

Initial condition for the Balitsky-Kovchegov equation at next-to-leading order

Carlisle Casuga^{1,2,*}, Henri Hänninen^{3,†} and Heikki Mäntysaari^{1,2,‡}

¹*Department of Physics, University of Jyväskylä, P.O. Box 35, 40014 University of Jyväskylä, Finland*

²*Helsinki Institute of Physics, P.O. Box 64, 00014 University of Helsinki, Finland*

³*Department of Mathematics and Statistics, University of Jyväskylä, P.O. Box 35, 40014 University of Jyväskylä, Finland*



(Received 17 June 2025; accepted 15 July 2025; published 4 August 2025)

We determine the initial condition of the Balitsky-Kovchegov (BK) evolution equation at next-to-leading order (NLO) accuracy using HERA deep inelastic scattering data. We use Bayesian inference to compare model values, derived from NLO impact factors and the BK evolution equation with kinematical constraint resummation of most important higher order corrections, to precise data and to extract posterior distributions characterizing the initial condition. The total cross section and charm quark production data from HERA are found to provide stringent constraints on the posterior. These distributions quantify the uncertainty in the initial condition and serve as necessary input for propagating uncertainties to all NLO calculations within the color glass condensate framework.

DOI: [10.1103/54zd-hyvg](https://doi.org/10.1103/54zd-hyvg)

I. INTRODUCTION

The color glass condensate (CGC) [1] effective field theory provides a natural framework for describing the gluon-dense, high-energy structure of protons and nuclei. As the collision energy increases or the Bjorken- x decreases at fixed probe virtuality Q^2 , one unlocks the saturation regime with emergent nonlinear quantum chromodynamics (QCD) effects. This regime is characterized by the saturation scale, Q_s^2 , which scales as $Q_s^2 \sim A^{1/3}$, where A is the nuclear mass number. In this context, the upcoming Electron Ion Collider (EIC) [2] (or other potential future nuclear-DIS facilities [3,4]) will feature heavy nuclear targets and provide access to the saturated state of gluonic matter.

In the leading order (LO) dipole picture of deep inelastic scattering (DIS), the incoming virtual photon splits into a quark-antiquark pair which interacts with the target nucleus [5]. The necessary ingredients in calculations of DIS structure functions in the CGC framework are the DIS impact factor and the small- x Balitsky-Kovchegov (BK) [6,7] evolution equation at the desired order in perturbation theory.

First principle calculations of impact factors at NLO in α_s have been performed in the light cone perturbation theory for light [8–13] and massive [14–16] quarks. These impact factors, which describe the photon splitting into partonic states, are coupled to the high-energy BK evolution equation that resums contributions enhanced by large logarithms of energy $\sim \alpha_s \ln(1/x)$ (or $\alpha_s^2 \ln(1/x)$ at NLO). The full NLO BK evolution equation is derived in Ref. [17] with numerical solutions implemented in Refs. [18,19]. Prescriptions that resum dominant single and double logarithmic contributions, that provide an accurate approximation to the NLO BK equation [19], have been obtained in Refs. [20–23]. For this study, we adopt the BK evolution equation that resums the large double logarithmic terms to approximate the full NLO BK equation. Finite- N_c corrections to the NLO BK equation, found to have a small effect on dipole amplitudes [24], and other corrections like the resummation of single logarithmic terms and higher order terms not enhanced by transverse logarithms, are omitted from the present analysis. Furthermore, we do not include dependence on the impact parameter, as studied in Ref. [25].

The BK equation describes the energy dependence of the dipole-target scattering amplitude. This perturbative evolution equation takes as a nonperturbative input the dipole-target scattering amplitude at an initial energy scale, and predicts its energy evolution toward higher energies. Models like the McLerran-Venugopalan (MV) [26] model or the Golec-Biernat-Wüsthoff (GBW) model [27] provide reasonable parametrizations for this initial condition. Substantial efforts have been made in extracting the initial

*Contact author: carlisle.doc.casuga@jyu.fi

†Contact author: henri.j.hanninen@jyu.fi

‡Contact author: heikki.mantysaari@jyu.fi

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

condition from available data at LO [28–30], and at NLO with light [31] and heavy [32] quarks. In particular, Ref. [32] shows the feasibility of a global fit at NLO accuracy by demonstrating that it is possible to simultaneously describe both the total cross section [33] and heavy quark production [34] data from HERA at small x . These previous NLO dipole fits, however, do not include the heavy quark contribution in the actual fit procedure, and neglect correlated experimental uncertainties. Furthermore, they also lack a systematic quantification of the uncertainty of the determined nonperturbative BK initial condition, which has been obtained in a leading order analysis in Ref. [35]. An accurate extraction of this nonperturbative input, among with well-defined uncertainty estimates, is essential in order to promote the precision in theoretical calculations to the level that is comparable to current and future high-precision measurement sensitive to saturation effects at the EIC or at the LHC [36].

The main objective of this work is to extend our previous analysis [35] to infer the initial condition for the BK equation in NLO accuracy, combining NLO impact factors with dipole amplitudes evolved using an evolution equation of partial NLO accuracy to compute the DIS structure functions. For the first time, we perform a global analysis including both the HERA total cross section and the charm production data as constraints. To extract the model parameters in the initial condition, we use Bayesian inference, which not only identifies the best-fit values but also probes its neighboring area in the parameter space. This approach determines posterior distributions that provide a streamlined method to propagate the initial condition uncertainty to calculations of other observables within the CGC framework at NLO.

The paper is structured as follows. Section II provides a brief summary of the DIS cross section calculation at the next-to-leading order accuracy in the CGC approach. This involves a two-part discussion of the NLO impact factors in Sec. II A and the approximated NLO evolution equation in Sec. II B. Details of the Bayesian inference setup used to find the probability distribution describing the initial condition to the BK evolution equation are presented in Sec. III. The obtained posterior distributions, best-fit values, and other results are then shown in Sec. IV. Conclusions and main findings are summarized in Sec. V.

II. DEEP INELASTIC SCATTERING AT NLO IN DIPOLE PICTURE

A. NLO cross section

The leading order dipole picture of DIS consists of a photon fluctuating to a dipole and interacting instantaneously with the target hadron by exchanging low- x gluons. The separation between the lifetime of the dipole and the timescale of the interaction enables one to factorize the leading order γ^*A cross section, as a function of Bjorken- x

and photon virtuality, Q^2 , as [37]

$$\sigma_{T,L}^{\gamma^*p}(x, Q^2) = 2 \sum_f \int d^2\mathbf{b} d^2\mathbf{r} dz |\psi_{T,L}^{\gamma^* \rightarrow q\bar{q}}(\mathbf{r}, Q^2, z)|^2 \times N_{\text{LO}}(\mathbf{r}, \mathbf{b}, x). \quad (1)$$

The light-cone wave function, $\psi_{T,L}^{\gamma^* \rightarrow q\bar{q}}$, describes the splitting of the virtual photon to the quark-antiquark pair with a transverse separation $\mathbf{r}$, where the quark carries a fraction, z , of the large photon plus momentum. Here T and L refer to the transverse and longitudinal photon polarization, respectively. The dipole interacts with the target color field eikinally, so that $\mathbf{r}$ does not change during the interaction. In this work we further neglect the impact parameter dependence from the dipole-target scattering amplitude N_{LO} and, following, e.g., Refs. [29,30], replace $\int d^2\mathbf{b} \rightarrow \sigma_0/2$ where $\sigma_0/2$ is the proton transverse area and $\mathbf{b}$ is the impact parameter in the dipole-target scattering.

The total DIS cross section data is typically reported as a reduced cross section defined as

$$\sigma_r(y, x, Q^2) = F_2(x, Q^2) - \frac{y^2}{1 + (1 - y)^2} F_L(x, Q^2). \quad (2)$$

Here the inelasticity is defined as $y = Q^2/(sx)$, where $\sqrt{s}$ is the center-of-mass energy in the electron-proton scattering. The structure functions F_2 and F_L are related to the virtual photon-proton cross section as

$$F_2 = \frac{Q^2}{4\pi^2\alpha_{\text{em}}} (\sigma_T^{\gamma^*A} + \sigma_L^{\gamma^*A}), \quad (3)$$

$$F_L = \frac{Q^2}{4\pi^2\alpha_{\text{em}}} \sigma_L^{\gamma^*A}. \quad (4)$$

The NLO corrections to the $\sigma_{L,T}^{\gamma^*A}$ cross section include the tree-level contribution describing the photon fluctuating to a $q\bar{q}g$ system, where the additional gluon also interacts with the target. Furthermore, NLO corrections also include virtual one-loop corrections to the light-cone wave function describing the $\gamma^* \rightarrow q\bar{q}$ splitting. The $q\bar{q}g$ and the one-loop $q\bar{q}$ contributions in the dipole factorization have been first calculated in the massless limit in Refs. [10–12], and later results with a finite quark mass enabling also calculations of heavy quark production were published in Refs. [14–16].

At NLO accuracy, a rapidity divergence appears when the gluon longitudinal momentum fraction $z_2 \rightarrow 0$. This divergence needs to be absorbed into the Balitsky-Kovchegov evolution of the dipole-target scattering amplitude, and there are different scheme choices that can be employed. In this work, we apply the “unsubtracted” scheme following Refs. [11,38,39], which has been previously applied in small- x phenomenology, e.g., in Refs. [31,32,40]. In the unsubtracted scheme the gluon

longitudinal momentum fraction z_2 is limited from below by requiring that the invariant mass of the $q\bar{q}g$ system cannot exceed the center-of-mass energy of the photon-nucleon system W .

In the unsubtracted scheme, the NLO cross section for the γ^* -target scattering can be written as [31]

$$\sigma_{L,T}^{\text{NLO}} = \sigma_{L,T}^{\text{IC}} + \sigma_{L,T}^{\text{dip}} + \sigma_{L,T}^{qg,\text{unsub}}. \quad (5)$$

Here $\sigma_{L,T}^{\text{IC}}$ is the lowest order contribution from the $q\bar{q}$ -target scattering without including the BK evolution, $\sigma_{L,T}^{qg,\text{unsub}}$ describes the contribution where the $q\bar{q}g$ system interacts with the target, and the dipole contribution $\sigma_{L,T}^{\text{dip}}$ contains loop corrections to the $q\bar{q}$ -target scattering. Note that when the cross section is written in this form, UV divergences have been canceled between $\sigma_{L,T}^{\text{dip}}$ and $\sigma_{L,T}^{qg,\text{unsub}}$ and as such this division is not unique.

The lowest order contribution $\sigma_{L,T}^{\text{IC}}$ can be evaluated using Eq. (1) together with the dipole amplitude evaluated at the initial condition of the BK evolution. This is different from the leading order dipole picture result, as in the small- x power counting $\alpha_s \ln 1/x \sim 1$, and terms of this form are resummed by having an x -dependent dipole amplitude satisfying the BK equation. In the unsubtracted scheme applied here, the BK evolution dynamics is included in the term $\sigma_{L,T}^{qg,\text{unsub}}$.

The term describing the $q\bar{q}g$ -target interaction can be written as

$$\sigma_{L,T}^{qg,\text{unsub}} = K_{q\bar{q}g} \otimes N_{012}. \quad (6)$$

Similarly for the $q\bar{q}$ -target scattering at NLO we have

$$\sigma_{L,T}^{\text{dip}} = K_{q\bar{q}} \otimes N_{01}. \quad (7)$$

The convolutions in Eqs. (6) and (7) are over the transverse coordinates $\mathbf{x}_i$ and the parton longitudinal momentum fractions z_i . Explicit expressions for the hard factors $K_{q\bar{q}g}$ and $K_{q\bar{q}}$ with nonzero quark masses employed in this work can be found from Refs. [14–16], for brevity we do not repeat these expressions in this work.

Eikonal quark(gluon)-target interaction is described in terms of Wilson lines in fundamental (adjoint) representation. The amplitude for the quark-antiquark dipole to scatter off the dense CGC target is $N_{01} = 1 - S_{01}$, where

$$S_{01} = \frac{1}{N_c} \langle \text{Tr}\{V(\mathbf{x}_0)V^\dagger(\mathbf{x}_1)\} \rangle. \quad (8)$$

The average $\langle \cdot \rangle$ refers to an average over the target color field configurations. Here $\mathbf{x}_0$ and $\mathbf{x}_1$ are the quark and antiquark transverse coordinates, and $N_{01} = N(|\mathbf{x}_{01}|)$ with $\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j$. As mentioned above, the impact parameter dependence of the dipole amplitude N_{01} is neglected in this

work (see also Refs. [25,41,42] for recent discussions of the impact parameter dependence in the BK evolution). Similarly, the scattering amplitude for the $q\bar{q}g$ system $N_{012} = 1 - S_{012}$ is [10]

$$S_{012} = \frac{N_c}{2C_F} \left(S_{02}S_{21} - \frac{1}{N_c^2} S_{01} \right). \quad (9)$$

In the unsubtracted scheme, as already mentioned above, one introduces a lower limit for the gluon longitudinal momentum fraction z_2 when evaluating Eq. (6). Following Ref. [31], this limit is

$$z_2 > z_{2,\min} = \frac{Q_0^2}{W^2}. \quad (10)$$

Again following [31] we set the nonperturbative target momentum scale $Q_0^2 = 1 \text{ GeV}^2$.

When the gluon loop corrections to the $q\bar{q}$ -target scattering, Eq. (7), are calculated in Refs. [12,14–16], the integral over z_2 has been calculated analytically, integrating down to $z_2 = 0$. This is possible as, unlike when evaluating $\sigma_{L,T}^{qg,\text{unsub}}$, the z_2 integral in $\sigma_{L,T}^{\text{dip}}$ is not logarithmically divergent and the contribution from $z_2 < z_{2,\min}$ can be expected to be small.

B. Approximative BK evolution at next-to-leading order

The dipole-target scattering amplitude depends on the energy (or Bjorken- x), and in the large- N_c limit this evolution is described in terms of the BK equation. The BK equation at next-to-leading order has been derived in Ref. [17], and the most important higher-order contributions enhanced by large transverse logarithms have been resummed to all orders using different prescriptions in Refs. [20–23].

In this work, we approximate the next-to-leading order BK equation by the “kinematically constrained BK equation” (KCBK) from Ref. [23] (following the terminology of Ref. [31]). The KCBK equation resums the most important contributions enhanced by double transverse logarithms, which in practice is obtained by imposing the required time ordering between the consecutive gluon emissions in the evolution. It has been shown in Ref. [19] that when such higher order contributions are resummed to the leading order equation, a good approximation for the NLO BK equation is obtained. We have also explored a setup that additionally resums contributions enhanced by single transverse logarithms [22], but such setup did not result in good fits as we will discuss in more detail in Sec. IV. The inclusion of numerically more demanding NLO corrections not enhanced by transverse logarithms to the evolution is left for future work. In that case the resummation of single transverse logarithms will also be crucial [22]).

The KCBK evolution equation reads

$$\partial_Y S_{01}(Y) = \int d^2 \mathbf{x}_2 K_{\text{BK}}(\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2) \theta(Y - \Delta_{012}) \times [S_{02}(Y - \Delta_{012}) S_{21}(Y - \Delta_{012}) - S_{01}(Y)]. \quad (11)$$

Here the evolution rapidity Y is defined using the gluon plus momentum $k^+ = z_2 q^+$ where q^+ is the large photon plus momentum

$$Y = \ln \frac{k^+}{P^+}. \quad (12)$$

Here $P^+ = Q_0^2/(2P^-)$ is the target plus momentum.

In this work we apply two different running coupling prescriptions. First, the coordinate space strong coupling constant in the BK evolution and in the impact factor can be evaluated at the distance scale set by the parent dipole $|\mathbf{x}_{01}|$. In this case the BK kernel, which describes the probability density to emit a gluon to the transverse position $\mathbf{x}_2$, reads

$$K_{\text{BK}}(\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2) = \frac{N_c \alpha_s(\mathbf{x}_{01}^2)}{2\pi^2} \frac{\mathbf{x}_{01}^2}{\mathbf{x}_{12}^2 \mathbf{x}_{02}^2} \quad (13)$$

We refer to this as the parent dipole prescription. The other option, referred to as the Balitsky + smallest dipole prescription [43], corresponds to the BK kernel

$$K_{\text{BK}}(\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2) = \frac{N_c \alpha_s(\mathbf{x}_{01}^2)}{2\pi^2} \left[\frac{\mathbf{x}_{01}^2}{\mathbf{x}_{12}^2 \mathbf{x}_{02}^2} + \frac{1}{\mathbf{x}_{02}^2} \left(\frac{\alpha_s(\mathbf{x}_{02}^2)}{\alpha_s(\mathbf{x}_{12}^2)} - 1 \right) + \frac{1}{\mathbf{x}_{12}^2} \left(\frac{\alpha_s(\mathbf{x}_{12}^2)}{\alpha_s(\mathbf{x}_{02}^2)} - 1 \right) \right]. \quad (14)$$

In the limit where one of the daughter dipoles $\mathbf{x}_{02}, \mathbf{x}_{12}$ is very small, the running coupling scale is set by the smallest dipole, and the Balitsky prescription typically results in slower evolution compared to the parent dipole one. As such, in this prescription we evaluate the running coupling in the impact factor at the scale set by the smallest dipole size: $\alpha_s = \alpha_s(\min\{\mathbf{x}_{01}^2, \mathbf{x}_{02}^2, \mathbf{x}_{12}^2\})$.

The strong coupling in the coordinate space is evaluated as

$$\alpha_s(\mathbf{x}_{ij}^2) = \frac{4\pi}{\beta_0 \ln \left[\left(\frac{\mu_0^2}{\Lambda_{\text{QCD}}^2} \right)^{1/c} + \left(\frac{4C^2}{\mathbf{x}_{ij}^2 \Lambda_{\text{QCD}}^2} \right)^{1/c} \right]}. \quad (15)$$

Here $\Lambda_{\text{QCD}} = 0.241 \text{ GeV}$, $\beta = (11N_c - 2N_f)/3$, and we use $N_f = 3$ in this work. Based on [44,45], the expected value for C^2 is $C^2 = e^{-2\gamma_E} \approx 0.32$. This work, however, takes C^2 as a fit parameter to describe the uncertainty related to the running coupling prescriptions and to account

for higher order corrections. Note that although we also include the charm quark contribution, we use $N_f = 3$ at all transverse distance scales for simplicity (as in Ref. [32]) as it is not obvious how N_f should depend on this coordinate space scale.¹

The KCBK equation (11) is nonlocal in rapidity, with the rapidity shift

$$\Delta_{012} = \max \left\{ 0, \ln \frac{\min\{\mathbf{x}_{02}^2, \mathbf{x}_{21}^2\}}{\mathbf{x}_{01}^2} \right\}. \quad (16)$$

Following again Refs. [31,32], we initialize the BK evolution at the kinematical limit $Y = 0$ (corresponding to the “ $Y_{0,\text{BK}} = 0$ ” setup of [31]), and solve the equation numerically using the Euler method.

The initial dipole is parametrized using a McLerran-Venugopalan (MV) model [26] inspired parametrization, similarly as in Refs. [29–31]

$$S_{01}(Y = 0) = \exp \left[-\frac{(\mathbf{x}_{ij}^2 Q_{s,0}^2)^\gamma}{4} \ln \left(\frac{1}{|\mathbf{x}_{ij}| \Lambda_{\text{QCD}}} + e_c \cdot e \right) \right]. \quad (17)$$

Here the free parameters are the anomalous dimension, γ , infrared regulator, e_c , and $Q_{s,0}^2$ controlling the initial saturation scale.

In principle it would be preferable to take $\gamma = 1$ as in the MV model, as that forces the momentum space dipole amplitude (unintegrated gluon distribution) to be positive definite [46], and enables a more natural generalization for the dipole-nucleus scattering in the optical Glauber model [30]. However, with such a parametrization for the initial condition we did not find any good fit. As such, we use the so-called MV^γ parametrization from now on, fixing $e_c = 1$ and considering γ as a free parameter.

The dipole amplitudes N_{01} and N_{012} in convolutions (6) and (7) satisfy the BK equation. Following the previous NLO fit of Ref. [31], when calculating the dipole term $\sigma_{L,T}^{\text{dip}}$ the dipole amplitude is evaluated at $Y = \ln(1/x)$. In the case of $q\bar{q}g$ -target interaction in $\sigma_{L,T}^{q\bar{q},\text{unsub}}$, stability of the factorization scheme requires the evolution rapidity to depend also on the gluon momentum fraction z_2 [23,38,39,47,48], and it becomes

$$Y \equiv \ln \frac{k^+}{P^+} = \ln \frac{W^2 z_2}{Q_0^2}. \quad (18)$$

Here $k^+ = z_2 q^+$ is the gluon plus momentum, q^+ the large photon plus momentum and $P^+ = Q_0^2/(2P^-)$. Note that as already discussed above, when the $\sigma_{L,T}^{\text{dip}}$ term is derived, an integral over z_2 from $z_2 = 0$ to $z_2 = 1$ is already performed

¹One potential prescription has been suggested in Ref. [29].

analytically, and as such it is not possible to use the same z_2 dependent evolution rapidity in that term.²

III. BAYESIAN INFERENCE SETUP

The inverse problem is a common challenge in high-energy particle physics phenomenology, where the underlying cause is extracted from observed effects, instead of the usual forward mindset of using theory to predict observations. Given the nonperturbative nature of the initial dipole-proton amplitude, it can only be constrained using available experimental data. The experimental datasets used in this work are the combined reduced total cross section [33] and the reduced charm production cross section [34] measured in deep inelastic ep scattering at HERA. In this analysis, we include data from the kinematical domain $2.0 \text{ GeV}^2 \leq Q^2 < 50.0 \text{ GeV}^2$ and $x \leq 0.01$ which gives us a total of 404 data points for the total reduced cross section and 34 data points for the charm contribution. The lower limit in Q^2 ensures the validity of the perturbative calculation, and the upper limit excludes the part of the phase space where Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) dynamics would dominate.

Bayesian inference is a statistical approach that addresses the inverse problem in a probabilistic manner. The inference setup used in this work is equivalent to what we used in Ref. [35] and is summarized here for completeness.

The posterior function $P(\mathbf{y}(\boldsymbol{\theta})|\mathbf{y}_{\text{exp}})$ is the probability that a model $\mathbf{y}(\boldsymbol{\theta})$, at parameter point $\boldsymbol{\theta}$, is true in light of experimental data $\mathbf{y}_{\text{exp}}$. Bayes' theorem inverts the likelihood function, $P(\mathbf{y}_{\text{exp}}|\mathbf{y}(\boldsymbol{\theta}))$, to find the posterior

$$P(\mathbf{y}(\boldsymbol{\theta})|\mathbf{y}_{\text{exp}}) \propto P(\mathbf{y}_{\text{exp}}|\mathbf{y}(\boldsymbol{\theta}))P(\boldsymbol{\theta}). \quad (19)$$

The likelihood function, $P(\mathbf{y}_{\text{exp}}|\mathbf{y}(\boldsymbol{\theta}))$, is the degree that the model describes the observed data at a given parameter point. The log likelihood function can be written as

$$\begin{aligned} \log P(\mathbf{y}_{\text{exp}}|\mathbf{y}(\boldsymbol{\theta})) \\ = -\frac{1}{2} [\Delta\mathbf{y}(\boldsymbol{\theta})^T \Sigma^{-1}(\boldsymbol{\theta}) \Delta\mathbf{y}(\boldsymbol{\theta}) + \log(2\pi \det(\Sigma))], \end{aligned} \quad (20)$$

where $\Delta\mathbf{y}(\boldsymbol{\theta}) = \mathbf{y}(\boldsymbol{\theta}) - \mathbf{y}_{\text{exp}}$, while $\Sigma = \Sigma_{\text{exp}} + \Sigma_{\text{model}}(\boldsymbol{\theta})$. This work accounts for both uncorrelated and correlated systematic uncertainties present in the experimental data, where Σ_{exp} is the full experimental covariance matrix. The theoretical covariance matrix provided by the Gaussian process emulator described below is represented by Σ_{model} .

The prior $P(\boldsymbol{\theta})$ encodes knowledge, like bounds of the parameter space, about the model without the bias introduced by the experiment. In this work we consider a

uniform prior, i.e., $P(\boldsymbol{\theta}) = 1$ in the allowed region of the parameter space and 0 outside.

Sampling algorithms belonging to the Markov chain Monte Carlo (MCMC) class estimate the posterior and find the target distribution once the Markov chain has reached equilibrium. An example of such is the Metropolis-Hastings (MH) algorithm. Starting from an initial position $\boldsymbol{\theta}_0$, a random point in the vicinity of the original point within the parameter space is selected as the proposed next step $\boldsymbol{\theta}_1$. The probability that the proposal is accepted is based on the ratio $P(\mathbf{y}(\boldsymbol{\theta}_{i+1})|\mathbf{y}_{\text{exp}})/P(\mathbf{y}(\boldsymbol{\theta}_i)|\mathbf{y}_{\text{exp}})$. Eventually, the chain will converge to areas of high posterior and samples from this chain will make up the target posterior distribution.

The MCMC algorithm, in this study, evaluates the posterior at around $\sim 10^6$ points or more in the parameter space. Hence, a Gaussian process emulator (GPE) is used in place of the model (the NLO cross section calculation) as a more efficient substitute. An estimate from the GPE is a function drawn from a Gaussian process that is characterized by a covariance kernel. We use a combination of the radial basis function and the Matérn kernels to form a more flexible covariance kernel. The hyperparameters describing the kernel are then optimized by training the emulator. The training data comprises of model calculations for parameter vectors chosen through a latin hypercube algorithm, ensuring an even spread throughout the parameter space. Principal component analysis further reduces the dimensionality of the training data and enables independent training of a GPE for each principal component. The covariance matrix of the GPE Σ_{emulator} , which describes the GPE's estimated uncertainty, will represent the theoretical uncertainty of the model, $\Sigma_{\text{model}} = \Sigma_{\text{emulator}}$.

This work employs the EMCEE package [50] that implements an affine-invariant ensemble sampler of MH walkers, whose positions are updated based on the posterior evaluated by the other walkers. The GPEs are implemented using the SKLEARN package [51]. We train separate emulators for the reduced cross section and for the charm production using training data at 500 parameter points, and 6 principal components that capture more than 99% of the model variance. The GPE and MCMC sampler work hand in hand to explore the parameter space and extract the target posterior distribution.

To validate the emulators, we quantify the emulator accuracies and determine if the uncertainty estimates are realistic. First, comparisons between the emulator and the model when calculating the total reduced cross section σ_r and the charm production cross section $\sigma_{r,c}$ are shown for the Balitsky + smallest dipole and the parent dipole setups in Figs. 1 and 2, respectively. Overall we find an excellent agreement, and the average relative uncertainty δ_{rel} shown in the figures in all cases is clearly smaller than the typical uncertainty in the HERA data.

The uncertainty estimate provided by the emulator is validated by computing the z score, which is a distribution of the difference between the emulator prediction and the

²The z_2 dependence to the $\sigma_{L,T}^{\text{dip}}$ has been restored in the massless limit in Ref. [49].

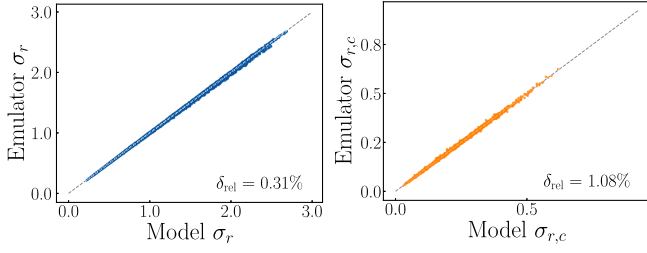


FIG. 1. Comparison between the emulator prediction and the model calculation for the setup using the Balitsky + smallest dipole running coupling prescription.

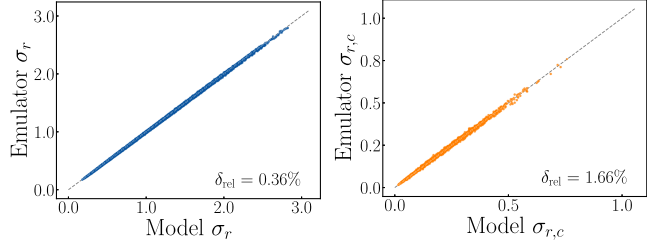


FIG. 2. Comparison between the emulator prediction and the model calculation for the setup using the parent dipole running coupling.

actual model output, normalized by the emulator uncertainty estimate. The z -score distribution for both running coupling schemes is shown in Fig. 3. For a perfect emulator this distribution has a zero mean and unit width. In our case, we observe that our GPEs slightly overestimate the

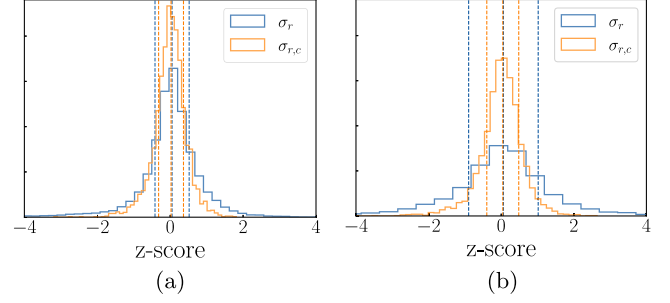


FIG. 3. The z -score distribution over 100 validation points sampled from the prior. The vertical lines indicate the one-standard-deviation interval. (a) Balitsky + smallest dipole. (b) Parent dipole.

emulator uncertainty, as reflected by the standard deviations indicated by dashed vertical lines being typically less than 1. Overall we conclude that the emulators are very accurate, and the emulator uncertainties are also realistic and can be used in the MCMC sampling phase.

Finally we validate the MCMC sampler and the overall Bayesian inference framework by performing a closure test. This test constrains the model to an artificially generated data constructed from known input parameters, here considered as the *true* values. Each data point is then assigned an uncertainty such that the relative uncertainty matches the HERA data. The established GPE and MCMC sampler should then be able to recover a posterior distribution that is aligned with the true values. We do such a test for 4 distinct parameter points chosen, and subsequently removed, from the emulator training set. Figure 4 demonstrates a close

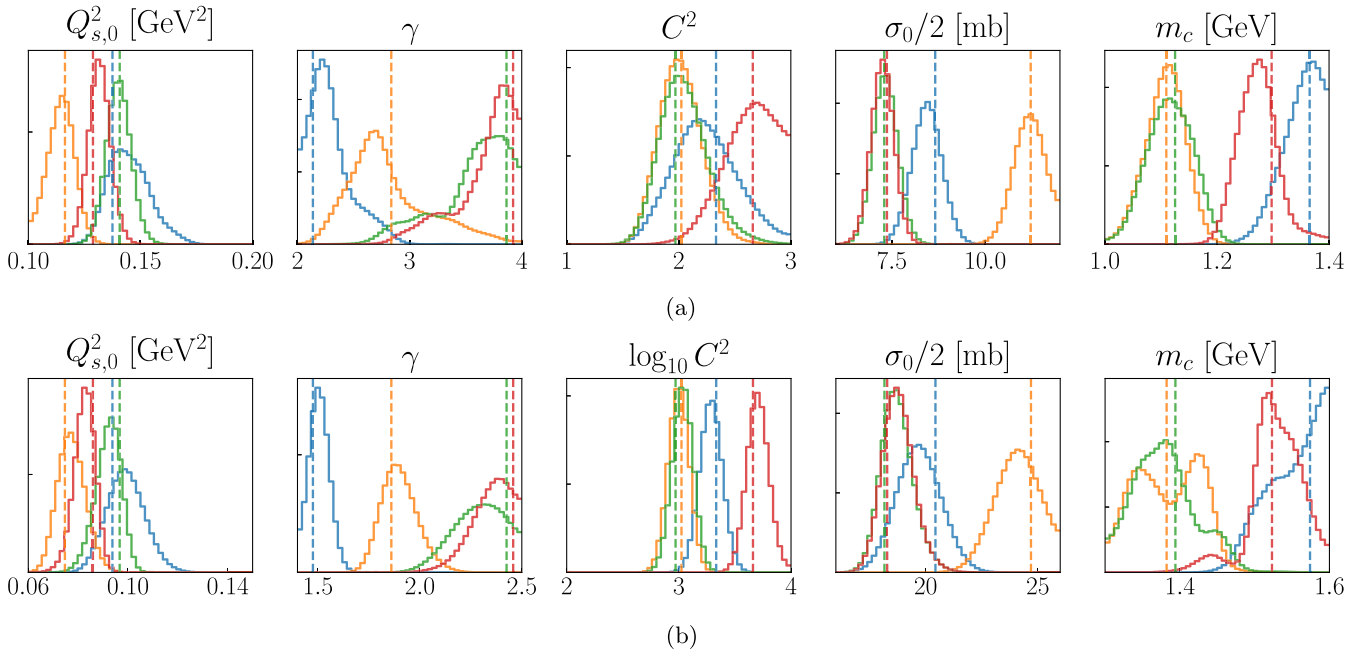


FIG. 4. Results from the closure test, comparison between extracted posterior distributions (solid lines) and corresponding true parameter values (dashed lines). (a) Balitsky + smallest dipole coupling. (b) Parent dipole coupling.

alignment between the peaks of the posterior distributions and the known true values (vertical dashed lines). Such agreement allows us to be confident on the robustness and reliability of our own Bayesian setup in extracting target model parameters.

IV. RESULTS

The model parameters $Q_{s,0}^2$, γ , C^2 , $\sigma_0/2$, and m_c are simultaneously constrained through Bayesian inference

using HERA combined data for the total DIS cross section and for the charm production. The charm mass, m_c , in the pole mass scheme used in the calculation of Ref. [16], is also considered to be a fit parameter as in Ref. [32].

The posterior distribution of the model parameters obtained in the setup using the Balitsky + smallest dipole running coupling prescription is shown in Fig. 5, and similar results obtained using the parent dipole coupling are presented in Fig. 6. Here the diagonal panels show the marginal distribution, and off-diagonal panels illustrate

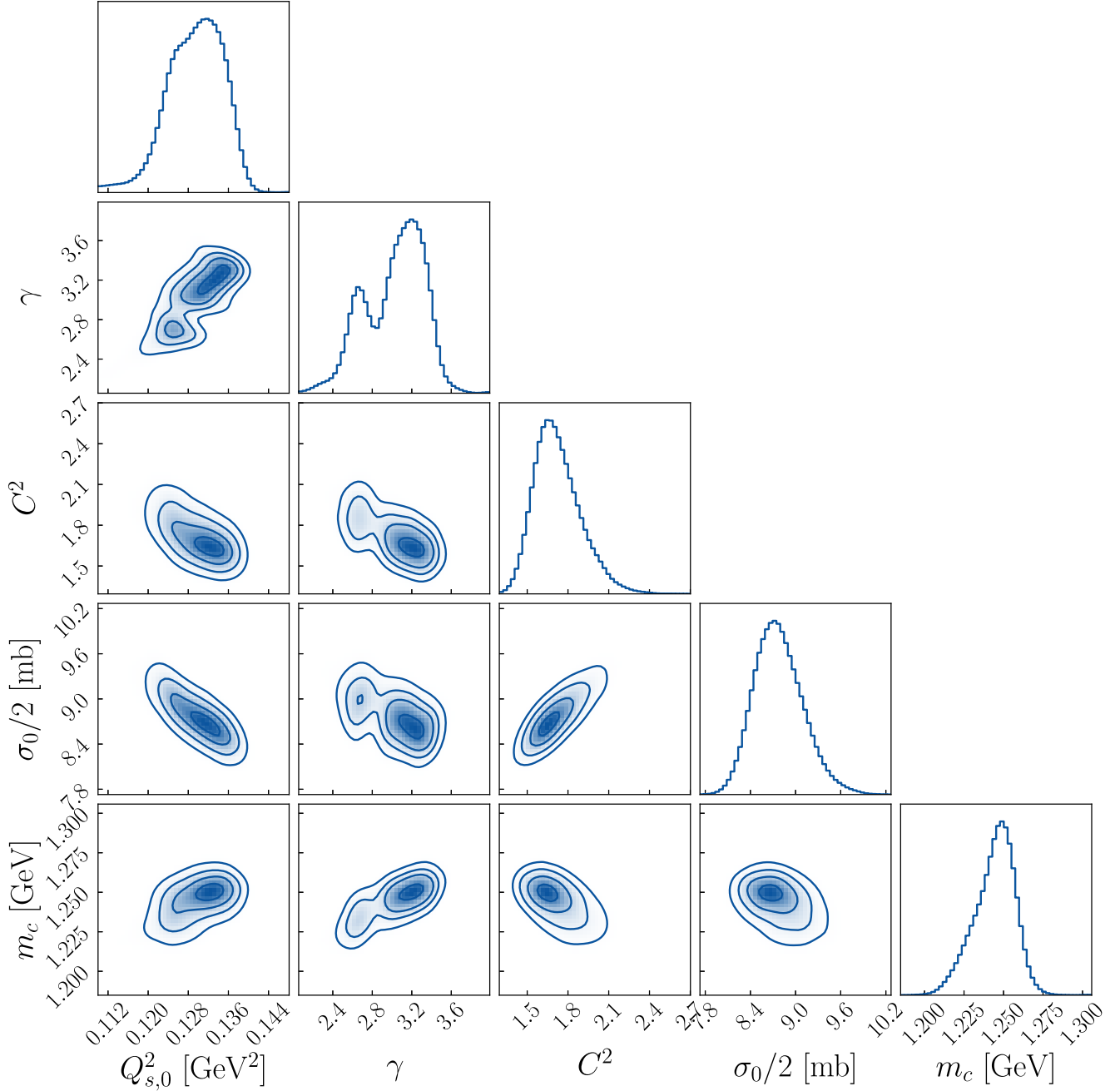


FIG. 5. 1- and 2-dimensional projections of the posterior probability distribution for the Balitsky + smallest dipole running coupling setup.

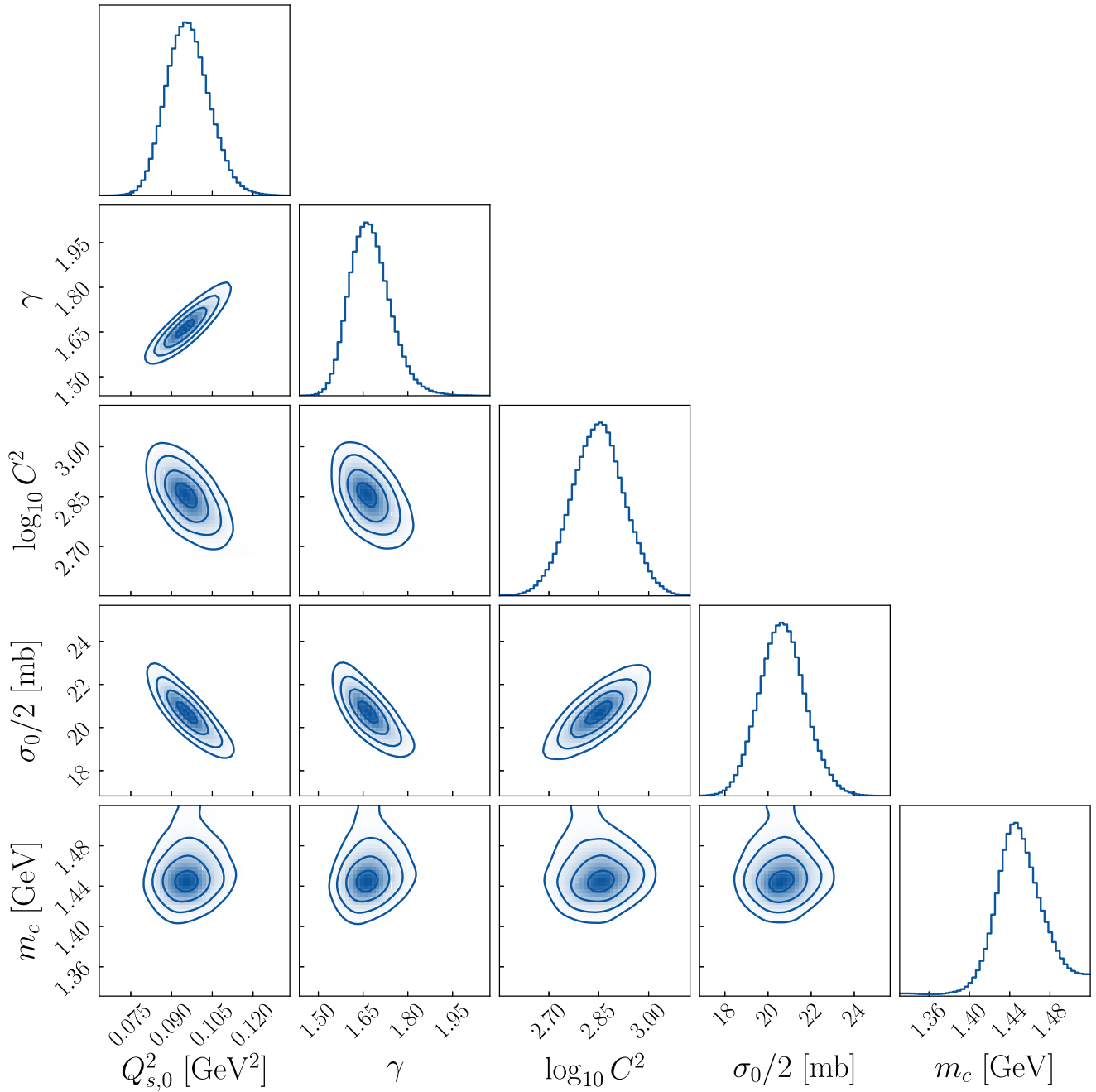


FIG. 6. 1- and 2-dimensional projections of the posterior probability distribution for the parent dipole running coupling setup.

correlations between the two model parameters. In both setups all model parameters are tightly constrained by the HERA data.

The maximum *a posteriori* (MAP) parametrizations resulting in the highest likelihoods are shown in Table I, along with the $\chi^2/\text{d.o.f.}$ values of $\mathcal{O}(1)$ demonstrating that the description of the precise HERA data obtained with only five parameters is overall excellent. We also report the posterior median with uncertainty values covering the 95% credible interval. Reported also are the corresponding values for the model-independent proton saturation scale,

Q_s^2 , at $Y = \ln(1/0.01)$ obtained by solving $N(r^2 = 2/Q_s^2) = 1 - e^{-1/2}$. Samples from the posterior distribution are available in Ref. [52]. These model parameter vectors can be used to define the initial condition for the BK evolution which is a necessary input for all NLO calculations within the CGC framework, enabling one to propagate the initial condition uncertainty to the cross section level. For all the posterior samples, data files containing the corresponding evolved dipole amplitudes are provided as well.

Similarly as in Ref. [32], a better simultaneous description of both the total and charm production datasets is

TABLE I. MAP and median values for the BK initial condition parametrization obtained using the Balitsky + smallest dipole (labeled as Bal + SD) and parent dipole running coupling schemes. The uncertainty estimates within the 95% credible intervals are also shown, as well as $\chi^2/\text{d.o.f.}$ values when compared different datasets and when including all HERA data, and saturation scales at $Y = \ln \frac{1}{0.01}$ defined as $N(r^2 = 2/Q_s^2) = 1 - e^{-1/2}$.

Parameter description	KCBK + Bal+SD		KCBK + parent dipole	
	Median $\pm 2\sigma$	MAP	Median $\pm 2\sigma$	MAP
Initial scale, $Q_{s,0}^2$ [GeV ²]	$0.130^{+0.008}_{-0.012}$	0.124	$0.096^{+0.016}_{-0.014}$	0.090
Anomalous dimension, γ	$3.08^{+0.38}_{-0.64}$	3.23	$1.67^{+0.15}_{-0.11}$	1.60
Running coupling scale, C^2	$1.70^{+0.41}_{-0.25}$	1.74	705^{+289}_{-209}	663
Proton transverse area, $\sigma_0/2$ [mb]	$8.75^{+0.72}_{-0.53}$	9.08	$20.7^{+2.1}_{-2.0}$	20.7
Charm mass, m_c [GeV]	$1.25^{+0.02}_{-0.03}$	1.24	$1.45^{+0.06}_{-0.05}$	1.40
Saturation scale, Q_s^2 , at $Y = \ln(1/0.01)$ [GeV ²]	0.203	0.196	0.208	0.199
$\chi^2/\text{d.o.f.}$ values				
σ_r data only	1.18	1.18	1.12	1.09
$\sigma_{r,c}$ data only	3.96	4.37	2.84	3.12
All data	1.35	1.38	1.22	1.21
Average $\chi^2/\text{d.o.f.}$ over 100 samples				
σ_r data only	1.19		1.17	
$\sigma_{r,c}$ data only	3.80		2.68	
All data	1.38		1.25	

obtained if the parent dipole running coupling scheme is used. The total cross section is described approximatively equally well with both setups. When the theoretically better motivated Balitsky running coupling scheme [43] is used, we have $\chi^2/\text{d.o.f.} \sim 4$ when the best-fit values are compared with the charm production data. However, also in that case the agreement with the very precise world data can be considered to be good as illustrated by the combined $\chi^2/\text{d.o.f.} \sim 1.4$ values (averaged over 100 samples).

The different running coupling schemes result in somewhat different parametrizations. In particular, the proton transverse area $\sigma_0/2$ depends heavily on this scheme choice. The value obtained for $\sigma_0/2$ with the parent dipole prescription is more than twice the value obtained when the Balitsky + smallest dipole prescription is used. Similar systematics has been previously observed in Ref. [31] when fitting the light quark contribution only.

While both setups give similar results for γ^* -proton scattering, the differing proton transverse areas can significantly affect predictions for nuclear targets. In the optical Glauber model, the nuclear saturation scale as a function of the impact parameter $\mathbf{b}$ behaves as $Q_{s,A}^2(\mathbf{b}) \sim \frac{\sigma_0}{2} A T_A(\mathbf{b}) Q_{s,0}^{2\gamma}$ [30]. As such the magnitude of the nonlinear effects in dipole-nucleus scattering depends rather heavily on the extracted proton size, see, e.g., Ref. [53] in the case of inclusive hadron production in proton-lead collisions. Here A is the nuclear mass number and $T_A(\mathbf{b})$ is the transverse thickness profile normalized as $\int d^2\mathbf{b} T_A(\mathbf{b}) = 1$. As such, future nuclear DIS data, or other scattering processes

measured, e.g., at the LHC, can be expected to provide complementary constraints for the nonperturbative initial condition.

Complementary constraints for the proton size can also be obtained from exclusive processes that directly probe the target shape [54], or from diffractive structure function measurements [55]. Depending on the form assumed for the proton geometry, exclusive J/ψ production data from HERA [56,57] prefers $\sigma_0/2 \sim 10 \dots 20$ mb, with the lower value obtained with a Gaussian shape and the upper limit when a step function profile (which is not fully compatible with the measured spectra [58]) is used. As such, we conclude (as in Ref. [31]) that both setups are compatible with the vector meson production data, although those measurements prefer the value obtained with the (more realistic) Balitsky + smallest dipole prescription.

The other main difference between the two running coupling schemes is on the value obtained for the running coupling scale, parametrized by C^2 . As discussed in Sec. II B, the Balitsky prescription generically results in slower evolution speed. For this reason, the fit with the parent dipole scheme features larger values of $C^2 \sim 700$, as a larger C^2 slows down the evolution, see Eq. (15). When the Balitsky prescription is used, one obtains $C^2 \sim 1.5$ (note that dependence on C^2 is logarithmic). It is still significantly larger than the $C^2 = e^{-2\gamma_E}$ estimate discussed in Sec. II B. However, we also note that the full NLO BK equation is known to result in even slower evolution than

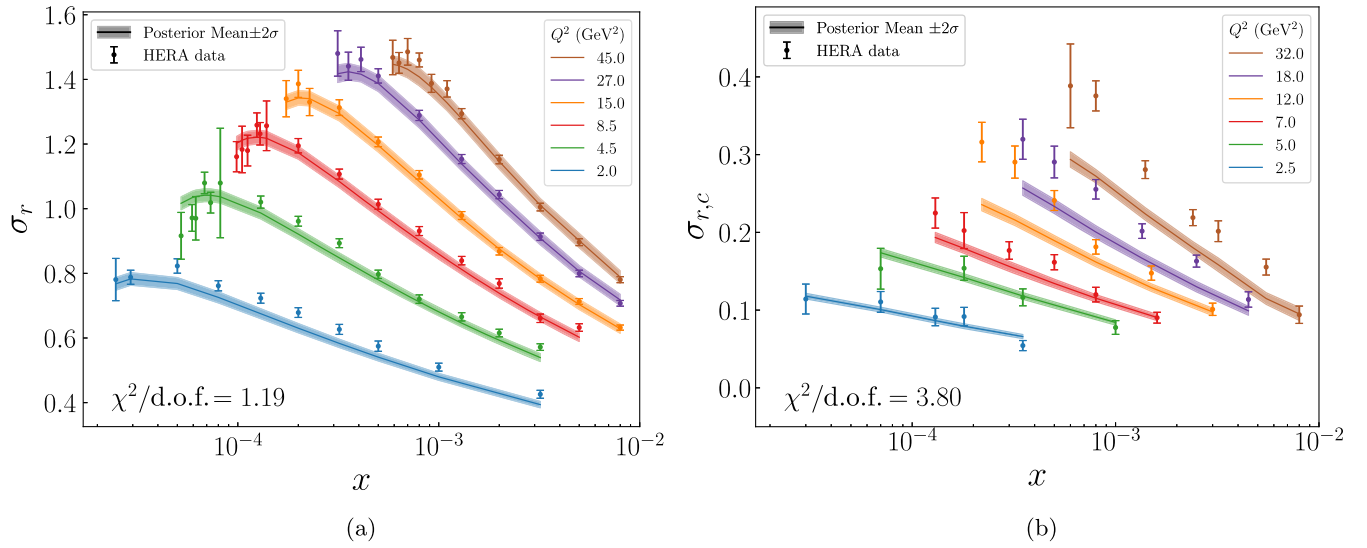


FIG. 7. Reduced total cross section and charm production as a function of x , in selected Q^2 bins at $\sqrt{s} = 318$ GeV, calculated for posterior samples from fit using Balitsky + smallest dipole running coupling scheme compared to HERA data. (a) Total cross section. (b) Charm production.

the resummation-only approximation employed here [19], which should reduce the optimal value for C^2 further.

The charm mass is weakly correlated with other model parameters. This is expected, as it is most directly constrained by the charm production data. The values for the charm quark mass m_c obtained using both running coupling schemes are similar to what has been reported in Ref. [32].

The proton transverse area $\sigma_0/2$ (that acts as a normalization factor) is correlated with C^2 and anticorrelated with $Q_{s,0}^2$. These are the same correlations observed in the leading order posterior distribution in Ref. [35]. The

anticorrelation between $Q_{s,0}^2$ and $\sigma_0/2$ is because nonlinear effects are weak in dipole-proton scattering, and as such the cross section scales as $\sim \frac{\sigma_0}{2} Q_{s,0}^{2\gamma}$.

The quality of our fit is demonstrated in Figs. 7 and 8 where the reduced cross section and charm production are calculated using samples from both posterior distributions, and compared to the HERA data. In addition to central (average) values, the 2σ uncertainty estimates are shown as well. As seen from the good $\chi^2/\text{d.o.f.}$ values shown in the figure and quoted in Table I, as well, an excellent description of the total cross section data is obtained.

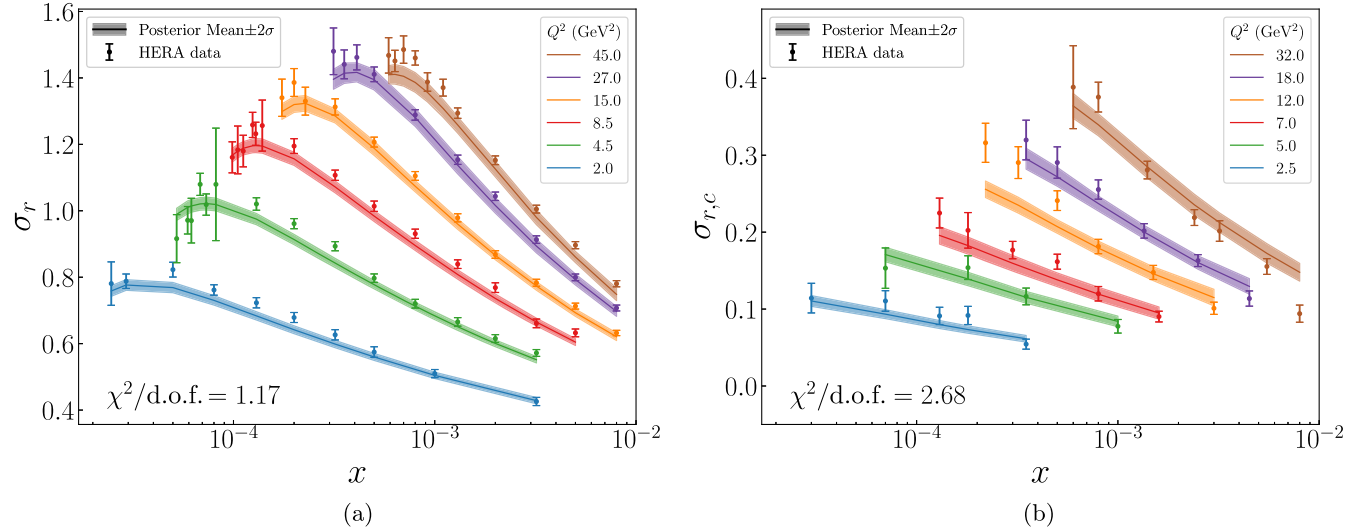


FIG. 8. Reduced total cross section and charm production as a function of x , in selected Q^2 bins at $\sqrt{s} = 318$ GeV, calculated for posterior samples from fit using parent dipole running coupling scheme compared to HERA data. (a) Total cross section. (b) Charm production.

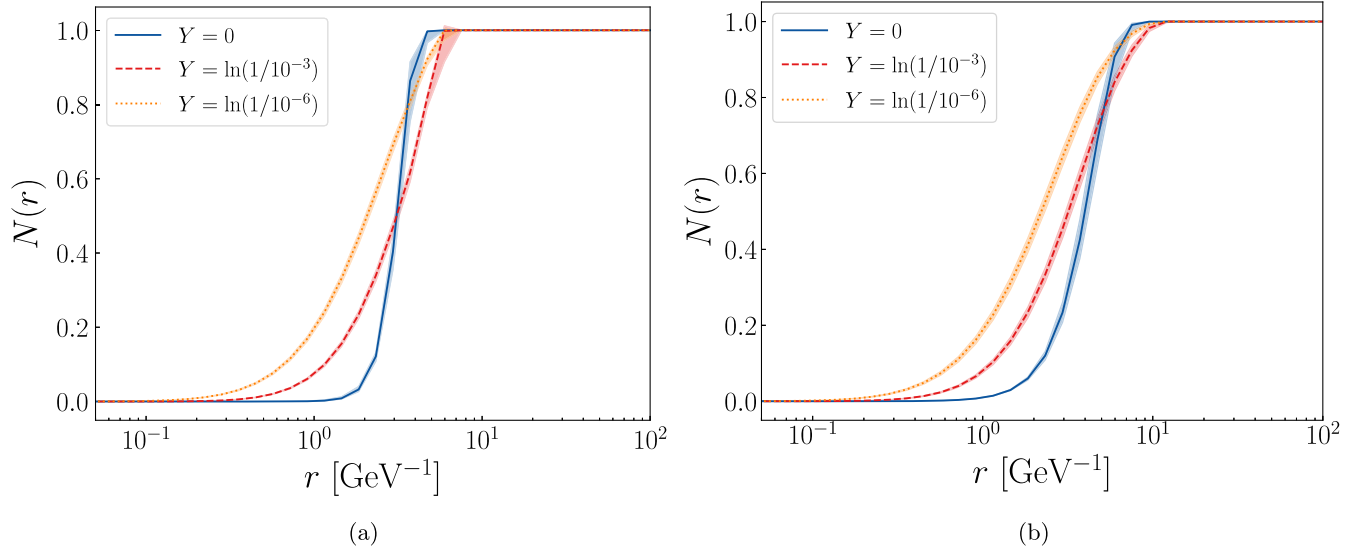


FIG. 9. The dipole amplitude at initial condition, $Y = 0$ and evolved toward $x = 10^{-3}$ and $x = 10^{-6}$ calculated for posterior samples from fit using Balitsky + smallest dipole or parent dipole schemes. The band describes the 2σ uncertainty. (a) Balitsky + smallest dipole running coupling scheme. (b) Parent dipole running coupling scheme.

The description of the charm production data is somewhat worse, with the cross section typically underestimated at small- x and large Q^2 . This tension between the two datasets is observed using both running coupling schemes, as the charm production data would prefer a faster Q^2 dependence. That would be obtained by having a smaller γ , which would then result in a worse description of the Q^2 dependence in the total cross section data. This tension is especially clear when the Balitsky + smallest dipole running coupling prescription is used. However, in both setups the global $\chi^2/\text{d.o.f.}$ (calculated including all data) remains good, see Table I. We also note that inclusion of the correlated experimental uncertainties allows for some model deviation and the agreement with the data may not look as good as the $\chi^2/\text{d.o.f.}$ value would suggest.

The dipole amplitudes at the initial condition and after $Y = \ln(1/10^{-3})$ and $Y = \ln(1/10^{-6})$ units of BK evolution are shown in Fig. 9 with uncertainty estimates. The anomalous dimension γ controls the slope of the dipole amplitude at small r where $N \sim (r^2 Q_{s,0}^2)^\gamma$, and due to the large γ values obtained here (similarly as in previous analyses [28–30]), the dipole amplitude is steeply falling at $r \lesssim 1/Q_s$. Although the resummed evolution does not significantly alter the shape of the dipole at small r [19], there is a clear transition toward a traveling-wave like solution at intermediate r . Especially with the Balitsky + smallest dipole prescription where the initial γ is larger, this transition effect is still visible in the LHC kinematics corresponding to $Y = \ln(1/10^{-6})$.

In the MV-model based parametrizations, $\gamma > 1$ results in negative unintegrated gluon distribution [46]. This is demonstrated in Fig. 10 for the 2-dimensional Fourier transform (2DFT) of the dipole amplitude

$$\tilde{S}_p(\mathbf{k}) = \int d^2\mathbf{r} e^{i\mathbf{k}\cdot\mathbf{r}} [1 - N(\mathbf{r})], \quad (21)$$

which is directly proportional to the unintegrated gluon distribution and to the leading order quark-target cross section. After the BK evolution, the 2D Fourier transform remains positive up to larger values of transverse momenta. However, given the rather steep initial dipole amplitude (large γ), especially in the Balitsky + smallest dipole setup, a very long evolution is necessary to render the Fourier transform positive definite in the phenomenologically relevant transverse momentum range.

We have also considered the “ResumBK” equation (following the terminology of Ref. [31]), which resums single transverse logarithms [22] on top of the double logs [21] included also in the KCBK equation. Using the same MV^γ parametrization, the setup produced fits that were not able to describe the HERA data, the best-fit parametrization resulting in global $\chi^2/\text{d.o.f.} \sim 1.9$ (and $\chi^2/\text{d.o.f.} \sim 5.5$ when comparing to the charm data only). This setup also required an anomalous dimensions $\gamma \gtrsim 4$ at the initial condition, which would also result in negative dipole amplitudes at small- r as a result of the evolution.³ We also considered the MV^e initial condition parametrization from Ref. [30] with the ResumBK evolution, which corresponds to Eq. (17) with fixed $\gamma = 1$, but it did not result in better fit quality as well (total $\chi^2/\text{d.o.f.} \gtrsim 2.5$). For these reasons, in this work we only report results obtained with the KCBK evolution. In the future, we plan to couple these single and double log resummations included in the ResumBK equation, and the full next-to-leading order BK

³In Ref. [31] this problem was avoided by enforcing $N(r) \geq 0$.

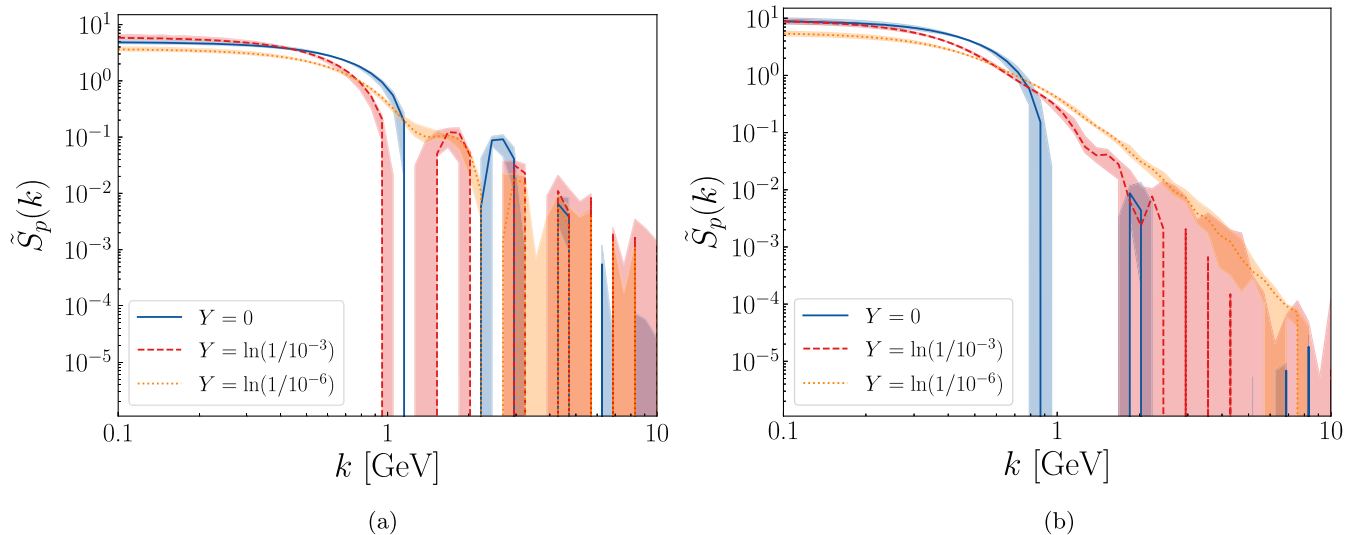


FIG. 10. 2-dimensional Fourier transform of the initial and evolved, to $x = 10^{-3}$ and $x = 10^{-6}$, dipole calculated for posterior samples from fit using Balitsky + smallest dipole or parent dipole schemes. The band describes the 2σ uncertainty. (a) Balitsky + smallest dipole running coupling scheme. (b) Parent dipole running coupling scheme.

equation [59], into the global analysis setup developed in this work.

V. CONCLUSIONS AND OUTLOOK

We have constrained the model parameters describing the nonperturbative initial condition of the small- x BK equation at next-to-leading order accuracy through a global analysis including the total cross section and charm production data from HERA. These datasets are found to provide tight constraints for the model parameters, as demonstrated in the posterior distributions extracted using Bayesian inference and shown in Figs. 5 and 6. Compared to the previous NLO fits [31,32] in the dipole picture, this work corresponds to the first global analysis with heavy quark contribution included, provides uncertainty estimates for the nonperturbative parameters and accounts for the correlated systematic uncertainties in the HERA data. The resulting posterior distributions can be used to propagate uncertainties in the initial condition of the BK equation to all CGC calculations at next-to-leading order accuracy, including also heavy quark contributions. Parametrization vectors sampled from the posterior distributions, as well as BK evolved dipole amplitudes, are available at [52].

In this work we combined the NLO impact factors to the approximated version of the NLO BK evolution equation. Results with two different running coupling prescriptions, parent dipole and Balitsky + smallest dipole, are reported. The Balitsky + smallest dipole prescription is physically better motivated and it results in model parameters that are also favored by exclusive vector meson production data. On the other hand, using both running coupling setups there is some tension between the total cross section and the

charm production data, and this tension is stronger when the Balitsky + smallest dipole running coupling scheme is used.

An analysis using the full NLO BK equation is the natural next step for this work. This will include, in the BK equation, contributions not enhanced by large transverse logarithms captured by resummation prescriptions, and additionally a resummation of single transverse logarithms not included in the KCBK equation used in this work. This can be achieved following the setup developed in Ref. [19]. Furthermore, it will also be important to repeat the analysis using the recently derived BK equation formulated in terms of the target rapidity evolution [20], and to include other observables such as diffractive cross section [60] to the global analysis.

ACKNOWLEDGMENTS

C. C. and H. M. are supported by the Research Council of Finland, the Centre of Excellence in Quark Matter, and Projects 338263, 346567 and 359902, and by the European Research Council (ERC, Grant Agreements No. ERC-2023-101123801 GlueSatLight and No. ERC-2018-ADG-835105 YoctoLHC). C. C. acknowledges the support of the Vilho, Yrjö and Kalle Väisälä Foundation. H. H. is supported by the Research Council of Finland (Flagship of Advanced Mathematics for Sensing Imaging and Modelling grant 359208; Centre of Excellence of Inverse Modelling and Imaging grant 353092), and the Vilho, Yrjö and Kalle Väisälä Foundation. Computing resources from CSC—IT Center for Science in Espoo, Finland and the Finnish Grid and Cloud Infrastructure (persistent identifier urn:nbn:fi:research-infras-2016072533) were used in this work.

The content of this article does not reflect the official opinion of the European Union and responsibility for the information and views expressed therein lies entirely with the authors.

DATA AVAILABILITY

The data that support the findings of this article are openly available [52].

-
- [1] E. Iancu and R. Venugopalan, The color glass condensate and high-energy scattering in QCD, in *Quark-Gluon Plasma 4*, edited by R. C. Hwa and X.-N. Wang (World Scientific, Singapore, 2003), pp. 249–3363.
 - [2] R. Abdul Khalek *et al.*, Science requirements and detector concepts for the electron-ion collider: EIC yellow report, *Nucl. Phys. A* **1026**, 122447 (2022).
 - [3] P. Agostini *et al.* (LHeC Collaboration and FCC-he Study Group), The large hadron-electron collider at the HL-LHC, *J. Phys. G* **48**, 110501 (2021).
 - [4] D. P. Anderle *et al.*, Electron-ion collider in China, *Front. Phys. (Beijing)* **16**, 64701 (2021).
 - [5] A. H. Mueller and B. Patel, Single and double BFKL pomeron exchange and a dipole picture of high-energy hard processes, *Nucl. Phys. B* **442**, 471 (1994).
 - [6] Y. V. Kovchegov, Small x F_2 structure function of a nucleus including multiple pomeron exchanges, *Phys. Rev. D* **60**, 034008 (1999).
 - [7] I. Balitsky, Operator expansion for high-energy scattering, *Nucl. Phys. B* **463**, 99 (1996).
 - [8] I. Balitsky and G. A. Chirilli, Photon impact factor in the next-to-leading order, *Phys. Rev. D* **83**, 031502 (2011).
 - [9] I. Balitsky and G. A. Chirilli, Photon impact factor and k_T -factorization for DIS in the next-to-leading order, *Phys. Rev. D* **87**, 014013 (2013).
 - [10] H. Hänninen, T. Lappi, and R. Paatelainen, One-loop corrections to light cone wave functions: The dipole picture DIS cross section, *Ann. Phys. (N.Y.)* **393**, 358 (2018).
 - [11] G. Beuf, Dipole factorization for DIS at NLO: Combining the $q\bar{q}$ and $q\bar{q}g$ contributions, *Phys. Rev. D* **96**, 074033 (2017).
 - [12] G. Beuf, Dipole factorization for DIS at NLO: Loop correction to the $\gamma_{T,L}^* \rightarrow q\bar{q}$ light-front wave functions, *Phys. Rev. D* **94**, 054016 (2016).
 - [13] G. Beuf, NLO corrections for the dipole factorization of DIS structure functions at low x , *Phys. Rev. D* **85**, 034039 (2012).
 - [14] G. Beuf, T. Lappi, and R. Paatelainen, Massive quarks in NLO dipole factorization for DIS: Transverse photon, *Phys. Rev. D* **106**, 034013 (2022).
 - [15] G. Beuf, T. Lappi, and R. Paatelainen, Massive quarks in NLO dipole factorization for DIS: Longitudinal photon, *Phys. Rev. D* **104**, 056032 (2021).
 - [16] G. Beuf, T. Lappi, and R. Paatelainen, Massive quarks at one loop in the dipole picture of deep inelastic scattering, *Phys. Rev. Lett.* **129**, 072001 (2022).
 - [17] I. Balitsky and G. A. Chirilli, Next-to-leading order evolution of color dipoles, *Phys. Rev. D* **77**, 014019 (2008).
 - [18] T. Lappi and H. Mäntysaari, Direct numerical solution of the coordinate space Balitsky-Kovchegov equation at next to leading order, *Phys. Rev. D* **91**, 074016 (2015).
 - [19] T. Lappi and H. Mäntysaari, Next-to-leading order Balitsky-Kovchegov equation with resummation, *Phys. Rev. D* **93**, 094004 (2016).
 - [20] B. Ducloué, E. Iancu, A. H. Mueller, G. Soyez, and D. N. Triantafyllopoulos, Non-linear evolution in QCD at high-energy beyond leading order, *J. High Energy Phys.* **04** (2019) 081.
 - [21] E. Iancu, J. D. Madrigal, A. H. Mueller, G. Soyez, and D. N. Triantafyllopoulos, Resumming double logarithms in the QCD evolution of color dipoles, *Phys. Lett. B* **744**, 293 (2015).
 - [22] E. Iancu, J. D. Madrigal, A. H. Mueller, G. Soyez, and D. N. Triantafyllopoulos, Collinearly-improved BK evolution meets the HERA data, *Phys. Lett. B* **750**, 643 (2015).
 - [23] G. Beuf, Improving the kinematics for low- x QCD evolution equations in coordinate space, *Phys. Rev. D* **89**, 074039 (2014).
 - [24] T. Lappi, H. Mäntysaari, and A. Ramnath, Next-to-leading order Balitsky-Kovchegov equation beyond large N_c , *Phys. Rev. D* **102**, 074027 (2020).
 - [25] J. Cepila, J. G. Contreras, M. Matas, and M. Vaculciak, Impact-parameter-dependent solutions to the Balitsky-Kovchegov equation at next-to-leading order, *Phys. Rev. D* **111**, 096015 (2025).
 - [26] L. D. McLerran and R. Venugopalan, Computing quark and gluon distribution functions for very large nuclei, *Phys. Rev. D* **49**, 2233 (1994).
 - [27] K. J. Golec-Biernat and M. Wusthoff, Saturation effects in deep inelastic scattering at low Q^2 and its implications on diffraction, *Phys. Rev. D* **59**, 014017 (1998).
 - [28] J. L. Albacete, N. Armesto, J. G. Milhano, and C. A. Salgado, Non-linear QCD meets data: A global analysis of lepton-proton scattering with running coupling BK evolution, *Phys. Rev. D* **80**, 034031 (2009).
 - [29] J. L. Albacete, N. Armesto, J. G. Milhano, P. Quiroga-Arias, and C. A. Salgado, AAMQS: A non-linear QCD analysis of new HERA data at small- x including heavy quarks, *Eur. Phys. J. C* **71**, 1705 (2011).
 - [30] T. Lappi and H. Mäntysaari, Single inclusive particle production at high energy from HERA data to proton-nucleus collisions, *Phys. Rev. D* **88**, 114020 (2013).
 - [31] G. Beuf, H. Hänninen, T. Lappi, and H. Mäntysaari, Color glass condensate at next-to-leading order meets HERA data, *Phys. Rev. D* **102**, 074028 (2020).
 - [32] H. Hänninen, H. Mäntysaari, R. Paatelainen, and J. Penttala, Proton structure functions at next-to-leading order in the

- dipole picture with massive quarks, *Phys. Rev. Lett.* **130**, 192301 (2023).
- [33] H. Abramowicz *et al.* (H1 and ZEUS Collaborations), Combination of measurements of inclusive deep inelastic $e^\pm p$ scattering cross sections and QCD analysis of HERA data, *Eur. Phys. J. C* **75**, 580 (2015).
- [34] H. Abramowicz *et al.* (H1 and ZEUS Collaborations), Combination and QCD analysis of charm and beauty production cross-section measurements in deep inelastic ep scattering at HERA, *Eur. Phys. J. C* **78**, 473 (2018).
- [35] C. Casuga, M. Karhunen, and H. Mäntysaari, Inferring the initial condition for the Balitsky-Kovchegov equation, *Phys. Rev. D* **109**, 054018 (2024).
- [36] A. Morreale and F. Salazar, Mining for gluon saturation at colliders, *Universe* **7**, 312 (2021).
- [37] Y. V. Kovchegov and E. Levin, *Quantum Chromodynamics at High Energy* (Oxford University Press, New York, 2013), Vol. 33.
- [38] E. Iancu, A. H. Mueller, and D. N. Triantafyllopoulos, CGC factorization for forward particle production in proton-nucleus collisions at next-to-leading order, *J. High Energy Phys.* **12** (2016) 041.
- [39] B. Ducloué, H. Hänninen, T. Lappi, and Y. Zhu, Deep inelastic scattering in the dipole picture at next-to-leading order, *Phys. Rev. D* **96**, 094017 (2017).
- [40] H. Mäntysaari and J. Penttala, Exclusive heavy vector meson production at next-to-leading order in the dipole picture, *Phys. Lett. B* **823**, 136723 (2021).
- [41] D. Bendova, J. Cepila, J. G. Contreras, and M. Matas, Solution to the Balitsky-Kovchegov equation with the collinearly improved kernel including impact-parameter dependence, *Phys. Rev. D* **100**, 054015 (2019).
- [42] H. Mäntysaari, J. Penttala, F. Salazar, and B. Schenke, Finite-size effects on small- x evolution and saturation in proton and nuclear targets, *Phys. Rev. D* **111**, 054033 (2025).
- [43] I. Balitsky, Quark contribution to the small- x evolution of color dipole, *Phys. Rev. D* **75**, 014001 (2007).
- [44] Y. V. Kovchegov and H. Weigert, Triumvirate of running couplings in small- x evolution, *Nucl. Phys. A* **784**, 188 (2007).
- [45] T. Lappi and H. Mäntysaari, On the running coupling in the JIMWLK equation, *Eur. Phys. J. C* **73**, 2307 (2013).
- [46] B. G. Giraud and R. Peschanski, Fourier-positivity constraints on QCD dipole models, *Phys. Lett. B* **760**, 26 (2016).
- [47] B. Ducloué, T. Lappi, and Y. Zhu, Single inclusive forward hadron production at next-to-leading order, *Phys. Rev. D* **93**, 114016 (2016).
- [48] B. Ducloué, T. Lappi, and Y. Zhu, Implementation of NLO high energy factorization in single inclusive forward hadron production, *Phys. Rev. D* **95**, 114007 (2017).
- [49] H. Hänninen, Deep inelastic scattering in the dipole picture at next-to-leading order, Ph.D. thesis, Jyväskylä University, 2021.
- [50] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, EMCEE: The MCMC hammer, *Publ. Astron. Soc. Pac.* **125**, 306 (2013).
- [51] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, SCIKIT-LEARN: Machine learning in python, *J. Mach. Learn. Res.* **12**, 2825 (2011).
- [52] C. Casuga, H. Hänninen, and H. Mäntysaari, Posterior samples from Bayesian fit of initial condition to the BK equation, repo: Camcasuga/bayesian-nlodsfit-dipole: V1.0.0, 10.5281/zenodo.15552940.
- [53] H. Mäntysaari and Y. Tawabutr, Complete next-to-leading order calculation of single inclusive π^0 production in forward proton-nucleus collisions, *Phys. Rev. D* **109**, 034018 (2024).
- [54] S. R. Klein and H. Mäntysaari, Imaging the nucleus with high-energy photons, *Nat. Rev. Phys.* **1**, 662 (2019).
- [55] T. Lappi, A. D. Le, and H. Mäntysaari, Rapidity gap distribution of diffractive small- x_p events at HERA and at the EIC, *Phys. Rev. D* **108**, 114023 (2023).
- [56] C. Alexa *et al.* (H1 Collaboration), Elastic and proton-dissociative photoproduction of J/ψ mesons at HERA, *Eur. Phys. J. C* **73**, 2466 (2013).
- [57] S. Chekanov *et al.* (ZEUS Collaboration), Exclusive photoproduction of J/ψ mesons at HERA, *Eur. Phys. J. C* **24**, 345 (2002).
- [58] H. Kowalski, L. Motyka, and G. Watt, Exclusive diffractive processes at HERA within the dipole picture, *Phys. Rev. D* **74**, 074016 (2006).
- [59] I. Balitsky and G. A. Chirilli, Next-to-leading order evolution of color dipoles, *Phys. Rev. D* **77**, 014019 (2008).
- [60] G. Beuf, T. Lappi, H. Mäntysaari, R. Paatelainen, and J. Penttala, Diffractive deep inelastic scattering at NLO in the dipole picture, *J. High Energy Phys.* **05** (2024) 024.