

# Unpolarized transverse momentum distributions from a global fit of Drell-Yan and semi-inclusive deep-inelastic scattering data

## The MAP Collaboration<sup>1</sup>

Alessandro Bacchetta <sup>a,b</sup> Valerio Bertone <sup>c</sup> Chiara Bissolotti <sup>a,d</sup>  
 Giuseppe Bozzi <sup>e,f</sup> Matteo Cerutti <sup>a,b</sup> Fulvio Piacenza,<sup>a</sup> Marco Radici <sup>b</sup>  
 and Andrea Signori <sup>a,b,2</sup>

<sup>a</sup>*Dipartimento di Fisica, Università di Pavia, via Bassi 6, I-27100 Pavia, Italy*

<sup>b</sup>*INFN — Sezione di Pavia, via Bassi 6, I-27100 Pavia, Italy*

<sup>c</sup>*IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France*

<sup>d</sup>*HEP Division, Argonne National Laboratory, 9700 S. Cass Avenue, Lemont, IL, 60439 U.S.A.*

<sup>e</sup>*Dipartimento di Fisica, Università di Cagliari, Cittadella Universitaria, I-09042 Monserrato (CA), Italy*

<sup>f</sup>*INFN — Sezione di Cagliari, Cittadella Universitaria, I-09042 Monserrato (CA), Italy*

*E-mail: [alessandro.bacchetta@unipv.it](mailto:alessandro.bacchetta@unipv.it), [valerio.bertone@cea.fr](mailto:valerio.bertone@cea.fr),  
[cbissolotti@anl.gov](mailto:cbissolotti@anl.gov), [giuseppe.bozzi@unica.it](mailto:giuseppe.bozzi@unica.it),  
[matteo.cerutti@pv.infn.it](mailto:matteo.cerutti@pv.infn.it), [fu.piacenza@gmail.com](mailto:fu.piacenza@gmail.com),  
[marco.radici@pv.infn.it](mailto:marco.radici@pv.infn.it), [andrea.signori@unipv.it](mailto:andrea.signori@unipv.it)*

**ABSTRACT:** We present an extraction of unpolarized transverse-momentum-dependent parton distribution and fragmentation functions based on more than two thousand data points from several experiments for two different processes: semi-inclusive deep-inelastic scattering and Drell-Yan production. The baseline analysis is performed using the Monte Carlo replica method and resumming large logarithms at N<sup>3</sup>LL accuracy. The resulting description of the data is very good ( $\chi^2/N_{\text{dat}} = 1.06$ ). For semi-inclusive deep-inelastic scattering, predictions for multiplicities are normalized by factors that cure the discrepancy with data introduced by higher-order perturbative corrections.

**KEYWORDS:** Parton Distributions, Deep Inelastic Scattering or Small-x Physics, Resummation, Specific QCD Phenomenology

ARXIV EPRINT: [2206.07598](https://arxiv.org/abs/2206.07598)

<sup>1</sup>The MAP acronym stands for “Multi-dimensional Analyses of Partonic distributions”. It refers to a collaboration aimed at studying the three-dimensional structure of hadrons. The public codes released by the collaboration are available at <https://github.com/MapCollaboration>.

<sup>2</sup>Currently at Università di Torino, I-10125 Torino, Italy.

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Formalism</b>	<b>3</b>
2.1	Drell-Yan	3
2.2	Semi-Inclusive Deep-Inelastic Scattering (SIDIS)	7
2.3	Transverse Momentum Distributions (TMDs)	10
2.4	Normalization factors for SIDIS	13
<b>3</b>	<b>Data selection</b>	<b>16</b>
3.1	Drell-Yan	16
3.2	SIDIS	17
3.3	Error treatment	20
<b>4</b>	<b>Results</b>	<b>21</b>
4.1	Fit quality	21
4.1.1	SIDIS	22
4.1.2	Drell-Yan	26
4.2	TMD distributions	30
4.2.1	Collins-Soper kernel	34
4.2.2	Average squared transverse momenta	35
4.3	Variations on the fit configurations	37
4.3.1	Global fit at NNLL	37
4.3.2	Global fit at NLL	39
4.3.3	Cut in $ q_T /Q$	39
<b>5</b>	<b>Conclusions and outlook</b>	<b>41</b>
<b>A</b>	<b>Explicit expression of <math>C</math> coefficients at order <math>\alpha_s</math></b>	<b>43</b>

---

## 1 Introduction

In this paper, we present an extraction of Transverse Momentum Distributions (TMDs) [1–4] using more than two thousand data points from several experiments for two different kinds of processes: Semi-Inclusive Deep Inelastic Scattering (SIDIS) and production of Drell-Yan (DY) lepton pairs, significantly improving our previous analysis [5].

Building maps of the internal partonic structure of nucleons is a crucial step towards understanding the interactions between quarks and gluons and the phenomenon of confinement. A steady progress in the last decades has led to more and more refined versions

of such maps. The TMDs extracted in this article encode information about the three-dimensional distributions of quarks in momentum space. The level of sophistication of a TMD extraction essentially depends on two ingredients: the amount of analyzed data from different processes, namely how “global” the experimental information from which TMDs are extracted, and the perturbative accuracy reached in the theoretical formalism.

The extraction of TMDs is based on TMD factorization theorems, which provide a precise definition of the objects to be extracted, establishing their universality and evolution equations. In this case, the accuracy of the calculation is defined by the amount of large logarithms being resummed, that in turn defines the powers of the strong coupling  $\alpha_s$  to be included in the perturbative quantities [6–9].

Of particular relevance is also the combination of analyzed data from different processes. Similarly to the extraction of collinear Parton Distribution Functions (PDFs), we can talk about a global fit, i.e., a fit that leverages the universality of the parton distributions and uses different processes to constrain them. However, in SIDIS two types of TMDs enter the cross section: the TMD PDFs, describing how partons are arranged in the nucleon, and TMD Fragmentation Functions (FFs), describing how a parton produces a final-state hadron. The knowledge of TMDs, in particular of TMD FFs, would be greatly improved by using data from electron-positron annihilations into two almost back-to-back hadrons [10]. Unfortunately, this data is presently not available. There are measurements for the inclusive production of single hadrons [11] but, in this case, transverse momenta need to be defined with respect to the thrust axis: a careful description of the latter is non trivial, and a rigorous factorization theorem for this process has been discussed only recently (see, e.g., refs. [12–14] and references therein). Therefore, for TMD extractions we currently talk about a global fit when data from SIDIS and DY processes are included.

In the last ten years, several extractions of TMDs have been presented [5, 7, 15–24]. Most of them suffered some shortcomings: they were either obtained in a parton-model framework without QCD corrections, or they took into account only a limited set of data, or did not perform a full fit. TMDs were also studied in a different framework, the so-called parton-branching approach [25–27].

At present, only two works have reached the stage of combining SIDIS and DY data in a full-fledged global TMD fit: the above mentioned extraction of ref. [5], henceforth named PV17, and the extraction of ref. [22], henceforth named SV19.

The PV17 extraction reached the NLL accuracy, was based on the calculation of observables at mean kinematics in each bin, and did not manage to describe the normalization of all datasets. In this work, we push the accuracy of the analysis to what we will refer to as  $N^3LL^-$  (only NNLO collinear FFs are currently missing in order to reach full  $N^3LL$  accuracy).<sup>1</sup> Apart from the increase in perturbative accuracy, in this work we also include many measurements published after 2017.

The SV19 extraction reached the same perturbative accuracy and included essentially the same datasets. Our work has crucial differences in the selection of specific data points

---

<sup>1</sup>While this work was being completed, further efforts, including a work by the MAP collaboration, were made to include part of the NNLO corrections [28, 29] in the extraction of FFs.

(we include a much larger number of points), the implementation of TMD evolution, the choice of nonperturbative components, and the handling of normalization for SIDIS data.

As we will describe in detail, we found it particularly difficult to describe in a satisfactory way the normalization of SIDIS data obtained in fixed-target experiments at moderate to low scales. When the analysis is performed at NLL accuracy, we can describe well shape and normalization of SIDIS data, but the description of high-energy DY data is very poor. Going to N<sup>2</sup>LL and N<sup>3</sup>LL<sup>-</sup>, the description of DY data significantly improves, but we fail to reproduce the normalization of SIDIS data, mainly due to the  $\mathcal{O}(\alpha_s)$  corrections to the hard factor. As a possible solution, we choose to adjust the normalization of the TMD predictions by comparing their integral upon transverse momentum and the corresponding collinear formula. With this procedure, we fix the normalization a priori, in a way that is independent of the results of the fit.

Our baseline fit is performed at N<sup>3</sup>LL<sup>-</sup>, using 2031 data points and obtaining  $\chi^2/N_{\text{dat}} = 1.06$ . We also discuss variations of this baseline fit by changing the theoretical accuracy and the selected data.

The paper is organized as follows. In section 2 we describe the theoretical framework used in the analysis. In section 3 we explain how experimental data have been selected. Section 4 presents the results of our extraction. Finally, in section 5 we draw our conclusions.

## 2 Formalism

### 2.1 Drell-Yan

In the Drell-Yan (DY) process

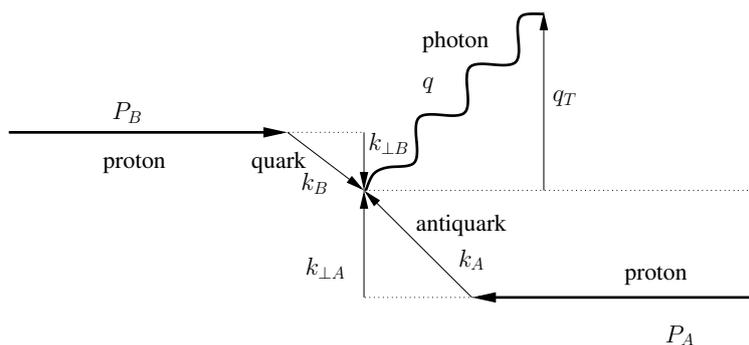
$$h_A(P_A) + h_B(P_B) \longrightarrow \gamma^*/Z(q) + X \longrightarrow \ell^+(l) + \ell^-(l') + X, \quad (2.1)$$

two hadrons  $h_A$  and  $h_B$  with four-momenta  $P_A$  and  $P_B$ , respectively, collide with center-of-mass energy squared  $s = (P_A + P_B)^2$ , producing a neutral vector boson  $\gamma^*/Z$  with four-momentum  $q$  and large invariant mass  $Q = \sqrt{q^2}$ . The vector boson eventually decays into a lepton and an antilepton with four-momenta constrained by momentum conservation,  $q = l + l'$ . The involved momenta and the respective transverse components are summarized in figure 1.

The hadronic four momenta  $P_A$  and  $P_B$  can be chosen to identify the longitudinal direction  $z$  and define the transverse momentum  $\mathbf{q}_T$  of the  $\gamma^*/Z$ . The rapidity of the neutral boson (or, equivalently, of the lepton pair) is defined as

$$y = \frac{1}{2} \ln \left( \frac{q_0 + q_z}{q_0 - q_z} \right). \quad (2.2)$$

For our purposes, we need the cross section for this process initiated by unpolarized hadrons and integrated over the azimuthal angle of the exchanged boson. That cross section can be written in terms of two structure functions,  $F_{UU}^1$  and  $F_{UU}^2$  [30, 31]. In the



**Figure 1.** Diagram describing the relevant momenta involved in a DY event. Two partons from two hadrons collide. They have transverse momenta  $\mathbf{k}_{\perp A}$  and  $\mathbf{k}_{\perp B}$  (not measured). They produce a virtual photon with (measured) transverse momentum  $\mathbf{q}_T = \mathbf{k}_{\perp A} + \mathbf{k}_{\perp B}$  with respect to the hadron collision axis.

limit  $M^2 \ll Q^2$  (with  $M$  the mass of the incoming hadrons) and  $\mathbf{q}_T^2 \ll Q^2$ , the  $F_{UU}^2$  is suppressed. Accordingly, the cross section reads

$$\frac{d\sigma^{\text{DY/Z}}}{d|\mathbf{q}_T| dy dQ} = \frac{16\pi^2 \alpha^2}{9Q^3} |\mathbf{q}_T| \mathcal{P} F_{UU}^1(x_A, x_B, |\mathbf{q}_T|, Q), \quad (2.3)$$

where  $\alpha$  is the electromagnetic coupling,  $\mathcal{P}$  is the phase space factor to account for potential cuts on the lepton kinematics, which turns out to have a relevant impact when high-precision data are taken into account (see, e.g., a recent analysis in ref. [32]).<sup>2</sup> At low transverse momentum  $\mathbf{q}_T^2 \ll Q^2$ , the structure function can be expressed as a convolution over the partonic transverse momenta of two TMD PDFs:

$$\begin{aligned} & F_{UU}^1(x_A, x_B, |\mathbf{q}_T|, Q) \\ &= x_A x_B \mathcal{H}^{\text{DY}}(Q, \mu) \sum_a c_a(Q^2) \int d^2\mathbf{k}_{\perp A} d^2\mathbf{k}_{\perp B} f_1^a(x_A, \mathbf{k}_{\perp A}^2; \mu, \zeta_A) f_1^{\bar{a}}(x_B, \mathbf{k}_{\perp B}^2; \mu, \zeta_B) \\ & \quad \delta^{(2)}(\mathbf{k}_{\perp A} + \mathbf{k}_{\perp B} - \mathbf{q}_T) \\ &= \frac{x_A x_B}{2\pi} \mathcal{H}^{\text{DY}}(Q, \mu) \sum_a c_a(Q^2) \int_0^{+\infty} d|\mathbf{b}_T| |\mathbf{b}_T| J_0(|\mathbf{b}_T| |\mathbf{q}_T|) \hat{f}_1^a(x_A, \mathbf{b}_T^2; \mu, \zeta_A) \\ & \quad \hat{f}_1^{\bar{a}}(x_B, \mathbf{b}_T^2; \mu, \zeta_B). \end{aligned} \quad (2.4)$$

In the above equation,  $\mathcal{H}^{\text{DY}}$  is the hard factor, which can be computed order by order in the strong coupling  $\alpha_s$  and is equal to 1 at leading order.<sup>3</sup> This function encodes the virtual part of the hard scattering and depends on the hard scale  $Q$  and on the renormalisation scale  $\mu$ . The unpolarized TMDs are denoted by  $f_1$ . They depend on the renormalization

<sup>2</sup>In the presence of cuts on single-lepton variables, an additional parity-violating term contributes to the cross section [33]. However, in ref. [7] it was shown that this contribution is negligible in the experimental conditions considered in this analysis.

<sup>3</sup>In the present work, we follow the definition of ref. [34].

scale  $\mu$  and the rapidity scale  $\zeta$ . The rapidity scales must obey the relation  $\zeta_A \zeta_B = Q^4$ . Throughout the paper, we will set  $\mu^2 = \zeta_A = \zeta_B = Q^2$ .

The following definition of the Fourier transform of the TMD PDFs has been used:<sup>4</sup>

$$\begin{aligned} \hat{f}_1^a(x, |\mathbf{b}_T|; \mu, \zeta) &= \int d^2 \mathbf{k}_\perp e^{i \mathbf{b}_T \cdot \mathbf{k}_\perp} f_1^a(x, \mathbf{k}_\perp^2; \mu, \zeta) \\ &= 2\pi \int_0^\infty d|\mathbf{k}_\perp| |\mathbf{k}_\perp| J_0(|\mathbf{b}_T| |\mathbf{k}_\perp|) f_1^a(x, \mathbf{k}_\perp^2; \mu, \zeta). \end{aligned} \quad (2.5)$$

The structure of the TMD PDFs will be addressed in details in section 2.3. The transverse momentum of the active quark and antiquark are denoted as  $\mathbf{k}_{\perp A, B}$ . At low transverse momenta, the two variables  $x_{A, B}$  take the values:

$$x_A = \frac{Q}{\sqrt{s}} e^y, \quad x_B = \frac{Q}{\sqrt{s}} e^{-y}. \quad (2.6)$$

The summation over  $a$  in eq. (2.4) runs over the active quarks and antiquarks at the scale  $Q$ , and  $c_a(Q^2)$  are the respective electroweak charges given by

$$c_a(Q^2) = e_a^2 - 2e_a V_a V_\ell \chi_1(Q^2) + (V_\ell^2 + A_\ell^2) (V_a^2 + A_a^2) \chi_2(Q^2), \quad (2.7)$$

with

$$\chi_1(Q^2) = \frac{1}{4 \sin^2 \theta_W \cos^2 \theta_W} \frac{Q^2(Q^2 - M_Z^2)}{(Q^2 - M_Z^2)^2 + M_Z^2 \Gamma_Z^2}, \quad (2.8)$$

$$\chi_2(Q^2) = \frac{1}{16 \sin^4 \theta_W \cos^4 \theta_W} \frac{Q^4}{(Q^2 - M_Z^2)^2 + M_Z^2 \Gamma_Z^2}, \quad (2.9)$$

where  $e_a$ ,  $V_a$ , and  $A_a$  are the electric, vector, and axial charges of the flavor  $a$ , respectively;  $V_\ell$  and  $A_\ell$  are the vector and axial charges of the lepton  $\ell$ ;  $\sin \theta_W$  is the weak mixing angle;  $M_Z$  and  $\Gamma_Z$  are mass and width of the  $Z$  boson.

As discussed in section 3 and summarized in table 2, for DY production the observable provided by the experimental collaborations is the (normalized) cross section differential with respect to  $|\mathbf{q}_T|$ . For each bin delimited by the initial ( $i$ ) and final ( $f$ ) values of kinematical variables, the experimental values are compared with the following theoretical quantity:

$$\mathcal{O}_{\text{DY}, 1}^{th}(|\mathbf{q}_T|_{i, f}, y_{i, f}, Q_{i, f}) = \int_{|\mathbf{q}_T|_i}^{|\mathbf{q}_T|_f} d|\mathbf{q}_T| \int_{y_i}^{y_f} dy \int_{Q_i}^{Q_f} dQ \frac{d\sigma^{\text{DY}/Z}}{d|\mathbf{q}_T| dy dQ}, \quad (2.10)$$

where the  $\int$  symbol represents the integral divided by the width of the integration range. Hence, eq. (2.10) corresponds to the cross section in eq. (2.3) averaged over the transverse momentum and integrated over rapidity and invariant mass of the exchanged boson. The normalized cross section is obtained by dividing both sides of eq. (2.10) by the appropriate fiducial cross section, which is computed by employing the DYNL0 code [35, 36].<sup>5</sup>

<sup>4</sup>Notice that in ref. [5] the Fourier transform was defined with an extra  $1/(2\pi)$  factor.

<sup>5</sup>See <https://www.physik.uzh.ch/en/groups/grazzini/research/Tools.html>.

The low-energy fixed-target experiments included in this analysis (E288, E605, E772, see table 2) measure the following cross section

$$E \frac{d\sigma^{\text{DY}}}{d^3\mathbf{q}} = \frac{1}{2\pi|\mathbf{q}_T|} \frac{d\sigma^{\text{DY}}}{d|\mathbf{q}_T| dy}, \quad (2.11)$$

where  $E$  and  $\mathbf{q}$  are the energy and the three-momentum of the photon, respectively.

Given eq. (2.11), in principle the experimental value in a given bin needs to be compared against the following theoretical quantity:

$$\mathcal{O}_{\text{DY,ft}}^{\text{th}}(|\mathbf{q}_T|_{i,f}, y_{i,f}, Q_{i,f}) = \int_{|\mathbf{q}_T|_i}^{|\mathbf{q}_T|_f} d|\mathbf{q}_T| \int_{y_i}^{y_f} dy \int_{Q_i}^{Q_f} dQ \frac{1}{2\pi|\mathbf{q}_T|} \frac{d\sigma^{\text{DY}}}{d|\mathbf{q}_T| dy dQ}. \quad (2.12)$$

However, since all the considered fixed-target experiments do not provide bins of  $|\mathbf{q}_T|$  but just the average transverse momentum values  $|\overline{\mathbf{q}_T}|$ , the integration over  $|\mathbf{q}_T|$  is not considered. Moreover, the E288 provides only the average value  $\bar{y}$  for the rapidity. Accordingly, the theoretical quantity considered for that experiment reads

$$\mathcal{O}_{\text{DY,E288}}^{\text{th}}(|\overline{\mathbf{q}_T}|, \bar{y}, Q_{i,f}) = \frac{1}{2\pi|\overline{\mathbf{q}_T}|} \int_{Q_i}^{Q_f} dQ \frac{d\sigma^{\text{DY}}}{d|\mathbf{q}_T| dy dQ} \Big|_{y=\bar{y}, |\mathbf{q}_T|=|\overline{\mathbf{q}_T}|}. \quad (2.13)$$

The E605 and E772 low-energy fixed-target experiments (see table 2) use, in place of the rapidity  $y$ , the variable  $x_F$ , which is connected to the other kinematic variables as follows:

$$y(x_F, Q) = \sinh^{-1} \left( \frac{\sqrt{s} x_F}{Q} \right), \quad x_A = \sqrt{\frac{Q^2}{s} + \frac{x_F^2}{4}} + \frac{x_F}{2}, \quad x_B = x_A - x_F. \quad (2.14)$$

Using eq. (2.14), one obtains

$$E \frac{d\sigma^{\text{DY}}}{d^3\mathbf{q}} = \frac{2E}{\pi\sqrt{s}} \frac{d\sigma^{\text{DY}}}{d\mathbf{q}_T^2 dx_F}. \quad (2.15)$$

The E772 collaboration provides bins in  $x_F$  and average transverse momentum values  $|\overline{\mathbf{q}_T}|$ . Accordingly, in that case the experimental values are compared against the following theoretical quantity:

$$\begin{aligned} \mathcal{O}_{\text{DY,E772}}^{\text{th}}(|\overline{\mathbf{q}_T}|, x_{F_i,f}, Q_{i,f}) &= \int_{Q_i}^{Q_f} dQ \int_{x_{F_i}}^{x_{F_f}} dx_F \frac{2E}{\pi\sqrt{s}} \frac{d\sigma^{\text{DY}}}{d\mathbf{q}_T^2 dx_F dQ} \Big|_{|\mathbf{q}_T|=|\overline{\mathbf{q}_T}|} \\ &\approx \frac{\overline{Q} \cosh(\bar{y})}{\pi|\overline{\mathbf{q}_T}| \sqrt{s} (x_{F_f} - x_{F_i})} \int_{y(x_{F_i}, \overline{Q})}^{y(x_{F_f}, \overline{Q})} dy \int_{Q_i}^{Q_f} dQ \frac{d\sigma^{\text{DY}}}{d|\mathbf{q}_T| dy dQ} \Big|_{|\mathbf{q}_T|=|\overline{\mathbf{q}_T}|}, \end{aligned} \quad (2.16)$$

where

$$\overline{Q} = (Q_i + Q_f)/2, \quad \bar{y} = [y(x_{F_i}, \overline{Q}) + y(x_{F_f}, \overline{Q})]/2. \quad (2.17)$$

For the sake of simplicity, we replaced  $y$  and  $Q$  with  $\bar{y}$  and  $\overline{Q}$  in the prefactor in front of the cross section and pull it out of the integral.

The E605 experiment, instead, provides average values for both transverse momentum and  $x_F$  and its data are compared against the following theoretical quantity:

$$\mathcal{O}_{\text{DY, E605}}^{th}(|\bar{\mathbf{q}}_T|, \bar{x}_F, Q_{i,f}) \approx \frac{\bar{Q} \cosh(\bar{y})}{\pi |\bar{\mathbf{q}}_T| \sqrt{s}} \int_{Q_i}^{Q_f} dQ \frac{d\sigma^{\text{DY}}}{d|\mathbf{q}_T| dy dQ} \Big|_{|\mathbf{q}_T|=|\bar{\mathbf{q}}_T|, y=\bar{y}}, \quad (2.18)$$

where, in this case,  $\bar{y} = y(\bar{x}_F, \bar{Q})$  with  $\bar{x}_F = (x_{Fi} + x_{Ff})/2$ .

## 2.2 Semi-Inclusive Deep-Inelastic Scattering (SIDIS)

In SIDIS, a lepton with momentum  $l$  scatters off a hadron target  $N$  with mass  $M$  and four momentum  $P$ . In the final state, the scattered lepton momentum  $l'$  is measured together with one hadron  $h$  with mass  $M_h$  and four momentum  $P_h$ . The other products of the scattering are undetected. Thus the reaction reads

$$\ell(l) + N(P) \rightarrow \ell(l') + h(P_h) + X. \quad (2.19)$$

The (space-like) four-momentum transfer is  $q = l - l'$ , with  $Q^2 \equiv -q^2 > 0$ . We use the standard SIDIS variables [37, 38]:

$$x = \frac{Q^2}{2P \cdot q}, \quad y = \frac{P \cdot q}{P \cdot l}, \quad z = \frac{P \cdot P_h}{P \cdot q}, \quad \gamma = \frac{2Mx}{Q}. \quad (2.20)$$

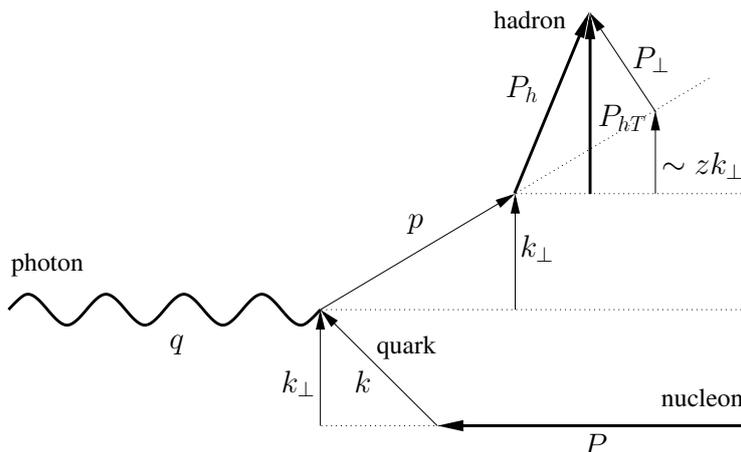
For transverse momenta, we will follow the definitions and notations discussed in refs. [39, 40] (see also figure 2). In particular, we define  $\mathbf{P}_{hT}$  as the hadron transverse momentum in the Breit frame, where  $P$  and  $q$  form a light-cone basis; as a consequence,  $P_h$  has only transverse components and  $|\mathbf{P}_{hT}|^2 = -P_{hT}^2$ , which is frame independent. Similarly, we define  $\mathbf{q}_T$  as the photon transverse momentum in the hadron frame, where  $P$  and  $P_h$  form a light-cone basis; as a consequence,  $q_T$  has only transverse components and  $|\mathbf{q}_T|^2 = -q_T^2$ , which is also frame independent. The two momenta are related by [41, 42]

$$q_T^\mu = -\frac{P_{hT}^\mu}{z} - 2x \frac{|\mathbf{q}_T|^2}{Q^2} P^\mu. \quad (2.21)$$

In the following, we will always work assuming that the invariant mass of the photon is large compared to the target and hadron masses ( $M^2, M_h^2 \ll Q^2$ ) and to the transverse momenta  $\mathbf{q}_T$  and  $\mathbf{P}_{hT}$  ( $q_T^2, P_{hT}^2 \ll Q^2$ ). We neglect any power corrections that vanish in this limit, both kinematic and dynamical (higher twist), apart from some modifications to the normalization of the SIDIS observables (that could be seen as the effect of power corrections, see section 2.4 for more details). In this limit, eq. (2.21) reduces to

$$\mathbf{q}_T \approx -\frac{\mathbf{P}_{hT}}{z}. \quad (2.22)$$

Several studies have been made concerning higher-twist corrections of various origin (see, e.g., refs. [38, 43, 44] for recent works). A careful study of the impact of power corrections to the case of unpolarized TMDs has been discussed in ref. [22]. Lately, important advances in the study of higher-twist TMD factorization have been published [42, 45–47], but they do not directly affect the observables we consider here.



**Figure 2.** Diagram describing the relevant momenta involved in a SIDIS event in the Breit (nucleon-photon) frame. A virtual photon with momentum  $q$  (defining the reference axis) strikes a parton with momentum  $k$  inside a nucleon with momentum  $P$ . The parton has a transverse momentum  $\mathbf{k}_\perp$  (not measured). The struck parton with momentum  $p = k + q$  fragments into a hadron with momentum  $P_h$ , which acquires a further transverse momentum  $\mathbf{P}_\perp$  (not measured) with respect to the fragmenting quark axis. The total measured transverse-momentum of the final hadron is  $\mathbf{P}_{hT}$ . When  $Q^2$  is very large, the longitudinal components are all much larger than the transverse components. In this regime,  $\mathbf{P}_{hT} \approx z\mathbf{k}_\perp + \mathbf{P}_\perp$ .

The differential cross section for SIDIS can be written in terms of two structure functions,  $F_{UU,T}$  and  $F_{UU,L}$  [37]. The subscripts refer to the lepton, the target, and the photon polarization, respectively. The second structure function is formally a twist four contribution and is suppressed in the limit considered here, thus we neglect it. The differential cross section at small transverse momentum [5, 37] reads

$$\frac{d\sigma^{\text{SIDIS}}}{dx dz d|\mathbf{q}_T| dQ} = \frac{8\pi^2 \alpha^2 z^2 |\mathbf{q}_T|}{x Q^3} \frac{y^2}{2(1-\epsilon)} \left(1 + \frac{\gamma^2}{2x}\right) F_{UU,T}(x, z, |\mathbf{q}_T|, Q), \quad (2.23)$$

where  $\alpha$  is the QED coupling constant and  $\epsilon$  is the photon flux factor [37]. Neglecting target mass corrections  $\mathcal{O}(\gamma)$  and  $\mathcal{O}(\gamma^2)$ , the prefactors can be approximated as

$$y \approx \frac{Q^2}{xs}, \quad \frac{y^2}{(1-\epsilon)} \left(1 + \frac{\gamma^2}{2x}\right) \approx Y_+ = 1 + \left(1 - \frac{Q^2}{xs}\right)^2. \quad (2.24)$$

Since we are interested only in the small-transverse-momentum limit, in eq. (2.23) we have neglected the contributions from fixed-order calculations at high  $|\mathbf{q}_T|$  [48, 49] and the matching of these on TMD factorization [50–52].

The unpolarized SIDIS structure function  $F_{UU,T}$  is defined as [37]

$$\begin{aligned}
F_{UU,T}(x, z, |\mathbf{q}_T|, Q) &= x \mathcal{H}^{SIDIS}(Q, \mu) \sum_a e_a^2 \int d^2 \mathbf{k}_\perp \int \frac{d^2 \mathbf{P}_\perp}{z^2} f_1^a(x, \mathbf{k}_\perp^2; \mu, \zeta_A) D_1^{a \rightarrow h}(z, \mathbf{P}_\perp^2; \mu, \zeta_B) \\
&\quad \delta^{(2)}(\mathbf{k}_\perp + \mathbf{P}_\perp/z + \mathbf{q}_T) \\
&= \frac{x}{2\pi} \mathcal{H}^{SIDIS}(Q, \mu) \sum_a e_a^2 \int_0^{+\infty} d|\mathbf{b}_T| |\mathbf{b}_T| J_0(|\mathbf{b}_T| |\mathbf{q}_T|) \hat{f}_1^a(x, \mathbf{b}_T^2; \mu, \zeta_A) \\
&\quad \hat{D}_1^{a \rightarrow h}(z, \mathbf{b}_T^2; \mu, \zeta_B),
\end{aligned} \tag{2.25}$$

where the sum runs over quarks and antiquarks  $a$ . The hard factor  $\mathcal{H}^{SIDIS}$  can be computed order by order in the strong coupling  $\alpha_s$  and is equal to 1 at leading order.<sup>6</sup> The variable  $\mathbf{k}_\perp$  is the transverse momentum of the struck quark with respect to the nucleon axis, whereas  $\mathbf{P}_\perp$  is the transverse momentum of the produced hadron  $h$  with respect to the fragmenting quark axis (see figure 2).

The variable  $\mathbf{b}_T$  is conjugated via Fourier transform to the transverse momentum  $\mathbf{q}_T$ .  $f_1^a(x, \mathbf{k}_\perp^2; \mu, \zeta_A)$  and  $D_1^{a \rightarrow h}(z, \mathbf{P}_\perp^2; \mu, \zeta_B)$  are the unpolarized TMD PDF for a quark  $a$  in a nucleon and the unpolarized TMD FF for a quark with flavor  $a$  fragmenting into a hadron with flavor  $h$ , respectively;  $\hat{f}_1^a(x, \mathbf{b}_T^2; \mu, \zeta_A)$  and  $\hat{D}_1^{a \rightarrow h}(z, \mathbf{b}_T^2; \mu, \zeta_B)$  are their Fourier transforms. The former is defined in eq. (2.5), the latter is defined as

$$\begin{aligned}
\hat{D}_1^{a \rightarrow h}(z, \mathbf{b}_T^2; \mu, \zeta) &= \int \frac{d^2 \mathbf{P}_\perp}{z^2} e^{-i \mathbf{b}_T \cdot \mathbf{P}_\perp / z} D_1^a(z, \mathbf{P}_\perp^2; \mu, \zeta) \\
&= 2\pi \int_0^\infty \frac{d|\mathbf{P}_\perp|}{z^2} |\mathbf{P}_\perp| J_0(|\mathbf{b}_T| |\mathbf{P}_\perp| / z) D_1^a(z, \mathbf{P}_\perp^2; \mu, \zeta).
\end{aligned} \tag{2.26}$$

Their structure will be discussed in details in section 2.3.

The observable provided by the HERMES and COMPASS collaborations is the multiplicity, namely the ratio of the one-hadron inclusive cross section as a function of the transverse momentum of the hadron  $|\mathbf{P}_{hT}|$  over the fully inclusive one:

$$M(x, z, |\mathbf{P}_{hT}|, Q) = \frac{d\sigma^{SIDIS}}{dx dz d|\mathbf{P}_{hT}| dQ} \bigg/ \frac{d\sigma^{DIS}}{dx dQ} = \frac{1}{z} \frac{d\sigma^{SIDIS}}{dx dz d|\mathbf{q}_T| dQ} \bigg/ \frac{d\sigma^{DIS}}{dx dQ}. \tag{2.27}$$

The cross section for unpolarized DIS in the denominator of the multiplicities reads

$$\frac{d\sigma^{DIS}}{dx dQ} = \frac{8\pi \alpha_s^2}{x Q^3} \frac{y^2}{2(1-\epsilon)} \left[ F_T(x, Q^2) + \epsilon F_L(x, Q^2) \right] \approx \frac{4\pi \alpha_s^2}{x Q^3} \left[ Y_+ F_2(x, Q^2) - y^2 F_L(x, Q^2) \right], \tag{2.28}$$

where the approximation is justified by neglecting the target mass corrections. At the perturbative order considered in this analysis, the longitudinal DIS structure function  $F_L$  cannot be neglected, at variance with, e.g., refs. [5, 15].

---

<sup>6</sup>In the present work, we follow the definition of ref. [34].

The experimental values in each bin are compared against the quantity built by *separately* averaging the numerator and denominator of the multiplicity in eq. (2.27) over the respective kinematics:

$$\mathcal{O}_{\text{SIDIS}}^{\text{th}}(x_{i,f}, z_{i,f}, |\mathbf{P}_{hT}|_{i,f}, Q_{i,f}) = \int_{Q_i}^{Q_f} dQ \int_{x_i}^{x_f} dx \int_{z_i}^{z_f} dz \int_{|\mathbf{P}_{hT}|_i}^{|\mathbf{P}_{hT}|_f} d|\mathbf{P}_{hT}| \frac{d\sigma^{\text{SIDIS}}}{dx dz d|\mathbf{P}_{hT}| dQ}, \quad (2.29)$$

$$\mathcal{O}_{\text{DIS}}^{\text{th}}(x_{i,f}, Q_{i,f}) = \int_{Q_i}^{Q_f} dQ \int_{x_i}^{x_f} dx \frac{d\sigma^{\text{DIS}}}{dx dQ},$$

$$\mathcal{M}^{\text{th}}(x_{i,f}, z_{i,f}, |\mathbf{P}_{hT}|_{i,f}, Q_{i,f}) = \mathcal{O}_{\text{SIDIS}}^{\text{th}}(x_{i,f}, z_{i,f}, |\mathbf{P}_{hT}|_{i,f}, Q_{i,f}) / \mathcal{O}_{\text{DIS}}^{\text{th}}(x_{i,f}, Q_{i,f}). \quad (2.30)$$

The HERMES collaboration provides multiplicities in bins of  $|\mathbf{P}_{hT}|$ , whereas the COMPASS collaboration in bins of  $\mathbf{P}_{hT}^2$  (see also table 3). In both cases, the observable can be calculated as in eq. (2.29), but in the COMPASS case the average is on  $\mathbf{P}_{hT}^2$ . Moreover, both collaborations introduce a cut on the invariant mass of the hadronic final states  $W^2 = (P + q)^2$  (see table 3), which makes the upper integration limit  $x_f$  a  $Q$ -dependent quantity.

### 2.3 Transverse Momentum Distributions (TMDs)

As a consequence of the renormalization of ultraviolet and rapidity divergences [6, 53, 54], TMD PDFs and FFs acquire a dependence on the renormalization scale  $\mu$  and on the rapidity scale  $\zeta$ . The evolution of TMDs from some initial values of the scales  $\mu_i, \zeta_i$ , to some final values  $\mu_f, \zeta_f$ , is given by

$$\hat{f}_1^a(x, \mathbf{b}_T^2; \mu_f, \zeta_f) = \hat{f}_1^a(x, \mathbf{b}_T^2; \mu_i, \zeta_i) \exp \left\{ \int_{\mu_i}^{\mu_f} \frac{d\mu}{\mu} \gamma(\mu, \zeta_f) \right\} \left( \frac{\zeta_f}{\zeta_i} \right)^{K(|\mathbf{b}_T|, \mu_i)/2}. \quad (2.31)$$

The anomalous dimension  $\gamma$  for the renormalization-group evolution in  $\mu$  reads:

$$\gamma(\mu, \zeta) = \gamma_F(\alpha_s(\mu)) - \gamma_K(\alpha_s(\mu)) \ln \frac{\sqrt{\zeta}}{\mu}, \quad (2.32)$$

where  $\gamma_K$  is the cusp anomalous dimension and  $\gamma_F(\alpha_s(\mu)) = \gamma(\mu, \mu^2)$  is the boundary condition [7]. The Collins-Soper kernel  $K$ , instead, is the anomalous dimension for the evolution in  $\zeta$  [6]. The same structure holds for the TMD FF. In order to avoid the insurgence of large logarithms, the scales  $\mu_i$  and  $\zeta_i$  are conveniently fixed as  $\mu_i = \sqrt{\zeta_i} = \mu_b = 2e^{-\gamma_E} / |\mathbf{b}_T|$ . Since the coupling  $\alpha_s$  is computed at this scale (see eq. (2.32)) the evolution of the TMD is perturbatively meaningful only at low values of  $|\mathbf{b}_T|$  such that the scale  $\mu_b$  is sufficiently larger than the Landau pole  $\Lambda_{\text{QCD}}$ . This condition can be implemented by replacing the scale  $\mu_b$  with  $\mu_{b_*} = 2e^{-\gamma_E} / b_*$ , where [5]

$$b_*(|\mathbf{b}_T|, b_{\min}, b_{\max}) = b_{\max} \left( \frac{1 - e^{-|\mathbf{b}_T|^4 / b_{\max}^4}}{1 - e^{-|\mathbf{b}_T|^4 / b_{\min}^4}} \right)^{1/4}, \quad (2.33)$$

with

$$b_{\max} = 2e^{-\gamma_E} \text{ GeV}^{-1} \approx 1.123 \text{ GeV}^{-1}, \quad b_{\min} = 2e^{-\gamma_E} / \mu_f. \quad (2.34)$$

As suggested by the CSS formalism [6],  $b_*$  saturates to  $b_{\max}$  at large  $|\mathbf{b}_T|$  guaranteeing that  $\mu_{b_*}$  never enters the nonperturbative regime. However, this has also the effect of introducing power corrections scaling like  $(\Lambda_{\text{QCD}}/|\mathbf{q}_T|)^k$  [55], with  $k > 0$ , that in the region  $|\mathbf{q}_T| \simeq \Lambda_{\text{QCD}}$  need to be accounted for by a nonperturbative function. At small  $|\mathbf{b}_T|$ ,  $b_*$  saturates to  $b_{\min}$ . Since  $\mu_f$  is of the order of the boson virtuality  $Q$ , this introduces subleading power corrections scaling like  $(|\mathbf{q}_T|/Q)^n$ , with  $n > 0$ . Such a procedure has the advantage of facilitating a possible matching of the TMD formula, valid for  $|\mathbf{q}_T| \ll Q$ , onto the fixed-order calculation valid for  $|\mathbf{q}_T| \simeq Q$  [8, 56, 57]. Accordingly, in the limit  $|\mathbf{b}_T| \rightarrow 0$  the Sudakov exponent vanishes.

Performing at the input scales the Operator Product Expansion (OPE) of the TMD PDFs (TMD FFs) around  $|\mathbf{b}_T| = 0$  one gets:

$$\hat{f}_1^a(x, b_*; \mu_{b_*}, \mu_{b_*}^2) = \sum_b \int_x^1 \frac{dx'}{x'} C^{ab}(x', b_*; \mu_{b_*}, \mu_{b_*}^2) f_1^b\left(\frac{x}{x'}; \mu_{b_*}\right) \equiv [C \otimes f_1](x, b_*; \mu_{b_*}, \mu_{b_*}^2), \quad (2.35)$$

where the sum runs over quarks, antiquarks, and the gluon. The matching coefficients  $C$  are calculated as a perturbative expansion in powers of  $\alpha_s$ .

In view of the power corrections introduced by the  $b_*$  prescription, both the Collins-Soper kernel  $K$  [54] and the OPE in eq. (2.35) need to be modified to account for nonperturbative effects. For the Collins-Soper kernel  $K$ , this results in a nonperturbative correction term,  $g_K(\mathbf{b}_T^2)$ , for which we choose a specific functional form:

$$K(|\mathbf{b}_T|, \mu_{b_*}) = K(b_*, \mu_{b_*}) + g_K(|\mathbf{b}_T|), \quad g_K(\mathbf{b}_T^2) = -g_2^2 \frac{\mathbf{b}_T^2}{2}. \quad (2.36)$$

This correction gives rise in the evolution to a nonperturbative factor that goes like  $(\zeta_f/Q_0^2)^{g_K/2}$  where  $Q_0$  is an arbitrary scale at which this correction is parameterised; we set  $Q_0 = 1 \text{ GeV}$ . In order not to affect the perturbative calculation at small  $|\mathbf{b}_T|$ , the term  $g_K$  needs to vanish in the limit  $|\mathbf{b}_T| \rightarrow 0$ . The nonperturbative corrections to the OPE can also be parameterised by a multiplicative function that generally depends on  $x$  or  $z$  and  $\mathbf{b}_T$ . The net result of the inclusion of the nonperturbative corrections into the evolved TMD PDF reads:

$$\begin{aligned} & \hat{f}_1^a(x, \mathbf{b}_T^2; \mu_f, \zeta_f) \\ &= [C \otimes f_1](x, b_*; \mu_{b_*}, \mu_{b_*}^2) \exp \left\{ \int_{\mu_{b_*}}^{\mu_f} \frac{d\mu}{\mu} \gamma(\mu, \zeta_f) \right\} \left( \frac{\zeta_f}{\mu_{b_*}^2} \right)^{K(b_*, \mu_{b_*})/2} f_{1NP}(x, \mathbf{b}_T^2; \zeta_f, Q_0), \end{aligned} \quad (2.37)$$

and the same holds for the TMD FF where one introduces  $D_{1NP}(z, \mathbf{b}_T^2; \zeta, Q_0)$ . Note that the number of active flavors  $n_f$  in the perturbative quantities  $\gamma$ ,  $C$ , and the hard function  $\mathcal{H}$  of eqs. (2.4) and (2.25), is separately determined by the scales  $\mu$ ,  $\mu_{b_*}$  and  $Q$ , respectively. To be more precise, given a set of quark thresholds  $\{m_1, m_2, m_3, \dots\}$ , the  $n_f$  associated to each of the three scales above is computed by requiring that the scale lies between  $m_{n_f}$  and  $m_{n_f+1}$ . Analogously, since the collinear distributions involved in the matching formula in eq. (2.35) are computed at the scale  $\mu_{b_*}$ , the value of  $n_f$  for PDFs and FFs is chosen accordingly, i.e. it is the same used for the matching functions  $C$ .

The  $f_{1NP}$  and  $D_{1NP}$  factors (which we assume to be flavor-independent) incorporate both the correction to the evolution associated to the  $g_K$  function and the correction to the respective OPE. For the TMD PDF we define

$$\begin{aligned}
 & f_{1NP}(x, \mathbf{b}_T^2; \zeta, Q_0) \\
 &= \frac{g_1(x) e^{-g_1(x) \frac{\mathbf{b}_T^2}{4}} + \lambda^2 g_{1B}^2(x) \left[ 1 - g_{1B}(x) \frac{\mathbf{b}_T^2}{4} \right] e^{-g_{1B}(x) \frac{\mathbf{b}_T^2}{4}} + \lambda_2^2 g_{1C}(x) e^{-g_{1C}(x) \frac{\mathbf{b}_T^2}{4}}}{g_1(x) + \lambda^2 g_{1B}^2(x) + \lambda_2^2 g_{1C}(x)} \left[ \frac{\zeta}{Q_0^2} \right]^{g_K(\mathbf{b}_T^2)/2},
 \end{aligned} \tag{2.38}$$

and for the TMD FF the form is

$$D_{1NP}(z, \mathbf{b}_T^2; \zeta, Q_0) = \frac{g_3(z) e^{-g_3(z) \frac{\mathbf{b}_T^2}{4z^2}} + \frac{\lambda_F}{z^2} g_{3B}^2(z) \left[ 1 - g_{3B}(z) \frac{\mathbf{b}_T^2}{4z^2} \right] e^{-g_{3B}(z) \frac{\mathbf{b}_T^2}{4z^2}}}{g_3(z) + \frac{\lambda_F}{z^2} g_{3B}^2(z)} \left[ \frac{\zeta}{Q_0^2} \right]^{g_K(\mathbf{b}_T^2)/2}. \tag{2.39}$$

The nonperturbative factors  $f_{1NP}$ ,  $D_{1NP} \rightarrow 1$  for  $\mathbf{b}_T \rightarrow 0$ . The  $g_i$  functions describe the dependence of the widths of the distributions on  $x$  and  $z$ :

$$g_{\{1,1B,1C\}}(x) = N_{\{1,1B,1C\}} \frac{x^{\sigma_{\{1,2,3\}}} (1-x)^{\alpha_{\{1,2,3\}}^2}}{\hat{x}^{\sigma_{\{1,2,3\}}} (1-\hat{x})^{\alpha_{\{1,2,3\}}^2}}, \tag{2.40}$$

$$g_{\{3,3B\}}(z) = N_{\{3,3B\}} \frac{(z^{\beta_{\{1,2\}}} + \delta_{\{1,2\}}^2)(1-z)^{\gamma_{\{1,2\}}^2}}{(\hat{z}^{\beta_{\{1,2\}}} + \delta_{\{1,2\}}^2)(1-\hat{z})^{\gamma_{\{1,2\}}^2}}, \tag{2.41}$$

where  $\hat{x} = 0.1$ ,  $\hat{z} = 0.5$ .

In total, the default configuration for the fit involves 21 free parameters: one associated to the nonperturbative part of the Collins-Soper kernel (eq. (2.36)), 11 related to the nonperturbative part of the TMD PDF (eqs. (2.38), (2.40)), and 9 for the nonperturbative part of the TMD FF (eqs. (2.39), (2.41)).

The functional forms in eqs. (2.38)–(2.41) are largely arbitrary. However, an important feature is that they are the Fourier transforms of the sum of a Gaussian, a weighted Gaussian (multiplied by  $\mathbf{k}_\perp^2$ ) and, in the case of the TMD PDFs, a third Gaussian. They are therefore positive definite for all values of  $\mathbf{k}_\perp^2$ .<sup>7</sup> The parameters  $\lambda$  and  $\lambda_2$  in eq. (2.38) are squared in order to avoid negative contributions (there is no need to square the parameter  $\lambda_F$  in eq. (2.39) because the fit always selects positive values for this parameter). The widths of the Gaussians, expressed by eqs. (2.40), (2.41), are  $x$  (or  $z$ ) dependent and vanish as  $x$  (or  $z$ ) approaches 1. Our choice of the functional form is also inspired by model calculations of TMD PDFs (see, e.g., [58–65]) and TMD FFs (see, e.g., [66, 67]). Many of these models predict the existence of terms that behave similarly to Gaussians and

<sup>7</sup>Note, however, that the evolved TMD PDF, eq. (2.37), can become negative at large values of transverse momentum. The same holds true for TMD FFs.

Accuracy	$H$ and $C$	$K$ and $\gamma_F$	$\gamma_K$	PDF and $\alpha_s$ evolution	FF evolution
NLL	0	1	2	LO	LO
N <sup>2</sup> LL	1	2	3	NLO	NLO
N <sup>3</sup> LL <sup>-</sup>	2	3	4	NNLO	NLO
N <sup>3</sup> LL	2	3	4	NNLO	NNLO

**Table 1.** Truncation orders in the expansions of the perturbative ingredients of TMDs relevant to the logarithmic counting considered in this paper (see text). The last column refers to the order used for the evolution of the collinear FFs.

weighted Gaussians. The details of the functional dependence predicted by the models are related to the correlation between the spin of the quarks and their transverse momentum. In the case of fragmentation functions, a different role can be played by different production channels (e.g., direct production vs. production through the decay of hadronic resonances).

Finally, for the logarithmic ordering we use the same convention adopted in ref. [7]. In particular, the orders of truncation of the perturbative ingredients relevant to the present analysis are summarized in table 1. At the time of this analysis, the full N<sup>3</sup>LL accuracy could not be achieved because NNLO collinear FFs were not available. Very recently, two analyses of collinear FFs were presented in refs. [28, 29] making an extraction of TMDs at full N<sup>3</sup>LL possible. We leave this study to a future publication.

## 2.4 Normalization factors for SIDIS

In ref. [5] it was demonstrated that TMD factorization at NLL accuracy is able to successfully reproduce the normalization and shape of HERMES SIDIS multiplicities and the shape of the available COMPASS multiplicities. More recently, the COMPASS collaboration published a reanalysis of their data [68]. The NLL TMD predictions correctly reproduce normalization and shape of the new data [69]. However, when increasing the accuracy to N<sup>2</sup>LL or higher, the TMD formula severely underestimates the measurements [69, 70] by nearly constant factors in each bin. Note that tensions between the TMD cross sections and the associated measurements exist also at large transverse momentum in SIDIS [48], DY [71], and electron-positron annihilation into two hadrons [72].

In the present study, as will be shown in section 4, we confirm that we obtain an excellent description of both normalization and shape of the SIDIS multiplicities at NLL and that the N<sup>2</sup>LL results are much smaller. At N<sup>3</sup>LL the results increase slightly, but they are still far from the NLL ones and, therefore, from data. At average kinematics of the COMPASS measurements, we obtain the following ratios of multiplicities:

$$\frac{M_{\text{NLL}}}{M_{\text{N}^2\text{LL}}} \gtrsim 2, \quad \frac{M_{\text{NLL}}}{M_{\text{N}^3\text{LL}}} \gtrsim 1.5. \quad (2.42)$$

The reason for the difference between the logarithmic orders is almost entirely due to the hard factor in eq. (2.25) [34]. If we look at the explicit expression for the hard factor<sup>8</sup>

<sup>8</sup>There are different definitions for the hard factor in the literature, which are compensated by different definitions of the matching coefficients  $C$  in eq. (2.35). Here we follow the definition of ref. [34].

with the standard choice  $\mu = Q$

$$\mathcal{H}^{\text{SIDIS}}(Q, Q) = 1 + \frac{\alpha_s(Q)}{4\pi} C_F \left( -16 + \frac{\pi^2}{3} \right), \quad (2.43)$$

we can immediately see that just by introducing  $\mathcal{O}(\alpha_s)$  corrections, at  $Q = 2 \text{ GeV}$  with  $\alpha_s \approx 0.3$ , we reduce the structure function to about 60% of its original value. This change is not compensated by a similar reduction in the denominator of eq. (2.27): the differences between the LO and NLO expressions of the inclusive DIS cross section are typically below 5% and the NLO results are actually larger than the LO ones.

If the NLL expression is much larger than the N<sup>2</sup>LL and N<sup>3</sup>LL ones, we may suspect that it should overshoot the data by a factor 1.5 at least. However, several works before the present one have shown a good agreement with data using a parton-model approach [15, 16] or at NLL [17, 18]. Moreover, the integral of the structure function over  $\mathbf{q}_T$  is equal to the value of its Fourier transform at  $\mathbf{b}_T = 0$ . Using any  $b^*$  prescription with  $b_{\min} = 2e^{-\gamma_E}/Q$ , the integral of the NLL expression by construction corresponds to the LO expression of the collinear SIDIS structure function, independent of the TMD nonperturbative parameters:<sup>9</sup>

$$\begin{aligned} & \int d^2\mathbf{q}_T F_{UU,T}(x, z, |\mathbf{q}_T|, Q) \\ &= x \sum_a e_a^2 \mathcal{H}^{\text{SIDIS}}(Q, Q) \left( \hat{f}_1^a(x, \mathbf{b}_T^2; Q, Q^2) \hat{D}_1^{a \rightarrow h}(z, \mathbf{b}_T^2; Q, Q^2) \right) \Big|_{|\mathbf{b}_T|=0} \\ &\stackrel{\text{NLL}}{=} x \sum_a e_a^2 f_1^a(x; Q) D_1^{a \rightarrow h}(z; Q). \end{aligned} \quad (2.44)$$

The LO collinear SIDIS predictions are known to describe the data reasonably well and, if anything, they seem to be lower than the data [73, 74]. Therefore, the integral of our NLL expression is in good agreement with data, which also indicates the absence of a large normalization error.

If the NLL predictions describe the data well and are a factor 2 or 1.5 above the N<sup>2</sup>LL and N<sup>3</sup>LL predictions, we propose to modify the normalization of the latter to recover a good agreement with data. An extended discussion of this issue can be found in ref. [69]. We observe that the integral of the TMD formula, valid at low  $\mathbf{q}_T$ , should reproduce only part of the full collinear cross section. The only exception is the order  $\mathcal{O}(\alpha_s^0)$  case, as we have seen above, since at that order there is no contribution from gluon radiation at high transverse momentum, beyond the TMD region.

However, at N<sup>2</sup>LL or higher, in the kinematics of fixed-target SIDIS experiments, the integral of the TMD region (i.e., the integral of the so-called  $W$  term in the language of ref. [6]) is much smaller than the corresponding collinear cross section. The missing contribution to the integral should be recovered by the terms in the fixed-order calculation that are not included in the TMD resummed expression (the so-called  $Y$  term). Ideally, the  $Y$  term should be negligible in the low- $\mathbf{q}_T$  region. This is not the case in the experimental regions under consideration: the  $Y$  term is finite but relatively large, even at  $\mathbf{q}_T = 0$ .

---

<sup>9</sup>Note that in the absence of a  $b_{\min}$  prescription, the integral would vanish.

If we consider the contribution to the integral of the  $W$  term (i.e., the integral of eq. (2.23)), with our  $b_{\min}$  prescription, at order  $\alpha_s$  we obtain, schematically

$$\begin{aligned} & \int d^2 \mathbf{q}_T W \Big|_{\mathcal{O}(\alpha_s)} \\ &= \sigma_0 \frac{\alpha_s}{4\pi} \sum_q e_q^2 \left[ D_1^{q \rightarrow h} \otimes C_{\text{TMD}}^{qq} \otimes f_1^q + D_1^{q \rightarrow h} \otimes C_{\text{TMD}}^{gg} \otimes f_1^g + D_1^{g \rightarrow h} \otimes C_{\text{TMD}}^{gg} \otimes f_1^q \right] (x, z, Q), \end{aligned} \quad (2.45)$$

where

$$\sigma_0 = \frac{4\pi^2 \alpha^2 z^2 |\mathbf{q}_T|}{x Q^3} Y_+. \quad (2.46)$$

The double convolution over both  $x$  and  $z$  is defined as

$$\left[ D_1^{a \rightarrow h} \otimes C^{ab} \otimes f_1^b \right] (x, z, Q) = \frac{1}{z^2} \int_x^1 \frac{dx'}{x'} \int_z^1 \frac{dz'}{z'} D_1^{a \rightarrow h}(z'; Q) C^{ab} \left( \frac{x}{x'}, \frac{z}{z'} \right) f_1^b(x'; Q) \quad (2.47)$$

and the  $C_{\text{TMD}}$  coefficients can be found in appendix A.

The integral in eq. (2.45) should be compared to the collinear expression at the same order (see, e.g., ref. [75])

$$\begin{aligned} & \frac{d\sigma^{\text{SIDIS}}}{dx dQ dz} \Big|_{\mathcal{O}(\alpha_s)} \\ &= \sigma_0 \frac{\alpha_s}{4\pi} \sum_q e_q^2 \left\{ \left[ D_1^{q \rightarrow h} \otimes C_1^{qq} \otimes f_1^q + D_1^{q \rightarrow h} \otimes C_1^{gg} \otimes f_1^g + D_1^{g \rightarrow h} \otimes C_1^{gg} \otimes f_1^q \right] (x, z, Q) \right. \\ & \quad \left. + \frac{1-y}{1+(1-y)^2} \left[ D_1^{q \rightarrow h} \otimes C_L^{qq} \otimes f_1^q + D_1^{q \rightarrow h} \otimes C_L^{gg} \otimes f_1^g + D_1^{g \rightarrow h} \otimes C_L^{gg} \otimes f_1^q \right] (x, z, Q) \right\}. \end{aligned} \quad (2.48)$$

The QCD coefficients  $C_1^{ab}$  and  $C_L^{ab}$  are calculated in perturbation theory. The former can be written as

$$C_1^{ab}(x, z; Q, \mu) = C_{\text{nomix}}^{ab}(x, z, Q, \mu) + C_{\text{mix}}^{ab}(x, z). \quad (2.49)$$

The coefficient  $C_{\text{nomix}}$  is the sum of all those terms that contain either a  $\delta(1-x)$  or a  $\delta(1-z)$  (or both). Some of these terms are present in  $C_{\text{TMD}}$ , but not all. This definition holds at all orders in  $\alpha_s$ . The  $C_L$  matching coefficients, instead, only contain ‘‘mixed’’ contributions. For convenience, we reproduce all coefficients at order  $\alpha_s$  in appendix A.

In order to increase the size of the TMD component, we consider the contribution of all the ‘‘nonmixed’’ terms  $C_{\text{nomix}}$ . The reason behind this choice is that it might be possible to include such terms into a redefinition of the individual TMDs. Hence, we define

$$\begin{aligned} & \frac{d\sigma^{\text{nomix}}}{dx dQ dz} \Big|_{\mathcal{O}(\alpha_s)} \\ &= \sigma_0 \frac{\alpha_s}{4\pi} \sum_q e_q^2 \left[ D_1^{q \rightarrow h} \otimes C_{\text{nomix}}^{qq} \otimes f_1^q + D_1^{q \rightarrow h} \otimes C_{\text{nomix}}^{gg} \otimes f_1^g + D_1^{g \rightarrow h} \otimes C_{\text{nomix}}^{gg} \otimes f_1^q \right] (x, z, Q), \end{aligned} \quad (2.50)$$

and similarly for higher orders, and we introduce the following normalization factor:

$$\omega(x, z, Q) = \frac{d\sigma^{\text{nomix}}}{dx dz dQ} \bigg/ \int d^2\mathbf{q}_T W. \quad (2.51)$$

We stress that these normalization factors depend only on the collinear PDFs and FFs, are independent of the parametrization of the nonperturbative part of the TMDs, and can be computed before performing a fit of the latter.

At NLL,  $\omega(x, z, Q) = 1$ . Beyond NLL, the prefactor becomes larger than one and guarantees that the integral of the TMD part of the cross section reproduces most of the collinear cross section, as suggested by the data. On the contrary, without the enhancement due to the normalization factor, the integral of the TMD part of the cross section would be too small, requiring a large role of the high-transverse-momentum tail, which is not observed in the data. The impact of the normalization factor defined in eq. (2.51) will be addressed in detail in section 4.

As a consequence of our procedure, the theoretical expression for the SIDIS cross section in eq. (2.23) becomes

$$\frac{d\sigma_{\omega}^{\text{SIDIS}}}{dx dz d|\mathbf{q}_T| dQ} = \omega(x, z, Q) \frac{d\sigma^{\text{SIDIS}}}{dx dz d|\mathbf{q}_T| dQ}. \quad (2.52)$$

### 3 Data selection

In this section we describe the experimental data included in our global analysis. We consider a large number of datasets related to DY lepton pair production and SIDIS, for the observables discussed in section 2.1 and section 2.2. The coverage in the  $x$ - $Q^2$  plane spanned by these datasets is illustrated in figure 3.

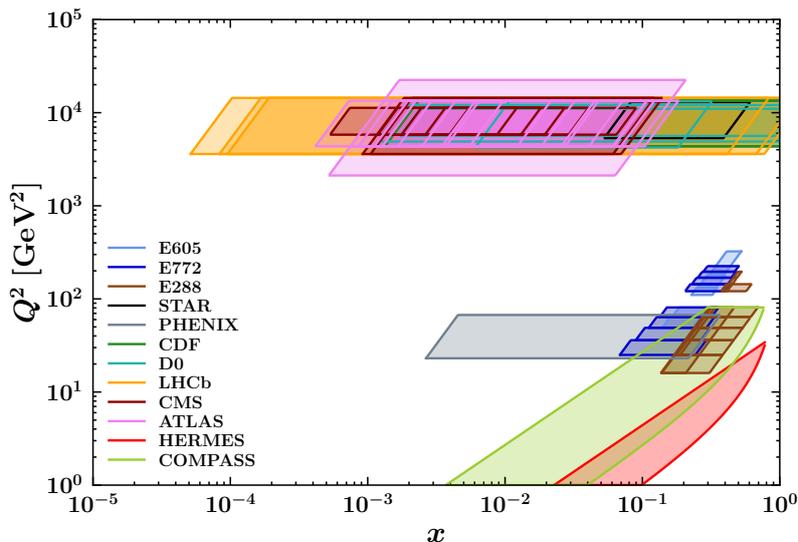
The majority of datasets analyzed in the present work was already included in the global analysis of SIDIS and DY data in ref. [5] and in the fit of DY data discussed in ref. [7]. For more details, we refer the reader to those references. The new datasets included in the present analysis are:

- DY di-muon production from the collision of a proton beam with an energy of 800 GeV on a  $^2H$  fixed target from E772 ( $\sqrt{s} = 38.8$  GeV) [76];
- DY di-muon production from the PHENIX Collaboration [77];
- DY data at 13 TeV from the CMS Collaboration [78] and the ATLAS Collaboration [79].

#### 3.1 Drell-Yan

Our analysis is based on TMD factorization, which is applicable only in the region  $|\mathbf{q}_T| \ll Q$ . Therefore, in agreement with the choices of refs. [7, 22] we impose the following cut

$$|\mathbf{q}_T| < 0.2 Q. \quad (3.1)$$



**Figure 3.** The  $x$  vs.  $Q^2$  coverage spanned by the experimental data considered in this analysis (see also table 2 and table 3).

Table 2 summarizes all the DY datasets included in our analysis. For some DY datasets the experimental observable is given within a fiducial region. This means that kinematic cuts on transverse momentum  $p_{T\ell}$  and pseudo-rapidity  $\eta_\ell$  of the single final-state leptons are enforced (values reported in the next-to-last column of table 2). For more details we refer the reader to ref. [7]. The second column of table 2 reports, for each experiment, the number of data points ( $N_{\text{dat}}$ ) that survive the kinematic cuts. The total number of DY data points considered in this work is 484. Note that for E605 and E288 at 400 GeV we have excluded the bin in  $Q$  containing the  $\Upsilon$  resonance ( $Q \simeq 9.5$  GeV).

As can be seen in table 2, the cross sections are released in different forms: some of them are normalized to the total (fiducial) cross section while others are not. When necessary, the required total cross section  $\sigma$  is computed using the code DYNLLO [35, 36] with the MMHT14 collinear PDF set, consistently with the perturbative order of the differential cross section (see also table 1). More precisely, the total cross section is computed at NLO for NNLL accuracy, and NNLO for  $N^3\text{LL}^-$  accuracy. The values of the total cross sections at different orders can be found in table 3 of ref. [7]. For the ATLAS dataset at 13 TeV, the value of the fiducial cross section is 694.3 pb at NLO and 707.3 pb at NNLO.

### 3.2 SIDIS

The identification of the TMD region in SIDIS is not a trivial task and may be subject to revision as new data appears and the theoretical description is improved, as discussed in dedicated studies [38, 94, 95].

First of all, a cut in the virtuality  $Q$  of the exchanged photon is necessary to respect the condition  $Q \gg \Lambda_{\text{QCD}}$  needed for perturbation theory to be applicable. In this way also mass corrections and higher twist corrections can be neglected. In this work, we require that  $Q > 1.4$  GeV. Studies of SIDIS in collinear kinematics employ similar cuts [29, 96].

Experiment	$N_{\text{dat}}$	Observable	$\sqrt{s}$ [GeV]	$Q$ [GeV]	$y$ or $x_F$	Lepton cuts	Ref.
E605	50	$Ed^3\sigma/d^3\mathbf{q}$	38.8	7 - 18	$x_F = 0.1$	-	[80]
E772	53	$Ed^3\sigma/d^3\mathbf{q}$	38.8	5 - 15	$0.1 < x_F < 0.3$	-	[76]
E288 200 GeV	30	$Ed^3\sigma/d^3\mathbf{q}$	19.4	4 - 9	$y = 0.40$	-	[81]
E288 300 GeV	39	$Ed^3\sigma/d^3\mathbf{q}$	23.8	4 - 12	$y = 0.21$	-	[81]
E288 400 GeV	61	$Ed^3\sigma/d^3\mathbf{q}$	27.4	5 - 14	$y = 0.03$	-	[81]
STAR 510	7	$d\sigma/d q_T $	510	73 - 114	$ y  < 1$	$p_{T\ell} > 25 \text{ GeV}$ $ \eta_\ell  < 1$	-
PHENIX200	2	$d\sigma/d q_T $	200	4.8 - 8.2	$1.2 < y < 2.2$	-	[77]
CDF Run I	25	$d\sigma/d q_T $	1800	66 - 116	Inclusive	-	[82]
CDF Run II	26	$d\sigma/d q_T $	1960	66 - 116	Inclusive	-	[83]
D0 Run I	12	$d\sigma/d q_T $	1800	75 - 105	Inclusive	-	[84]
D0 Run II	5	$(1/\sigma)d\sigma/d q_T $	1960	70 - 110	Inclusive	-	[85]
D0 Run II ( $\mu$ )	3	$(1/\sigma)d\sigma/d q_T $	1960	65 - 115	$ y  < 1.7$	$p_{T\ell} > 15 \text{ GeV}$ $ \eta_\ell  < 1.7$	[86]
LHCb 7 TeV	7	$d\sigma/d q_T $	7000	60 - 120	$2 < y < 4.5$	$p_{T\ell} > 20 \text{ GeV}$ $2 < \eta_\ell < 4.5$	[87]
LHCb 8 TeV	7	$d\sigma/d q_T $	8000	60 - 120	$2 < y < 4.5$	$p_{T\ell} > 20 \text{ GeV}$ $2 < \eta_\ell < 4.5$	[88]
LHCb 13 TeV	7	$d\sigma/d q_T $	13000	60 - 120	$2 < y < 4.5$	$p_{T\ell} > 20 \text{ GeV}$ $2 < \eta_\ell < 4.5$	[89]
CMS 7 TeV	4	$(1/\sigma)d\sigma/d q_T $	7000	60 - 120	$ y  < 2.1$	$p_{T\ell} > 20 \text{ GeV}$ $ \eta_\ell  < 2.1$	[90]
CMS 8 TeV	4	$(1/\sigma)d\sigma/d q_T $	8000	60 - 120	$ y  < 2.1$	$p_{T\ell} > 15 \text{ GeV}$ $ \eta_\ell  < 2.1$	[91]
CMS 13 TeV	70	$d\sigma/d q_T $	13000	76 - 106	$ y  < 0.4$ $0.4 <  y  < 0.8$ $0.8 <  y  < 1.2$ $1.2 <  y  < 1.6$ $1.6 <  y  < 2.4$	$p_{T\ell} > 25 \text{ GeV}$ $ \eta_\ell  < 2.4$	[78]
ATLAS 7 TeV	6 6 6	$(1/\sigma)d\sigma/d q_T $	7000	66 - 116	$ y  < 1$ $1 <  y  < 2$ $2 <  y  < 2.4$	$p_{T\ell} > 20 \text{ GeV}$ $ \eta_\ell  < 2.4$	[92]
ATLAS 8 TeV on-peak	6 6 6 6 6	$(1/\sigma)d\sigma/d q_T $	8000	66 - 116	$ y  < 0.4$ $0.4 <  y  < 0.8$ $0.8 <  y  < 1.2$ $1.2 <  y  < 1.6$ $1.6 <  y  < 2$ $2 <  y  < 2.4$	$p_{T\ell} > 20 \text{ GeV}$ $ \eta_\ell  < 2.4$	[93]
ATLAS 8 TeV off-peak	4 8	$(1/\sigma)d\sigma/d q_T $	8000	46 - 66 116 - 150	$ y  < 2.4$	$p_{T\ell} > 20 \text{ GeV}$ $ \eta_\ell  < 2.4$	[93]
ATLAS 13 TeV	6	$(1/\sigma)d\sigma/d q_T $	13000	66 - 113	$ y  < 2.5$	$p_{T\ell} > 27 \text{ GeV}$ $ \eta_\ell  < 2.5$	[79]
Total	484						

**Table 2.** Breakdown of the DY datasets considered in this analysis. For each dataset, the table includes information on: the number of data points ( $N_{\text{dat}}$ ) that survive the nominal cut on  $|q_T|$  (see eq. (3.1)), the observable delivered, the center-of-mass energy  $\sqrt{s}$ , the range(s) in invariant mass  $Q$ , the angular variable (either  $y$  or  $x_F$ ), possible cuts on the single final-state leptons, and the published reference (when available). The total number of DY data points amounts to 484.

Experiment	$N_{\text{dat}}$	Observable	Channels	$Q$ [GeV]	$x$	$z$	Phase space cuts	Ref.
HERMES	344	$M(x, z,  \mathbf{P}_{hT} , Q)$	$p \rightarrow \pi^+$ $p \rightarrow \pi^-$ $p \rightarrow K^+$ $p \rightarrow K^-$ $d \rightarrow \pi^+$ $d \rightarrow \pi^-$ $d \rightarrow K^+$ $d \rightarrow K^-$	$1 - \sqrt{15}$	$0.023 < x < 0.6$ (6 bins)	$0.1 < z < 1.1$ (8 bins)	$W^2 > 10 \text{ GeV}^2$ $0.1 < y < 0.85$	[74]
COMPASS	1203	$M(x, z, \mathbf{P}_{hT}^2, Q)$	$d \rightarrow h^+$ $d \rightarrow h^-$	1 - 9 (5 bins)	$0.003 < x < 0.4$ (8 bins)	$0.2 < z < 0.8$ (4 bins)	$W^2 > 25 \text{ GeV}^2$ $0.1 < y < 0.9$	[68]
Total	1547							

**Table 3.** Breakdown of the SIDIS datasets included in this analysis. For each dataset, the table includes information on: the number of data points ( $N_{\text{dat}}$ ) surviving the nominal cut on  $|\mathbf{P}_{hT}|$ , the observable delivered, the SIDIS channel, the range(s) in photon invariant mass  $Q$ , the ranges in the kinematic variables  $x$  and  $z$ , possible cuts on the single final-state lepton, and the public reference (when available). The total number of SIDIS data points amounts to 1547.

In order to restrict ourselves to the SIDIS current fragmentation region and interpret the observables in terms of parton distribution and fragmentation functions, we apply a cut in the kinematic variable  $z$  by requiring  $0.2 < z < 0.7$ . The lower limit is the same used in the study of collinear fragmentation functions [29, 96]. We used a slightly more restrictive upper limit, to avoid contributions from exclusive channels and to focus on a region where the collinear fragmentation functions have small relative uncertainties.

For what concerns the cut on transverse momentum, our baseline choice is

$$|\mathbf{P}_{hT}| < \min[\min[c_1 Q, c_2 zQ] + c_3 \text{ GeV}, zQ], \quad (3.2)$$

with fixed parameters  $c_1 = 0.2$ ,  $c_2 = 0.5$  and  $c_3 = 0.3$ . This choice is more restrictive than a similar one made in ref. [5], but less restrictive than the one made in ref. [22]. It allows for many data points with  $|\mathbf{P}_{hT}| \ll Q$  but also with  $0.2Q < |\mathbf{q}_T| < Q$ . In section 4, we will discuss variations of the baseline SIDIS cut in eq. (3.2) that give phenomenological support to our choice.

As for the datasets included in the present analysis, the main difference with ref. [5] is that we include the new release of COMPASS data [68]. In this dataset, the vector-boson contributions have been subtracted. For the HERMES dataset we consistently select the vector-meson-subtracted dataset (.vmsub set). Moreover, we select the `zxpt-3D`-binning for HERMES multiplicities, since it presents a finer binning in  $|\mathbf{P}_{hT}|$ . The breakdown of the entire SIDIS dataset included in the present analysis is reported in table 3.

The second column of table 3 shows the number of data points ( $N_{\text{dat}}$ ) that respect the kinematic cuts for each dataset, with a total number of 1547 data points.

In conclusion, the total number of DY and SIDIS data points surviving our kinematic cuts is 2031.

### 3.3 Error treatment

The considered experimental datasets are released with a set of systematic and statistical uncertainties. As already pointed out in ref. [7], a proper treatment of the experimental uncertainties is extremely important in order to obtain a reliable extraction of TMDs. Thus, we choose to treat systematic uncertainties as fully correlated only if, in the corresponding publication, it is explicitly specified that they are correlated. The statistical uncertainties, instead, are always considered as uncorrelated.

At variance with ref. [7], in this analysis we do not make use of the iterative  $t_0$ -prescription [97] for the treatment of correlated normalization uncertainties. This prescription is usually introduced to avoid the underestimation of the predictions caused by the so-called D’Agostini bias [98]. After performing the fit with and without the  $t_0$ -prescription, we found that our analysis is not affected by the D’Agostini bias and therefore we saw no reason to introduce the  $t_0$ -prescription in the computation of the  $\chi^2$ .

On top of experimental systematic uncertainties, there can be several sources of systematic theoretical errors. The first one comes from the choice of the underlying collinear parton distribution and fragmentation functions. In our case we choose the MMHT2014 [99] collinear PDFs and the DSS collinear FFs. Since the HERMES collaboration provides multiplicities for pions and kaons separately (see table 3), we use DSS14 [100] for  $\pi^\pm$  and DSS17 [96] for  $K^\pm$ . Given the nature of the PDF and FF set used in this analysis, their uncertainties are computed using the Hessian method [100–102]. We observed that PDF and FF uncertainties are significantly correlated across bins. In order to account for this correlation, we decomposed the corresponding uncertainties into a fully correlated part that amounts to 80% of the total, while we treated the remaining 60% as uncorrelated.<sup>10</sup>

Moreover, for the observables measured by COMPASS one needs a collinear set of FFs for unidentified charged hadrons. Since the dedicated DSS07 FF set for unidentified charged hadrons [103] does not provide an estimate of the uncertainties, we computed the COMPASS multiplicities by using the sum of the DSS14 and DSS17 sets for pions and kaons.<sup>11</sup> The associated uncertainty is calculated by propagating to the multiplicity the Hessian errors associated to each of the two hadronic component. As pointed out in ref. [22], the choice of specific sets for the collinear distributions may have a sizeable impact on the final result. In our analysis, we did not consider alternative collinear sets and postpone this study to a future publication. Likewise, we leave for future work the study of other sources of theoretical uncertainties such as higher-twist corrections, TMD flavor dependence and the choice of the perturbative scales.

---

<sup>10</sup>Notice that, using this decomposition, the sum in quadrature of correlated and uncorrelated parts reproduce the original uncertainties.

<sup>11</sup>We thus assumed that the yield due to other hadronic species, such as protons,  $\Lambda$ , etc., is negligible as compared to the sum of kaons and pions. As argued in ref. [104], the contribution to the total yield due to hadrons heavier than pions and kaons is indeed marginal.

## 4 Results

In this section, we present the results obtained for the extraction of unpolarized quark TMDs from a global analysis including both DY and SIDIS data (see section 3). This work represents an important upgrade with respect to ref. [7]. In fact, in the SIDIS process a single hadron is measured in the final state, allowing us to extract information about fragmentation functions. Therefore, the final result of this work is the simultaneous extraction of both TMD PDFs and FFs at  $N^3LL^-$ . At the moment, this is the most precise extraction of TMDs that has been achieved on more than two thousand data points. In section 4.1 we present the quality of the fit, in section 4.2 we discuss the extracted TMD distributions, and in section 4.3 we investigate the effect of variations on the baseline fit configuration.

### 4.1 Fit quality

In this section, we discuss the quality of the baseline fit performed at  $N^3LL^-$  imposing the cuts on  $|\mathbf{q}_T|/Q$  as discussed in section 3. The error analysis is performed with the so-called bootstrap method, which consists in fitting an ensemble of 250 Monte Carlo replicas of the experimental data (see ref. [5] for more details).

The most complete statistical information about the TMDs is given by the full ensemble of 250 replicas. For some purposes, however, it is useful to define a single, representative result instead of the full replica ensemble. In order to estimate the quality of our fit, the most appropriate indicator is the  $\chi^2$  value of the best fit to the central (not fluctuated) experimental data ( $\chi_0^2$ ). We refer to this fit as the “central replica.”

It is possible to analyze also the average of the  $\chi^2$  over all replicas ( $\langle\chi^2\rangle$ ) as well as the  $\chi^2$  of the mean replica ( $\chi_m^2$ ), constructed as the average of all replicas [7]. These values should be very close to each other.

Table 4 reports the breakdown of the  $\chi_0^2$  values normalized to the number of data points ( $N_{\text{dat}}$ ) for DY and SIDIS datasets. As already discussed in ref. [7], in the presence of bin-by-bin correlated uncertainties the total  $\chi^2$  can be expressed as the sum of two contributions

$$\chi^2 = \sum_i^N \left( \frac{\text{exp}_i - \overline{\text{th}}_i}{\sigma_i} \right)^2 + \chi_\lambda^2 = \chi_D^2 + \chi_\lambda^2, \quad (4.1)$$

where  $\chi_D^2$  is given by the standard formula for  $N$  experimental data points  $\text{exp}_i$  and statistical and uncorrelated uncertainties  $\sigma_i^2 = \sigma_{i,\text{stat}}^2 + \sigma_{i,\text{uncor}}^2$ , but involving theoretical predictions shifted by the correlated uncertainties,

$$\overline{\text{th}}_i = \text{th}_i + \sum_{\alpha=1}^k \lambda_\alpha \sigma_{i,\text{corr}}^{(\alpha)}, \quad (4.2)$$

where  $\sigma_{i,\text{corr}}^{(\alpha)}$  is the  $\alpha$ -th (100%) correlated uncertainty associated with the  $i$ -th experimental data point and  $\lambda_\alpha$  is the so-called nuisance parameter. In eq. (4.1), the  $\chi_\lambda^2$  is a penalty term due to the presence of correlated uncertainties and is completely determined by the

nuisance parameters,

$$\chi_\lambda^2 = \sum_{\alpha=1}^k \lambda_\alpha^2. \quad (4.3)$$

The optimal value of the nuisance parameters is obtained by minimizing the total  $\chi^2$  in eq. (4.1) with respect to them. Because the shifted predictions in eq. (4.2) provide a better visual assessment of the fit quality, we will consistently display them for all observables used in the fit.

From table 4, the global  $\chi_0^2$  value is 1.06, indicating that the description of the whole dataset is very good.<sup>12</sup> This means that the fit is able to simultaneously describe experimental data coming from two different processes over a wide kinematic range. As can be seen in tables 2 and 3, the low-energy dataset comes from fixed-target DY experiments and SIDIS observables, while the high-energy dataset comes from collider experiments at the LHC and Tevatron at energies higher by more than two orders of magnitude.

It is important to notice that the correlated penalty term  $\chi_\lambda^2$  gives a significant contribution to the total  $\chi_0^2$  value. This means that the shifts induced by correlated uncertainties are often large. The  $\chi_\lambda^2/N_{\text{dat}} = 0.29$  obtained in this analysis is larger than the one in ref. [7], mainly because of the different treatment of the theoretical uncertainties related to collinear PDFs and FFs, which are considered here as 80% correlated.

#### 4.1.1 SIDIS

For HERMES and COMPASS multiplicities, which represent about 75% of the total number of data points considered in this work, the description obtained by our global analysis is very good. From table 4, the values of  $\chi_0^2/N_{\text{dat}}$  are almost always smaller than 1.

In the case of HERMES multiplicities, we note that the largest contribution to the  $\chi_0^2$  comes from the  $\pi^+$  channel, for both proton and deuteron targets. This result is consistent with the findings in both refs. [5, 15] and [22]. Since kaon multiplicities are affected by larger statistical errors, and the collinear FFs for kaons display large uncertainties, the corresponding  $\chi_0^2$  value is lower.

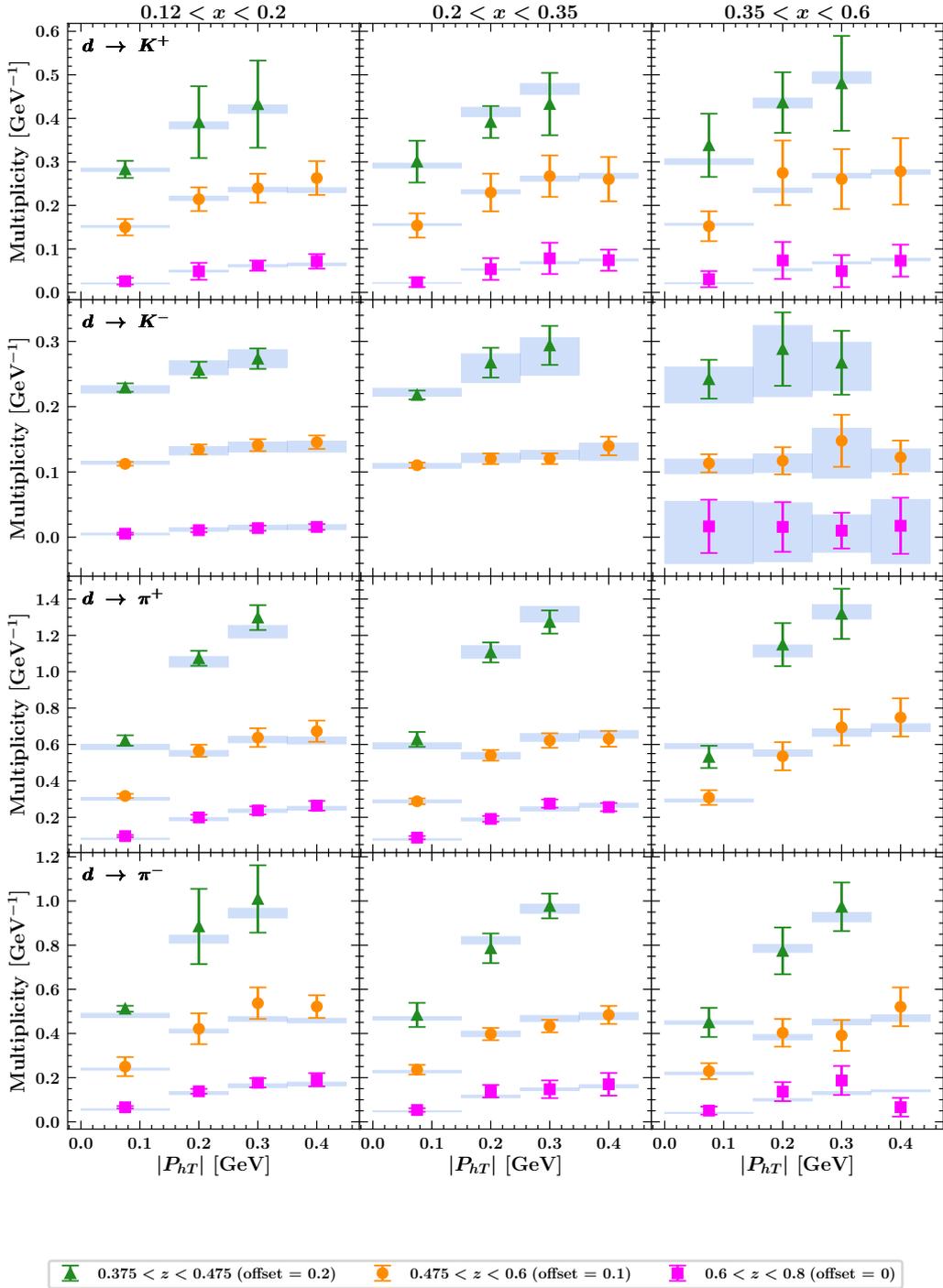
The comparison between theoretical results for the SIDIS multiplicities of eq. (2.30) and HERMES data for the production of charged pions and kaons off a deuteron target is shown in figure 4. Each column corresponds to a specific  $x$  bin. Each row corresponds to a specific final-state channel. The results are displayed as functions of the transverse momentum  $|\mathbf{P}_{hT}|$  of the measured final-state hadron. Points with different markers and colors correspond to different representative  $z$  bins, and are offset for a better visualization as indicated in the plot legend. The light blue rectangles are the theoretical results and correspond to the 68% Confidence-Level (CL) band (namely excluding the largest and smallest 16% of the replicas).

We note that for  $K^-$  production the central column, corresponding to the  $0.2 < x < 0.35$  bin, does not include the magenta points for the highest  $0.6 < z < 0.8$  bin because of the kinematic cut in  $z$ . Similarly, in all panels there are only three green points (for the lowest  $z$  bin) because of the  $z$ -dependent cut in eq. (3.2), which leads to the exclusion

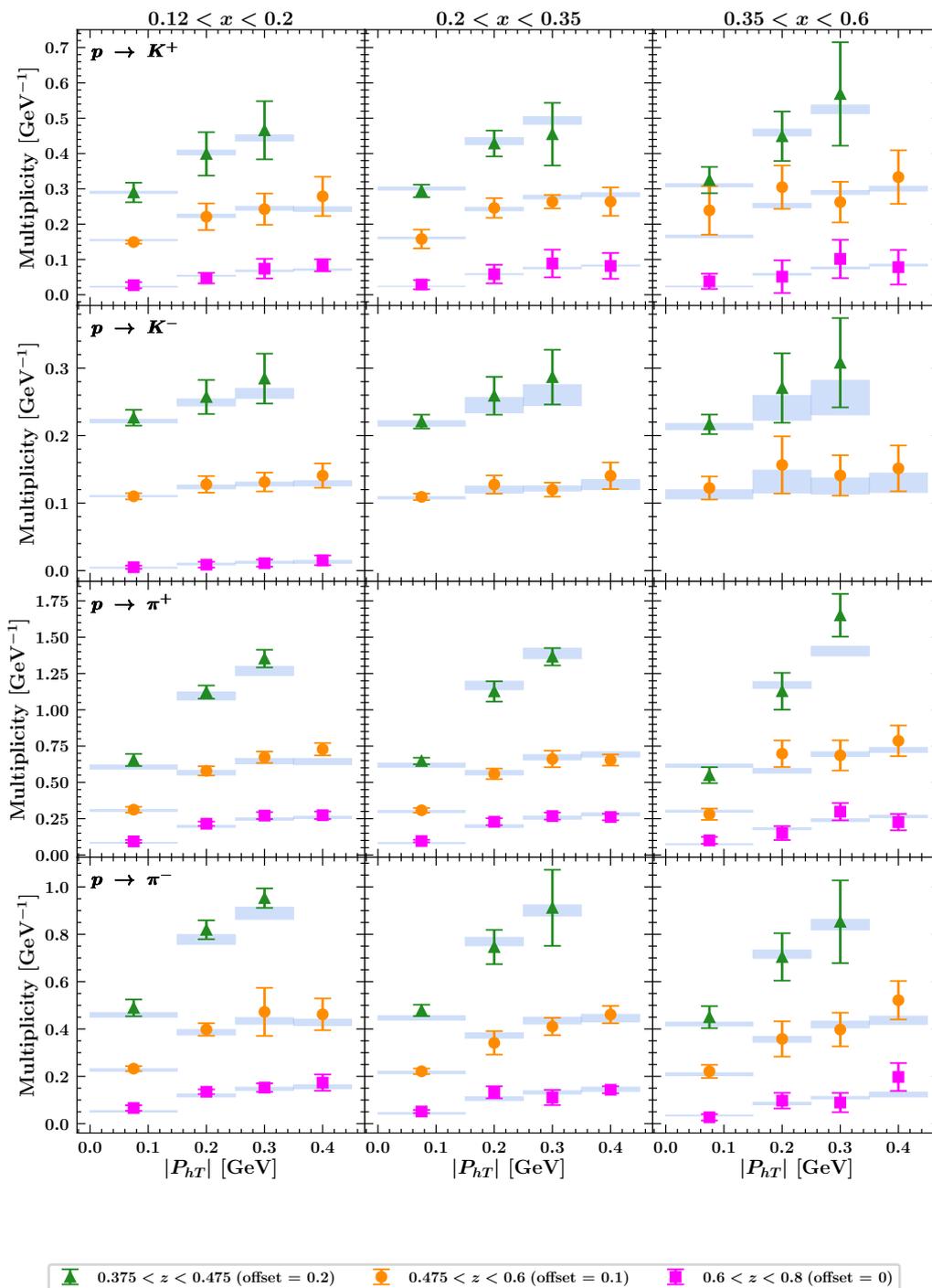
<sup>12</sup>We also obtain  $\langle \chi^2 \rangle = 1.08 \pm 0.01$  and  $\chi_m^2 = 1.07$ .

Data set	$N^3LL^-$			
	$N_{\text{dat}}$	$\chi_D^2$	$\chi_\lambda^2$	$\chi_0^2$
CDF Run I	25	0.45	0.09	0.54
CDF Run II	26	0.995	0.004	1.0
D0 Run I	12	0.67	0.01	0.68
D0 Run II	5	0.89	0.21	1.10
D0 Run II ( $\mu$ )	3	3.96	0.28	4.2
<i>Tevatron total</i>	71	0.87	0.06	0.93
LHCb 7 TeV	7	1.24	0.49	1.73
LHCb 8 TeV	7	0.78	0.36	1.14
LHCb 13 TeV	7	1.42	0.06	1.48
<i>LHCb total</i>	21	1.15	0.3	1.45
ATLAS 7 TeV	18	6.43	0.92	7.35
ATLAS 8 TeV	48	3.7	0.32	4.02
ATLAS 13 TeV	6	5.9	0.5	6.4
<i>ATLAS total</i>	72	4.56	0.48	5.05
CMS 7 TeV	4	2.21	0.10	2.31
CMS 8 TeV	4	1.938	0.001	1.94
CMS 13 TeV	70	0.36	0.02	0.37
<i>CMS total</i>	78	0.53	0.02	0.55
PHENIX 200	2	2.21	0.88	3.08
STAR 510	7	1.05	0.10	1.15
DY collider total	251	1.86	0.2	2.06
E288 200 GeV	30	0.35	0.19	0.54
E288 300 GeV	39	0.33	0.09	0.42
E288 400 GeV	61	0.5	0.11	0.61
E772	53	1.52	1.03	2.56
E605	50	1.26	0.44	1.7
DY fixed-target total	233	0.85	0.4	1.24
HERMES ( $p \rightarrow \pi^+$ )	45	0.86	0.42	1.28
HERMES ( $p \rightarrow \pi^-$ )	45	0.61	0.31	0.92
HERMES ( $p \rightarrow K^+$ )	45	0.49	0.04	0.53
HERMES ( $p \rightarrow K^-$ )	37	0.18	0.13	0.31
HERMES ( $d \rightarrow \pi^+$ )	41	0.68	0.45	1.13
HERMES ( $d \rightarrow \pi^-$ )	45	0.63	0.35	0.97
HERMES ( $d \rightarrow K^+$ )	45	0.2	0.02	0.22
HERMES ( $d \rightarrow K^-$ )	41	0.14	0.08	0.22
<i>HERMES total</i>	344	0.48	0.23	0.71
COMPASS ( $d \rightarrow h^+$ )	602	0.55	0.31	0.86
COMPASS ( $d \rightarrow h^-$ )	601	0.68	0.3	0.98
<i>COMPASS total</i>	1203	0.62	0.3	0.92
SIDIS total	1547	0.59	0.28	0.87
<b>Total</b>	<b>2031</b>	<b>0.77</b>	<b>0.29</b>	<b>1.06</b>

**Table 4.** Breakdown of the values of  $\chi^2$  normalized to the number of data points  $N_{\text{dat}}$  that survive the kinematic cuts for all datasets considered in our baseline fit. The  $\chi_D^2$  refers to uncorrelated uncertainties,  $\chi_\lambda^2$  is the penalty term due to correlated uncertainties (see eq. (4.1)),  $\chi_0^2$  is the sum of  $\chi_D^2$  and  $\chi_\lambda^2$ . All  $\chi^2$  values refer to the central replica (see text).



**Figure 4.** Comparison between data and theoretical predictions for the HERMES multiplicities for the production of charged pions and kaons off a deuteron target for different  $x$  and  $z$  bins as a function of the transverse momentum  $|P_{hT}|$  of the final-state hadron. For better visualization, each  $z$  bin is shifted by the indicated offset.



**Figure 5.** Same conventions and notation as in previous figure but for charged pions and kaons off proton target.

of a larger number of  $|q_T|$  bins at lower values of  $z$ . The theoretical results display larger uncertainties for  $K^-$  production (second row) because of the combined effect of larger experimental errors and larger uncertainties in the kaon collinear FFs.

Figure 5 refers to the same HERMES multiplicities with same conventions and notation as in figure 4 but off a proton target. We remark that for  $K^-$  production the kinematic cuts have a more drastic effect because the magenta points for the  $|\mathbf{P}_{hT}|$  distributions at the largest  $z$  bin are excluded for the two largest  $x$  bins considered (central and rightmost panels of second row from top).

In figure 6, we show the result of our fit for the COMPASS SIDIS multiplicities for the production of unidentified negatively charged hadrons off a deuteron target. For each  $Q$  and  $x$  bin, each panel displays the multiplicity on a logarithmic scale as a function of  $\mathbf{P}_{hT}^2/Q^2$ . Again, points with different markers and colors correspond to different representative  $z$  bins, as indicated in the plot legend. As before, the light-blue rectangles correspond to the 68% CL theoretical results. The results for unidentified negatively charged hadrons  $h^-$  are obtained by simply adding the results for negatively charged pions and kaons,  $h^- \sim \pi^- + K^-$ .

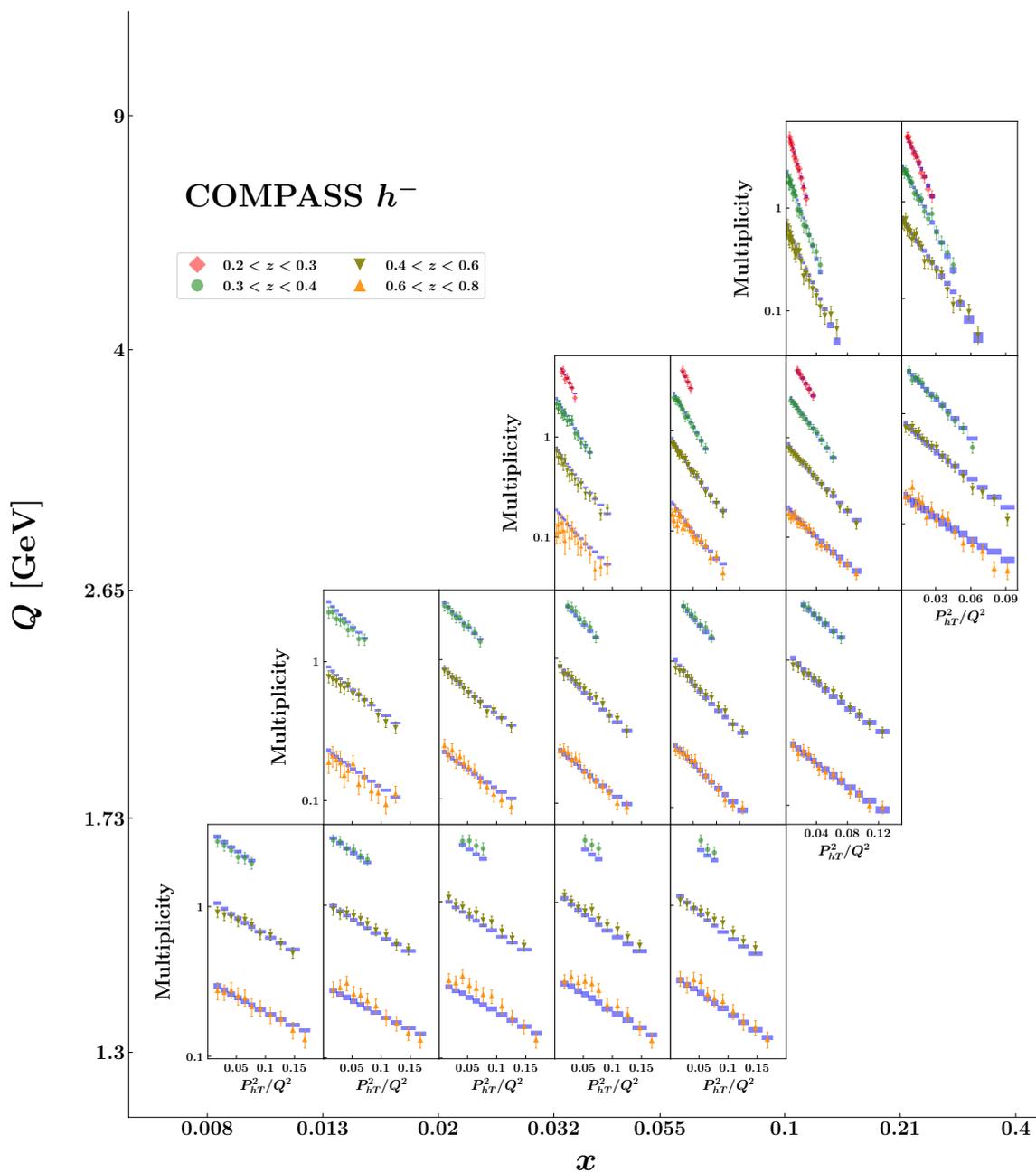
We note that the agreement is good for almost all bins, which is reflected in small  $\chi_0^2$  values in table 4. The situation worsens for the lowest  $Q$  bin ( $1.3 < Q < 1.73$  GeV), particularly for  $x \gtrsim 0.02$ . We also remark that for some  $Q$  and  $x$  bins the theoretical uncertainties for the largest  $z$  bin compatible with our kinematic cuts ( $0.6 < z < 0.8$ ) are significantly larger than for other  $z$  bins because of much larger uncertainties in the collinear FFs. Finally, looking at the table of panels from top to bottom, we realize that the lowest  $z$ -bin distributions (red diamonds) are present only for the largest  $Q$  bins, and vice-versa, because of the kinematic cut in eq. (3.2)

Figure 7 refers to the same COMPASS multiplicities with same conventions and notation as in figure 6 but for unidentified positively charged hadrons  $h^+$ . Again, the light-blue rectangles correspond to the 68% CL theoretical results, and are obtained by adding the results for positively charged pions and kaons,  $h^+ \sim \pi^+ + K^+$ . Comments similar to figure 6 can be made about the agreement between data and theory.

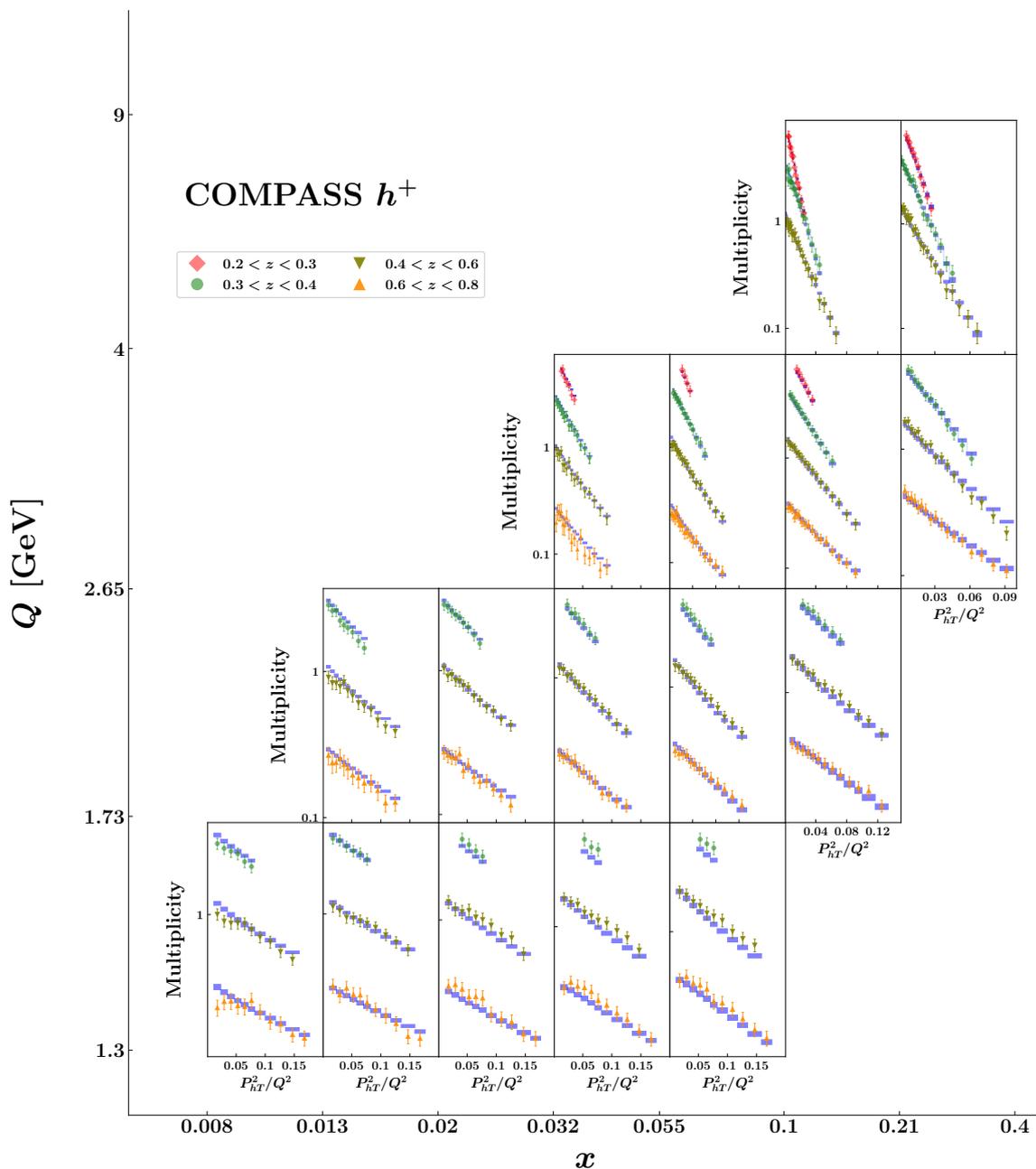
#### 4.1.2 Drell-Yan

DY data represents approximately 25% of the full set of analyzed data. From table 4 it is evident that most of low-energy DY data from fixed-target experiments (E605, E288, E772), but also from PHENIX and STAR, can be fitted with low  $\chi^2$  values, much lower than high-energy DY data from collider experiments like, e.g., those at the LHC. As already pointed out in ref. [7], this most likely originates from the fact that low-energy DY data are affected by larger errors and collinear PDFs at these kinematics have larger uncertainties.

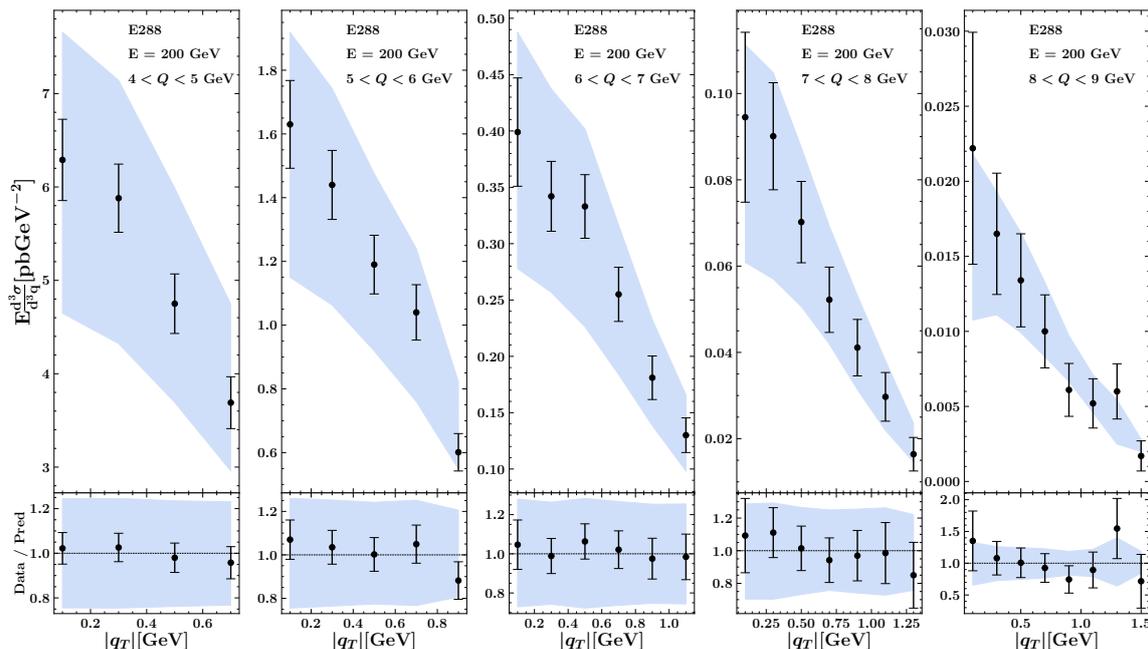
From table 4, we also note that the quality of our fit for the ATLAS datasets is poor. In particular, the description worsens for the first two low-rapidity bins of both ATLAS 7 TeV and ATLAS 8 TeV datasets, the worst case being at  $|y| < 1$  for ATLAS 7 TeV. Several effects might be responsible for this result. Since the experimental observable is a normalized cross section, systematic errors cancel in the ratio producing measurements with very small error bars. Fitting these data is very difficult, also because small theoretical effects can give significant contributions to the  $\chi^2$ . Moreover, different implementations of phase-space cuts on the final-state leptons could lead to modifications in both the shape and the normalization of the theoretical observable (see, e.g., refs. [32, 105, 106]). We



**Figure 6.** Comparison between data and theoretical predictions for the COMPASS multiplicities for the production of negatively charged hadrons off a deuteron target. For each  $Q, x$  bin, the multiplicities are displayed as functions of  $P_{hT}^2/Q^2$  for different  $z$  bins surviving kinematic cuts, as indicated in the legend.



**Figure 7.** Same conventions and notation as in previous figure but for unidentified positively charged hadrons off deuteron target.



**Figure 8.** Upper panel: comparison between data and theoretical predictions for the DY cross section differential in  $|q_T|$  for the E288 dataset at  $E_{beam} = 200$  GeV for different  $Q$  bins; uncertainty bands correspond to the 68% CL. Lower panel: ratio between experimental data and theoretical cross section.

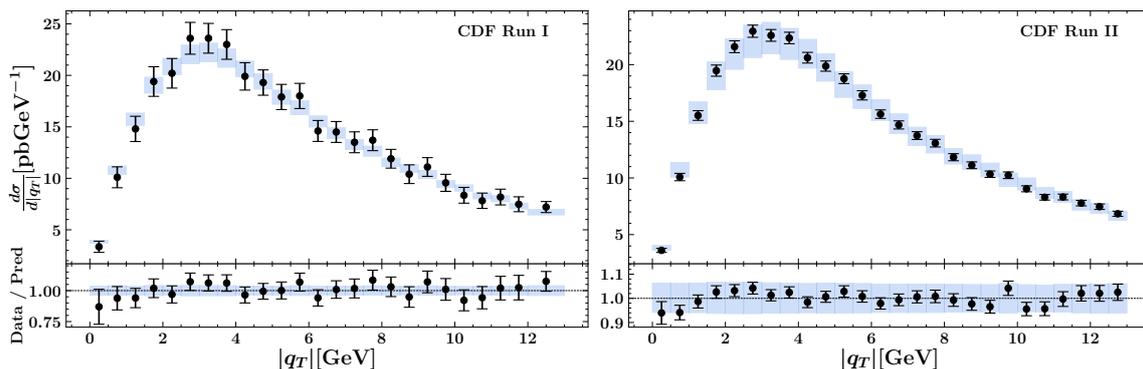
leave this issue for future studies. At variance with ref. [22], we obtain our results without excluding any extra data points on top of the ones exceeding the maximum value of  $|q_T|/Q$  in eq. (3.1).

It is interesting to comment the results of the fit for those datasets that were not included in the previous analysis of ref. [7] (see section 3). For E772, we are able to obtain a good description only for data points above the peak of the  $\Upsilon$  resonance. For  $Q < 9$  GeV, the quality of the fit worsens. At variance with ref. [22], we keep the  $Q < 9$  GeV bins because there is no evident motivation to exclude them.

The new CMS dataset at  $\sqrt{s} = 13$  TeV is important because it extends the kinematic coverage considered in ref. [7]. This dataset is very nicely described, even better than the ones at the smaller center-of-mass energies  $\sqrt{s} = 7, 8$  TeV. This is probably due to the fact that the CMS 13 TeV dataset is densely binned in rapidity.

In order to visualize the quality of our fit, in figures 8–11 we present the comparison between experimental data and theoretical results for a representative selection of the DY dataset. In the upper panels of each plot, we display the cross section differential in  $|q_T|$ , while in the lower panels we show the ratio of data to theory. As for the SIDIS case, the light-blue bands are the 68% CL theoretical results.

Figure 8 displays the DY cross section for the E288 dataset at beam energy  $E_{beam} = 200$  GeV for different bins in  $Q$ . For DY fixed-target observables, we calculate the cross section at mean values of  $|q_T|$  (not integrating upon the bin limits). Hence, we display the 68% CL uncertainty as a band rather than a series of rectangles. We remark that



**Figure 9.** Upper panels: comparison between experimental data and theoretical predictions for the cross section differential in  $|q_T|$  for  $Z$  bosons produced in  $p\bar{p}$  collisions at the Tevatron from CDF Run I (left panel) and run II (right panel); uncertainty bands correspond to the 68% CL. Lower panel: ratio between experimental data and theoretical results.

the uncertainty band is larger for lower  $Q$  bins; this trend is induced by larger correlated uncertainties for smaller invariant masses of the lepton pair.

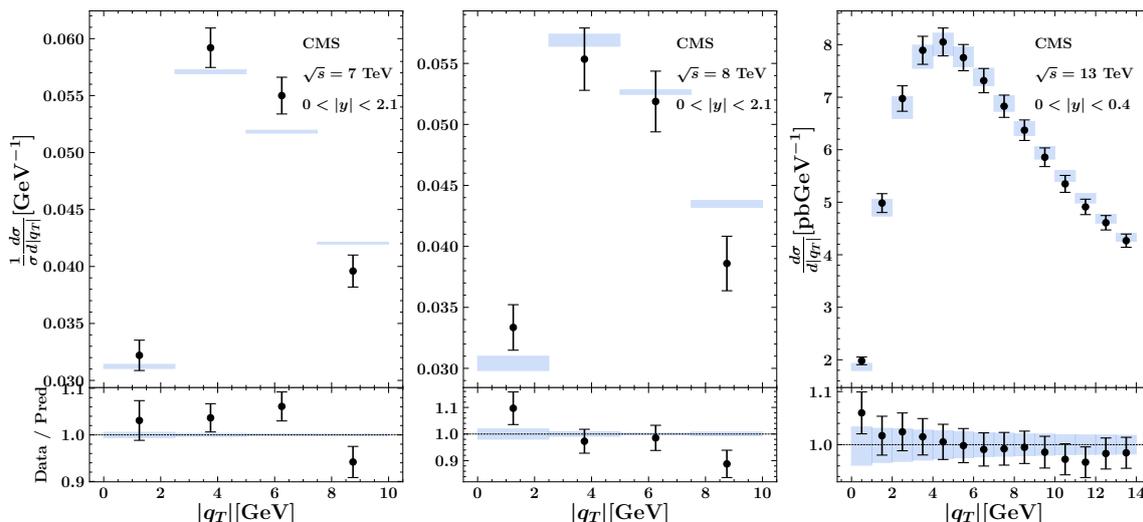
In figure 9, we compare the theoretical results for the cross section for DY in  $p\bar{p}$  collisions at the Tevatron. Black data points in the left panel correspond to the results of Run I of the CDF experiment, while in the right panel the results for Run II are reported. The lower panels show the corresponding ratio of experimental data to theoretical results. The latter are displayed as light-blue rectangles, each one corresponding to the integral of the cross section within the corresponding bin limits. The size of the rectangle is given by the 68% CL. The quality of the fit for CDF data is comparable to the one in ref. [7].

In figure 10, we compare the theoretical results for the DY cross section in  $pp$  collisions at the LHC. From left to right, the black data points refer to the measurements by the CMS Collaboration at increasing  $\sqrt{s} = 7, 8, 13$  TeV. For  $\sqrt{s} = 7, 8$  TeV, the results are normalized to the fiducial cross section. As in previous figures, the lower panels display the ratio of experimental data to the 68% CL theoretical results. For  $\sqrt{s} = 7, 8$  TeV, the quality of the fit is comparable to that in ref. [7]. For the new dataset at  $\sqrt{s} = 13$  TeV, the agreement in the displayed rapidity bin is excellent, but remains very good also for higher rapidities.

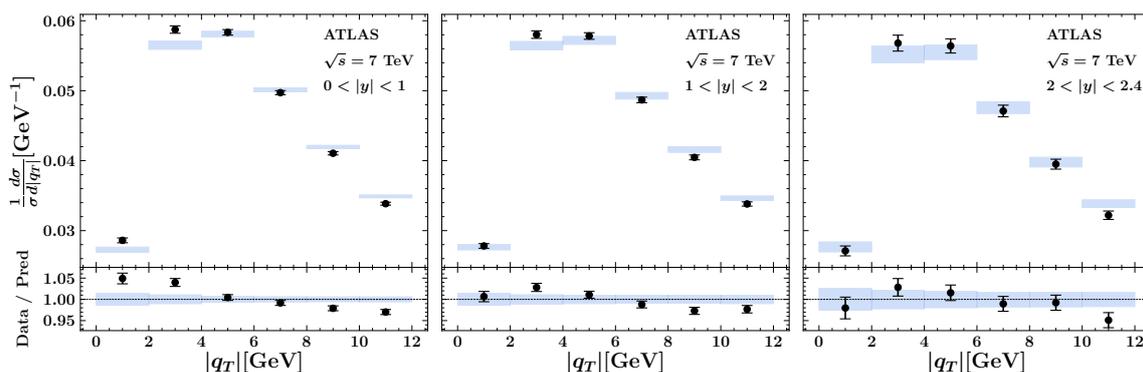
In figure 11, we compare the theoretical results for the DY cross section in  $pp$  collisions normalized to the fiducial cross section for the ATLAS data at  $\sqrt{s} = 7$  TeV. From left to right, we consider three representative bins at increasing rapidity  $|y|$ . The leftmost one corresponds to the worst described bin in our global fit, with  $\chi^2/N_{\text{data}} = 13.5$ . As in previous figures, the lower part of each panel displays the ratio of experimental data to the 68% CL theoretical results. The quality of the fit increases at more forward rapidities (from left to right). The same trend is observed at  $\sqrt{s} = 8$  TeV, but not for CMS at 13 TeV.

## 4.2 TMD distributions

We now discuss the TMD distributions extracted from our baseline fit with  $N^3\text{LL}^-$  accuracy. Table 5 displays the list of our 21 fitting parameters with their mean value and



**Figure 10.** Same as in previous figure but for  $Z$  boson production in  $pp$  collisions measured by the CMS Collaboration. From left to right: increasing  $\sqrt{s} = 7, 8, 13$  TeV, respectively. For  $\sqrt{s} = 7, 8$  TeV, the results are normalized to the fiducial cross section.



**Figure 11.** Same as in the left and central panels of previous figure, but for ATLAS kinematics at  $\sqrt{s} = 7$  TeV. From left to right, results at increasing rapidity.

standard deviation. The majority of the parameters is well constrained. The only parameter that is compatible with zero is  $\gamma_2$ .

The  $\lambda$  parameter measures the relative weight of the first Gaussian and the weighted-Gaussian in the nonperturbative part of the TMD PDF in eq. (2.38). The value of this parameter is close to 2, indicating that the contribution of the weighted-Gaussian component is important. In eq. (2.38), the parameter  $\lambda_2$  measures the relative weight of the first Gaussian and the third Gaussian; this parameter is small but not compatible with zero, which means that also this component of the TMD PDF is important to reach a good description of experimental data.

Our parametrization of the nonperturbative part of TMD FFs in eq. (2.39) contains just the combination of a Gaussian and a weighted Gaussian: this is sufficient to describe the data in an accurate way.

Parameter	Average over replicas
$g_2$ [GeV]	$0.248 \pm 0.008$
$N_1$ [GeV <sup>2</sup> ]	$0.316 \pm 0.025$
$\alpha_1$	$1.29 \pm 0.19$
$\sigma_1$	$0.68 \pm 0.13$
$\lambda$ [GeV <sup>-1</sup> ]	$1.82 \pm 0.29$
$N_3$ [GeV <sup>2</sup> ]	$0.0055 \pm 0.0006$
$\beta_1$	$10.23 \pm 0.29$
$\delta_1$	$0.0094 \pm 0.0012$
$\gamma_1$	$1.406 \pm 0.084$
$\lambda_F$ [GeV <sup>-2</sup> ]	$0.078 \pm 0.011$
$N_{3B}$ [GeV <sup>2</sup> ]	$0.2167 \pm 0.0055$
$N_{1B}$ [GeV <sup>2</sup> ]	$0.134 \pm 0.017$
$N_{1C}$ [GeV <sup>2</sup> ]	$0.0130 \pm 0.0069$
$\lambda_2$ [GeV <sup>-1</sup> ]	$0.0215 \pm 0.0058$
$\alpha_2$	$4.27 \pm 0.31$
$\alpha_3$	$4.27 \pm 0.13$
$\sigma_2$	$0.455 \pm 0.050$
$\sigma_3$	$12.71 \pm 0.21$
$\beta_2$	$4.17 \pm 0.13$
$\delta_2$	$0.167 \pm 0.006$
$\gamma_2$	$0.0007 \pm 0.0110$

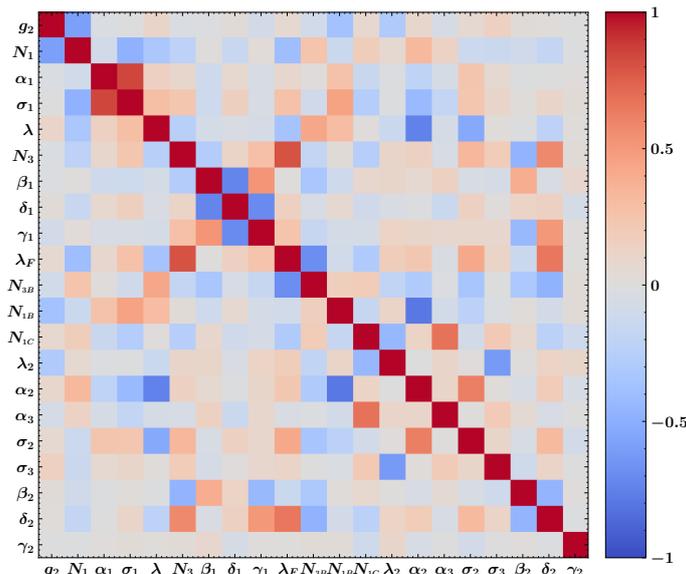
**Table 5.** Average and standard deviation over the Monte Carlo replicas of the free parameters fitted to the data.

The  $\lambda_F$  parameter measures the relative weight of the two components; its value is close to 0.1, indicating that the contribution of the weighted Gaussian is small. Nevertheless, it has non-trivial consequences on the tail of the TMD FF, as we will show below.

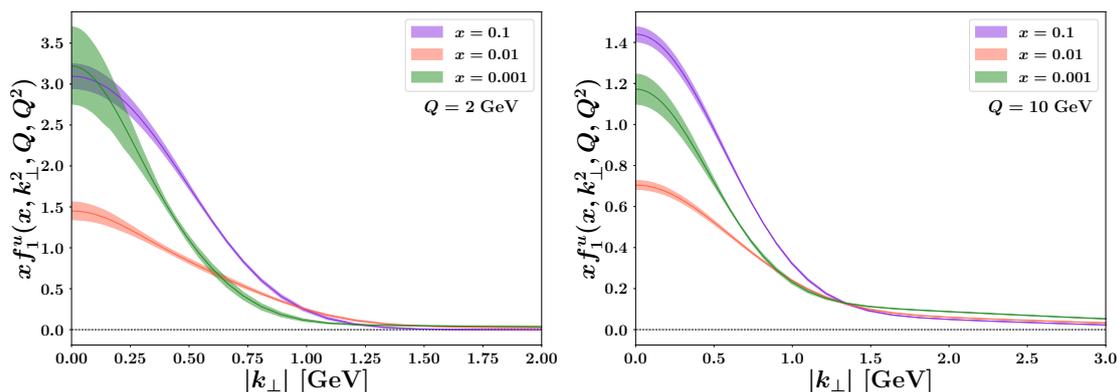
The  $g_2$  parameter is a key ingredient to the extraction of the Collins-Soper kernel, discussed in section 4.2.1. The same parameter was used in the analysis of ref. [5]. It is interesting to observe that the value obtained in the present global fit is smaller by almost a factor of 4 with respect to ref. [5]. This may be due to the higher theoretical accuracy of the present analysis and to the role of the very precise high-energy Drell-Yan measurements, which also determine the very small standard deviation of  $g_2$ .

In figure 12, we show a graphical representation of the correlations among the 21 fitting parameters. Using the color code indicated in the legend, it is easy to realize that the nondiagonal elements are very small except for some (anti-)correlation among the  $\beta_1$ ,  $\delta_1$  and  $\gamma_1$  parameters that control the  $z$ -dependent width of the Gaussians in the TMD FF (see eqs. (2.39), (2.41)). The overall absence of large correlations suggests that the model parametrization of the non perturbative parts of TMDs is appropriate.

In figure 13, we show the unpolarized TMD PDF for the up quark in the proton at  $\mu = \sqrt{\zeta} = Q = 2$  GeV (left panel) and 10 GeV (right panel) as a function of the quark



**Figure 12.** Graphical representation of the correlation matrix for the fitted parameters.

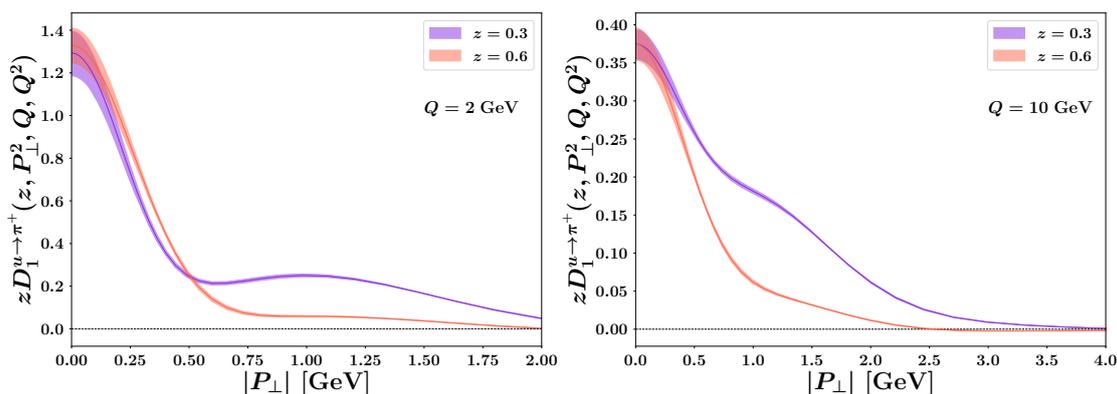


**Figure 13.** The TMD PDF of the up quark in a proton at  $\mu = \sqrt{\zeta} = Q = 2$  GeV (left panel) and 10 GeV (right panel) as a function of the partonic transverse momentum  $|\mathbf{k}_\perp|$  for  $x = 0.001, 0.01$  and  $0.1$ . The uncertainty bands represent the 68% CL.

transverse momentum  $|\mathbf{k}_\perp|$  for three different values of  $x$ , namely  $x = 0.001, 0.01$ , and  $0.1$ . The bands correspond to the 68% CL.

The TMD seems to be wider at intermediate  $x = 0.01$ , but has also a high tail at  $x = 0.001$ . As already mentioned, a significant role is played by the weighted Gaussian and by the second Gaussian in eq. (2.38). This may be a sign of the presence of contributions from different quark flavors and/or from different spin configurations (see section 2.3).

It is worth noticing that in both left and right panels the TMD PDF at  $x = 0.001$  shows the largest error band, particularly at low  $|\mathbf{k}_\perp|$ . This is due to the lack of experimental points in that kinematic region (see figure 3). Future data from the Electron-Ion Collider



**Figure 14.** The TMD FF for an up quark fragmenting into a  $\pi^+$  at  $\mu = \sqrt{\zeta} = Q = 2$  GeV (left panel) and 10 GeV (right panel) as a function of the hadron transverse momentum  $|\mathbf{P}_\perp|$  for  $z = 0.3$  and 0.6. The uncertainty bands represent the 68% CL.

(EIC) are expected to play an important role in getting a better description of the TMD PDFs at low  $x$  [107, 108].

In figure 14, we show the TMD FF for the up quark fragmenting into a  $\pi^+$  at  $\mu = \sqrt{\zeta} = Q = 2$  GeV (left panel) and 10 GeV (right panel) as a function of the pion transverse momentum  $|\mathbf{P}_\perp|$  (with respect to the fragmenting quark axis) for two different values of  $z = 0.3$  and 0.6. As in the previous figure, the uncertainty bands correspond to the 68% CL. In both left and right panels, an additional structure clearly emerges at intermediate  $P_\perp$ , especially at  $z = 0.3$ , which is induced by the weighted Gaussian in eq. (2.39). Further investigations on this topic are needed, and data from electron-positron annihilations would be valuable to better explore these features.

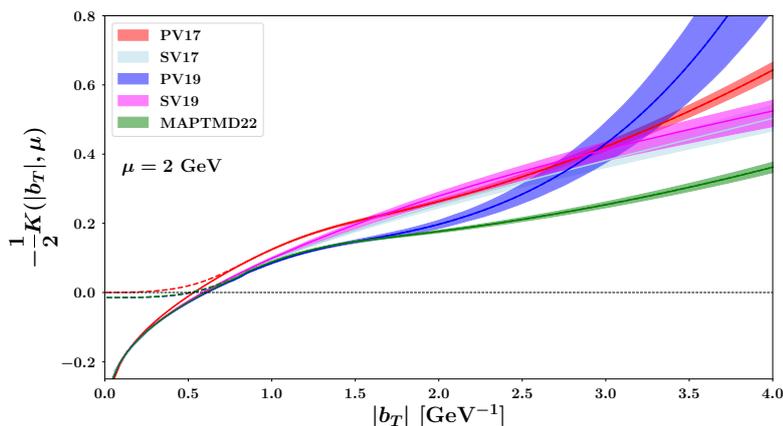
We stress that the error bands displayed in figures 13–14 reflect the uncertainty on the fitted parameters (see eqs. (2.38)–(2.39)) that are determined by taking into account the uncertainty on the collinear PDFs and FFs as discussed in section 3.3. However, since the fits are performed using the central set of the collinear distributions, all TMD replicas have the same integral in  $\mathbf{k}_\perp$  (i.e., their values at  $\mathbf{b}_T = 0$  are the same). As a consequence, the plots in figures 13–14 only partially account for the error of the collinear distributions.

### 4.2.1 Collins-Soper kernel

It is interesting to study the Collins-Soper kernel [6, 109] that drives the evolution of TMDs in terms of the rapidity scale  $\zeta$ . Recent discussions of this crucial component of the TMD formalism have been presented in refs. [110, 111] and estimates based on lattice QCD have been proposed in refs. [112–114].

The Collins-Soper kernel, as written in eq. (2.36), is composed of two parts. The first part can be calculated perturbatively at  $N^k\text{LL}$  accuracy, and is computed at  $b_*$ :

$$K(b_*(|\mathbf{b}_T|), \mu) = \sum_{n=0}^{k-1} \left( \frac{\alpha_s(\mu_{b_*})}{4\pi} \right)^{n+1} K^{(n,0)} - \sum_{n=0}^k \gamma_K^{(n)} \int_{\mu_{b_*}}^\mu \frac{d\mu'}{\mu'} \left( \frac{\alpha_s(\mu')}{4\pi} \right)^{n+1}, \quad (4.4)$$



**Figure 15.** The Collins-Soper kernel as a function of  $|\mathbf{b}_T|$  at a scale  $\mu = 2$  GeV from the present analysis (MAPTMD22), compared with the PV17 [5], SV17 [20], PV19 [7], and SV19 [22] analyses. For the MAPTMD22, PV17, and PV19 curves, the uncertainty bands represent the 68% CL. The corresponding dashed lines show the effect of including the  $b_{\min}$ -prescription (see text).

where  $K^{(n,0)}$  and  $\gamma_K^{(n)}$  are coefficients of the perturbative expansion (see, e.g., ref. [34]). Note that the integral on the r.h.s. is directly computed by means of numerical integration, thus providing a fully resummed result. The second part, denoted as  $g_K$ , cannot be computed in perturbation theory and is one of the results of our fit. Only the full Collins-Soper kernel can be compared with other works.

In figure 15, we show the Collins-Soper kernel as a function of  $|\mathbf{b}_T|$  by conventionally keeping the scale  $\mu$  fixed at 2 GeV, for our present analysis (MAPTMD22, green band) and for four other analyses in the literature [5, 7, 20, 22]. The solid lines at low  $|\mathbf{b}_T|$  follow the perturbative result. For MAPTMD22, PV19 [7] and PV17 [5], they correspond to setting  $b_{\min} = 0$  for sake of comparison with the other SV19 [22], SV17 [20] results. The slight differences between the curves are due to the different logarithmic accuracies of the perturbative calculations: the PV17 analysis was performed at NLL, the SV17 analysis at N<sup>2</sup>LL, the PV19, SV19 and MAPTMD22 at N<sup>3</sup>LL. The size of the bands around the solid lines corresponds to one standard deviation of the parameter  $g_2$  around its best-fit value. The  $b_*$  prescription modifies the curves starting from  $|\mathbf{b}_T| \approx 1$  GeV<sup>-1</sup>. The behavior at high  $|\mathbf{b}_T|$  is driven by  $g_K$  and is different for the various analyses.

The dashed curves show the effect of using our prescription  $b_{\min} = 2e^{-\gamma_E}/\mu \approx 1.123/\mu$  in MAPTMD22, PV19 and PV17. This implies that at low  $|\mathbf{b}_T|$  the Collins-Soper kernel saturates to a finite value, as indicated by the dashed lines. As the scale increases, this modification occurs at lower and lower values of  $|\mathbf{b}_T|$  and becomes less relevant.

#### 4.2.2 Average squared transverse momenta

The average squared transverse momenta  $\langle \mathbf{k}_\perp^2 \rangle(x, Q)$ ,  $\langle \mathbf{P}_\perp^2 \rangle(z, Q)$  are calculated with the Bessel weighting technique suggested in refs. [115, 116].

In the case of the TMD PDF for a quark  $q$  in the proton at  $\mu = \sqrt{\zeta} = Q$ , one has [115, 116]:

$$\langle \mathbf{k}_\perp^2 \rangle^q(x, Q) = \frac{\int d^2 \mathbf{k}_\perp \mathbf{k}_\perp^2 f_1^q(x, \mathbf{k}_\perp^2, Q, Q^2)}{\int d^2 \mathbf{k}_\perp f_1^q(x, \mathbf{k}_\perp^2, Q, Q^2)} = \frac{2M^2 \hat{f}_1^{q(1)}(x, |\mathbf{b}_T|, Q, Q^2)}{\hat{f}_1^q(x, |\mathbf{b}_T|, Q, Q^2)} \Big|_{|\mathbf{b}_T|=0}, \quad (4.5)$$

where the Fourier transform  $\hat{f}_1^q$  of the TMD PDF has been defined in eq. (2.5) and the first Bessel moment of the TMD PDF  $\hat{f}_1^{q(1)}$  is defined as [115]:

$$\begin{aligned} \hat{f}_1^{q(1)}(x, |\mathbf{b}_T|, Q, Q^2) &= \frac{2\pi}{M^2} \int_0^{+\infty} d|\mathbf{k}_\perp| \frac{\mathbf{k}_\perp^2}{|\mathbf{b}_T|} J_1(|\mathbf{k}_\perp| |\mathbf{b}_T|) f_1^q(x, \mathbf{k}_\perp^2, Q, Q^2) \\ &= -\frac{2}{M^2} \frac{\partial}{\partial \mathbf{b}_T^2} \hat{f}_1^q(x, |\mathbf{b}_T|, Q, Q^2). \end{aligned} \quad (4.6)$$

In order to obtain meaningful values for the average squared transverse momenta, i.e., finite, positive, and not dominated by the perturbative tails of the TMDs, we shift the value of  $|\mathbf{b}_T|$  from 0 to a value well inside the nonperturbative region [116]. In this way the Bessel functions  $J_{0,1}$  tame the power-law behavior of the TMD at large transverse momentum. The choice of the specific value for  $|\mathbf{b}_T|$  is of course arbitrary and has a significant effect on the associated numerical values. We choose  $|\mathbf{b}_T| = 1.5 b_{\max}$  that guarantees that the average squared transverse momenta are positive across the  $x, Q$  values considered in the fit. Accordingly, eq. (4.6) becomes:

$$\langle \mathbf{k}_\perp^2 \rangle_r^q(x, Q) = \frac{2M^2 \hat{f}_1^{q(1)}(x, |\mathbf{b}_T|, Q, Q^2)}{\hat{f}_1^q(x, |\mathbf{b}_T|, Q, Q^2)} \Big|_{|\mathbf{b}_T|=1.5 b_{\max}}, \quad (4.7)$$

where the subscript  $r$  stands for *regularized*. We have checked that the results are consistent when choosing either the integral or differential expressions in eq. (4.6).

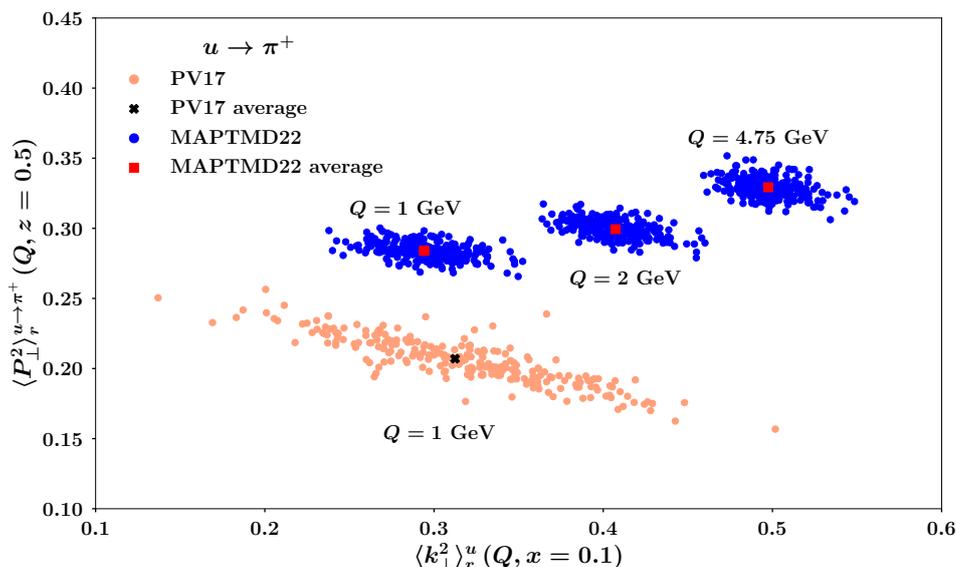
The same arguments apply to the *regularized* average squared transverse momentum produced during the hadronization of the quark  $q$  into the final state hadron  $h$  [42, 115, 116]:

$$\langle \mathbf{P}_\perp^2 \rangle_r^{q \rightarrow h}(z, Q) = \frac{2z^2 M_h^2 \hat{D}_1^{q \rightarrow h(1)}(z, |\mathbf{b}_T|, Q, Q^2)}{\hat{D}_1^{q \rightarrow h}(z, |\mathbf{b}_T|, Q, Q^2)} \Big|_{|\mathbf{b}_T|=1.5 b_{\max}}, \quad (4.8)$$

where the Fourier transform  $\hat{D}_1^{q \rightarrow h}$  of the TMD FF is defined in eq. (2.26) and the first Bessel moment of the TMD FF  $\hat{D}_1^{q \rightarrow h(1)}$  is defined as [42]:

$$\begin{aligned} \hat{D}_1^{q \rightarrow h(1)}(z, |\mathbf{b}_T|, Q, Q^2) &= \frac{2\pi}{M_h^2} \int_0^{+\infty} \frac{d|\mathbf{P}_\perp|}{z} \frac{|\mathbf{P}_\perp|}{z} \frac{|\mathbf{P}_\perp|}{z|\mathbf{b}_T|} J_1(|\mathbf{b}_T| |\mathbf{P}_\perp| / z) D_1^{q \rightarrow h}(z, \mathbf{P}_\perp^2, Q, Q^2) \\ &= -\frac{2}{M_h^2} \frac{\partial}{\partial \mathbf{b}_T^2} \hat{D}_1^{q \rightarrow h}(z, |\mathbf{b}_T|, Q, Q^2). \end{aligned} \quad (4.9)$$

In figure 16, we show the scatter plot of  $\langle \mathbf{k}_\perp^2 \rangle_r^u$  (up quark contribution) at  $x = 0.1$  vs.  $\langle \mathbf{P}_\perp^2 \rangle_r^{u \rightarrow \pi^+}$  (“favored” fragmentation) at  $z = 0.5$ . The blue circles (denoted by MAPTMD22) correspond to the single replicas while the red square is the average over



**Figure 16.** Scatter plot of average squared transverse momenta for the TMD PDF of the up quark at  $x = 0.1$  and for the TMD FF of the  $u \rightarrow \pi^+$  fragmentation at  $z = 0.5$ . Orange circles for the PV17 analysis [5] at NLL at  $Q = 1$  GeV; the black cross represents the average. Blue circles for this analysis (MAPTMD22) at  $N^3LL^-$  and at  $Q = 1, 2, 4.75$  GeV; the red squares represent the average values for each considered  $Q$  value.

all replicas for the  $N^3LL^-$  analysis. Three different values of  $Q = 1, 2, 4.75$  GeV are included to show the evolution of the average transverse momenta with the scale. The orange circles indicate the PV17 replicas [5] at NLL and  $Q = 1$  GeV with the black cross being the average. No regularization is needed for the values extracted in the PV17 analysis, since the involved TMDs at  $Q = 1$  reduce entirely to their nonperturbative components. By comparing MAPTMD22 at  $Q = 1$  GeV to PV17, we observe that the former produces much less anti-correlation between  $\langle k_{\perp}^2 \rangle$  and  $\langle P_{\perp}^2 \rangle$  than the latter, probably because of the inclusion of very precise DY data.

### 4.3 Variations on the fit configurations

In this subsection we discuss the results obtained by modifying some of the baseline settings. In sections 4.3.1 and 4.3.2 we present fits at NNLL and NLL accuracy, respectively, comparing them to the baseline  $N^3LL^-$ . Finally, in section 4.3.3 we study the impact of adopting different cuts in  $|q_T|$  on the SIDIS dataset.

#### 4.3.1 Global fit at NNLL

The baseline fit presented in the previous section is performed at  $N^3LL^-$  (see table 1). As already emphasized in ref. [7], the inclusion of perturbative corrections up to  $N^3LL$  is crucial to achieve an optimal description of some of the most recent experimental measurements, such as those by the LHC. However, it might be useful to extract unpolarized TMDs at

Data set	N <sup>3</sup> LL <sup>-</sup>		NNLL		NLL	
	$N_{\text{dat}}$	$\langle\chi^2\rangle \pm \delta\langle\chi^2\rangle$	$N_{\text{dat}}$	$\langle\chi^2\rangle \pm \delta\langle\chi^2\rangle$	$N_{\text{dat}}$	$\langle\chi^2\rangle \pm \delta\langle\chi^2\rangle$
ATLAS	72	$5.01 \pm 0.26$	/	/	/	/
PHENIX 200	2	$3.26 \pm 0.31$	2	$0.81 \pm 0.11$	/	/
STAR 510	7	$1.16 \pm 0.04$	7	$0.99 \pm 0.03$	/	/
Other sets	170	$0.83 \pm 0.01$	170	$2.37 \pm 0.11$	/	/
DY collider	251	$2.06 \pm 0.07$	179	$2.3 \pm 0.1$	/	/
E772	53	$2.48 \pm 0.12$	53	$2.05 \pm 0.22$	/	/
Other sets	180	$0.87 \pm 0.04$	180	$0.71 \pm 0.04$	180	$0.81 \pm 0.04$
DY fixed-target	233	$1.24 \pm 0.04$	233	$1.01 \pm 0.05$	180	$0.81 \pm 0.04$
HERMES	344	$0.71 \pm 0.04$	344	$1.1 \pm 0.06$	344	$0.51 \pm 0.02$
COMPASS	1203	$0.95 \pm 0.02$	1203	$0.6 \pm 0.06$	1203	$0.41 \pm 0.01$
SIDIS	1547	$0.89 \pm 0.02$	1547	$0.71 \pm 0.05$	1547	$0.43 \pm 0.01$
Total	2031	$1.08 \pm 0.01$	1959	$0.89 \pm 0.01$	1727	$0.47 \pm 0.01$

**Table 6.** Comparison of  $\chi^2$  values normalised to the number of data points  $N_{\text{dat}}$  for fits at different perturbative accuracies. The  $\langle\chi^2\rangle$  and  $\delta\langle\chi^2\rangle$  are the average and standard deviation of the  $\chi^2$  values of all replicas.

lower perturbative orders. One of the reasons is that such sets can be used in global analyses of polarized TMDs where it is not possible to reach the same level of accuracy.

This is the case of the Sivers TMDs where the computation of the polarized cross section for the Sivers effect presently cannot go beyond the NNLL level [117–119], hence demanding unpolarised TMDs at the same level of accuracy.

To this aim, we perform a new global fit at NNLL. However, when lowering the perturbative accuracy, it is possible to obtain acceptably good fits only by excluding those datasets whose precision requires the highest theoretical accuracy. Specifically, we found that only by removing the ATLAS dataset we were able to achieve an acceptable global description at NNLL accuracy. As a matter of fact, in table 6 the value of  $\chi^2$  in this configuration, namely for fixed-target DY and SIDIS, is lower than at N<sup>3</sup>LL<sup>-</sup> where ATLAS data is included.

Because of the difference in the perturbative accuracy as well as in the dataset, we do not expect to get compatible values for the best fit parameters between the NNLL and N<sup>3</sup>LL<sup>-</sup> fits. For instance, we obtain  $\lambda = 12 \pm 10 \text{ GeV}^{-1}$  and  $\lambda_F = 340 \pm 280 \text{ GeV}^{-2}$  at NNLL, to be compared to  $\lambda = 1.8 \pm 0.3 \text{ GeV}^{-1}$  and  $\lambda_F = 0.08 \pm 0.01 \text{ GeV}^{-2}$  at N<sup>3</sup>LL<sup>-</sup>.

The  $\lambda$  and  $\lambda_F$  parameters control the relative weight of the weighted Gaussian in the non perturbative part of the TMD PDF and FF, respectively, and control the size of the DY and SIDIS spectrum at middle to large values of  $|\mathbf{q}_T|$ . The large values obtained in the NNLL imply that the weighted Gaussian dominates for both TMD PDF and FF parametrizations. This behavior may be partially induced by the lack of perturbative corrections of the NNLL fit with respect to the N<sup>3</sup>LL<sup>-</sup> one, which are compensated by nonperturbative effects.

### 4.3.2 Global fit at NLL

We performed also a global analysis at NLL accuracy. Similarly to the NNLL fit discussed above, it might be useful to have also a global fit at NLL accuracy for contexts where it is not possible to reach the highest accuracy. However, in order to obtain acceptable  $\chi^2$  values at this order, we had to exclude all collider DY data and the E772 fixed-target DY dataset. Thus, we reduced the dataset to the SIDIS data and the remaining fixed-target DY data, i.e., E605 and E288.

We point out that in our approach the integral of the  $W$ -term at NLL is equal by construction to the SIDIS  $q_T$ -integrated collinear cross section. As a consequence, the value of the prefactor  $\omega$  in eq. (2.52) is automatically equal to 1. Moreover, for this fit we consistently used MMHT2014 at LO for the collinear PDFs and we used DSS at NLO for collinear FFs. As shown in table 1, at NLL both collinear PDFs and FFs should be evaluated at LO, but we choose FFs at NLO because no recent extractions at LO are currently available.

As can be seen in table 6, we obtain low  $\chi^2$  values for all included datasets. It is interesting to compare this result with that of the PV17 analysis [5]. In that work, the normalization of COMPASS data was fixed by the first bin in  $P_{hT}^2$ . Here we demonstrate that we can obtain an excellent description of the most recent COMPASS data without any adjustment of the normalization.

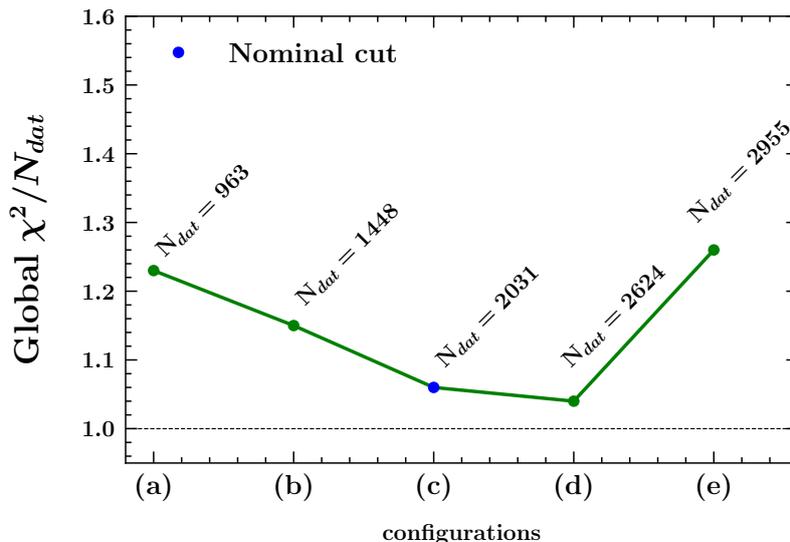
### 4.3.3 Cut in $|q_T|/Q$

A crucial ingredient of any phenomenological analysis of TMDs is the introduction of the kinematic cut  $|q_T|/Q \ll 1$  on the dataset to ensure that TMD factorization is valid. Our default choices are discussed in section 3. It is interesting to study how the global quality of our fit changes upon variations of this cut in order to get quantitative information on the range of validity of TMD factorization.

To this purpose, we changed the parameters  $c_1$ ,  $c_2$ , and  $c_3$  in eq. (3.2) devised for the SIDIS data, while keeping the constraint  $|q_T|/Q < 0.2$  for the DY data. We refer the reader to ref. [7] for an analogous study of the effect of the  $|q_T|/Q$  cut on the DY data.

We consider five different configurations for the cut on SIDIS data:

- (a) A first and most conservative cut is performed by fixing the  $z$ -independent upper value  $|q_T|/Q < 0.4$  which can be obtained by setting  $c_1 = c_2 = 0.4$  and  $c_3 = 0$  in eq. (3.2);
- (b) A second cut by setting  $c_1 = 0.15$ ,  $c_2 = 0.4$  and  $c_3 = 0.2$  in eq. (3.2);
- (c) The cut of our baseline fit with  $c_1 = 0.2$ ,  $c_2 = 0.5$  and  $c_3 = 0.3$ ;
- (d) A fourth cut with  $c_1 = 0.2$ ,  $c_2 = 0.6$  and  $c_3 = 0.4$  but without imposing  $|q_T| < Q$  (i.e. removing the outermost “min” in eq. (3.2));
- (e) A fifth cut inspired by the PV17 analysis [5], namely the same as in the previous case but with  $c_1 = 0.2$ ,  $c_2 = 0.7$  and  $c_3 = 0.5$ ; this is the least conservative choice.



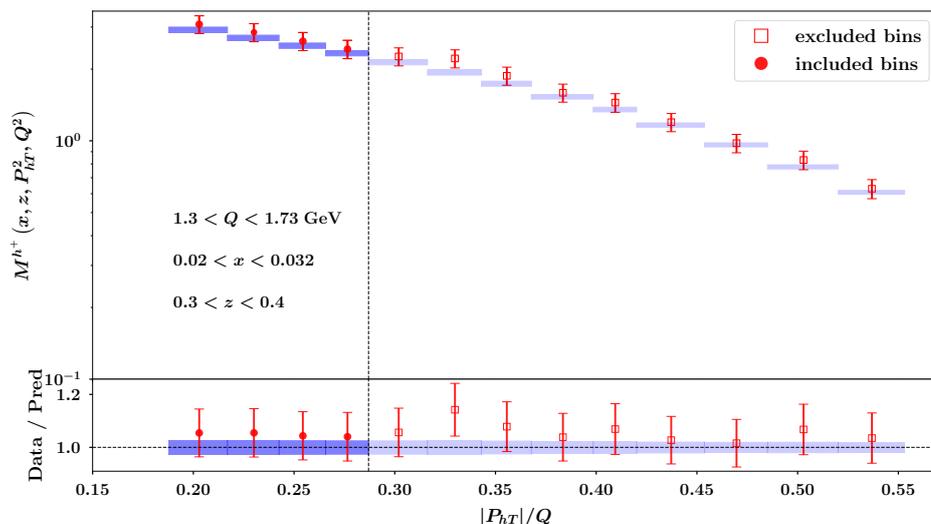
**Figure 17.** Global  $\chi^2/N_{\text{dat}}$  for different configurations of the kinematic cut on SIDIS data sets (see text). The blue point corresponds to the reference cut used in the present baseline fit.

In figure 17, we show how the global  $\chi^2/N_{\text{dat}}$  changes when considering the five configurations described above. By observing that more conservative choices do not necessarily correspond to better  $\chi^2$  values, we conclude that the TMD formalism is able to describe SIDIS data that fails to fulfil the formal requirement  $|\mathbf{q}_T|/Q \ll 1$ . In this respect, we notice that the global  $\chi^2/N_{\text{dat}}$  of cut (d) is smaller than the baseline fit, despite including a larger amount of data, some of which at  $|\mathbf{q}_T| \gtrsim Q$ , i.e., well outside the region where TMD factorization is valid.

In figure 18, we better illustrate the situation by showing the comparison between COMPASS data and theoretical predictions from our baseline fit for the SIDIS multiplicity for positively charged hadrons as a function of  $|\mathbf{P}_{hT}|/Q$  in the bin  $1.3 < Q < 1.73$  GeV,  $0.02 < x < 0.032$ ,  $0.3 < z < 0.4$ . The upper panel of the plot displays the multiplicity while the lower panel shows the ratio of experimental data to the theoretical predictions. The solid circles indicate data points included in the fit while empty squares refer to those that do not survive the cut. Remarkably, the agreement remains very good up to  $|\mathbf{P}_{hT}|/Q \simeq 0.5$ , well beyond the largest  $|\mathbf{P}_{hT}|$  allowed by the cut. We also stress that this behavior is not specific of the considered bin but is a general feature also of other bins, as well as of the HERMES dataset.

In conclusion, from our analysis it emerges that the validity of the TMD formalism in the kinematic region covered by COMPASS and HERMES seems to extend well beyond the customary cut  $|\mathbf{q}_T|/Q \ll 1$ .

This evidence justifies in a quantitative way our choice for the cut  $|\mathbf{q}_T|/Q$  in eq. (3.2) for the baseline fit, and explains why we obtain values of  $\chi^2/N_{\text{dat}}$  close to one also with less conservative cuts. Moreover, it suggests that the applicability of TMD factorization in SIDIS might be defined in terms of  $|\mathbf{P}_{hT}|$  rather than  $|\mathbf{q}_T|$ , calling for more extensive studies in this direction.



**Figure 18.** Comparison between COMPASS multiplicities and theoretical results for the SIDIS production of unidentified positively charged hadrons off a deuteron target at  $1.3 < Q < 1.73$  GeV,  $0.02 < x < 0.032$  and  $0.3 < z < 0.4$  as a function of  $|P_{hT}|/Q$ . Upper panel: light-blue rectangles for baseline fit at 68% CL, empty squares for data points not included in the baseline fit. Lower panel: ratio between experimental data and theoretical results.

## 5 Conclusions and outlook

In this article, we presented an extraction of unpolarized Transverse-Momentum Dependent Parton Distribution Functions and Fragmentation Functions (TMD PDFs and TMD FFs, respectively), which we refer to as MAPTMD22.

We analyzed 2031 data points collected by several experiments: 251 data points from Drell-Yan (DY) production measured at Tevatron, LHC and RHIC, 233 points from fixed-target DY (see table 2) and 1547 data points from Semi-Inclusive Deep Inelastic Scattering (SIDIS) measured by the HERMES and COMPASS collaborations (see table 3).

Our description of the experimental observables is based on TMD factorization at a perturbative accuracy defined as  $N^3LL^-$ , in the sense that we use TMD evolution at the  $N^3LL$  level, hard factor and matching coefficients at order  $\alpha_s^2$ , collinear PDFs at NNLO, and collinear FFs at NLO (see table 1).

We constructed the full TMDs by combining the perturbative components, regularized by means of the  $b_*$  prescription defined in eq. (2.33), and nonperturbative parts described as a sum of Gaussians and weighted Gaussians (see eqs. (2.38) and (2.39)). We used 21 free parameters: 11 related to the TMD PDF, 9 for the TMD FF, and one for the Collins-Soper kernel driving the TMD evolution. We assumed these parameters to be the same for all quark flavors.

The TMD formalism is applicable only if observed transverse momenta are much smaller than the hard scale  $Q$  of the considered process. The details of our data selection

are explained in section 3. In the DY case, for the final lepton pair transverse momentum  $\mathbf{q}_T$  we adopted a simple criterion,  $|\mathbf{q}_T| < 0.2Q$ , which is common to other works in the literature. In the SIDIS case, our selection criterion is more restrictive than the one made in ref. [5], but less restrictive than the one made in ref. [22]. For SIDIS data, where the detected hadron momentum is  $\mathbf{P}_{hT} \approx -z\mathbf{q}_T$ , our cut includes many data points with  $|\mathbf{P}_{hT}| \ll Q$  but also with  $0.2Q < |\mathbf{q}_T| < Q$ .

We also tested the quality of our fit by exploring other criteria for the  $|\mathbf{q}_T|/Q$  cut. Surprisingly, we find that fits of quality comparable to the baseline result can be obtained with less conservative cuts such that  $|\mathbf{q}_T| \gtrsim Q$  for several data points. We believe that the definition of the range of applicability of the TMD factorization formula deserves further studies.

We found it difficult to reproduce the normalization of SIDIS data measured in fixed-target experiments at moderate to low scales. The formalism works well at NLL order, but severely underestimates the data at the N<sup>2</sup>LL order and above. The reason can be traced back to the fact that the integral upon transverse momentum of the TMD cross section coincides with the known collinear result only at NLL, while it misses other contributions at higher orders. A rigorous solution of the problem would imply the calculation of all these contributions (including the so-called  $Y$ -term in the language of ref. [6]), which are currently not under control, and are beyond the scope of this work. We decided to modify our predictions by including pre-computed normalization factors that are independent of the fit, and that are identified by comparing the integral upon transverse momenta of the TMD formula to the corresponding collinear calculation, as explained in section 2.4.

The fit is performed with the bootstrap method, leading to an ensemble of 250 replicas. We reach a very good overall agreement with data, with  $\chi^2/N_{\text{dat}} = 1.06$  for the central replica (defined in section 4.1). The description of all individual datasets is satisfactory, except for a few cases, in particular for the ATLAS data (see table 4 and figure 11). The discrepancy with the ATLAS data may be due to corrections not fully included in our analysis (see, e.g., ref. [32]) which, in spite of being small, can have a large impact in the comparison with very high precision data.

The resulting TMDs are shown in figures 13 and 14. It is interesting to note that they deviate from a simple Gaussian shape (especially for the FFs) and the TMD shape changes in a nontrivial way as a function of  $x$  or  $z$ . Tables with grids of the obtained TMDs will be made publicly available at the NANGAPARBAT website<sup>13</sup> and the TMDLIB website<sup>14</sup> [120].

The availability of TMD analyses at increasing precision will be essential in guiding detector design and feasibility studies for the future Electron Ion Collider, and will also have a strong impact on precision measurements of observables particularly sensitive to the hadron structure, such as  $W$  mass measurements at hadron colliders (see, e.g., refs. [121–123]).

Our work can be extended by improving the perturbative accuracy of the analysis (see, e.g., refs. [32, 124]), including theoretical uncertainties in terms of suitable scale variations

<sup>13</sup><https://github.com/MapCollaboration/NangaParbat>.

<sup>14</sup><https://tmdlib.hepforge.org>.

(along the lines of ref. [125]), improving the treatment of correlated uncertainties, and investigating possible flavor dependence of TMDs [15, 24].

## A Explicit expression of $C$ coefficients at order $\alpha_s$

For completeness, in this appendix we report the explicit expression of the coefficients mentioned section 2.4. The  $C_{\text{TMD}}^{ab}$  coefficients of eq. (2.45) are given by:

$$C_{\text{TMD}}^{qq}(x, z) = 2C_F \left( -8\delta(1-x)\delta(1-z) + \delta(1-x) \left[ 2L_2(z) + (1-z) \right] + \delta(1-z)(1-x) \right), \quad (\text{A.1})$$

$$C_{\text{TMD}}^{gq}(x, z) = 2C_F \left( P_{gq}(z)\delta(1-x) \ln(z(1-z)) + z\delta(1-x) \right), \quad (\text{A.2})$$

$$C_{\text{TMD}}^{qq}(x, z) = 2T_F \left( \delta(1-z)2x(1-x) \right). \quad (\text{A.3})$$

They correspond to the combination of the matching coefficients  $C^{ab}$  included in the TMD PDFs and FFs in section 2.3. The following  $C_{\text{nomix}}^{ab}$  coefficients are the ones included in the computation of the SIDIS normalization factors of eq. (2.50):

$$C_{\text{nomix}}^{qq}(x, z; Q, \mu) = C_{\text{TMD}}^{qq}(x, z) + 2C_F \left[ \delta(1-x) \left( P_{qq}(z) \ln \frac{Q^2}{\mu^2} + L_1(z) - L_2(z) \right) + \delta(1-z) \left( P_{qq}(x) \ln \frac{Q^2}{\mu^2} + L_1(x) - L_2(x) \right) \right], \quad (\text{A.4})$$

$$C_{\text{nomix}}^{gq}(x, z; Q, \mu) = C_{\text{TMD}}^{gq}(x, z) + 2C_F \delta(1-x) P_{gq}(z) \ln \frac{Q^2}{\mu^2}, \quad (\text{A.5})$$

$$C_{\text{nomix}}^{qq}(x, z; Q, \mu) = C_{\text{TMD}}^{qq}(x, z) + 2T_F \delta(1-z) P_{qq}(x) \ln \left( \frac{Q^2}{\mu^2} \frac{1-x}{x} \right). \quad (\text{A.6})$$

Finally, the missing contributions to eq. (2.49),  $C_{\text{mixed}}^{ab}$ , are given by:

$$C_{\text{mixed}}^{qq}(x, z; Q, \mu) = 2C_F \left[ 2 \frac{1}{(1-x)_+} \frac{1}{(1-z)_+} - \frac{1+z}{(1-x)_+} - \frac{1+x}{(1-z)_+} + 2(1+xz) \right], \quad (\text{A.7})$$

$$C_{\text{mixed}}^{gq}(x, z; Q, \mu) = 2C_F \left[ P_{gq}(z) \frac{1}{(1-x)_+} + 2(1+x-xz) - \frac{1+x}{z} \right], \quad (\text{A.8})$$

$$C_{\text{mixed}}^{qq}(x, z; Q, \mu) = 2T_F P_{qq}(x) \left[ \frac{1}{(1-z)_+} + \frac{1}{z} - 2 \right]. \quad (\text{A.9})$$

In the above expressions  $\mu$  denotes the factorization scale for PDFs or FFs, while  $L_{1,2}$  are abbreviations for the following logarithmic functions:

$$L_1(\xi) \equiv \left( 1 + \xi^2 \right) \left( \frac{\ln(1-\xi)}{1-\xi} \right)_+, \quad L_2(\xi) \equiv \frac{1+\xi^2}{1-\xi} \ln \xi. \quad (\text{A.10})$$

## Acknowledgments

We thank Mariaelena Boglione, Osvaldo Gonzalez Hernandez, Emanuele Nocera, Ignazio Scimemi, Alexey Vladimirov for stimulating discussions. This work is supported by the European Union's Horizon 2020 programme under grant agreement No. 824093 (STRONG2020). AS acknowledges support from the European Commission through the Marie Skłodowska-Curie Action SQuHadron (grant agreement ID: 795475). CB is supported by the DOE contract DE-AC02-06CH11357.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited. SCOAP<sup>3</sup> supports the goals of the International Year of Basic Sciences for Sustainable Development.

## References

- [1] T.C. Rogers, *An overview of transverse-momentum-dependent factorization and evolution*, *Eur. Phys. J. A* **52** (2016) 153 [[arXiv:1509.04766](https://arxiv.org/abs/1509.04766)] [[INSPIRE](#)].
- [2] M. Diehl, *Introduction to GPDs and TMDs*, *Eur. Phys. J. A* **52** (2016) 149 [[arXiv:1512.01328](https://arxiv.org/abs/1512.01328)] [[INSPIRE](#)].
- [3] R. Angeles-Martinez et al., *Transverse Momentum Dependent (TMD) parton distribution functions: status and prospects*, *Acta Phys. Polon. B* **46** (2015) 2501 [[arXiv:1507.05267](https://arxiv.org/abs/1507.05267)] [[INSPIRE](#)].
- [4] I. Scimemi, *A short review on recent developments in TMD factorization and implementation*, *Adv. High Energy Phys.* **2019** (2019) 3142510 [[arXiv:1901.08398](https://arxiv.org/abs/1901.08398)] [[INSPIRE](#)].
- [5] A. Bacchetta, F. Delcarro, C. Pisano, M. Radici and A. Signori, *Extraction of partonic transverse momentum distributions from semi-inclusive deep-inelastic scattering, Drell-Yan and Z-boson production*, *JHEP* **06** (2017) 081 [*Erratum ibid.* **06** (2019) 051] [[arXiv:1703.10157](https://arxiv.org/abs/1703.10157)] [[INSPIRE](#)].
- [6] J. Collins, *Foundations of perturbative QCD*, vol. 32. Cambridge University Press (2013).
- [7] A. Bacchetta et al., *Transverse-momentum-dependent parton distributions up to  $N^3LL$  from Drell-Yan data*, *JHEP* **07** (2020) 117 [[arXiv:1912.07550](https://arxiv.org/abs/1912.07550)] [[INSPIRE](#)].
- [8] G. Bozzi, S. Catani, D. de Florian and M. Grazzini, *The  $q(T)$  spectrum of the Higgs boson at the LHC in QCD perturbation theory*, *Phys. Lett. B* **564** (2003) 65 [[hep-ph/0302104](https://arxiv.org/abs/hep-ph/0302104)] [[INSPIRE](#)].
- [9] C. Muselli, S. Forte and G. Ridolfi, *Combined threshold and transverse momentum resummation for inclusive observables*, *JHEP* **03** (2017) 106 [[arXiv:1701.01464](https://arxiv.org/abs/1701.01464)] [[INSPIRE](#)].
- [10] A. Bacchetta, M.G. Echevarria, P.J.G. Mulders, M. Radici and A. Signori, *Effects of TMD evolution and partonic flavor on  $e^+e^-$  annihilation into hadrons*, *JHEP* **11** (2015) 076 [[arXiv:1508.00402](https://arxiv.org/abs/1508.00402)] [[INSPIRE](#)].
- [11] BELLE collaboration, *Transverse momentum dependent production cross sections of charged pions, kaons and protons produced in inclusive  $e^+e^-$  annihilation at  $\sqrt{s} = 10.58$  GeV*, *Phys. Rev. D* **99** (2019) 112006 [[arXiv:1902.01552](https://arxiv.org/abs/1902.01552)] [[INSPIRE](#)].

- [12] Z.-B. Kang, D.Y. Shao and F. Zhao, *QCD resummation on single hadron transverse momentum distribution with the thrust axis*, *JHEP* **12** (2020) 127 [[arXiv:2007.14425](#)] [[INSPIRE](#)].
- [13] Y. Makris, F. Ringer and W.J. Waalewijn, *Joint thrust and TMD resummation in electron-positron and electron-proton collisions*, *JHEP* **02** (2021) 070 [[arXiv:2009.11871](#)] [[INSPIRE](#)].
- [14] M. Boglione and A. Simonelli, *Kinematic regions in the  $e^+e^- \rightarrow hX$  factorized cross section in a 2-jet topology with thrust*, *JHEP* **02** (2022) [[arXiv:2109.11497](#)] [[INSPIRE](#)].
- [15] A. Signori, A. Bacchetta, M. Radici and G. Schnell, *Investigations into the flavor dependence of partonic transverse momentum*, *JHEP* **11** (2013) 194 [[arXiv:1309.3507](#)] [[INSPIRE](#)].
- [16] M. Anselmino, M. Boglione, J.O. Gonzalez Hernandez, S. Melis and A. Prokudin, *Unpolarised Transverse Momentum Dependent Distribution and Fragmentation Functions from SIDIS Multiplicities*, *JHEP* **04** (2014) 005 [[arXiv:1312.6261](#)] [[INSPIRE](#)].
- [17] M.G. Echevarria, A. Idilbi, Z.-B. Kang and I. Vitev, *QCD Evolution of the Sivers Asymmetry*, *Phys. Rev. D* **89** (2014) 074013 [[arXiv:1401.5078](#)] [[INSPIRE](#)].
- [18] P. Sun, J. Isaacson, C.P. Yuan and F. Yuan, *Nonperturbative functions for SIDIS and Drell–Yan processes*, *Int. J. Mod. Phys. A* **33** (2018) 1841006 [[arXiv:1406.3073](#)] [[INSPIRE](#)].
- [19] U. D’Alesio, M.G. Echevarria, S. Melis and I. Scimemi, *Non-perturbative QCD effects in  $qT$  spectra of Drell–Yan and Z-boson production*, *JHEP* **11** (2014) 098 [[arXiv:1407.3311](#)] [[INSPIRE](#)].
- [20] I. Scimemi and A. Vladimirov, *Analysis of vector boson production within TMD factorization*, *Eur. Phys. J. C* **78** (2018) 89 [[arXiv:1706.01473](#)] [[INSPIRE](#)].
- [21] V. Bertone, I. Scimemi and A. Vladimirov, *Extraction of unpolarized quark transverse momentum dependent parton distributions from Drell–Yan/Z-boson production*, *JHEP* **06** (2019) 028 [[arXiv:1902.08474](#)] [[INSPIRE](#)].
- [22] I. Scimemi and A. Vladimirov, *Non-perturbative structure of semi-inclusive deep-inelastic and Drell–Yan scattering at small transverse momentum*, *JHEP* **06** (2020) 137 [[arXiv:1912.06532](#)] [[INSPIRE](#)].
- [23] F. Hautmann, I. Scimemi and A. Vladimirov, *Non-perturbative contributions to vector-boson transverse momentum spectra in hadronic collisions*, *Phys. Lett. B* **806** (2020) 135478 [[arXiv:2002.12810](#)] [[INSPIRE](#)].
- [24] M. Bury, F. Hautmann, S. Leal-Gomez, I. Scimemi, A. Vladimirov and P. Zurita, *PDF bias and flavor dependence in TMD distributions*, [arXiv:2201.07114](#) [[INSPIRE](#)].
- [25] A. Bermudez Martinez, P. Connor, H. Jung, A. Lelek, R. Žlebčik, F. Hautmann et al., *Collinear and TMD parton densities from fits to precision DIS measurements in the parton branching method*, *Phys. Rev. D* **99** (2019) 074008 [[arXiv:1804.11152](#)] [[INSPIRE](#)].
- [26] A. Bermudez Martinez et al., *Production of Z-bosons in the parton branching method*, *Phys. Rev. D* **100** (2019) 074027 [[arXiv:1906.00919](#)] [[INSPIRE](#)].
- [27] A. Bermudez Martinez et al., *The transverse momentum spectrum of low mass Drell–Yan production at next-to-leading order in the parton branching method*, *Eur. Phys. J. C* **80** (2020) 598 [[arXiv:2001.06488](#)] [[INSPIRE](#)].

- [28] I. Borsa, R. Sassot, D. de Florian, M. Stratmann and W. Vogelsang, *Towards a Global QCD Analysis of Fragmentation Functions at Next-to-Next-to-Leading Order Accuracy*, *Phys. Rev. Lett.* **129** (2022) 012002 [[arXiv:2202.05060](#)] [[INSPIRE](#)].
- [29] R. Abdul Khalek, V. Bertone, A. Khoudli and E.R. Nocera, *Pion and kaon fragmentation functions at next-to-next-to-leading order*, *Phys. Lett. B* **834** (2022) 137456 [[arXiv:2204.10331](#)] [[INSPIRE](#)].
- [30] D. Boer and W. Vogelsang, *Drell-Yan lepton angular distribution at small transverse momentum*, *Phys. Rev. D* **74** (2006) 014004 [[hep-ph/0604177](#)] [[INSPIRE](#)].
- [31] S. Arnold, A. Metz and M. Schlegel, *Dilepton production from polarized hadron-hadron collisions*, *Phys. Rev. D* **79** (2009) 034005 [[arXiv:0809.2262](#)] [[INSPIRE](#)].
- [32] X. Chen et al., *Third-Order Fiducial Predictions for Drell-Yan Production at the LHC*, *Phys. Rev. Lett.* **128** (2022) 252001 [[arXiv:2203.01565](#)] [[INSPIRE](#)].
- [33] D. Boer, *Investigating the origins of transverse spin asymmetries at RHIC*, *Phys. Rev. D* **60** (1999) 014012 [[hep-ph/9902255](#)] [[INSPIRE](#)].
- [34] J. Collins and T.C. Rogers, *Connecting Different TMD Factorization Formalisms in QCD*, *Phys. Rev. D* **96** (2017) 054011 [[arXiv:1705.07167](#)] [[INSPIRE](#)].
- [35] S. Catani and M. Grazzini, *An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC*, *Phys. Rev. Lett.* **98** (2007) 222002 [[hep-ph/0703012](#)] [[INSPIRE](#)].
- [36] S. Catani, L. Cieri, G. Ferrera, D. de Florian and M. Grazzini, *Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO*, *Phys. Rev. Lett.* **103** (2009) 082001 [[arXiv:0903.2120](#)] [[INSPIRE](#)].
- [37] A. Bacchetta, M. Diehl, K. Goeke, A. Metz, P.J. Mulders and M. Schlegel, *Semi-inclusive deep inelastic scattering at small transverse momentum*, *JHEP* **02** (2007) 093 [[hep-ph/0611265](#)] [[INSPIRE](#)].
- [38] M. Boglione, A. Dotson, L. Gamberg, S. Gordon, J.O. Gonzalez-Hernandez, A. Prokudin et al., *Mapping the Kinematical Regimes of Semi-Inclusive Deep Inelastic Scattering*, *JHEP* **10** (2019) 122 [[arXiv:1904.12882](#)] [[INSPIRE](#)].
- [39] A. Bacchetta, U. D'Alesio, M. Diehl and C.A. Miller, *Single-spin asymmetries: The Trento conventions*, *Phys. Rev. D* **70** (2004) 117504 [[hep-ph/0410050](#)] [[INSPIRE](#)].
- [40] D. Boer et al., *Gluons and the quark sea at high energies: Distributions, polarization, tomography*, [arXiv:1108.1713](#) [[INSPIRE](#)].
- [41] A. Bacchetta, D. Boer, M. Diehl and P.J. Mulders, *Matches and mismatches in the descriptions of semi-inclusive processes at low and high transverse momentum*, *JHEP* **08** (2008) 023 [[arXiv:0803.0227](#)] [[INSPIRE](#)].
- [42] A. Bacchetta, G. Bozzi, M.G. Echevarria, C. Pisano, A. Prokudin and M. Radici, *Azimuthal asymmetries in unpolarized SIDIS and Drell-Yan processes: a case study towards TMD factorization at subleading twist*, *Phys. Lett. B* **797** (2019) 134850 [[arXiv:1906.07037](#)] [[INSPIRE](#)].
- [43] P.J. Mulders and C. Van Hulse, *Noncollinearity in dijet fragmentation in electron-positron scattering*, *Phys. Rev. D* **100** (2019) 034011 [[arXiv:1903.11467](#)] [[INSPIRE](#)].

- [44] A. Accardi and A. Signori, *On the connection between quark propagation and hadronization*, *Eur. Phys. J. C* **80** (2020) 825 [[arXiv:2005.11310](#)] [[INSPIRE](#)].
- [45] A. Vladimirov, V. Moos and I. Scimemi, *Transverse momentum dependent operator expansion at next-to-leading power*, *JHEP* **01** (2022) 110 [[arXiv:2109.09771](#)] [[INSPIRE](#)].
- [46] M.A. Ebert, A. Gao and I.W. Stewart, *Factorization for azimuthal asymmetries in SIDIS at next-to-leading power*, *JHEP* **06** (2022) 007 [[arXiv:2112.07680](#)] [[INSPIRE](#)].
- [47] S. Rodini and A. Vladimirov, *Definition and evolution of transverse momentum dependent distribution of twist-three*, *JHEP* **08** (2022) 031 [[arXiv:2204.03856](#)] [[INSPIRE](#)].
- [48] J.O. Gonzalez-Hernandez, T.C. Rogers, N. Sato and B. Wang, *Challenges with Large Transverse Momentum in Semi-Inclusive Deeply Inelastic Scattering*, *Phys. Rev. D* **98** (2018) 114005 [[arXiv:1808.04396](#)] [[INSPIRE](#)].
- [49] B. Wang, J.O. Gonzalez-Hernandez, T.C. Rogers and N. Sato, *Large Transverse Momentum in Semi-Inclusive Deeply Inelastic Scattering Beyond Lowest Order*, *Phys. Rev. D* **99** (2019) 094029 [[arXiv:1903.01529](#)] [[INSPIRE](#)].
- [50] P.B. Arnold and R.P. Kauffman, *W and Z production at next-to-leading order: From large  $q(t)$  to small*, *Nucl. Phys. B* **349** (1991) 381 [[INSPIRE](#)].
- [51] J. Collins, L. Gamberg, A. Prokudin, T.C. Rogers, N. Sato and B. Wang, *Relating Transverse Momentum Dependent and Collinear Factorization Theorems in a Generalized Formalism*, *Phys. Rev. D* **94** (2016) 034014 [[arXiv:1605.00671](#)] [[INSPIRE](#)].
- [52] M.G. Echevarria, T. Kasemets, J.-P. Lansberg, C. Pisano and A. Signori, *Matching factorization theorems with an inverse-error weighting*, *Phys. Lett. B* **781** (2018) 161 [[arXiv:1801.01480](#)] [[INSPIRE](#)].
- [53] M.G. Echevarria, A. Idilbi and I. Scimemi, *Factorization Theorem For Drell-Yan At Low  $q_T$  And Transverse Momentum Distributions On-The-Light-Cone*, *JHEP* **07** (2012) 002 [[arXiv:1111.4996](#)] [[INSPIRE](#)].
- [54] M. Grewal, Z.-B. Kang, J.-W. Qiu and A. Signori, *Predictive power of transverse-momentum-dependent distributions*, *Phys. Rev. D* **101** (2020) 114023 [[arXiv:2003.07453](#)] [[INSPIRE](#)].
- [55] S. Catani, M.L. Mangano, P. Nason and L. Trentadue, *The Resummation of soft gluons in hadronic collisions*, *Nucl. Phys. B* **478** (1996) 273 [[hep-ph/9604351](#)] [[INSPIRE](#)].
- [56] G. Bozzi, S. Catani, D. de Florian and M. Grazzini, *Transverse-momentum resummation and the spectrum of the Higgs boson at the LHC*, *Nucl. Phys. B* **737** (2006) 73 [[hep-ph/0508068](#)] [[INSPIRE](#)].
- [57] W. Bizoń et al., *Fiducial distributions in Higgs and Drell-Yan production at  $N^3LL+NNLO$* , *JHEP* **12** (2018) 132 [[arXiv:1805.05916](#)] [[INSPIRE](#)].
- [58] A. Bacchetta, F. Conti and M. Radici, *Transverse-momentum distributions in a diquark spectator model*, *Phys. Rev. D* **78** (2008) 074010 [[arXiv:0807.0323](#)] [[INSPIRE](#)].
- [59] B. Pasquini, S. Cazzaniga and S. Boffi, *Transverse momentum dependent parton distributions in a light-cone quark model*, *Phys. Rev. D* **78** (2008) 034025 [[arXiv:0806.2298](#)] [[INSPIRE](#)].
- [60] H. Avakian, A.V. Efremov, P. Schweitzer and F. Yuan, *The transverse momentum dependent distribution functions in the bag model*, *Phys. Rev. D* **81** (2010) 074035 [[arXiv:1001.5467](#)] [[INSPIRE](#)].

- [61] M. Burkardt and B. Pasquini, *Modelling the nucleon structure*, *Eur. Phys. J. A* **52** (2016) 161 [[arXiv:1510.02567](#)] [[INSPIRE](#)].
- [62] T. Gutsche, V.E. Lyubovitskij and I. Schmidt, *Nucleon parton distributions in a light-front quark model*, *Eur. Phys. J. C* **77** (2017) 86 [[arXiv:1610.03526](#)] [[INSPIRE](#)].
- [63] T. Maji and D. Chakrabarti, *Transverse structure of a proton in a light-front quark-diquark model*, *Phys. Rev. D* **95** (2017) 074009 [[arXiv:1702.04557](#)] [[INSPIRE](#)].
- [64] R. Alessandro, A. Del Dotto, E. Pace, G. Perna, G. Salmè and S. Scopetta, *Light-front transverse momentum distributions for  $J=1/2$  hadronic systems in valence approximation*, *Phys. Rev. C* **104** (2021) 065204 [[arXiv:2107.10187](#)] [[INSPIRE](#)].
- [65] A.I. Signal and F.G. Cao, *Transverse momentum and transverse momentum distributions in the MIT bag model*, *Phys. Lett. B* **826** (2022) 136898 [[arXiv:2108.12116](#)] [[INSPIRE](#)].
- [66] A. Bacchetta, L.P. Gamberg, G.R. Goldstein and A. Mukherjee, *Collins fragmentation function for pions and kaons in a spectator model*, *Phys. Lett. B* **659** (2008) 234 [[arXiv:0707.3372](#)] [[INSPIRE](#)].
- [67] H.H. Matevosyan, W. Bentz, I.C. Cloet and A.W. Thomas, *Transverse Momentum Dependent Fragmentation and Quark Distribution Functions from the NJL-jet Model*, *Phys. Rev. D* **85** (2012) 014021 [[arXiv:1111.1740](#)] [[INSPIRE](#)].
- [68] COMPASS collaboration, *Transverse-momentum-dependent Multiplicities of Charged Hadrons in Muon-Deuteron Deep Inelastic Scattering*, *Phys. Rev. D* **97** (2018) 032006 [[arXiv:1709.07374](#)] [[INSPIRE](#)].
- [69] F. Piacenza, *Perturbative and nonperturbative QCD regimes in transverse-momentum dependent observables*, PhD thesis, Università di Pavia, Pavia, Italy (2020).
- [70] J. Osvaldo Gonzalez-Hernandez, *Comments on the perturbative and non-perturbative contributions in unpolarized SIDIS*, *PoS DIS2019* (2019) 176 [[INSPIRE](#)].
- [71] A. Bacchetta, G. Bozzi, M. Lambertsen, F. Piacenza, J. Steiglechner and W. Vogelsang, *Difficulties in the description of Drell-Yan processes at moderate invariant mass and high transverse momentum*, *Phys. Rev. D* **100** (2019) 014018 [[arXiv:1901.06916](#)] [[INSPIRE](#)].
- [72] E. Moffat, T.C. Rogers, N. Sato and A. Signori, *Collinear factorization in wide-angle hadron pair production in  $e^+e^-$  annihilation*, *Phys. Rev. D* **100** (2019) 094014 [*Erratum ibid.* **104** (2021) 059904] [[arXiv:1909.02951](#)] [[INSPIRE](#)].
- [73] D. de Florian, R. Sassot and M. Stratmann, *Global analysis of fragmentation functions for pions and kaons and their uncertainties*, *Phys. Rev. D* **75** (2007) 114010 [[hep-ph/0703242](#)] [[INSPIRE](#)].
- [74] HERMES collaboration, *Multiplicities of charged pions and kaons from semi-inclusive deep-inelastic scattering by the proton and the deuteron*, *Phys. Rev. D* **87** (2013) 074029 [[arXiv:1212.5407](#)] [[INSPIRE](#)].
- [75] D. de Florian, M. Stratmann and W. Vogelsang, *QCD analysis of unpolarized and polarized Lambda baryon production in leading and next-to-leading order*, *Phys. Rev. D* **57** (1998) 5811 [[hep-ph/9711387](#)] [[INSPIRE](#)].
- [76] E772 collaboration, *Cross-sections for the production of high mass muon pairs from 800-GeV proton bombardment of H-2*, *Phys. Rev. D* **50** (1994) 3038 [*Erratum ibid.* **60** (1999) 119903] [[INSPIRE](#)].

- [77] PHENIX collaboration, *Measurements of  $\mu\mu$  pairs from open heavy flavor and Drell-Yan in  $p + p$  collisions at  $\sqrt{s} = 200$  GeV*, *Phys. Rev. D* **99** (2019) 072003 [[arXiv:1805.02448](#)] [[INSPIRE](#)].
- [78] CMS collaboration, *Measurements of differential Z boson production cross sections in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *JHEP* **12** (2019) 061 [[arXiv:1909.04133](#)] [[INSPIRE](#)].
- [79] ATLAS collaboration, *Measurement of the transverse momentum distribution of Drell-Yan lepton pairs in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *Eur. Phys. J. C* **80** (2020) 616 [[arXiv:1912.02844](#)] [[INSPIRE](#)].
- [80] G. Moreno et al., *Dimuon production in proton-copper collisions at  $\sqrt{s} = 38.8$ -GeV*, *Phys. Rev. D* **43** (1991) 2815 [[INSPIRE](#)].
- [81] A.S. Ito et al., *Measurement of the Continuum of Dimuons Produced in High-Energy Proton-Nucleus Collisions*, *Phys. Rev. D* **23** (1981) 604 [[INSPIRE](#)].
- [82] CDF collaboration, *The transverse momentum and total cross section of  $e^+e^-$  pairs in the Z boson region from  $p\bar{p}$  collisions at  $\sqrt{s} = 1.8$  TeV*, *Phys. Rev. Lett.* **84** (2000) 845 [[hep-ex/0001021](#)] [[INSPIRE](#)].
- [83] CDF collaboration, *Transverse momentum cross section of  $e^+e^-$  pairs in the Z-boson region from  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96$  TeV*, *Phys. Rev. D* **86** (2012) 052010 [[arXiv:1207.7138](#)] [[INSPIRE](#)].
- [84] D0 collaboration, *Measurement of the inclusive differential cross section for Z bosons as a function of transverse momentum in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.8$  TeV*, *Phys. Rev. D* **61** (2000) 032004 [[hep-ex/9907009](#)] [[INSPIRE](#)].
- [85] D0 collaboration, *Measurement of the shape of the boson transverse momentum distribution in  $p\bar{p} \rightarrow Z/\gamma^* \rightarrow e^+e^- + X$  events produced at  $\sqrt{s}=1.96$ -TeV*, *Phys. Rev. Lett.* **100** (2008) 102002 [[arXiv:0712.0803](#)] [[INSPIRE](#)].
- [86] D0 collaboration, *Measurement of the Normalized  $Z/\gamma^* \rightarrow \mu^+\mu^-$  Transverse Momentum Distribution in  $p\bar{p}$  Collisions at  $\sqrt{s} = 1.96$  TeV*, *Phys. Lett. B* **693** (2010) 522 [[arXiv:1006.0618](#)] [[INSPIRE](#)].
- [87] LHCb collaboration, *Measurement of the forward Z boson production cross-section in pp collisions at  $\sqrt{s} = 7$  TeV*, *JHEP* **08** (2015) 039 [[arXiv:1505.07024](#)] [[INSPIRE](#)].
- [88] LHCb collaboration, *Measurement of forward W and Z boson production in pp collisions at  $\sqrt{s} = 8$  TeV*, *JHEP* **01** (2016) 155 [[arXiv:1511.08039](#)] [[INSPIRE](#)].
- [89] LHCb collaboration, *Measurement of the forward Z boson production cross-section in pp collisions at  $\sqrt{s} = 13$  TeV*, *JHEP* **09** (2016) 136 [[arXiv:1607.06495](#)] [[INSPIRE](#)].
- [90] CMS collaboration, *Measurement of the Rapidity and Transverse Momentum Distributions of Z Bosons in pp Collisions at  $\sqrt{s} = 7$  TeV*, *Phys. Rev. D* **85** (2012) 032002 [[arXiv:1110.4973](#)] [[INSPIRE](#)].
- [91] CMS collaboration, *Measurement of the transverse momentum spectra of weak vector bosons produced in proton-proton collisions at  $\sqrt{s} = 8$  TeV*, *JHEP* **02** (2017) 096 [[arXiv:1606.05864](#)] [[INSPIRE](#)].
- [92] ATLAS collaboration, *Measurement of the  $Z/\gamma^*$  boson transverse momentum distribution in pp collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, *JHEP* **09** (2014) 145 [[arXiv:1406.3660](#)] [[INSPIRE](#)].

- [93] ATLAS collaboration, *Measurement of the transverse momentum and  $\phi_{\eta}^*$  distributions of Drell–Yan lepton pairs in proton-proton collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, *Eur. Phys. J. C* **76** (2016) 291 [[arXiv:1512.02192](#)] [[INSPIRE](#)].
- [94] M. Boglione, J. Collins, L. Gamberg, J.O. Gonzalez-Hernandez, T.C. Rogers and N. Sato, *Kinematics of Current Region Fragmentation in Semi-Inclusive Deeply Inelastic Scattering*, *Phys. Lett. B* **766** (2017) 245 [[arXiv:1611.10329](#)] [[INSPIRE](#)].
- [95] JEFFERSON LAB ANGULAR MOMENTUM (JAM) collaboration, *New tool for kinematic regime estimation in semi-inclusive deep-inelastic scattering*, *JHEP* **04** (2022) 084 [[arXiv:2201.12197](#)] [[INSPIRE](#)].
- [96] D. de Florian, M. Epele, R.J. Hernandez-Pinto, R. Sassot and M. Stratmann, *Parton-to-Kaon Fragmentation Revisited*, *Phys. Rev. D* **95** (2017) 094019 [[arXiv:1702.06353](#)] [[INSPIRE](#)].
- [97] NNPDF collaboration, *Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties*, *JHEP* **05** (2010) 075 [[arXiv:0912.2276](#)] [[INSPIRE](#)].
- [98] G. D’Agostini, *On the use of the covariance matrix to fit correlated data*, *Nucl. Instrum. Meth. A* **346** (1994) 306 [[INSPIRE](#)].
- [99] L.A. Harland-Lang, A.D. Martin, P. Motylinski and R.S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J. C* **75** (2015) 204 [[arXiv:1412.3989](#)] [[INSPIRE](#)].
- [100] D. de Florian, R. Sassot, M. Epele, R.J. Hernández-Pinto and M. Stratmann, *Parton-to-Pion Fragmentation Reloaded*, *Phys. Rev. D* **91** (2015) 014035 [[arXiv:1410.6027](#)] [[INSPIRE](#)].
- [101] J. Pumplin et al., *Uncertainties of predictions from parton distribution functions. 2. The Hessian method*, *Phys. Rev. D* **65** (2001) 014013 [[hep-ph/0101032](#)] [[INSPIRE](#)].
- [102] D. de Florian, R. Sassot, M. Stratmann and W. Vogelsang, *Extraction of Spin-Dependent Parton Densities and Their Uncertainties*, *Phys. Rev. D* **80** (2009) 034030 [[arXiv:0904.3821](#)] [[INSPIRE](#)].
- [103] D. de Florian, R. Sassot and M. Stratmann, *Global analysis of fragmentation functions for protons and charged hadrons*, *Phys. Rev. D* **76** (2007) 074033 [[arXiv:0707.1506](#)] [[INSPIRE](#)].
- [104] NNPDF collaboration, *Charged hadron fragmentation functions from collider data*, *Eur. Phys. J. C* **78** (2018) 651 [[arXiv:1807.03310](#)] [[INSPIRE](#)].
- [105] L. Buonocore, S. Kallweit, L. Rottoli and M. Wiesemann, *Linear power corrections for two-body kinematics in the  $qT$  subtraction formalism*, *Phys. Lett. B* **829** (2022) 137118 [[arXiv:2111.13661](#)] [[INSPIRE](#)].
- [106] S. Camarda, L. Cieri and G. Ferrera, *Fiducial perturbative power corrections within the  $q_T$  subtraction formalism*, *Eur. Phys. J. C* **82** (2022) 575 [[arXiv:2111.14509](#)] [[INSPIRE](#)].
- [107] R. Abdul Khalek et al., *Science Requirements and Detector Concepts for the Electron-Ion Collider: EIC Yellow Report*, *Nucl. Phys. A* **1026** (2022) 122447 [[arXiv:2103.05419](#)] [[INSPIRE](#)].
- [108] R. Abdul Khalek et al., *Snowmass 2021 White Paper: Electron Ion Collider for High Energy Physics*, in *2022 Snowmass Summer Study*, (2022) [[arXiv:2203.13199](#)] [[INSPIRE](#)].

- [109] J.C. Collins and D.E. Soper, *Back-To-Back Jets in QCD*, *Nucl. Phys. B* **193** (1981) 381 [Erratum *ibid.* **213** (1983) 545] [INSPIRE].
- [110] A.A. Vladimirov, *Self-contained definition of the Collins-Soper kernel*, *Phys. Rev. Lett.* **125** (2020) 192002 [arXiv:2003.02288] [INSPIRE].
- [111] A. Bermudez Martinez and A. Vladimirov, *Determination of Collins-Soper kernel from cross-sections ratios*, arXiv:2206.01105 [INSPIRE].
- [112] M. Schlemmer, A. Vladimirov, C. Zimmermann, M. Engelhardt and A. Schäfer, *Determination of the Collins-Soper Kernel from Lattice QCD*, *JHEP* **08** (2021) 004 [arXiv:2103.16991] [INSPIRE].
- [113] P. Shanahan, M. Wagman and Y. Zhao, *Lattice QCD calculation of the Collins-Soper kernel from quasi-TMDPDFs*, *Phys. Rev. D* **104** (2021) 114502 [arXiv:2107.11930] [INSPIRE].
- [114] LPC collaboration, *Nonperturbative determination of the Collins-Soper kernel from quasitransverse-momentum-dependent wave functions*, *Phys. Rev. D* **106** (2022) 034509 [arXiv:2204.00200] [INSPIRE].
- [115] D. Boer, L. Gamberg, B. Musch and A. Prokudin, *Bessel-Weighted Asymmetries in Semi Inclusive Deep Inelastic Scattering*, *JHEP* **10** (2011) 021 [arXiv:1107.5294] [INSPIRE].
- [116] D. Boer, *Average transverse momentum quantities approaching the lightfront*, *Few Body Syst.* **56** (2015) 439 [arXiv:1409.8317] [INSPIRE].
- [117] A. Bacchetta, F. Delcarro, C. Pisano and M. Radici, *The 3-dimensional distribution of quarks in momentum space*, *Phys. Lett. B* **827** (2022) 136961 [arXiv:2004.14278] [INSPIRE].
- [118] M.G. Echevarria, Z.-B. Kang and J. Terry, *Global analysis of the Sivers functions at NLO+NNLL in QCD*, *JHEP* **01** (2021) 126 [arXiv:2009.10710] [INSPIRE].
- [119] M. Bury, A. Prokudin and A. Vladimirov, *Extraction of the Sivers function from SIDIS, Drell-Yan, and  $W^\pm/Z$  boson production data with TMD evolution*, *JHEP* **05** (2021) 151 [arXiv:2103.03270] [INSPIRE].
- [120] N.A. Abdulov et al., *TMDlib2 and TMDplotter: a platform for 3D hadron structure studies*, *Eur. Phys. J. C* **81** (2021) 752 [arXiv:2103.09741] [INSPIRE].
- [121] A. Bacchetta, G. Bozzi, M. Radici, M. Ritzmann and A. Signori, *Effect of Flavor-Dependent Partonic Transverse Momentum on the Determination of the W Boson Mass in Hadronic Collisions*, *Phys. Lett. B* **788** (2019) 542 [arXiv:1807.02101] [INSPIRE].
- [122] G. Bozzi and A. Signori, *Nonperturbative Uncertainties on the Transverse Momentum Distribution of Electroweak Bosons and on the Determination of the Boson Mass at the LHC*, *Adv. High Energy Phys.* **2019** (2019) 2526897 [arXiv:1901.01162] [INSPIRE].
- [123] CDF collaboration, *High-precision measurement of the W boson mass with the CDF II detector*, *Science* **376** (2022) 170 [INSPIRE].
- [124] S. Camarda, L. Cieri and G. Ferrera, *Drell-Yan lepton-pair production:  $qT$  resummation at N3LL accuracy and fiducial cross sections at N3LO*, *Phys. Rev. D* **104** (2021) L111503 [arXiv:2103.04974] [INSPIRE].
- [125] V. Bertone, G. Bozzi and F. Hautmann, *Perturbative hysteresis and emergent resummation scales*, *Phys. Rev. D* **105** (2022) 096003 [arXiv:2202.03380] [INSPIRE].