. This kinematic region is described by the theory of Color Glass Condensates (CGC) [52,53] and there are also promising unified approaches which interpolate between the CGC at small x and collinear factorization at large x, see, e.g., Ref. [54] and references therein. Nevertheless, it is fair to say that for now there is no unambiguous microscopic picture of the inner workings of heavier nuclei.

It should be stressed again, that throughout this paper, our main underlying assumption is that the twist-2 collinear factorization remains valid also in the case of eA and pA collisions for the same observables. As it has been discussed in Refs. [55,56] this is reasonable, even if higher twist terms may be enhanced in the nuclear case up to higher hard scales ($\propto A^{1/3}$). We impose kinematic cuts on the data to effectively reduce the impact of these higher twist effects and confirm phenomenologically that all remaining data is well described. In the future, such higher twist effects could be modelled to extend the reach towards data with lower hard scales. One example is the effects due to fully coherent energy loss [57–59]. These contributions are formally higher twist (twist 3), but have been shown to be relevant for hard process data up to moderately large transverse momenta $p_T \approx 10$ GeV. It could therefore be interesting to include such effects in future global analyses, however more work would be needed both on the conceptual and the phenomenological side.

The next section provides an overview of the nCTEQ framework and the integration of the new data-driven approach. Following that, we perform and evaluate the fit of the proton-proton baseline for the theory in Sec. III. In Sec. IV we present the fits obtained using the HQ data and evaluate the compatibility between the new and old data. Finally, in Sec. V we summarize our findings and give an outlook for future work.

II. THEORETICAL APPROACH

A. The nCTEQ framework

The nCTEQ project expands upon the foundation of the proton PDF global fitting analysis by including the nuclear dimension. In early proton PDF analyses (e.g., Ref. [60]), the nuclear data was used to calculate correction factors which were then applied to the proton PDF fit without any uncertainties. In contrast, the nCTEQ framework provides full nuclear PDFs from dedicated global analyses with uncertainties given in the Hessian framework [61].

The details of the nCTEQ15 nPDFs are presented in Ref. [15]. The current analysis, along with the other recent nCTEQ analyses, such as nCTEQ15WZ [31], nCTEQ15HIX [62], and nCTEQ15WZ + SIH [32], is performed with a new C++-based code nCTEQ++. This allows us to easily interface external programs such as HOPPET [63], APPLgrid [64], and INCNLO [65]. In particular, we work at leading twist and next-to-leading order (NLO) of QCD for both the PDF and FF evolution equations as well as the hard-scattering coefficients. The calculation code for the quarkonia and open heavy quarks is a partial C++ adaption of HELAC-Onia 2.0 [66] and uses the data-driven

[2]The meaning of the word "inclusive" can be somewhat ambiguous in this context. Here, we use it to denote processes, where the hadron is produced together with an arbitrary number of other particles, as opposed to exclusive production. This is the case for all processes studied in this paper. Later we will therefore use the word "inclusive" to denote the sum of prompt and nonprompt production of a particle.

approach explained in Sec. II B instead of a pQCD calculation.

For the fits in this investigation, we use the same 19 parameters as for the nCTEQ15WZ(+SIH) sets. These 19 parameters include the 16 free parameters of the nCTEQ15 analysis, with an additional three open parameters for the strange distribution. For the nCTEQ15 set, the strange PDF was constrained by the relation $s = \bar{s} = (\kappa/2)(\bar{u} + \bar{d})$ at the initial scale $Q_0 = 1.3$ GeV, which forces it into the same form as the other sea quarks.

Our PDFs are parametrized at the initial scale $Q_0 = 1.3$ GeV as

$$xf_i^{p/A}(x, Q_0) = c_0 x^{c_1}(1-x)^{c_2} e^{c_3 x}(1 + e^{c_4} x)^{c_5}, \quad (2)$$

and the nuclear A dependence is encoded in the coefficients as

$$c_k \to c_k(A) \equiv p_k + a_k(1 - A^{-b_k}), \quad (3)$$

where $k = \{1, \ldots, 5\}$. The 16 free parameters used for the nCTEQ15 set describe the x dependence of the $\{g, u_v, d_v, \bar{d} + \bar{u}\}$ PDF combinations, and we do not vary the $\bar{d}/\bar{u}$ parameters; see Ref. [15] for details. As in the nCTEQ15WZ(+SIH) analysis, we have added three strange PDF parameters, $\{a_0^{s+\bar{s}}, a_1^{s+\bar{s}}, a_2^{s+\bar{s}}\}$; these parameters correspond to the nuclear modification of the overall normalization, the low-x exponent and the large-x exponent of the strange quark distribution, respectively.

In total, the 19 open parameters are

$$\{a_1^{u_v}, a_2^{u_v}, a_4^{u_v}, a_5^{u_v}, a_1^{d_v}, a_2^{d_v}, a_5^{d_v}, a_1^{\bar{u}+\bar{d}}, a_5^{\bar{u}+\bar{d}}, a_1^g, a_4^g, a_5^g, b_0^g, b_1^g, b_4^g, b_5^g, a_0^{s+\bar{s}}, a_1^{s+\bar{s}}, a_2^{s+\bar{s}}\}.$$

All the fixed parameters are kept as they were in nCTEQ15. This includes the proton PDF baseline, which is fixed to the parameters obtained in Ref. [67]. While the nCTEQ++ integrates both the proton and nuclear degrees of freedom into the same framework, in the present study we only vary the nuclear degrees of freedom. Although recent progress has significantly reduced the nuclear PDF uncertainties, these still dominate the proton uncertainties. For example, Ref. [68] has performed a comprehensive study investigating the interplay of the nuclear and proton uncertainties for the case of W production in pPb collisions. Using the methods of theoretical covariance matrices and Hessian PDF reweighting, they find that impact of the free-proton PDF uncertainties is small for the current data sets. However, given the improved statistics anticipated from the LHC Run 3, they expect it will be essential to include the proton uncertainties for future analyses.

B. The data-driven approach

Instead of performing the cross section calculations of the heavy mesons in perturbative QCD, we take the data-driven approach outlined initially in Ref. [69] and used for a reweighting study in Refs. [34,70]. In this approach, the cross section for two nuclei A and B scattering and producing a quarkonium or open heavy-flavor meson $\mathcal{Q}$ is calculated as the convolution integral of the two initial state gluon PDFs $f_{1,g}(x_1, \mu)$, $f_{2,g}(x_2, \mu)$ and a fitted effective scattering matrix element $\overline{|\mathcal{A}_{gg\to\mathcal{Q}+X}|^2}$ over the $AB \to \mathcal{Q}$ phase space

$$\sigma(AB \to \mathcal{Q} + X) = \int \mathrm{d}x_1\, \mathrm{d}x_2 f_{1,g}(x_1, \mu) f_{2,g}(x_2, \mu) \frac{1}{2\hat{s}} \overline{|\mathcal{A}_{gg\to\mathcal{Q}+X}|^2} \mathrm{dPS}.$$

The effective scattering matrix element is parametrized with the Crystal Ball function,

$$\overline{|\mathcal{A}_{gg\to\mathcal{Q}+X}|^2} = \frac{\lambda^2 \kappa \hat{s}}{M_{\mathcal{Q}}^2} e^{a|y|} \times \begin{cases} e^{-\kappa \frac{p_T^2}{M_{\mathcal{Q}}^2}} & \text{if } p_T \le \langle p_T \rangle \\ e^{-\kappa \frac{\langle p_T \rangle^2}{M_{\mathcal{Q}}^2}} \left(1 + \frac{\kappa}{n} \frac{p_T^2 - \langle p_T \rangle^2}{M_{\mathcal{Q}}^2}\right)^{-n} & \text{if } p_T > \langle p_T \rangle \end{cases}, \quad (4)$$

where the five parameters[3] λ, κ, $\langle p_T \rangle$, n, and a are then fitted for each final state $\mathcal{Q}$. We have introduced the fifth parameter a, which was not present in the original parametrization [71], to allow for a more accurate reproduction of the rapidity dependence [72]. The parameters are then fitted to $pp \to \mathcal{Q} + X$ data. Once the optimal parameters are found, we can also determine the uncertainty of our Crystal Ball fit via the same Hessian method used to calculate our PDF uncertainties. We can then account for these uncertainties by adding them in quadrature to the systematic uncertainties of the $pPb \to \mathcal{Q} + X$ data. The included final states in this analysis are D^0, J/ψ, $\Upsilon(1S)$, and $\psi(2S)$ mesons. Note, however, that prompt and nonprompt production of the same particle need to be considered as two different final states. Inclusive production is generally not fitted separately, but calculated as the sum of the other two. The exception to this is $\Upsilon(1S)$, where all available data is for inclusive production. Other final states, like $D^{\pm}$ or higher excitations of Υ, are excluded due to

[3]The parameter name "$\langle p_T \rangle$" is somewhat misleading. The parametrization was initially invented for a different purpose, where this parameter did have the physical meaning of the particle's average transverse momentum, but this interpretation is lost in the current context. However, we decided to keep the name to keep consistency with previous works.

TABLE I. Scale choices for the different particles.

	D^0	J/ψ	$B \to J/\psi$	$\Upsilon(1S)$	$\psi(2S)$	$B \to \psi(2S)$
μ_0^2	$4M_D^2 + p_{T,D}^2$	$M_{J/\psi}^2 + p_{T,J/\psi}^2$	$4M_B^2 + \frac{M_B^2}{M_{J/\psi}^2} p_{T,J/\psi}^2$	$M_{\Upsilon(1S)}^2 + p_{T,\Upsilon(1S)}^2$	$M_{\psi(2S)}^2 + p_{T,\psi(2S)}^2$	$4M_B^2 + \frac{M_B^2}{M_{\psi(2S)}^2} p_{T,\psi(2S)}^2$

insufficient data, but in principle the approach can be used for any measurement of single inclusive hadrons, as long as sufficient proton-proton data is available for the baseline and the cross section is dominated by the gluon-gluon channel.

The default scales $\mu = \mu_0$ that enter the PDFs are chosen as they were in the previous reweighting study, as shown in Table I. The entire procedure is also repeated once with the scale doubled and once with the scale halved to obtain an estimate of the impact that this choice has on the final result.

III. PROTON-PROTON BASELINE

In the first step of the analysis, we use the proton-proton data sets listed in Tables III–VI to determine the parameters of the Crystal Ball function. The largest portion of the proton-proton data comes from J/ψ production, followed by $\Upsilon(1S)$, then $\psi(2S)$, and finally D^0.

A. Cuts and excluded data

Both in the baseline fit and the PDF fit, we cut all heavy-quark data with $p_T < 3$ GeV and outside of the rapidity range $-4 < y_{\mathrm{cms}} < 4$. These cuts allow predictions with reasonable χ^2 for all remaining data points. Relaxing the cuts introduces data points with $\chi^2/N_{\mathrm{d.o.f.}} = \mathcal{O}(10)$. We also exclude the lowest p_T bin of the 2011 CMS J/ψ data sets, because the corresponding paper mentions that there may be large acceptance effects in this bin that are not included in the uncertainties. The lowest rapidity bin of the 7 TeV $\Upsilon(1S)$ production data set from LHCb is included in the fit, but the normalization is determined separately from the rest of the data set. This y-bin contains 20 p_T bins (17 after cuts), which qualitatively agree with the remaining fit, but the normalization is off by 25%. Therefore, keeping the normalization of this bin the same as the remainder of the data set causes the $\chi^2/N_{\mathrm{d.o.f.}}$ for the entire $\Upsilon(1S)$ fit to increase from 0.92 to 1.6. This particular bin is also not described by other models based on the color-octet mechanism [73].

We also limit the study to LHC data only, as the available RHIC data taken at very different energies is not guaranteed to work with the same fit parameters and would also likely not provide additional strong constraints.

B. Baseline fit

A comparison of the fits with the data is shown in Figs. 11–16. The uncertainties of the fit are determined via the same Hessian method that is used in our PDF fits. All predictions show very close agreement with the data across the region included in the fit. The large number of data points produces very small uncertainties for J/ψ and $\Upsilon(1S)$ (not visible on the logarithmic scale). While the uncertainties of D^0 and $\psi(2S)$ are somewhat larger, due to the lower number of data points, they are still small compared to the experimental uncertainties.

The obtained parameters and values of $\chi^2/N_{\mathrm{d.o.f.}}$ for each process are given in Table II. The parameters for J/ψ and $\psi(2S)$ are each obtained in a combined fit with prompt, nonprompt, and inclusive data, which results in only one value of $\chi^2/N_{\mathrm{d.o.f.}}$ for each. There is overall very good agreement between the data and the fitted theory with $\chi^2/N_{\mathrm{d.o.f.}}$ values slightly below one for J/ψ, $\psi(2S)$ and $\Upsilon(1S)$ and a particularly low $\chi^2/N_{\mathrm{d.o.f.}}$ value of 0.25 obtained for D^0 production.

Alternative baseline fits have been performed with a variety of parametrizations including more degrees of freedom, like the extended Crystal Ball function [90], which has a polynomial tail on both sides of the Gaussian for a total of seven parameters. In those fits, there is no significant improvement in the description of the currently included data and relaxing the cuts still leads to unreasonable χ^2 values for the data points that would be

TABLE II. Crystal Ball parameters and χ^2/d.o.f. values for the Crystal Ball function for the different processes.

	D^0	J/ψ	$B \to J/\psi$	$\Upsilon(1S)$	$\psi(2S)$	$B \to \psi(2S)$
κ	0.33457	0.47892	0.15488	0.94524	0.21589	0.45273
λ	1.82596	0.30379	0.12137	0.06562	0.07528	0.13852
$\langle p_T \rangle$	2.40097	5.29310	−7.65026	8.63780	8.98819	7.80526
n	2.00076	2.17366	1.55538	1.93239	1.07203	1.64797
a	−0.03295	0.02816	−0.08083	0.22389	−0.10614	0.06179
N_{points}	34	501		375	55	
$\chi^2/N_{\mathrm{d.o.f.}}$	0.25	0.88		0.92	0.77	

TABLE III. Overview of the available $pp \to D^0 + X$ production data sets and their number of data points.

Group	Year	Ref.	ID	Type	Points after/before cuts
ALICE	2012	[74]	3008	Prompt	7/9
LHCb	2013	[75]	3014	Prompt	22/38
ALICE	2016	[76]	3021	Prompt	7/10

TABLE IV. Overview of the available $pp \to J/\psi + X$ production data sets and their number of data points. Note that the ALICE 2017 data set is split into two for technical reasons (one part is taken at 5 TeV and one at 13 TeV).

Group	Year	Ref.	ID	Type	Points after/before cuts
ATLAS	2011	[77]	3003	Prompt	63/64
CMS	2011	[78]	3001	Prompt	39/44
LHCb	2011	[79]	3002	Prompt	50/66
CMS	2017	[80]	3020	Prompt	50/52
ATLAS	2018	[81]	3017	Prompt	33/33
CMS	2010	[82]	3007	Nonprompt	11/14
ATLAS	2011	[77]	3006	Nonprompt	63/64
CMS	2011	[78]	3005	Nonprompt	39/44
LHCb	2011	[79]	3004	Nonprompt	50/66
CMS	2017	[80]	3019	Nonprompt	50/52
ATLAS	2018	[81]	3018	Nonprompt	33/33
ALICE	2015	[83]	3022	Inclusive	10/13
ALICE	2017	[84]	3023	Inclusive	8/11
ALICE	2017	[84]	3024	Inclusive	14/18
ALICE	2019	[85]	3016	Inclusive	4/7

TABLE V. Overview of the available $pp \to \Upsilon(1S) + X$ production data sets and their number of data points. Note that the LHCb 2015 data set is split into two for technical reasons (one half is taken at 7 TeV and one at 8 TeV).

Group	Year	Ref.	ID	Type	Points after/before cuts
ATLAS	2012	[86]	3012	Inclusive	88/100
LHCb	2012	[87]	3025	Inclusive	55/75
CMS	2013	[88]	3013	Inclusive	30/42
ALICE	2014	[89]	3009	Inclusive	3/5
LHCb	2015	[73]	3011	Inclusive	89/109
LHCb	2015	[73]	3015	Inclusive	89/109
ALICE	2015	[83]	3031	Inclusive	3/5
ATLAS	2017	[81]	3031	Inclusive	18/24

TABLE VI. Overview of the available $pp \to \psi(2S) + X$ production data sets and their number of data points.

Group	Year	Ref.	ID	Type	Points after/before cuts
ATLAS	2017	[81]	3026	Prompt	15/15
ATLAS	2017	[81]	3027	Nonprompt	15/15
ALICE	2015	[83]	3028	Inclusive	6/9
CMS	2018	[104]	3029	Inclusive	9/10
ALICE	2017	[84]	3030	Inclusive	9/12

introduced. Therefore, we keep the parametrization as written in Eq. (4) as a good compromise between a reasonable number of parameters and accuracy of the data description.

The baseline fit naturally introduces a dependence on the used proton PDF, but the fit has also been repeated with different proton PDFs and in all cases very similar Crystal Ball fits were found.

C. Comparison with J/ψ production in nonrelativistic QCD

The predictions for J/ψ production from the data-driven method can also be compared to predictions made using nonrelativistic QCD (NRQCD). The NRQCD framework assumes the speed of the heavy quarks to be slow compared to the speed of light to derive a factorization theorem where the physics is separated into short- and long-distance factors [91]. The short-distance physics, i.e., the production of the heavy quark-antiquark pair, can then be calculated perturbatively, while the nonperturbative long distance matrix elements parametrizing the formation of the bound state need to be determined empirically. This has been done by various independent groups including Ma *et al.* [92,93] and Butenschoen *et al.* [94,95]. The latter group has provided us with predictions for prompt J/ψ production using their NLO calculation and our nCTEQ15 proton PDF, which is shown in Fig. 2. The uncertainties shown for NRQCD are calculated by varying all scales together by a factor of two around their default values $\mu_{r,0} = \mu_{f,0} = \sqrt{p_T^2 + 4m_c^2}$ and $m_{\mathrm{NRQCD},0} = m_c = 1.5$ GeV. There is a very good agreement between the two methods across the entire kinematic range, but the Crystal Ball method produces significantly smaller uncertainties. Varying the scales in the Crystal Ball fit produces only negligible differences, as the scale dependence is mostly absorbed into the parameters of the effective matrix element.

D. Comparison with D^0 production in the GMVFNS

The predictions for D^0 production can also be compared with perturbative calculations. These calculations can be carried out using the general-mass variable-flavor number scheme (GMVFNS) implementation of heavy quark production at NLO QCD by Kniehl *et al.* [37,38]. Figure 3 shows a comparison of the predictions obtained from the GMVFNS code, with those from our Crystal Ball fit for the data sets used in the fit for all the $pp \to D^0 + X$ data used in the baseline fit. In the input of GMVFNS we use the nCTEQ15 proton PDF set, we set the c quark mass to $m_c = 1.3$ GeV, and the renormalization and the initial/final factorization scales to $\mu_r = \mu_i = \mu_f = \sqrt{p_T^2 + 4m_c^2}$. As a fragmentation function we use the one with identifier 712 from the KKKS08 set of fragmentation functions [99] which was obtained in a global fit to Belle, CLEO, ALEPH, and OPAL data. The uncertainties of the GMVFNS

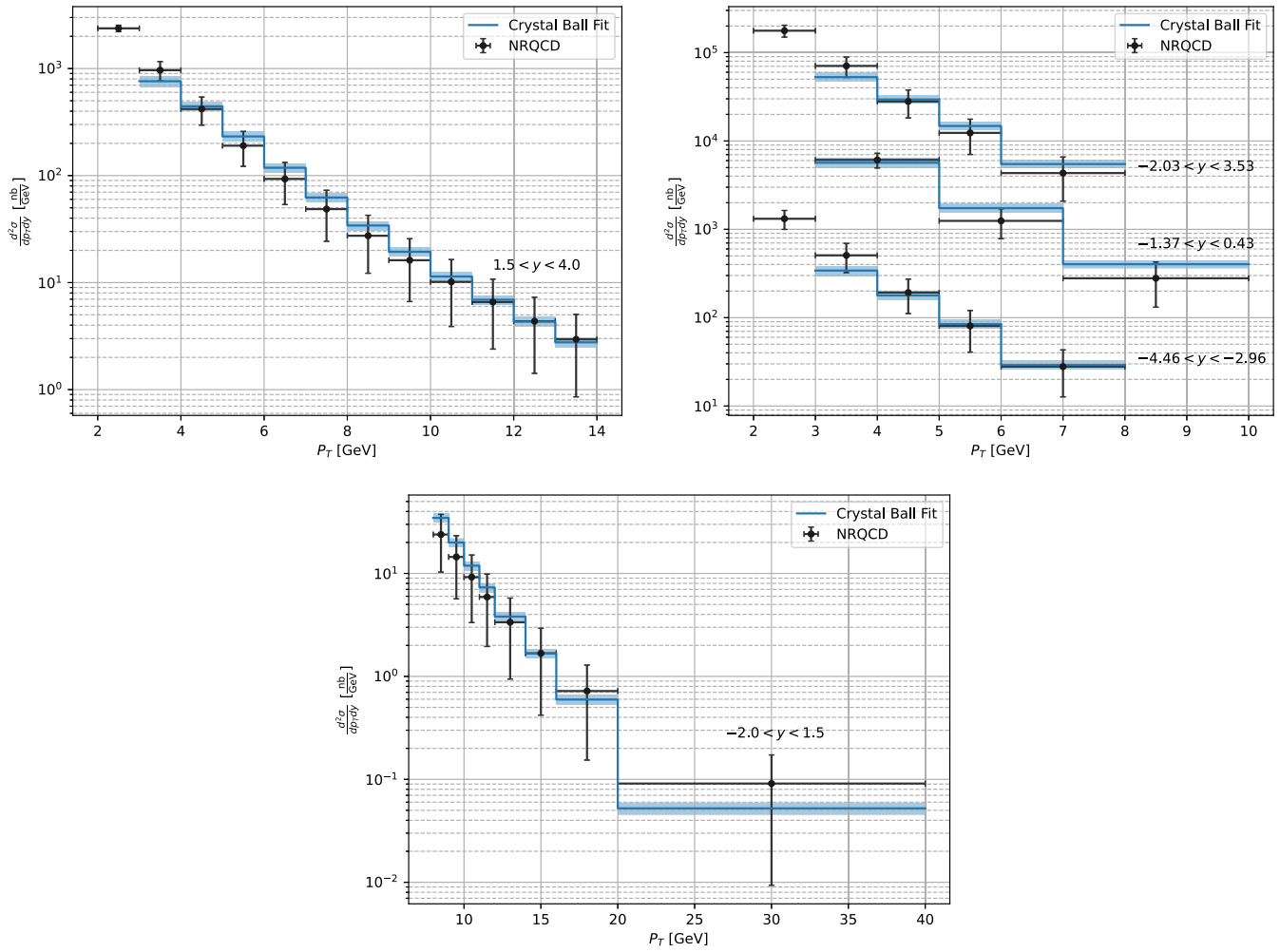

FIG. 2. Comparison between prompt J/ψ production in pp collisions for LHCb[96], ALICE[97], and ATLAS[98] kinematics as predicted by NRQCD and with the data-driven approach. The uncertainties of the NRQCD predictions come from scale variation $1/2 < \mu_r/\mu_{r,0} = \mu_f/\mu_{f,0} = \mu_{\text{NRQCD}}/\mu_{\text{NRQCD},0} < 2$ around the base scale $\mu_{r,0} = \mu_{f,0} = \sqrt{p_T^2 + 4m_c^2}$ and $m_{\text{NRQCD},0} = m_c$. Different rapidity bins are separated by multiplying the cross sections by powers of ten for visual clarity.

predictions are obtained by varying the three scales individually by a factor of two, such that there is never a factor four between two scales. These uncertainties are similar in size as the data uncertainty, except for the low-p_T region, where they are somewhat larger. Overall the central prediction of the GMVFNS calculation slightly overshoots the data. This can perhaps be attributed to the contribution from largely unconstrained gluon component of the fragmentation function, which contributes at almost 50%. However, there is always overlap between the data and GMVFNS theory uncertainty. The uncertainty of the Crystal Ball fit is similar in size as the GMVFNS one for large p_T, but contrary to the latter it decreases for lower p_T values. The central values are very close to the data points, as indicated by the low $\chi^2/N_{\text{d.o.f.}}$ value seen in Table II. Overall the two methods are in very good agreement with only minor discrepancies seen in the highest p_T bins. It should be noted, however, that Ref. [35] shows that the x-dependence of the GMVFNS and the Crystal Ball predictions do not match exactly. They show that the Crystal Ball approach depends on a more narrow x-region, but the increased flexibility in our updated Crystal Ball parametrization should mitigate this to some extent. Furthermore, constraints from the quarkonium data dominates over the ones from open heavy flavor data like D-meson production. At the same time in case of quarkonium data we do not have reliable perturbative calculations and we need to resort to the data driven method which further limits the impact of using the data driven or perturbative calculation for D meson data.

We also compared our Crystall Ball fit and GMVFNS predictions against more recent data of D^0 production in pp collisions from ALICE [100,101] and LHCb [102], which we have not been used in the present analysis. We do not show the comparisons here, but we report that both the Crystall Ball fit as well as GMVFNS reproduce the ALICE

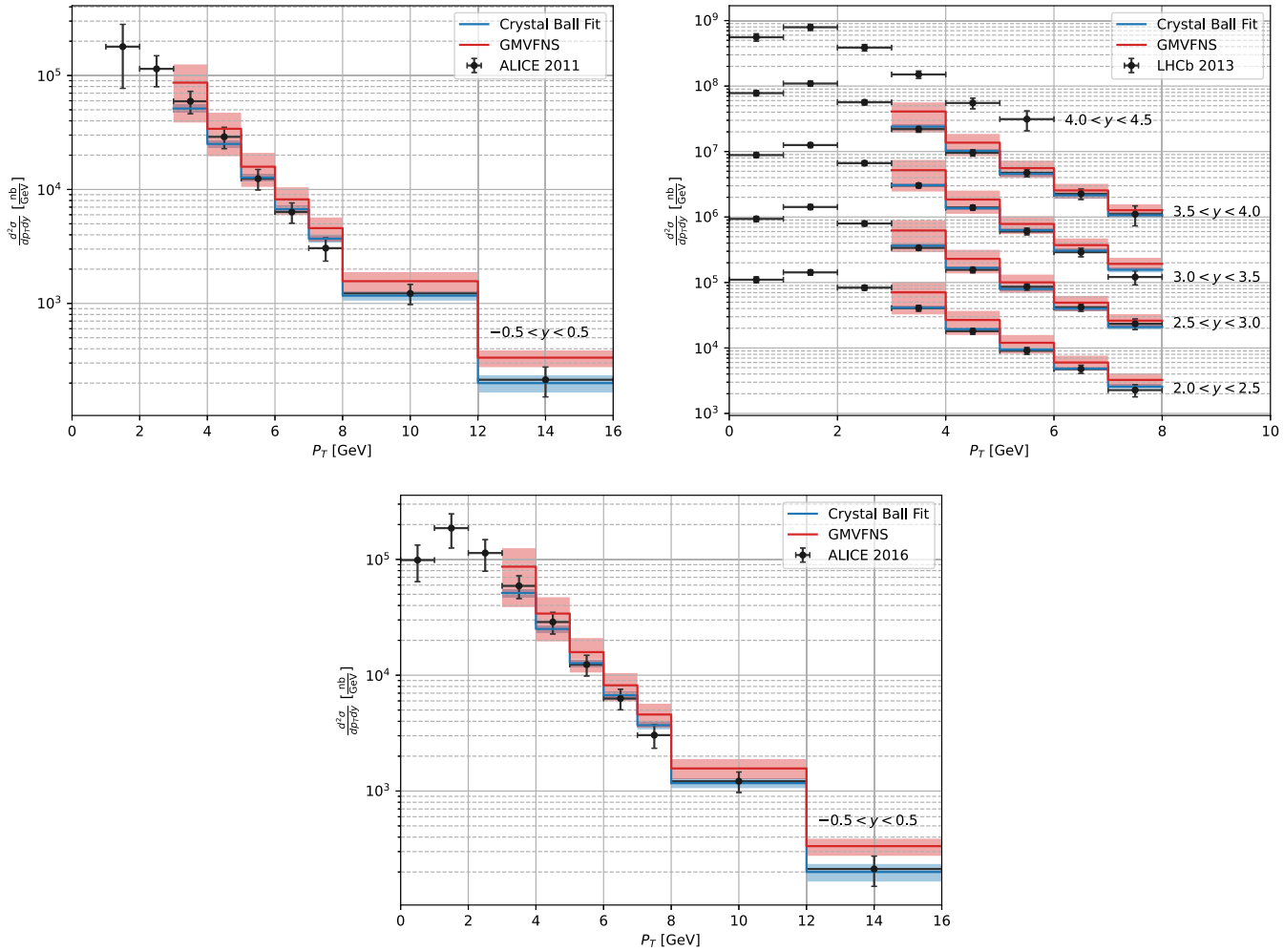

FIG. 3. Comparison between prompt D^0 production in pp collisions as predicted in the GMVFNS (red) and with the data-driven approach (blue). The uncertainties of the GMVFNS predictions come from varying the scales individually by a factor of 2, such that there is never a factor of 4 between the two scales. Different rapidity bins are separated by multiplying the cross sections by powers of ten for visual clarity.

and LHCb data well. This data could provide further constraints on the D^0 Crystall Ball parameters.

IV. IMPACT OF HQ DATA ON nPDF FITS

Using the Crystal Ball parameters determined in the previous section we can now perform a new global nPDF fit using the available heavy-quark data. The new fits are using the same framework as nCTEQ15WZ + SIH, including all settings like open parameters, scales and cuts for the previously included data. The Hessian tolerance is also kept at a value of $T = 35$. We do not include the changes made for nCTEQ15HIX [62] and nCTEQ15ν [103] as these developments are mostly orthogonal to those made in this study and do not affect the low-x gluon PDF. One minor change from the previous analyses is the treatment of normalizations. Previously, χ^2 penalties were assigned individually for each affected data set, whereas now they are applied only once per normalization parameter. This means that the data sets with fully correlated normalization uncertainties, e.g., the Run I ATLAS Z and $W^{\pm}$ production data sets now share a single penalty, instead of adding it multiple times.

A. Data selection

We add the heavy-quark data sets shown in Tables VII–X to the new PDF fit for a total of 1484 (548 new, 936 old) data points. Similar to the fragmentation function uncertainties of the SIH data in Ref. [32], we can compensate for the theoretical uncertainty of the data-driven approach by adding the uncertainty from the Crystal Ball fit as a systematic uncertainty to all new data sets.

For the new HQ data, we use the same cuts as in the proton-proton baseline and additionally exclude D^0 data points with $p_T > 15$ GeV, because there is no baseline data. Furthermore, we remove two individual points from the 2018 LHCb $\Upsilon(1S)$ data set that are described very

TABLE VII. Overview of the available $pPb \to D^0 + X$ production data sets and their number of data points.

Group	Year	Ref.	ID	Type	Points after/before cuts
ALICE	2014	[105]	3101	Prompt	8/10
ALICE	2016	[76]	3123	Prompt	8/11
LHCb	2017	[106]	3102	Prompt	53/92
ALICE	2019	[107]	3122	Prompt	13/21

TABLE VIII. Overview of the available $pPb \to J/\psi + X$ production data sets and their number of data points.

Group	Year	Ref.	ID	Type	Points after/before cuts
LHCb	2013	[108]	3108	Prompt	25/40
ATLAS	2015	[98]	3118	Prompt	10/10
LHCb	2017	[96]	3105	Prompt	88/140
CMS	2017	[80]	3120	Prompt	51/53
ATLAS	2018	[81]	3117	Prompt	8/8
LHCb	2013	[108]	3107	Nonprompt	25/40
ATLAS	2015	[98]	3119	Nonprompt	10/10
LHCb	2017	[96]	3106	Nonprompt	88/140
CMS	2017	[80]	3121	Nonprompt	51/53
ATLAS	2018	[81]	3116	Nonprompt	8/8
ALICE	2013	[109]	3103	Inclusive	0/12
ALICE	2015	[110]	3104	Inclusive	10/25
ALICE	2018	[97]	3112	Inclusive	9/24

TABLE IX. Overview of the available $pPb \to \Upsilon(1S) + X$ production data sets and their number of data points.

Group	Year	Ref.	ID	Type	Points after/before cuts
ALICE	2014	[111]	3110	Inclusive	0/4
LHCb	2014	[112]	3111	Inclusive	0/2
ATLAS	2018	[81]	3109	Inclusive	6/8
LHCb	2018	[113]	3113	Inclusive	36/66
ALICE	2019	[114]	3114	Inclusive	3/10

TABLE X. Overview of the available $pPb \to \psi(2S) + X$ production data sets and their number of data points.

Group	Year	Ref.	ID	Type	Points after/before cuts
ALICE	2014	[115]	3127	Inclusive	2/8
ALICE	2020	[116]	3126	Inclusive	3/10
ATLAS	2017	[81]	3124	Prompt	8/8
CMS	2018	[104]	3115	Prompt	17/17
ATLAS	2017	[81]	3125	Nonprompt	8/8

poorly with χ^2 values of 66 and 26, respectively. Both points are at the high-p_T edge of the experiment's kinematic range, which makes systematic errors a likely explanation, since the remaining 36 data points of the set can be well described.

To estimate the impact of the scale choice on the final PDFs, we repeat the entire procedure, including the proton-proton baseline fit, with the scale of the heavy quark production processes set to $\frac{1}{2}$ or 2 times their regular value.

B. Resulting PDFs

Figure 4 shows a comparison of the new fit, labeled nCTEQ15HQ,[4] with the PDFs obtained in previous nCTEQ analyses at a scale of $Q = 2$ GeV. The same comparison is shown in Fig. 5 in terms of nuclear modification factors. The central value of the gluon PDF retains a similar shape as the nCTEQ15WZ + SIH fit in the $x > 10^{-3}$ region. Below this point, the nCTEQ15WZ + SIH fit starts diverging towards higher values, while the new fit moves towards similar values as is nCTEQ15WZ. From $x \approx 0.2$ downwards the uncertainties of the new gluon PDF are reduced significantly compared to the previous analyses, particularly below $x < 10^{-4}$, where the previous uncertainties start to increase rapidly. The ratio plots underline that there are no significant changes to the up- and down-quark distributions, but there is a minor reduction in u_{valence} around $x \approx 0.1$. The uncertainties of the new fit still include the central values of all previous fits. The gluon ratio on the other hand shows a reduction in the nuclear modification for $x > 0.008$. In this region, the gluon ratio is mostly compatible with unity, but the central value goes from 0.8 at the lowest end of this region to 1.2 at its peak and then down to 0.6 at very high x. At lower x values, however, the reduced uncertainties clearly show a suppression of the gluon in lead between 20% and 40% compared to the proton case. The modification of the strange quark is similar to the other quark flavors in the new fit, but the uncertainties are still larger than the modification for all $x < 0.3$. The central values of the new up and down quark PDFs are mostly very close to those of nCTEQ15WZ + SIH, with a minor downward shift at low x and reduced uncertainties in the same region due to the better constraints on the gluon. The strange quark PDF stays similar to previous fits in the $x > 0.01$ region, but reverts to a low value similar to the original nCTEQ15, while nCTEQ15WZ and particularly nCTEQ15WZ + SIH start increasing. It is important to note however, that the relative uncertainty of the strange quark is still larger than 100% at low x, so that it is impossible to draw any strong conclusions from this analysis. We refer the reader to our recent neutrino analysis [103] for complementary information.

The impact of different scale choices is shown in Fig. 6, where each flavor i is shown as a ratio over the corresponding flavor i_{central} from the fit with the central scale choice. The central values of the up, down and gluon PDFs

[4]Since nCTEQ15WZ + SIH + HQ would be too long, we break this naming convention and shorten it to nCTEQ15HQ. This does not imply removal of the WZ and SIH data.

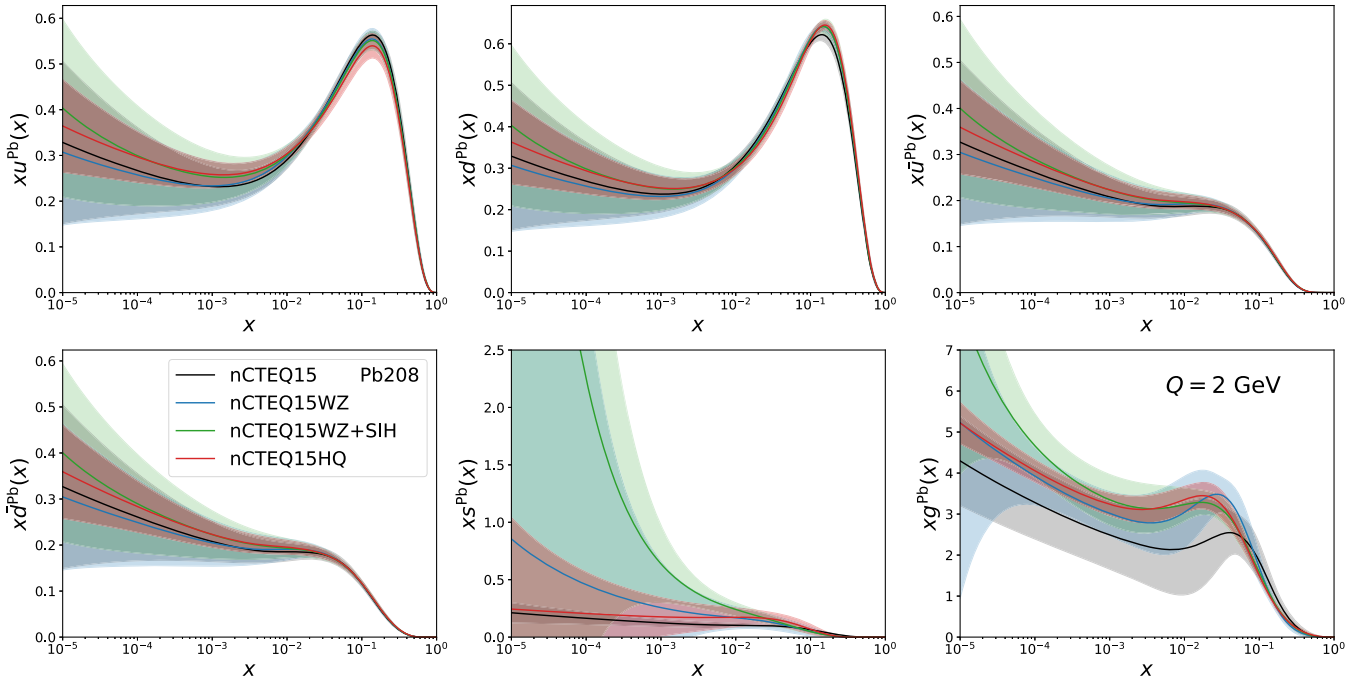

FIG. 4. Lead PDFs from different nCTEQ15 versions. The baseline nCTEQ15 fit is shown in black, nCTEQ15WZ in blue, nCTEQ15WZSIH in green, and the new fit in red.

with modified scales lie close together with differences no larger than a few percent. The only exception to this is the high-x region, where the gluon PDF vanishes and therefore the relative uncertainties blow up. The gluon PDFs cross each other at $x = 0.02$ and $x = 0.15$, with the lower scale being above the central scale in the high- and low-x regions and the upper scale showing the opposite behavior. The PDFs extracted with the modified scale stay well within the uncertainties of the regular fit and show slightly larger uncertainties themselves, which are likely due to a slightly

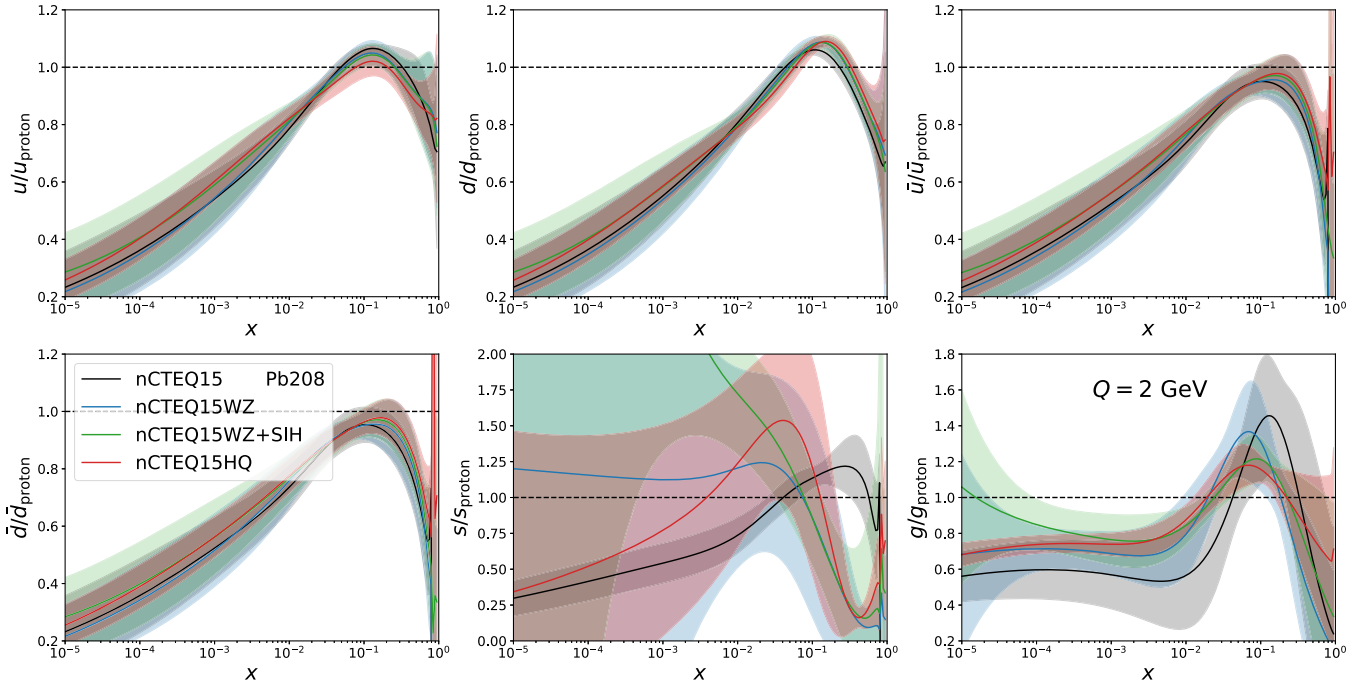

FIG. 5. Ratio of lead and proton PDF from different nCTEQ15 versions. The baseline nCTEQ15 fit is shown in black, nCTEQ15WZ in blue, nCTEQ15WZSIH in green, and the new fit in red.

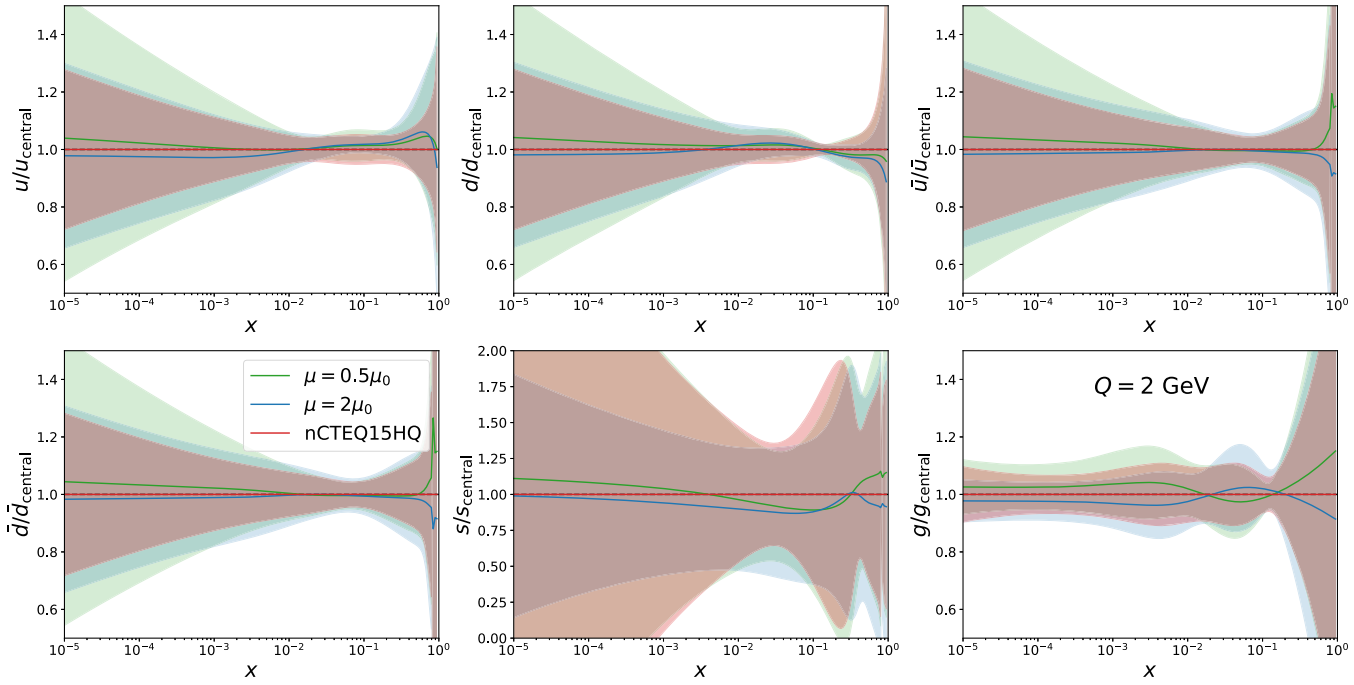

FIG. 6. Comparisons of fits, where the scale for heavy quark production is varied by a factor two around the central value μ_0.

worse fit. The strange quark sees variations of up to 15%, but given the very large uncertainties, this is still well within expectations. This conclusion is somewhat different from the one reached in the previous reweighting analysis [34,70], where the scale choice resulted in significant differences. Two updates in the methodology are responsible for this change: Firstly, the fact that normalizations of the data sets are now included as nuisance parameters in the fit, mitigates the scale dependence due to the large normalization uncertainties on many data sets. Secondly, the more restrictive kinematic cuts remove the data that is most sensitive to the scale choice.

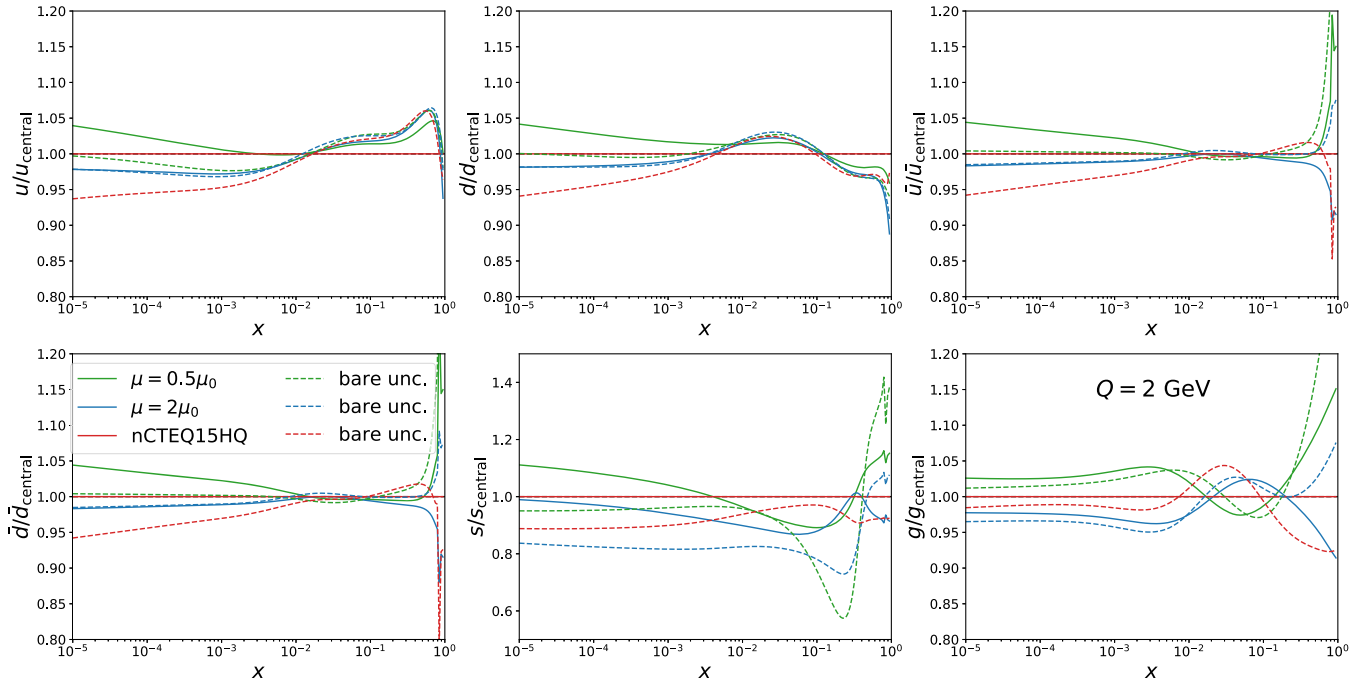

FIG. 7. Comparisons of PDF central values with Crystal Ball uncertainties included in the fit (solid lines) and without theory uncertainties (dashed lines).

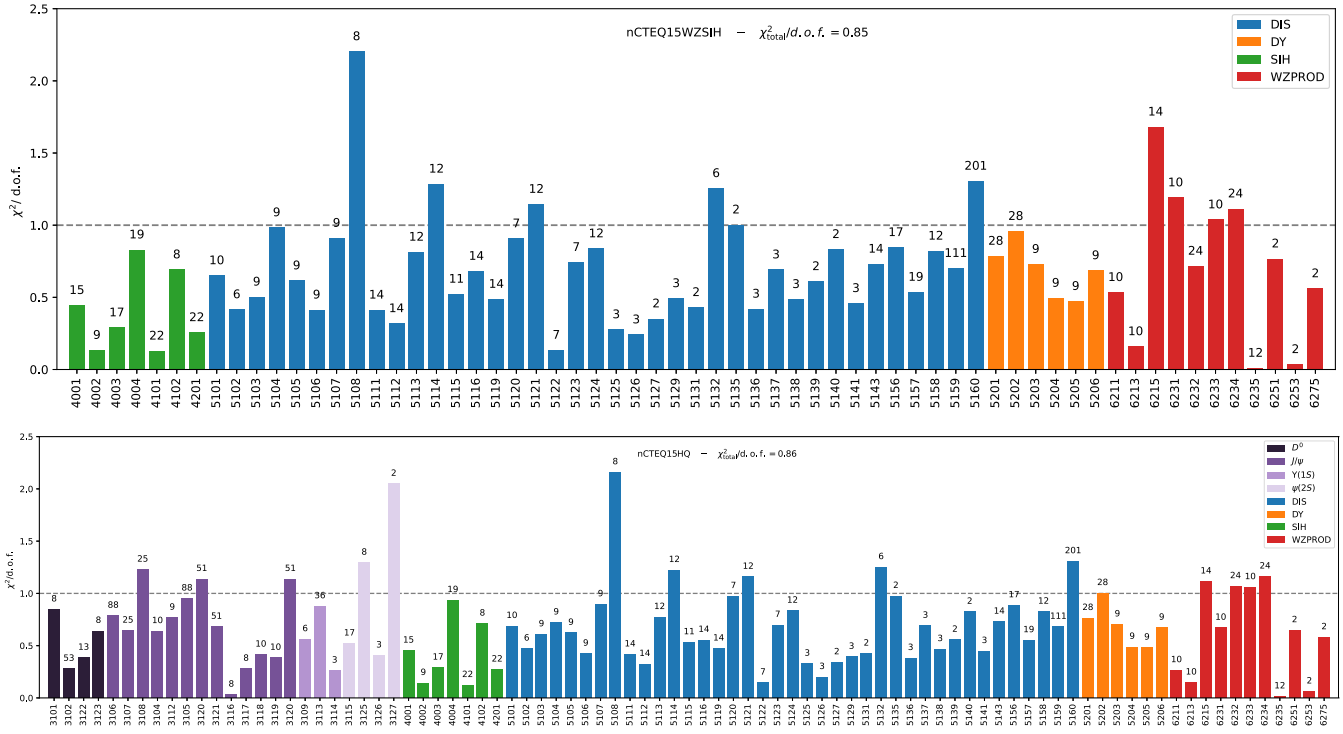

FIG. 8. $\chi^2/N_{\text{d.o.f.}}$ values for each data set in the previous nCTEQ15WZ + SIH fit (upper panel) and the new nCTEQ15HQ fit (lower panel).

A final set of alternative fits is shown in Fig. 7, where the previous three fits are compared to equivalent fits where the heavy-quark data sets keep their bare uncertainties, i.e., the uncertainties from the Crystal Ball fit are not added to the systematic uncertainties. For the up- and down-quark flavors this leads to shifts on a similar order as the scale variation, while the changes in the gluon are even smaller. The changes in the strange quark PDF are somewhat larger due to it's weakly constrained nature. The difference between the dashed and solid lines gives an estimate of the potential enhancement of the impact from the heavy-quark data if further proton-proton data is added to the baseline to reduce the Crystal Ball uncertainty.

C. Fit quality

Figures 18–21 show a comparison of the proton-lead data with predictions from the new nCTEQ15HQ PDFs including PDF uncertainties. For all four particle species, there is very close agreement across the entire kinematic range and the PDF uncertainties are barely visible. Note however, that the uncertainties from the Crystal Ball fit, which are not shown here, would still be of the same size as in the baseline predictions in Figs. 11–16.

For a more quantitative evaluation of the fit quality, one can take a look at Fig. 8, which shows the $\chi^2/N_{\text{d.o.f.}}$ for each data set of the previous nCTEQ15WZ + SIH and the new nCTEQ15HQ fits in the upper and lower panels, respectively. The comparison shows no significant rise in χ^2 for any of the established data sets. The χ^2 of the outlier 6215 (ATLAS Run I, Z production) among the vector boson production data sets is even reduced due to the new normalization treatment. This means that there are no incompatibilities between the previous and the new data sets. The new heavy-quark data sets themselves mostly show $\chi^2/N_{\text{d.o.f.}}$ values below or around one with a distribution similar to that of the DIS data sets. The only major outlier among the new sets is the $\psi(2S)$ production from a 2014 measurement by ALICE with just two data points.

Table XI shows a comparison of the $\chi^2/N_{\text{d.o.f.}}$ values for each process of the new fit and the preceding three nCTEQ15 generations. The original nCTEQ15 is the only set of PDFs that does not give a good description of the heavy-quark data, especially for J/ψ mesons. nCTEQ15WZ already displays significant improvement for D^0 and J/ψ, thus giving a reasonable $\chi^2/N_{\text{d.o.f.}}$ of 0.92 for the new data sets. nCTEQ15WZ + SIH retains a very similar overall χ^2, but does worse than nCTEQ15WZ for D^0 and $\Upsilon(1S)$, while improving the description of the J/ψ data. Finally, the new nCTEQ15HQ fit gives a good χ^2 for all processes with only minor increases for DIS and WZ production compared to both nCTEQ15WZ and nCTEQ15WZ + SIH, which are outweighed by a significant improvement in the large sample of new heavy-quark data. Interestingly, the $\chi^2/N_{\text{d.o.f.}}$ value for $\psi(2S)$ production barely changes between the four fits, but since it is the smallest of the new data sets with large

TABLE XI. $\chi^2/N_{\text{d.o.f.}}$ values for the individual heavy-quark final states, the individual processes DIS, DY, WZ, SIH, HQ, and the total. The shown χ^2 is the sum of regular χ^2 and normalization penalty. Excluded processes are shown in parentheses. Note that both nCTEQ15 AND nCTEQ15WZ included the neutral pions from STAR and PHENIX.

	D^0	J/ψ	$\Upsilon(1S)$	$\psi(2S)$	DIS	DY	WZ	SIH	HQ	Total
nCTEQ15	(0.56)	(2.50)	(0.82)	(1.06)	0.86	0.78	(2.19)	(0.78)	(1.96)	**1.23**
nCTEQ15WZ	(0.32)	(1.04)	(0.76)	(1.02)	0.91	0.77	0.63	(0.47)	(0.92)	**0.90**
nCTEQ15WZ + SIH	(0.46)	(0.84)	(0.90)	(1.07)	0.91	0.77	0.72	0.40	(0.93)	**0.92**
nCTEQ15HQ	0.35	0.79	0.79	1.06	0.93	0.77	0.78	0.40	0.77	**0.86**

theory uncertainties from the baseline fit, the small impact is to be expected.

To estimate the impact of the new data on the PDF parameters, it is instructive to look at the $\Delta\chi^2$ profiles along specific parameters. Figure 9 shows parameter scans for all of the seven gluon parameters included in the fit with the $\Delta\chi^2$ split into processes. It is immediately evident that the new heavy-quark data is the dominant constraint on most of the parameters with the WZ production always constraining one side, because its minimum is somewhat off center. This indicates some slight tensions with the WZ production data, but looking back at Table XI this tension largely comes from the SIH data and is still well within expectations. The only parameter that is not constrained mostly by vector boson and heavy quark production is b_0^g. This parameter determines the A dependence of the gluon normalization and minor shifts can therefore be compensated through the normalization parameters of the data sets. Additionally, since parameters with b indices are related to the nuclear A dependence, they may cancel out when looking only at a singular A as is the case in the LHC data, which is only taken in proton-lead collisions.

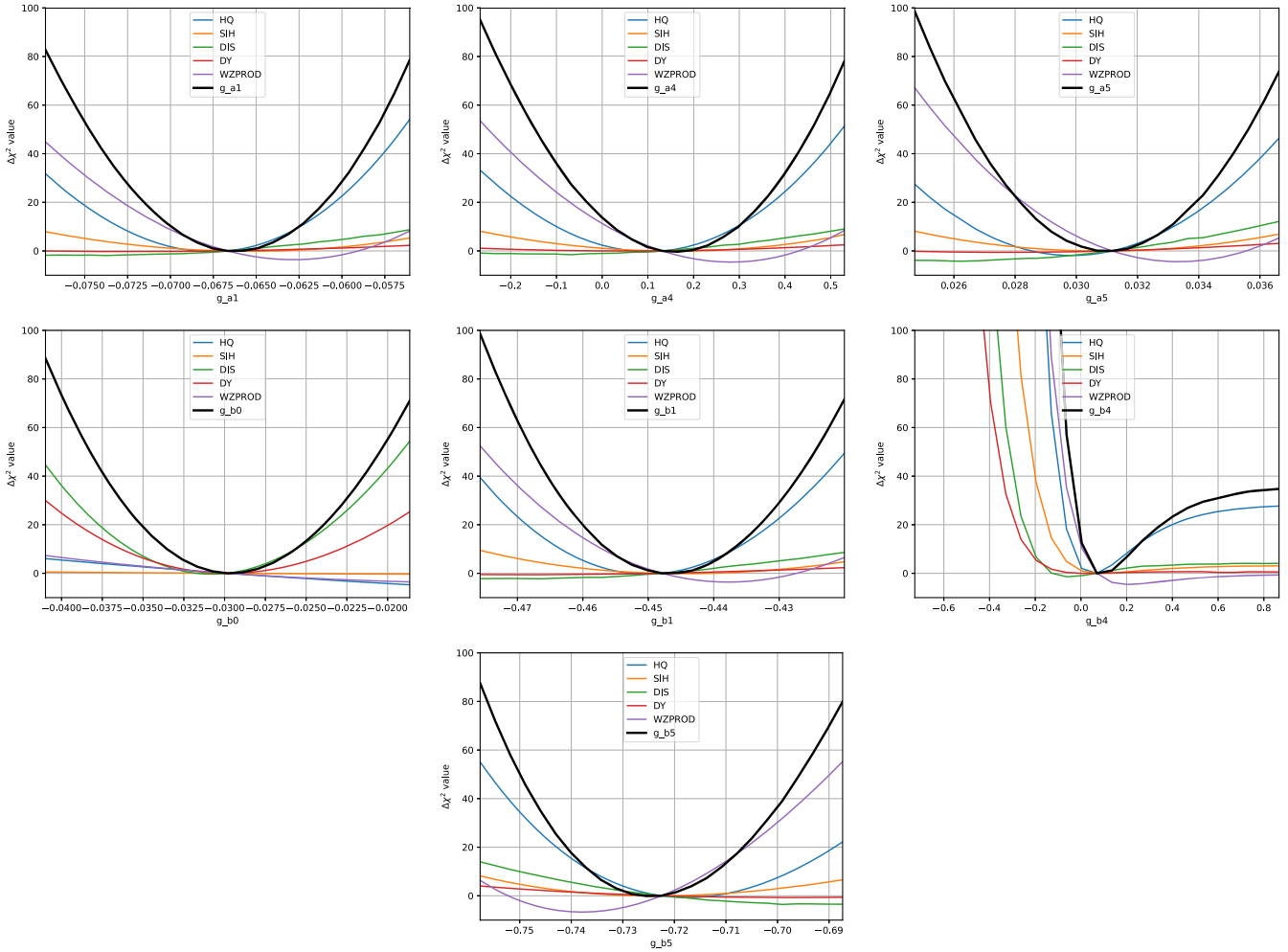

FIG. 9. Scans along the seven open gluon parameters, divided by experiment types with the total shown in black.

V. CONCLUSIONS

In conclusion, we have incorporated a large new data set of heavy quarkonium and open heavy-flavor production in the nCTEQ++ framework and extended our PDF analysis of the gluon to significantly lower x values than was previously possible.

We employed a data-driven approach to determine the theory and investigated the advantages and limitations of this approach in detail. In particular, we determined the kinematic range where the approach is applicable and verified the predictions with those from rigorous pQCD calculations for J/ψ and D^0 production.

We obtained good $\chi^2/N_{\text{d.o.f.}}$ values for the new data sets without compromising those of the established sets. The new data has a tremendous impact on the gluon PDF, especially in the region $x < 0.01$, and lowers the uncertainty considerably. Through the DGLAP evolution, this impact on the gluon also has consequences for the quark PDFs at low x, which remain at similar central values, but also experience a reduction of their uncertainties. Like the recent nNNPDF3.0 [28] and EPPS21 [24] analyses with D-meson production data, which use production of dijets instead of quarkonia as an additional constraint on the gluon PDF, we find similar qualitative results with strong evidence for gluon shadowing at low x.

The data-driven approach has proven to be both reliable with a reasonable estimation of its uncertainties, and also a powerful tool for the difficult task of constraining the nuclear gluon PDF. Therefore, the presented analysis and the PDFs obtained therein showcase a considerable advancement on the path towards a precise understanding of nuclear structure. Future studies can build upon the present work in a variety of ways. Most notably, the proton PDF baseline will be updated and the related uncertainties can be taken into account. Secondly, it would be interesting to use perturbative calculations in future analyses of the D-meson data to compare the resulting PDFs. Finally, additional data with complementary kinematics, e.g., the CMS dijet data or data on prompt photon production can be included to cross-check the impact of the heavy quark data.

ACKNOWLEDGMENTS

The work of P. D., T. J., M. K., and K. K. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project No. 273811115—SFB 1225. We thank Peter Braun-Munzinger and Johanna Stachel for stimulating this research and all our SFB Collaborators for very useful discussions. P. D., T. J., M. K., K. K., and K. F. M. also acknowledge support of the DFG through the Research Training Group GRK 2149. The work of I. S. was supported in part by the French National Centre for Scientific Research CNRS through IN2P3 Project GLUE@NLO. F. O. acknowledges support through US DOE Grant No. DE-SC0010129. A. K. acknowledges the support of Narodowe Centrum Nauki under Sonata Bis Grant No. 2019/34/E/ST2/00186. R. R. acknowledges the support of the Polska Akademia Nauk (Grant Agreement No. PAN.BFD.S.BDN. 613. 022. 2021—PASIFIC 1, POPSICLE).

APPENDIX: DATA COMPARISONS

This appendix contains comparisons between data and predictions made with the Crystal Ball fit. Figures 10–16 show the predictions for the fitted proton-proton data with the uncertainties of the Crystal Ball fit, while Figs. 17–21 show those for the proton-lead data with the uncertainties of the nCTEQ15HQ nuclear PDFs. Note that the proton-lead data is scaled by a factor $\frac{1}{208}$ to give the cross section per nucleon, which is used in the fit.

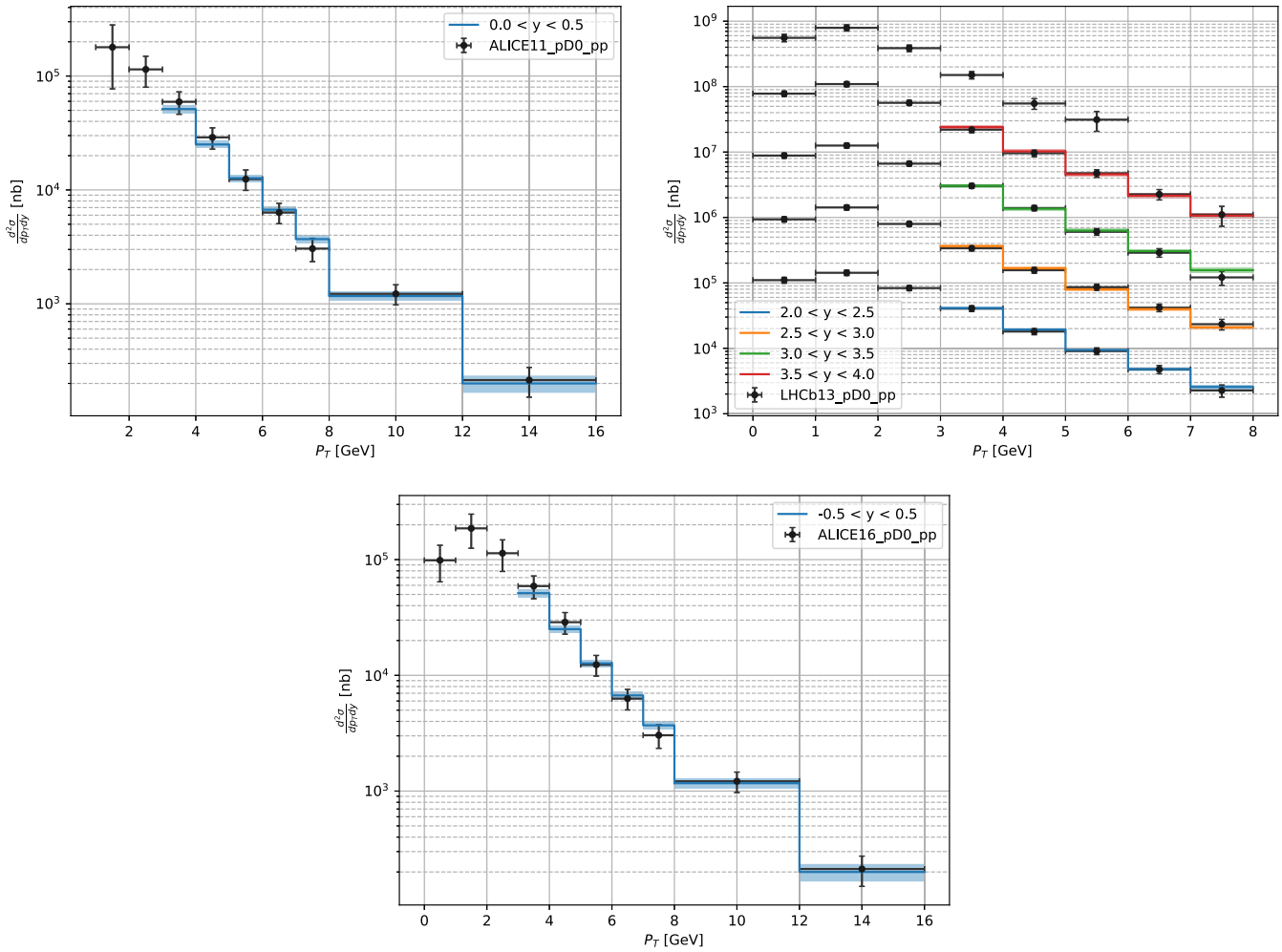

FIG. 10. Predictions for D^0 production in proton-proton collisions with uncertainties from the Crystal Ball fit. Different rapidity bins are separated by multiplying the cross sections by powers of ten for visual clarity.

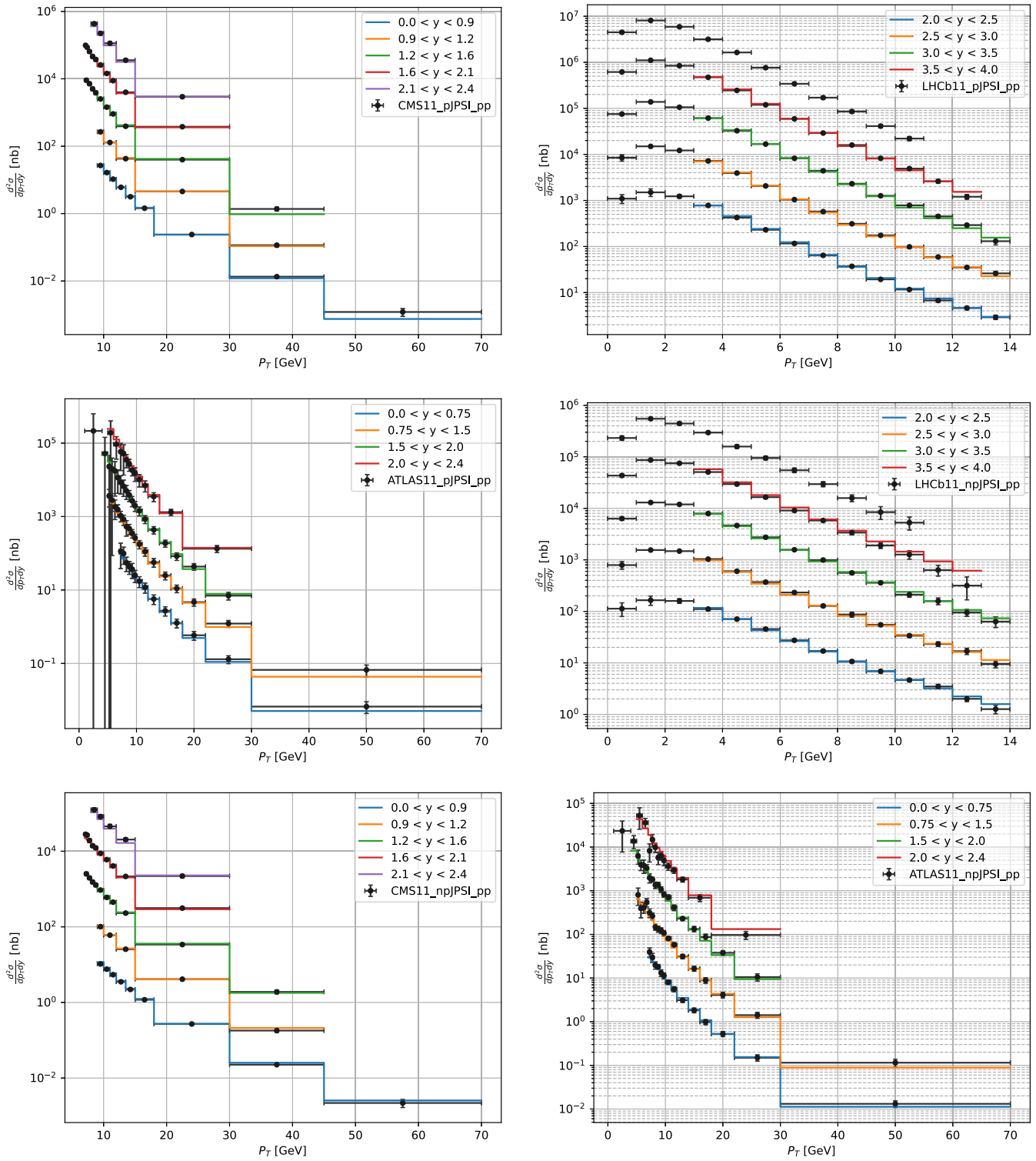

FIG. 11. Predictions for J/ψ production in proton-proton collisions with uncertainties from the Crystal Ball fit. Different rapidity bins are separated by multiplying the cross sections by powers of ten for visual clarity.

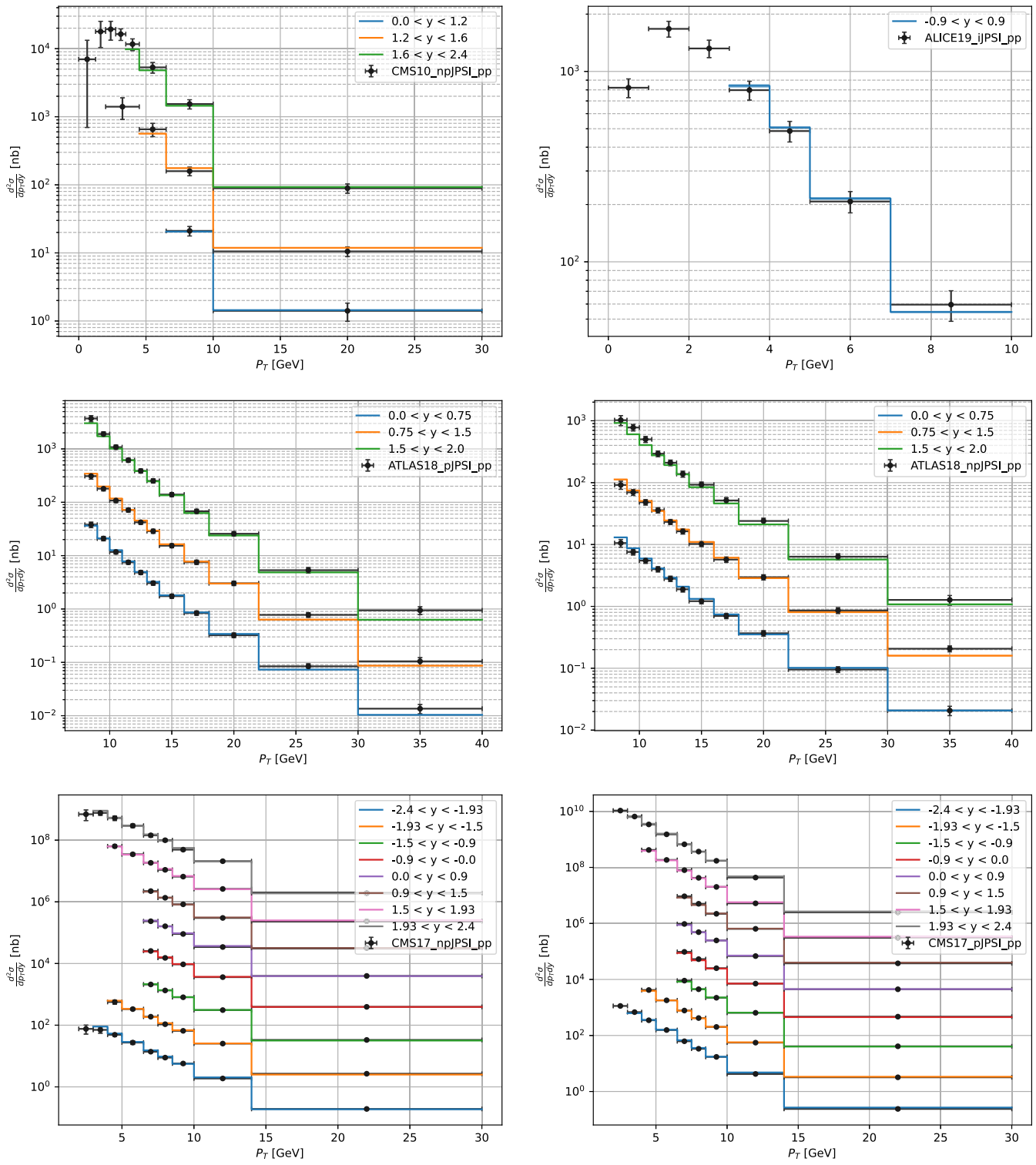

FIG. 12. Predictions for J/ψ production in proton-proton collisions with uncertainties from the Crystal Ball fit. Different rapidity bins are separated by multiplying the cross sections by powers of ten for visual clarity.

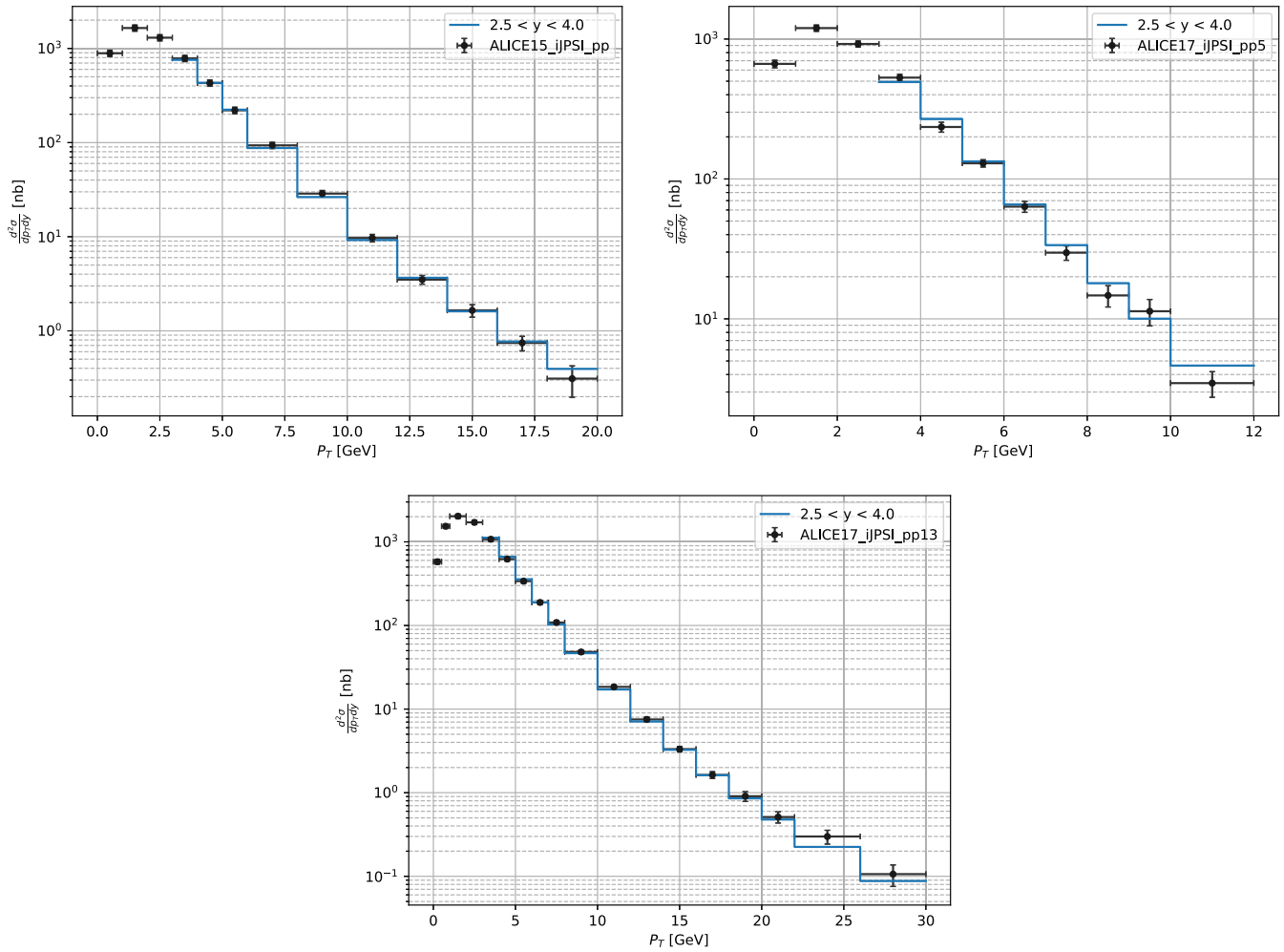

FIG. 13. Predictions for J/ψ production in proton-proton collisions with uncertainties from the Crystal Ball fit. Different rapidity bins are separated by multiplying the cross sections by powers of ten for visual clarity.

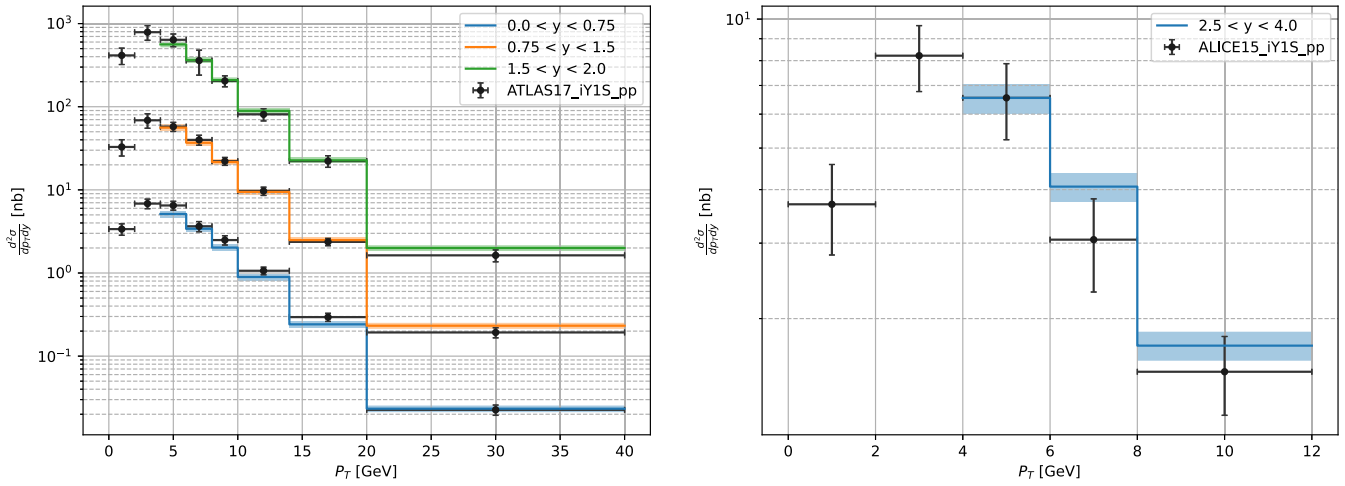

FIG. 14. Predictions for $\Upsilon(1S)$ production in proton-proton collisions with uncertainties from the Crystal Ball fit. Different rapidity bins are separated by multiplying the cross sections by powers of ten for visual clarity.

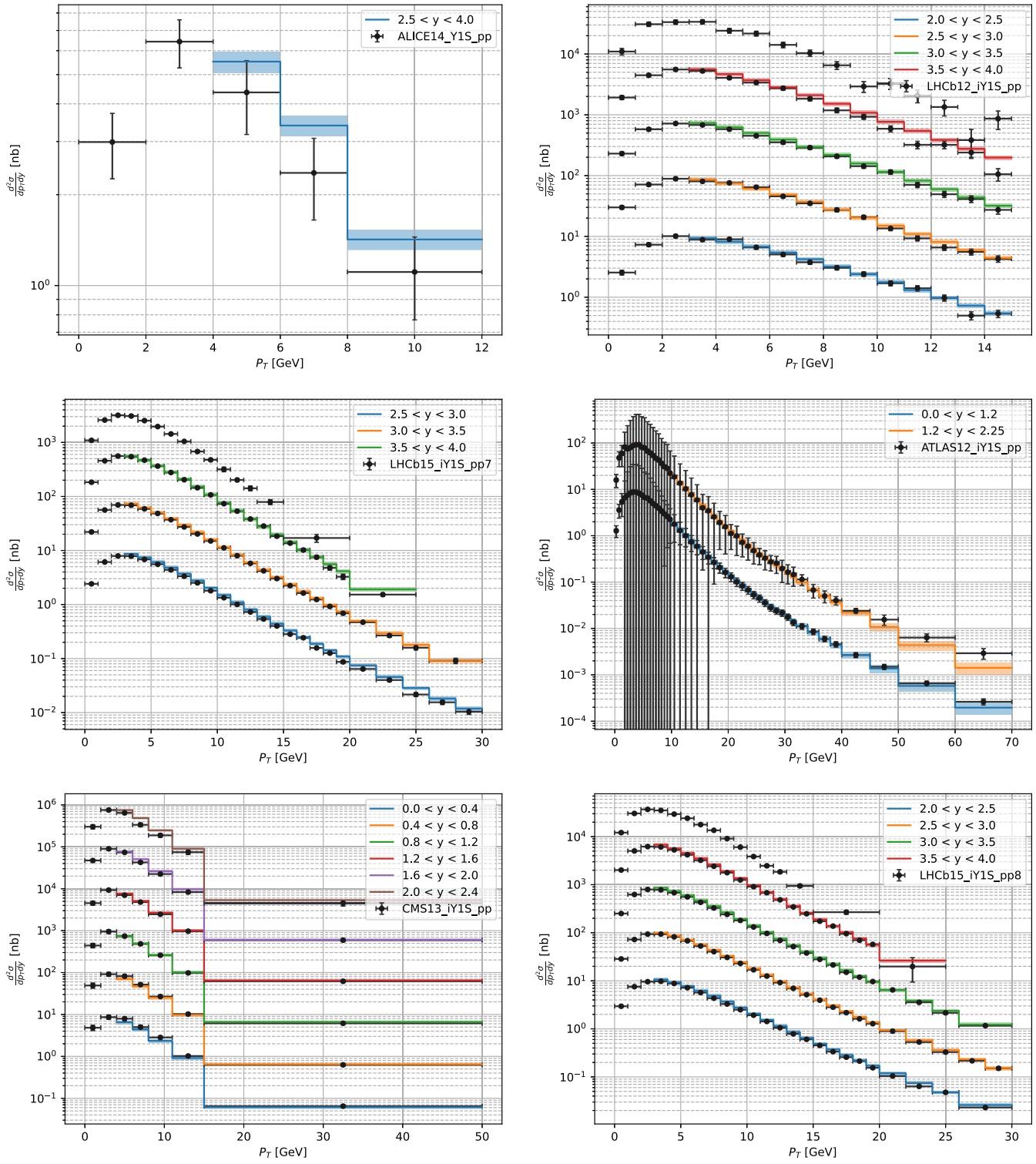

FIG. 15. Predictions for $\Upsilon(1S)$ production in proton-proton collisions with uncertainties from the Crystal Ball fit. Different rapidity bins are separated by multiplying the cross sections by powers of ten for visual clarity.

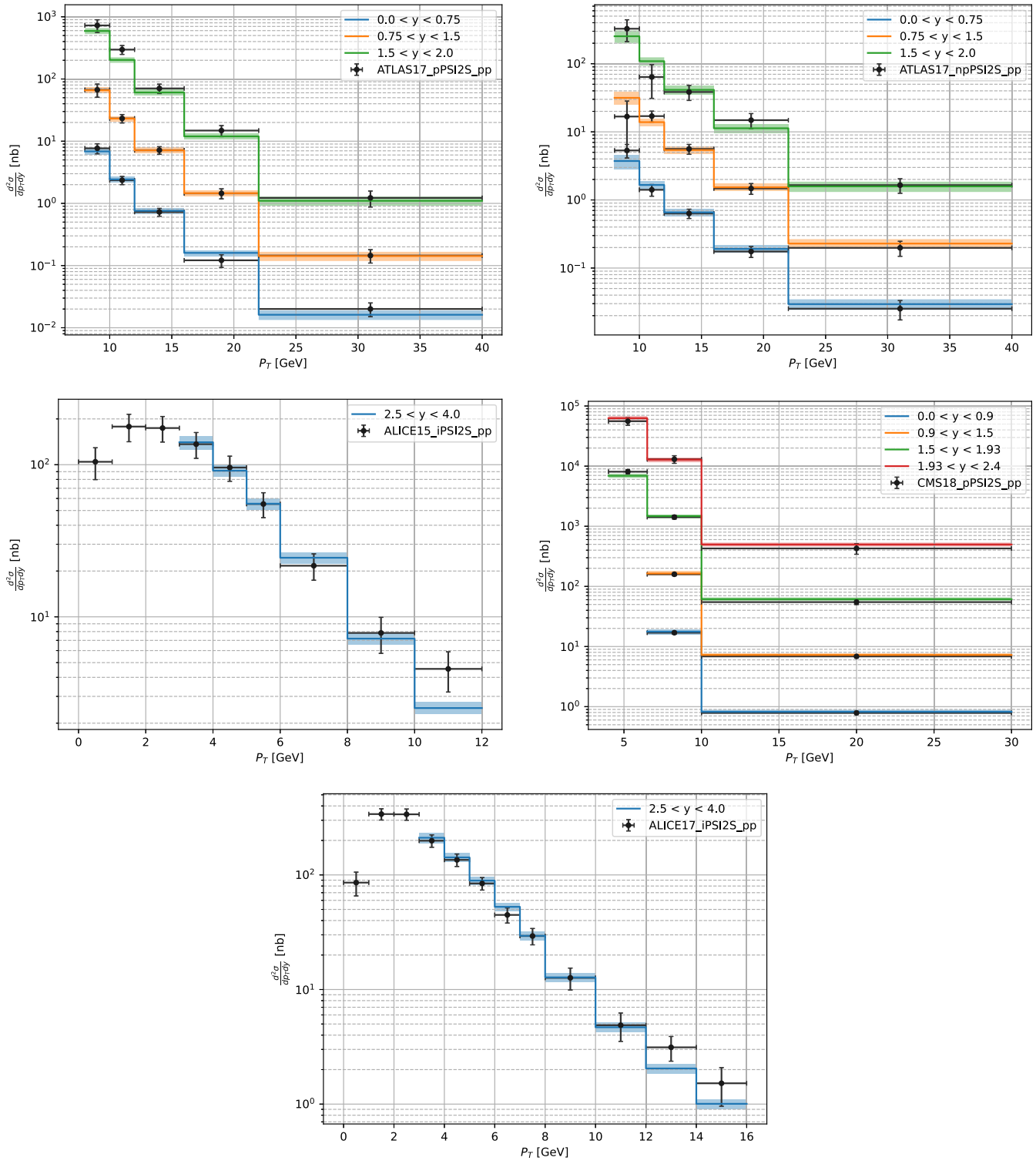

FIG. 16. Predictions for $\psi(2S)$ production in proton-proton collisions with uncertainties from the Crystal Ball fit. Different rapidity bins are separated by multiplying the cross sections by powers of ten for visual clarity.

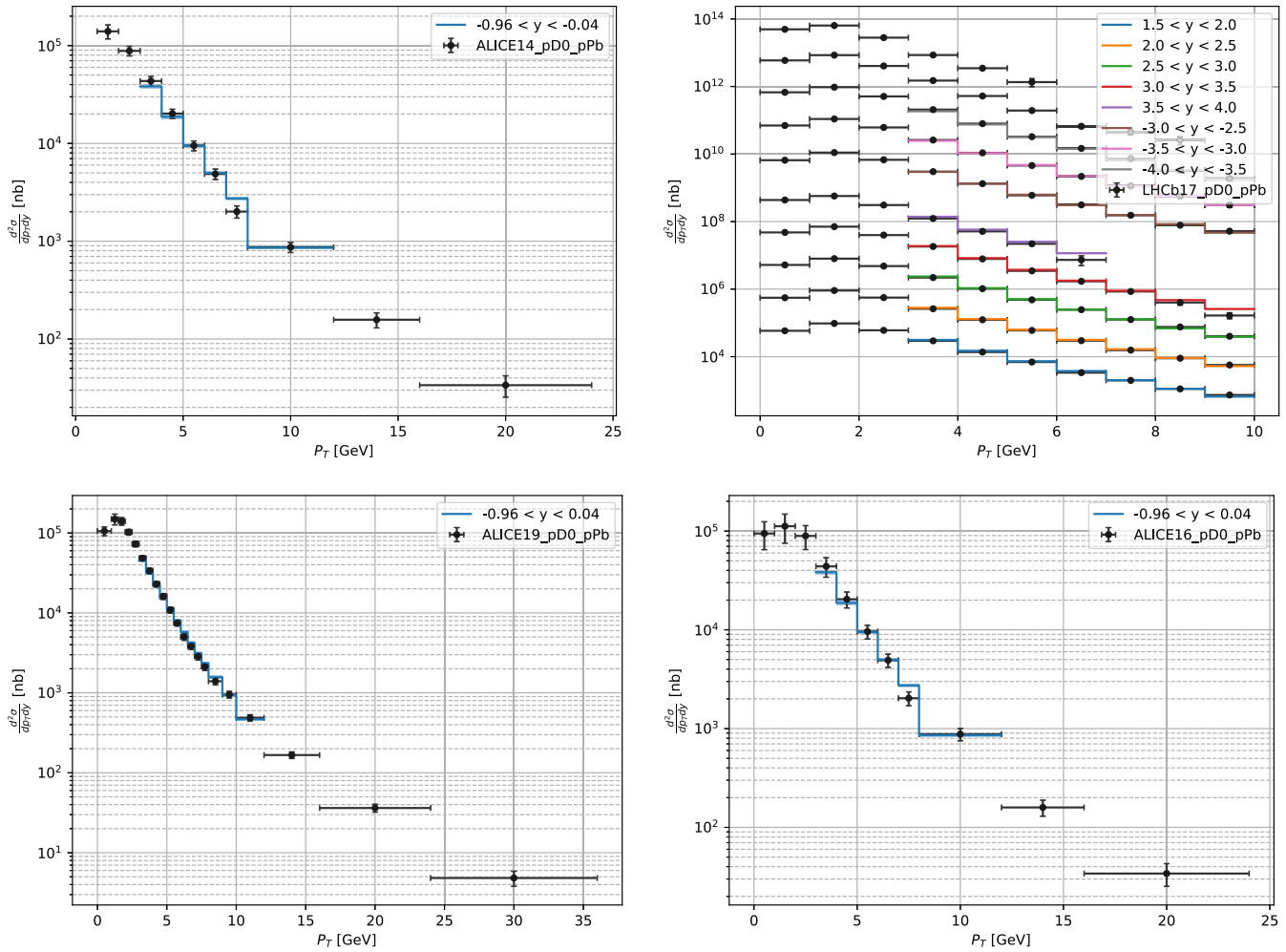

FIG. 17. Predictions for D^0 production in proton-lead collisions with PDF uncertainties of the nCTEQ15HQ fit. Different rapidity bins are separated by multiplying the cross sections by powers of ten for visual clarity.

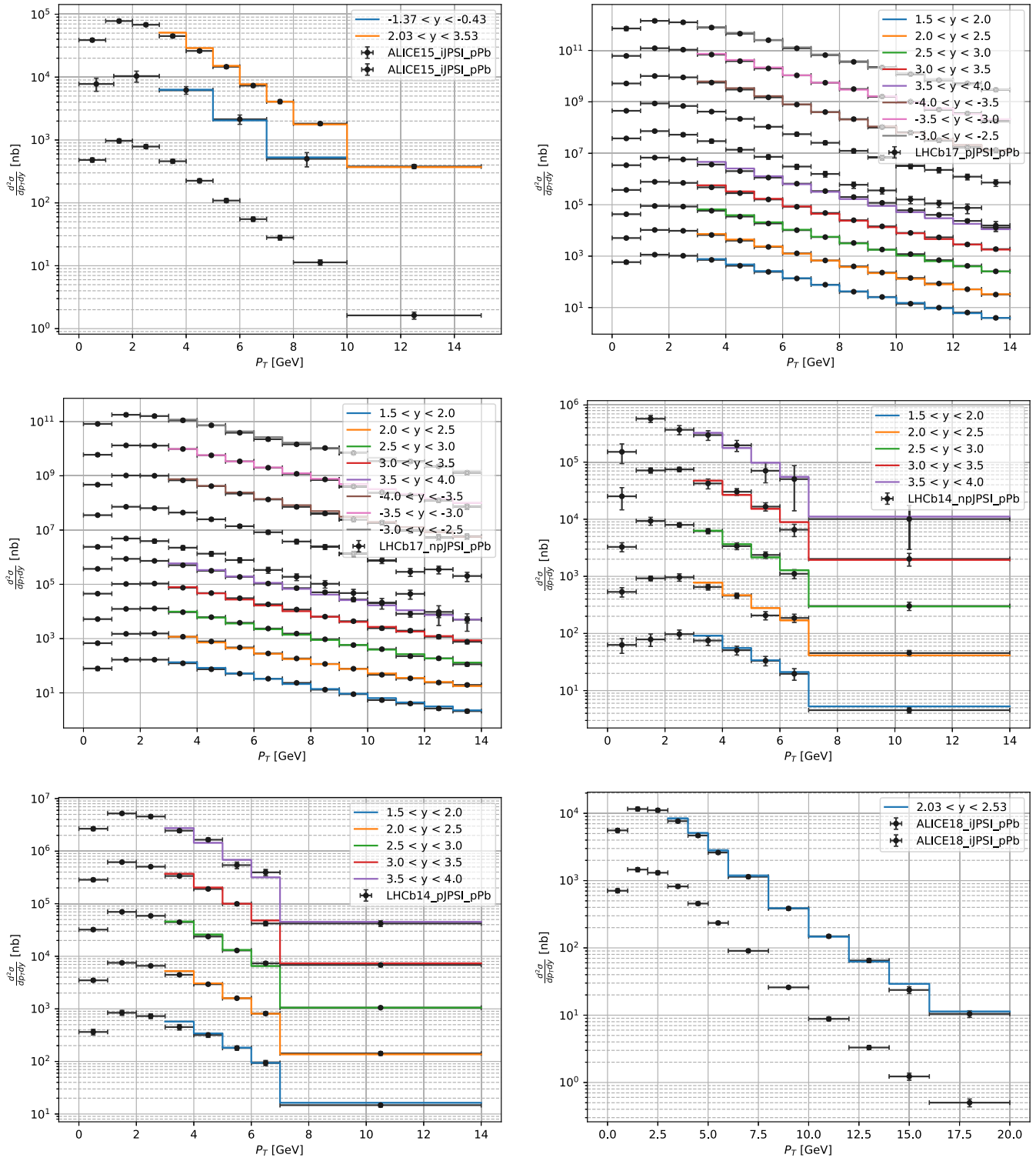

FIG. 18. Predictions for J/ψ production in proton-lead collisions with PDF uncertainties of the nCTEQ15HQ fit. Different rapidity bins are separated by multiplying the cross sections by powers of ten for visual clarity.

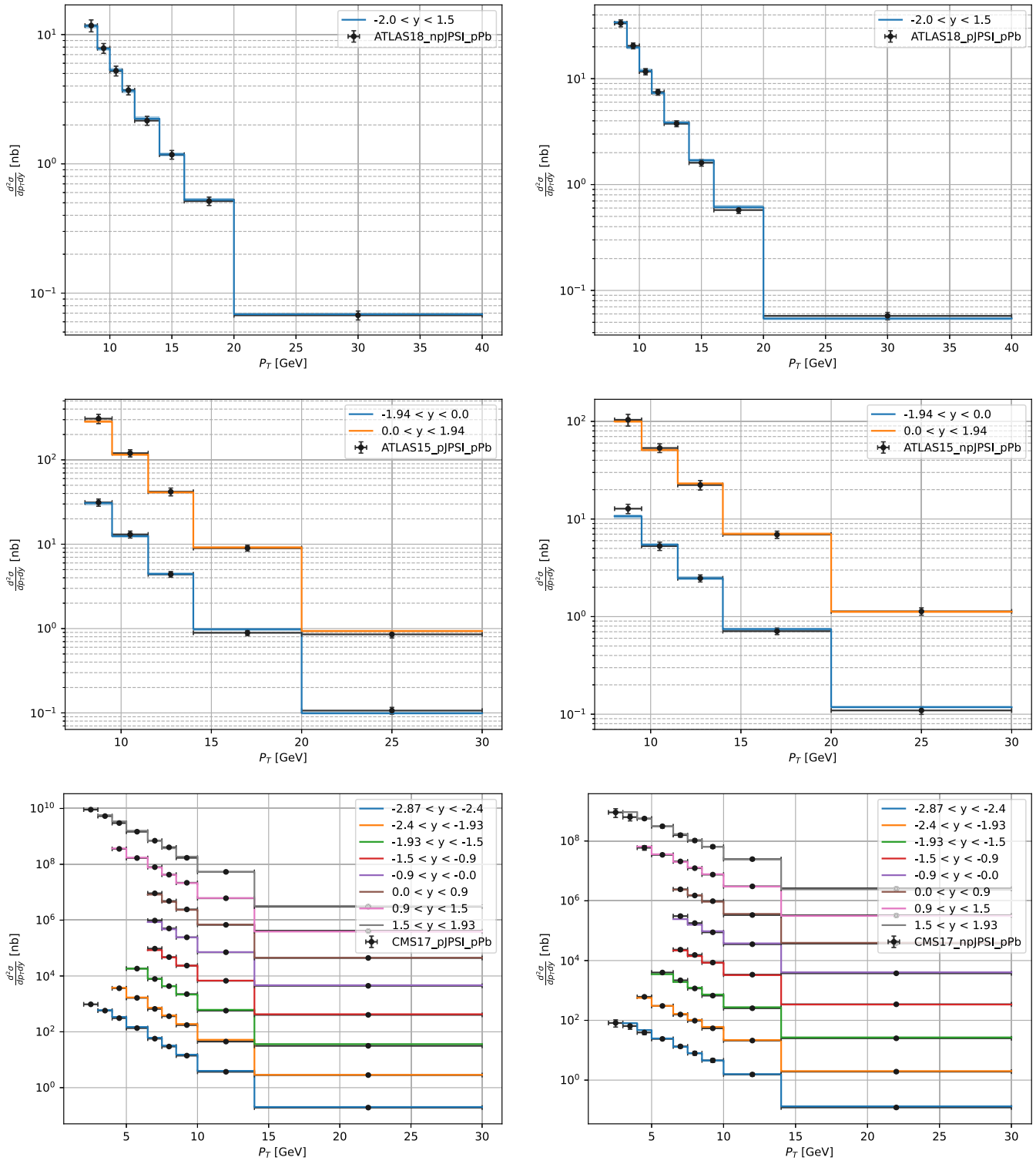

FIG. 19. Predictions for J/ψ production in proton-lead collisions with PDF uncertainties of the nCTEQ15HQ fit. Different rapidity bins are separated by multiplying the cross sections by powers of ten for visual clarity.

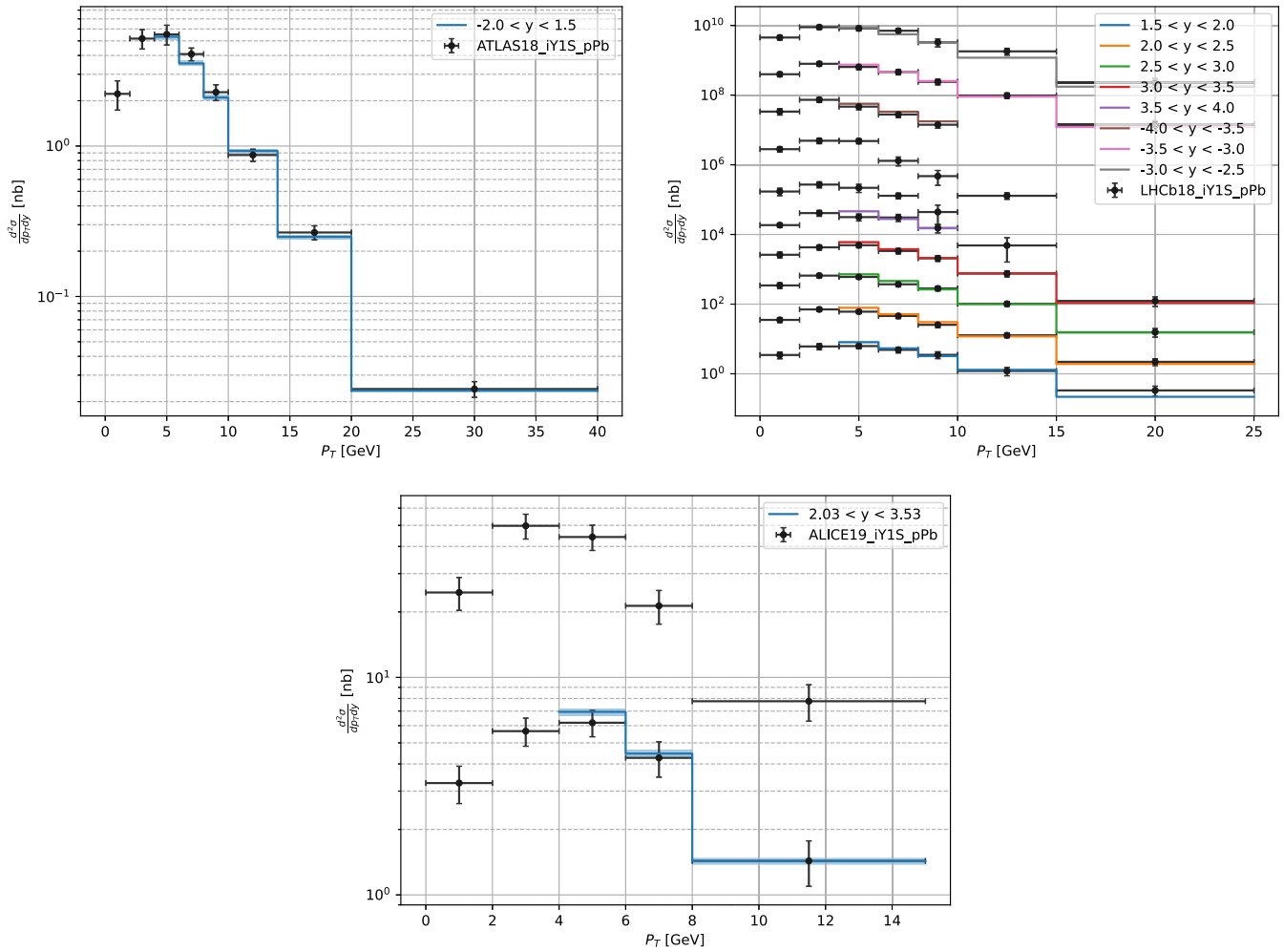

FIG. 20. Predictions for $\Upsilon(1S)$ production in proton-lead collisions with PDF uncertainties of the nCTEQ15HQ fit. Different rapidity bins are separated by multiplying the cross sections by powers of ten for visual clarity.

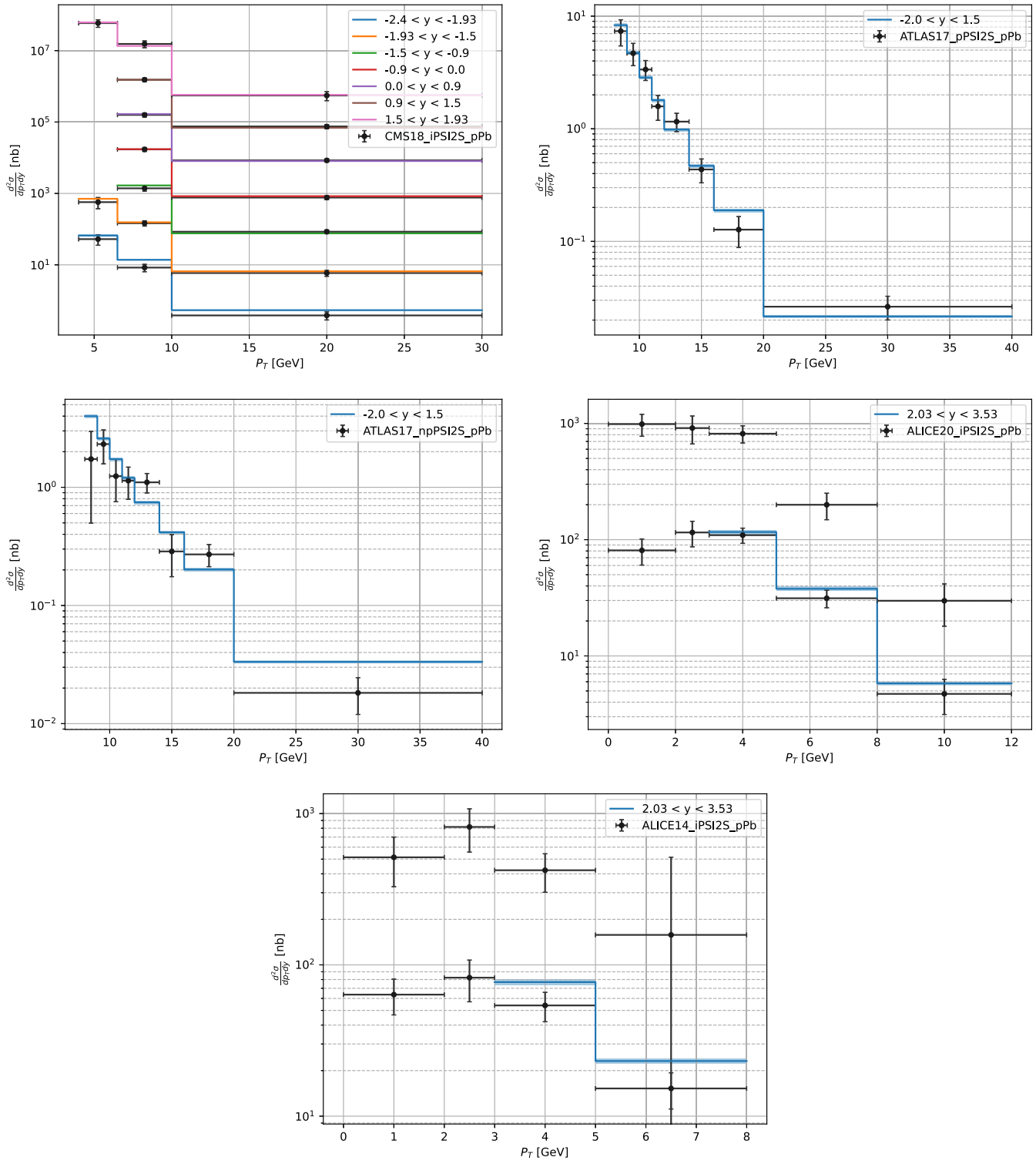

FIG. 21. Predictions for $\psi(2S)$ production in proton-lead collisions with PDF uncertainties of the nCTEQ15HQ fit. Different rapidity bins are separated by multiplying the cross sections by powers of ten for visual clarity.

[1] J. C. Collins, D. E. Soper, and G. F. Sterman, Factorization of hard processes in QCD, Adv. Ser. Dir. High Energy Phys. **5**, 1 (1989).

[2] T.-J. Hou *et al.*, New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC, Phys. Rev. D **103**, 014013 (2021).

[3] R. D. Ball *et al.* (NNPDF Collaboration), Parton distributions from high-precision collider data, Eur. Phys. J. C **77**, 663 (2017).

[4] R. Abdul Khalek, S. Bailey, J. Gao, L. Harland-Lang, and J. Rojo, Towards ultimate parton distributions at the high-luminosity LHC, Eur. Phys. J. C **78**, 962 (2018).

[5] J. Gao, L. Harland-Lang, and J. Rojo, The structure of the proton in the LHC precision era, Phys. Rep. **742**, 1 (2018).

[6] K. Kovarik, P. M. Nadolsky, and D. E. Soper, Hadron structure in high-energy collisions, Rev. Mod. Phys. **92**, 045003 (2020).

[7] S. Alekhin, J. Blümlein, and S. Moch, Strange sea determination from collider data, Phys. Lett. B **777**, 134 (2018).

[8] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, and C. P. Yuan, Implications of CTEQ global analysis for collider observables, Phys. Rev. D **78**, 013004 (2008).

[9] N. Sato, C. Andres, J. J. Ethier, and W. Melnitchouk (JAM Collaboration), Strange quark suppression from a simultaneous Monte Carlo analysis of parton distributions and fragmentation functions, Phys. Rev. D **101**, 074020 (2020).

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

[11] R. S. Thorne, S. Bailey, T. Cridge, L. A. Harland-Lang, A. Martin, and R. Nathvani, Updates of PDFs using the MMHT framework, *Proc. Sci.*, DIS2019 (2019) 036. [arXiv:1907.08147].

[12] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, A. Piccione, J. Rojo, and M. Ubiali (NNPDF Collaboration), Precision determination of electroweak parameters and the strange content of the proton from neutrino deep-inelastic scattering, Nucl. Phys. **B823**, 195 (2009).

[13] H.-W. Lin *et al.*, Parton distributions and lattice QCD calculations: A community white paper, Prog. Part. Nucl. Phys. **100**, 107 (2018).

[14] H.-W. Lin *et al.*, Parton distributions and lattice QCD calculations: Toward 3D structure, Prog. Part. Nucl. Phys. **121**, 103908 (2021).

[15] K. Kovarik *et al.*, nCTEQ15—Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework, Phys. Rev. D **93**, 085037 (2016).

[16] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, EPPS16: Nuclear parton distributions with LHC data, Eur. Phys. J. C **77**, 163 (2017).

[17] R. Abdul Khalek, J. J. Ethier, and J. Rojo (NNPDF Collaboration), Nuclear parton distributions from lepton-nucleus scattering and the impact of an electron-ion collider, Eur. Phys. J. C **79**, 471 (2019).

[18] R. Abdul Khalek, J. J. Ethier, J. Rojo, and G. van Weelden, nNNPDF2.0: Quark flavor separation in nuclei from LHC data, J. High Energy Phys. 09 (2020) 183.

[19] J. J. Ethier and E. R. Nocera, Parton distributions in nucleons and nuclei, Annu. Rev. Nucl. Part. Sci. **70**, 43 (2020).

[20] V. Guzey and M. Klasen, Constraints on nuclear parton distributions from dijet photoproduction at the LHC, Eur. Phys. J. C **79**, 396 (2019).

[21] M. Klasen, K. Kovarik, and J. Potthoff, Nuclear parton density functions from jet production in DIS at an EIC, Phys. Rev. D **95**, 094013 (2017).

[22] M. Klasen and K. Kovařík, Nuclear parton density functions from dijet photoproduction at the EIC, Phys. Rev. D **97**, 114013 (2018).

[23] N. Armesto, H. Paukkunen, J. M. Penín, C. A. Salgado, and P. Zurita, An analysis of the impact of LHC Run I proton–lead data on nuclear parton densities, Eur. Phys. J. C **76**, 218 (2016).

[24] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, EPPS21: A global QCD analysis of nuclear PDFs, Eur. Phys. J. C **82**, 413 (2022).

[25] H. Khanpour, M. Soleymaninia, S. Atashbar Tehrani, H. Spiesberger, and V. Guzey, Nuclear parton distribution functions with uncertainties in a general mass variable flavor number scheme, Phys. Rev. D **104**, 034010 (2021).

[26] K. J. Eskola, H. Paukkunen, and C. A. Salgado, EPS09: A new generation of NLO and LO nuclear parton distribution functions, J. High Energy Phys. 04 (2009) 065.

[27] D. de Florian, R. Sassot, P. Zurita, and M. Stratmann, Global analysis of nuclear parton distributions, Phys. Rev. D **85**, 074028 (2012).

[28] R. A. Khalek, R. Gauld, T. Giani, E. R. Nocera, T. R. Rabemananjara, and J. Rojo, nNNPDF3.0: Evidence for a modified partonic structure in heavy nuclei, Eur. Phys. J. C **82**, 507 (2022).

[29] M. Walt, I. Helenius, and W. Vogelsang, Open-source QCD analysis of nuclear parton distribution functions at NLO and NNLO, Phys. Rev. D **100**, 096015 (2019).

[30] I. Helenius, M. Walt, and W. Vogelsang, NNLO nuclear parton distribution functions with electroweak-boson production data from the LHC, Phys. Rev. D **105**, 094031 (2022).

[31] A. Kusina *et al.*, Impact of LHC vector boson production in heavy ion collisions on strange PDFs, Eur. Phys. J. C **80**, 968 (2020).

[32] P. Duwentäster, L. A. Husová, T. Ježo, M. Klasen, K. Kovařík, A. Kusina, K. F. Muzakka, F. I. Olness, I. Schienbein, and J. Y. Yu, Impact of inclusive hadron production data on nuclear gluon PDFs, Phys. Rev. D **104**, 094005 (2021).

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

[34] A. Kusina, J.-P. Lansberg, I. Schienbein, and H.-S. Shao, Gluon Shadowing in Heavy-Flavor Production at the LHC, Phys. Rev. Lett. **121**, 052004 (2018).

[35] K. J. Eskola, I. Helenius, P. Paakkinen, and H. Paukkunen, A QCD analysis of LHCb D-meson data in p + Pb collisions, J. High Energy Phys. 05 (2020) 037.

[36] I. Helenius and H. Paukkunen, Revisiting the D-meson hadroproduction in general-mass variable flavour number scheme, J. High Energy Phys. 05 (2018) 196.

[37] B. A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger, Inclusive $D^{*\pm}$ production in $p\bar{p}$ collisions with massive charm quarks, Phys. Rev. D **71**, 014018 (2005).

[38] B. A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger, Collinear subtractions in hadroproduction of heavy quarks, Eur. Phys. J. C **41**, 199 (2005).

[39] A. M. Sirunyan *et al.* (CMS Collaboration), Constraining Gluon Distributions in Nuclei Using Dijets in Proton-Proton and Proton-Lead Collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV, Phys. Rev. Lett. **121**, 062002 (2018).

[40] B. A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger, Open charm hadroproduction and the charm content of the proton, Phys. Rev. D **79**, 094009 (2009).

[41] N. Armesto, Nuclear shadowing, J. Phys. G **32**, R367 (2006).

[42] L. Frankfurt, V. Guzey, and M. Strikman, Leading twist nuclear shadowing phenomena in hard processes with nuclei, Phys. Rep. **512**, 255 (2012).

[43] B. Z. Kopeliovich, J. G. Morfin, and I. Schmidt, Nuclear shadowing in electro-weak interactions, Prog. Part. Nucl. Phys. **68**, 314 (2013).

[44] S. A. Kulagin and R. Petti, Global study of nuclear structure functions, Nucl. Phys. **A765**, 126 (2006).

[45] S. J. Brodsky and H. J. Lu, Shadowing and Antishadowing of Nuclear Structure Functions, Phys. Rev. Lett. **64**, 1342 (1990).

[46] S. J. Brodsky, I. Schmidt, and J.-J. Yang, Nuclear anti-shadowing in neutrino deep inelastic scattering, Phys. Rev. D **70**, 116003 (2004).

[47] D. F. Geesaman, K. Saito, and A. W. Thomas, The nuclear EMC effect, Annu. Rev. Nucl. Part. Sci. **45**, 337 (1995).

[48] P. R. Norton, The EMC effect, Rep. Prog. Phys. **66**, 1253 (2003).

[49] O. Hen, D. W. Higinbotham, G. A. Miller, E. Piasetzky, and L. B. Weinstein, The EMC effect and high momentum nucleons in nuclei, Int. J. Mod. Phys. E **22**, 1330017 (2013).

[50] S. Malace, D. Gaskell, D. W. Higinbotham, and I. Cloet, The challenge of the EMC effect: Existing data and future directions, Int. J. Mod. Phys. E **23**, 1430013 (2014).

[51] O. Hen, G. A. Miller, E. Piasetzky, and L. B. Weinstein, Nucleon-nucleon correlations, short-lived excitations, and the quarks within, Rev. Mod. Phys. **89**, 045002 (2017).

[52] E. Iancu, A. Leonidov, and L. D. McLerran, Nonlinear gluon evolution in the color glass condensate. I, Nucl. Phys. **A692**, 583 (2001).

[53] F. Gelis, E. Iancu, J. Jalilian-Marian, and R. Venugopalan, The color glass condensate, Annu. Rev. Nucl. Part. Sci. **60**, 463 (2010).

[54] J. Jalilian-Marian, From small to large x: Toward a unified description of high energy collisions, J. Phys. Conf. Ser. **1602**, 012024 (2020).

[55] J.-w. Qiu, QCD factorization and rescattering in proton nucleus collisions, arXiv:hep-ph/0305161.

[56] A. Accardi *et al.*, Hard probes in heavy ion collisions at the LHC: PDFs, shadowing and pa collisions, in *Proceedings of the 3rd Workshop on Hard Probes in Heavy Ion Collisions: 3rd Plenary Meeting* (2004), arXiv:hep-ph/0308248.

[57] F. Arleo and S. Peigne, J/ψ Suppression in $p^{-}A$ Collisions from Parton Energy Loss in Cold QCD Matter, Phys. Rev. Lett. **109**, 122301 (2012).

[58] F. Arleo and S. Peigne, Heavy-quarkonium suppression in p-A collisions from parton energy loss in cold QCD matter, J. High Energy Phys. 03 (2013) 122.

[59] F. Arleo, G. Jackson, and S. Peigné, Impact of fully coherent energy loss on heavy meson production in pA collisions, J. High Energy Phys. 01 (2022) 164.

[60] F. Olness, J. Pumplin, D. Stump, J. Huston, P. M. Nadolsky, H. L. Lai, S. Kretzer, J. F. Owens, and W. K. Tung, Neutrino dimuon production and the strangeness asymmetry of the nucleon, Eur. Phys. J. C **40**, 145 (2005).

[61] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung, Uncertainties of predictions from parton distribution functions. II. The Hessian method, Phys. Rev. D **65**, 014013 (2001).

[62] E. P. Segarra *et al.*, nCTEQ15HIX—Extending nPDF analyses into the high-x, low Q^2 region, Phys. Rev. D **103**, 114015 (2021).

[63] G. P. Salam and J. Rojo, A Higher Order Perturbative Parton Evolution Toolkit (HOPPET), Comput. Phys. Commun. **180**, 120 (2009).

[64] T. Carli, D. Clements, A. Cooper-Sarkar, C. Gwenlan, G. P. Salam, F. Siegert, P. Starovoitov, and M. Sutton, 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).

[65] M. Werlen, INCNLO-direct photon and inclusive hadron production code website, Version 1.4, http://lapth.cnrs.fr/PHOX_FAMILY.

[66] H.-S. Shao, HELAC-Onia 2.0: An upgraded matrix-element and event generator for heavy quarkonium physics, Comput. Phys. Commun. **198**, 238 (2016).

[67] J. F. Owens, J. Huston, C. E. Keppel, S. Kuhlmann, J. G. Morfin, F. Olness, J. Pumplin, and D. Stump, The impact of new neutrino DIS and Drell-Yan data on large-x parton distributions, Phys. Rev. D **75**, 054030 (2007).

[68] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, Proton-PDF uncertainties in extracting nuclear PDFs from $W^{\pm}$ production in p + Pb collisions, Eur. Phys. J. C **82**, 271 (2022).

[69] C. H. Kom, A. Kulesza, and W. J. Stirling, Pair Production of J/ψ as a Probe of Double Parton Scattering at LHCb, Phys. Rev. Lett. **107**, 082002 (2011).

[70] A. Kusina, J.-P. Lansberg, I. Schienbein, and H.-S. Shao, Reweighted nuclear PDFs using heavy-flavor production data at the LHC, Phys. Rev. D **104**, 014010 (2021).

[71] J. Gaiser, Charmonium spectroscopy from radiative decays of the J/ψ and $\psi(2s)$, https://www.slac.stanford.edu/cgi-bin/getdoc/slac-r-255.pdf.

[72] R. Ellis, W. Stirling, and B. Webber, *QCD and Collider Physics*, Cambridge Monographs on Particle Physics,

Nuclear Physics and Cosmology (Cambridge University Press, Cambridge, England, 2003).
[73] R. Aaij *et al.* (LHCb Collaboration), Forward production of Υ mesons in pp collisions at $\sqrt{s} = 7$ and 8 TeV, J. High Energy Phys. 11 (2015) 103.
[74] B. Abelev *et al.* (ALICE Collaboration), Measurement of charm production at central rapidity in proton-proton collisions at $\sqrt{s} = 7$ TeV, J. High Energy Phys. 01 (2012) 128.
[75] R. Aaij *et al.* (LHCb Collaboration), Prompt charm production in pp collisions at $\sqrt{s} = 7$ TeV, Nucl. Phys. **B871**, 1 (2013).
[76] J. Adam *et al.* (ALICE Collaboration), D-meson production in p-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV and in pp collisions at $\sqrt{s} = 7$ TeV, Phys. Rev. C **94**, 054908 (2016).
[77] G. Aad *et al.* (ATLAS Collaboration), Measurement of the differential cross-sections of inclusive, prompt and non-prompt J/ψ production in proton-proton collisions at $\sqrt{s} = 7$ TeV, Nucl. Phys. **B850**, 387 (2011).
[78] S. Chatrchyan *et al.* (CMS Collaboration), J/ψ and ψ_{2S} production in pp collisions at $\sqrt{s} = 7$ TeV, J. High Energy Phys. 02 (2012) 011.
[79] R. Aaij *et al.* (LHCb Collaboration), Measurement of J/ψ production in pp collisions at $\sqrt{s} = 7$ TeV, Eur. Phys. J. C **71**, 1645 (2011).
[80] A. M. Sirunyan *et al.* (CMS Collaboration), Measurement of prompt and nonprompt J/ψ production in pp and pPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV, Eur. Phys. J. C **77**, 269 (2017).
[81] M. Aaboud *et al.* (ATLAS Collaboration), Measurement of quarkonium production in proton–lead and proton–proton collisions at 5.02 TeV with the ATLAS detector, Eur. Phys. J. C **78**, 171 (2018).
[82] V. Khachatryan *et al.* (CMS Collaboration), Prompt and non-prompt J/ψ production in pp collisions at $\sqrt{s} = 7$ TeV, Eur. Phys. J. C **71**, 1575 (2011).
[83] J. Adam *et al.* (ALICE Collaboration), Inclusive quarkonium production at forward rapidity in pp collisions at $\sqrt{s} = 8$ TeV, Eur. Phys. J. C **76**, 184 (2016).
[84] S. Acharya *et al.* (ALICE Collaboration), Energy dependence of forward-rapidity J/ψ and $\psi(2S)$ production in pp collisions at the LHC, Eur. Phys. J. C **77**, 392 (2017).
[85] S. Acharya *et al.* (ALICE Collaboration), Inclusive J/ψ production at mid-rapidity in pp collisions at $\sqrt{s} = 5.02$ TeV, J. High Energy Phys. 10 (2019) 084.
[86] G. Aad *et al.* (ATLAS Collaboration), Measurement of upsilon production in 7 TeV pp collisions at ATLAS, Phys. Rev. D **87**, 052004 (2013).
[87] R. Aaij *et al.* (LHCb Collaboration), Measurement of upsilon production in pp collisions at $\sqrt{s} = 7$ TeV, Eur. Phys. J. C **72**, 2025 (2012).
[88] S. Chatrchyan *et al.* (CMS Collaboration), Measurement of the $\Upsilon(1S)$, $\Upsilon(2S)$, and $\Upsilon(3S)$ cross sections in pp collisions at $\sqrt{s} = 7$ TeV, Phys. Lett. B **727**, 101 (2013).
[89] B. B. Abelev *et al.* (ALICE Collaboration), Measurement of quarkonium production at forward rapidity in pp collisions at $\sqrt{s} = 7$ TeV, Eur. Phys. J. C **74**, 2974 (2014).
[90] ALICE Collaboration, Quarkonium signal extraction in ALICE, https://cds.cern.ch/record/2060096/files/QuarkoniumSignalExtraction_v5.pdf.
[91] G. T. Bodwin, E. Braaten, and G. P. Lepage, Rigorous QCD analysis of inclusive annihilation and production of heavy quarkonium, Phys. Rev. D **51**, 1125 (1995); **55**, 5853(E) (1997).
[92] Y.-Q. Ma, K. Wang, and K.-T. Chao, $J/\psi(\psi')$ Production at the Tevatron and LHC at $\mathcal{O}(\alpha_s^4 v^4)$ in Nonrelativistic QCD, Phys. Rev. Lett. **106**, 042002 (2011).
[93] Y.-Q. Ma, K. Wang, and K.-T. Chao, A complete NLO calculation of the J/ψ and ψ' production at hadron colliders, Phys. Rev. D **84**, 114001 (2011).
[94] M. Butenschoen and B. A. Kniehl, Reconciling J/ψ Production at HERA, RHIC, Tevatron, and LHC with NRQCD Factorization at Next-to-Leading Order, Phys. Rev. Lett. **106**, 022003 (2011).
[95] M. Butenschoen and B. A. Kniehl, World data of J/ψ production consolidate NRQCD factorization at NLO, Phys. Rev. D **84**, 051501 (2011).
[96] R. Aaij *et al.* (LHCb Collaboration), Prompt and non-prompt J/ψ production and nuclear modification in pPb collisions at $\sqrt{s_{\mathrm{NN}}} = 8.16$ TeV, Phys. Lett. B **774**, 159 (2017).
[97] S. Acharya *et al.* (ALICE Collaboration), Inclusive J/ψ production at forward and backward rapidity in p-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 8.16$ TeV, J. High Energy Phys. 07 (2018) 160.
[98] G. Aad *et al.*, Measurement of differential J/ψ production cross sections and forward-backward ratios in $\mathrm{p} + \mathrm{Pb}$ collisions with the ATLAS detector, Phys. Rev. C **92**, 034904 (2015).
[99] T. Kneesch, B. A. Kniehl, G. Kramer, and I. Schienbein, Charmed-meson fragmentation functions with finite-mass corrections, Nucl. Phys. **B799**, 34 (2008).
[100] S. Acharya *et al.* (ALICE Collaboration), Measurement of D-meson production at mid-rapidity in pp collisions at $\sqrt{s} = 7$ TeV, Eur. Phys. J. C **77**, 550 (2017).
[101] S. Acharya *et al.* (ALICE Collaboration), Measurement of D^0, D^+, D^{*+} and D_s^+ production in pp collisions at $\sqrt{s} =$ 5.02 TeV with ALICE, Eur. Phys. J. C **79**, 388 (2019).
[102] R. Aaij *et al.* (LHCb Collaboration), Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 5$ TeV, J. High Energy Phys. 06 (2017) 147.
[103] K. F. Muzakka, P. Duwentäster, T. Hobbs, T. Ježo, M. Klasen, K. Kovařík, A. Kusina, J. Morfin, F. I. Olness, R. Ruiz, I. Schienbein, and J. Y. Yu, Compatibility of neutrino DIS data and its impact on nuclear parton distribution functions, arXiv:2204.13157.
[104] A. M. Sirunyan *et al.* (CMS Collaboration), Measurement of prompt $\psi(2S)$ production cross sections in proton-lead and proton-proton collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV, Phys. Lett. B **790**, 509 (2019).
[105] B. B. Abelev *et al.* (ALICE Collaboration), Measurement of Prompt D-meson Production in $p - Pb$ Collisions at $\sqrt{s_{NN}} = 5.02$ TeV, Phys. Rev. Lett. **113**, 232301 (2014).
[106] R. Aaij *et al.* (LHCb Collaboration), Study of prompt D^0 meson production in pPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5$ TeV, J. High Energy Phys. 10 (**2017**) 090.

[107] S. Acharya *et al.* (ALICE Collaboration), Measurement of prompt D^0, D^+, D^{*+}, and D_S^+ production in p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, J. High Energy Phys. 12 (2019) 092.

[108] R. Aaij *et al.* (LHCb Collaboration), Study of J/ψ production and cold nuclear matter effects in pPb collisions at $\sqrt{s_{NN}} = 5$ TeV, J. High Energy Phys. 02 (2014) 072.

[109] B. B. Abelev *et al.* (ALICE Collaboration), J/ψ production and nuclear effects in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, J. High Energy Phys. 02 (2014) 073.

[110] J. Adam *et al.* (ALICE Collaboration), Rapidity and transverse-momentum dependence of the inclusive J/ψ nuclear modification factor in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, J. High Energy Phys. 06 (2015) 055.

[111] B. B. Abelev *et al.* (ALICE Collaboration), Production of inclusive $\Upsilon(1S)$ and $\Upsilon(2S)$ in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, Phys. Lett. B **740**, 105 (2015).

[112] R. Aaij *et al.* (LHCb Collaboration), Study of Υ production and cold nuclear matter effects in pPb collisions at $\sqrt{s_{NN}} = 5$ TeV, J. High Energy Phys. 07 (2014) 094.

[113] R. Aaij *et al.* (LHCb Collaboration), Study of Υ production in pPb collisions at $\sqrt{s_{NN}} = 8.16$ TeV, J. High Energy Phys. 11 (2018) 194; 02 (2020) 093(E).

[114] S. Acharya *et al.* (ALICE Collaboration), Υ production in p–Pb collisions at $\sqrt{s_{NN}} = 8.16$ TeV, Phys. Lett. B **806**, 135486 (2020).

[115] B. B. Abelev *et al.* (ALICE Collaboration), Suppression of $\psi(2S)$ production in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, J. High Energy Phys. 12 (2014) 073.

[116] S. Acharya *et al.* (ALICE Collaboration), Measurement of nuclear effects on $\psi(2S)$ production in p-Pb collisions at $\sqrt{s_{NN}} = 8.16$ TeV, J. High Energy Phys. 07 (2020) 237.