

# Determination of the initial condition for the Balitsky-Kovchegov equation with transformers

**Meisen Gao,<sup>a,b,h</sup> Zhong-Bo Kang,<sup>c,d,e</sup> Jani Penttala,<sup>c,d</sup> and Ding Yu Shao<sup>b,f,g</sup>**

<sup>a</sup>*School of Physics, East China University of Science and Technology, Shanghai 200237, China*

<sup>b</sup>*Department of Physics and Center for Field Theory and Particle Physics, Fudan University, Shanghai, 200433, China*

<sup>c</sup>*Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA*

<sup>d</sup>*Mani L. Bhaumik Institute for Theoretical Physics, University of California, Los Angeles, CA 90095, USA*

<sup>e</sup>*Center for Frontiers in Nuclear Science, Stony Brook University, Stony Brook, NY 11794, USA*

<sup>f</sup>*Key Laboratory of Nuclear Physics and Ion-beam Application (MOE), Fudan University, Shanghai, 200433, China*

<sup>g</sup>*Shanghai Research Center for Theoretical Nuclear Physics, NSFC and Fudan University, Shanghai 200438, China*

<sup>h</sup>*Shanghai Key Laboratory of Particle Physics and Cosmology, Shanghai 200240, China*

*E-mail:* [msgao@ecust.edu.cn](mailto:msgao@ecust.edu.cn), [zkang@physics.ucla.edu](mailto:zkang@physics.ucla.edu),

[janipenttala@physics.ucla.edu](mailto:janipenttala@physics.ucla.edu), [dingyu.shao@cern.ch](mailto:dingyu.shao@cern.ch)

**ABSTRACT:** In the high-energy limit of QCD, scattering off nucleons and nuclei can be described in terms of Wilson-line correlators whose energy dependence is perturbative. The energy dependence of the two-point correlator, called the dipole amplitude, is governed by the Balitsky–Kovchegov (BK) equation. The initial condition for the BK equation can be fitted to the experimental data, which requires evolving the dipole amplitude for a large set of different parameter values. In this work, we train a transformer model to learn the energy dependence of the dipole amplitude, skipping the time-consuming numerical evaluation of the BK equation. The transformer predicts the learned dipole amplitude and the leading order inclusive deep inelastic scattering cross section very accurately, allowing for efficient fitting of the initial condition to the experimental data. Using this setup, we fit the initial condition of the BK equation to the inclusive deep inelastic scattering data from HERA and consider two different starting points  $x_0$  for the evolution. We find better agreement with the experimental data for a smaller  $x_0$ . This work paves the way for future studies involving global fits of the dipole amplitude at leading order and beyond.

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Theoretical framework</b>	<b>2</b>
2.1	Inclusive DIS in the dipole picture	2
2.2	Balitsky–Kovchegov evolution	4
<b>3</b>	<b>Fitting procedure</b>	<b>4</b>
3.1	Sampling and solving the BK equation	5
3.2	Neural network emulation of the BK amplitude	5
<b>4</b>	<b>Application to DIS observables</b>	<b>7</b>
<b>5</b>	<b>Summary</b>	<b>13</b>

---

## 1 Introduction

Mapping the partonic structure of protons and nuclei in the high-energy regime is a foremost challenge that provides the primary scientific motivation for the future Electron-Ion Colliders [1–3]. At a very small Bjorken- $x$  variable, the density of gluons within a hadron grows rapidly, leading to the gluon saturation phenomenon where non-linear QCD dynamics become dominant. The Color Glass Condensate (CGC) effective field theory [4, 5] provides a powerful framework for describing this dense gluonic matter. Central to the CGC is the dipole–target scattering amplitude, which encodes the essential information about the nonperturbative scattering off the gluonic target. The energy evolution of the dipole amplitude is governed by perturbative evolution equations, most notably the Balitsky–Kovchegov (BK) equation [6, 7]. However, the starting point for this evolution—the initial condition at a moderately small  $x_0$ —is fundamentally nonperturbative and must be determined from experimental data.

A convenient initial condition for the high-energy evolution is given by the McLerran–Venugopalan (MV) model [8–10], which allows for a description of the dipole amplitude in terms of a few free parameter that can be extracted from the experimental data. The MV model, and its generalizations, have then be used in many successful extractions at both leading order [11–18] and next-to-leading order [19, 20] using the inclusive deep inelastic scattering (DIS) data from HERA [21, 22]. The extracted parameters contain the fundamental information about the target structure, including the saturation scale  $Q_s$  that describes the onset of the gluon saturation phenomenon.

However, to determine the initial condition from the experimental data, one has to run the BK evolution for a wide range of parameter values. Due to the nonlinear nature

of the BK evolution, its evaluation is much more demanding than the linear BFKL [23–27] and DGLAP [28–31] evolution equations, making the BK evolution the bottleneck for numerical studies even at leading order. At higher orders in perturbation theory, the cross section itself becomes also numerically demanding due to the large amount of integrals over transverse coordinates that are typical for the dipole picture used in small- $x$  calculations. For these reasons, it becomes important to study efficient methods for evaluating these quantities with different models for the dipole amplitude.

To overcome this limitation, we turn to the rapidly advancing field of machine learning and the neural network approach. We employ a large language model *transformer architecture*, a class of neural networks renowned for its ability to capture long-range correlations and complex patterns in sequential data. By treating both the model parameters and the kinematic variables as a sequence, the transformer’s self-attention mechanism can learn the intricate, non-local relationships within the experimental dataset (such as those from HERA) without preconceived notions of the underlying physics. This paper presents the first determination of the small- $x$  dipole amplitude using this novel transformer-based method. Moreover, we consider different starting points for the BK evolution, enabling us to study the bias in the initial condition and to assess the validity of the BK evolution for moderate values of the Bjorken- $x$  variable. It should be emphasized that the present analysis is restricted to leading-order accuracy, and our primary aim is to demonstrate the methodology rather than to obtain the most precise phenomenological fit. Gaussian-process emulators have been successfully applied in small- $x$  studies [17, 20], providing flexible surrogate models with quantified uncertainties. In this work, we employ a transformer-based approach, which can naturally accommodate larger training sets, capture correlations across the parameter space, and demonstrates stable performance even when modest extrapolation beyond the sampled region is required. This makes it well suited for global analyses requiring high precision over a wide kinematic range.

The paper is organized as follows. In Sec. 2, we briefly review the theoretical framework, including the dipole formalism in DIS and the BK equation. In Sec. 3, we detail our neural network architecture, the training methodology, and how we incorporate necessary physical constraints. In Sec. 4, we present the resulting dipole amplitude and perform a rigorous comparison against both the HERA data and the results from traditional parameterization-based fits. Finally, in Sec. 5, we summarize our conclusions and provide an outlook on how this data-driven approach can pave the way for precision studies at the future EIC.

## 2 Theoretical framework

### 2.1 Inclusive DIS in the dipole picture

Inclusive DIS is described by the structure functions  $F_2(x, Q^2)$  and  $F_L(x, Q^2)$ , which are related to the experimental reduced cross section  $\sigma_r$  by

$$\sigma_r(y, x, Q^2) = F_2(x, Q^2) - \frac{y^2}{1 + (1 - y)^2} F_L(x, Q^2), \quad (2.1)$$

with

$$F_2(x, Q^2) = F_L(x, Q^2) + F_T(x, Q^2), \quad (2.2)$$

where  $Q^2$  is the photon virtuality,  $x$  is the Bjorken scaling variable, and  $y$  is the inelasticity. The structure functions are defined in terms of the cross sections for the scattering of a transversely ( $T$ ) or longitudinally ( $L$ ) polarized virtual photon off a proton target

$$F_\lambda(x, Q^2) \equiv \frac{Q^2}{4\pi^2\alpha_{\text{em}}} \sigma^{\gamma_\lambda^* p}, \quad (2.3)$$

where  $\alpha_{\text{em}}$  is the fine-structure constant and  $\lambda = T$  or  $L$  denotes the polarization of the virtual photon.

In the small- $x$  regime, these cross sections can be calculated within the dipole picture [32], where the interaction factorizes into the splitting of the virtual photon into a quark–antiquark pair ( $q\bar{q}$ ), followed by the scattering of this color dipole off the proton target. The total cross section is given by an integral over the dipole’s transverse size  $\mathbf{r}$ , its impact parameter  $\mathbf{b}$ , and the quark’s longitudinal momentum fraction  $z$ , which reads

$$\sigma^{\gamma_\lambda^* p} = \frac{4\alpha_{\text{em}} N_c}{(2\pi)^2} \sum_f e_f^2 \int d^2\mathbf{r} d^2\mathbf{b} \int_0^1 dz \mathcal{K}_\lambda(\mathbf{r}, z) N(\mathbf{r}, \mathbf{b}, x), \quad (2.4)$$

where the sum runs over all active quark flavors  $f$ . The dynamics of the strong interaction are encoded in the nonperturbative dipole amplitude  $N(\mathbf{r}, \mathbf{b}, x)$ , which describes the probability for the dipole to scatter. The perturbative component is contained in the squared photon light-front wave functions,  $\mathcal{K}_\lambda$ , which are given by

$$\mathcal{K}_L = 4\bar{Q}^2 z(1-z) K_0(\bar{Q}|\mathbf{r}|)^2, \quad (2.5)$$

$$\mathcal{K}_T = \bar{Q}^2 [z^2 + (1-z)^2] K_1(\bar{Q}|\mathbf{r}|)^2, \quad (2.6)$$

and the argument  $\bar{Q}$  of the modified Bessel functions  $K_0$  and  $K_1$  is  $\bar{Q}^2 \equiv z(1-z)Q^2$ . Here, we have ignored quark mass corrections.

In Eq. (2.4), the full dipole–proton scattering amplitude  $N(\mathbf{r}, \mathbf{b}, x)$  depends on the transverse size of the dipole  $\mathbf{r}$ , Bjorken  $x$ , and the impact parameter  $\mathbf{b}$ . However, inclusive DIS data depends only on the dipole amplitude integrated over the impact parameter. We therefore simplify the calculation by integrating out the impact parameter as

$$\int d^2\mathbf{b} N(\mathbf{r}, \mathbf{b}, x) \equiv \frac{\sigma_0}{2} N(r, x), \quad (2.7)$$

and introducing a single normalization parameter,  $\sigma_0/2$ , which has the interpretation as the *proton transverse area* [11–13]. This parameter effectively sets the total strength of the interaction by replacing the impact parameter integral in the cross-section calculation. It is a free parameter in our model, constrained by the fit to HERA data. The integrated dipole amplitude  $N(r, x)$  now only depends on the dipole size  $r = |\mathbf{r}|$  and the Bjorken  $x$  variable.

## 2.2 Balitsky–Kovchegov evolution

The energy dependence of the dipole amplitude,  $N(r, x)$ , is governed by the BK evolution equation. This equation resums quantum corrections proportional to large logarithms of energy,  $\ln(1/x)$ , and incorporates the non-linear effects that lead to gluon saturation. For a dipole with quark and antiquark at transverse positions  $\mathbf{r}_0$  and  $\mathbf{r}_1$ , the equation reads

$$\frac{\partial N(|\mathbf{r}_{01}|, x)}{\partial \ln(1/x)} = \int d^2\mathbf{r}_2 K_{\text{BK}}(\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2) \times [N(|\mathbf{r}_{02}|, x) + N(|\mathbf{r}_{12}|, x) - N(|\mathbf{r}_{01}|, x) - N(|\mathbf{r}_{02}|, x)N(|\mathbf{r}_{12}|, x)], \quad (2.8)$$

where  $\mathbf{r}_{ij} \equiv \mathbf{r}_i - \mathbf{r}_j$  and we have ignored the impact-parameter dependence. The linear terms in the square brackets correspond to the BFKL equation, describing the emission of a gluon from the parent dipole. The final quadratic term,  $N(|\mathbf{r}_{02}|, x)N(|\mathbf{r}_{12}|, x)$ , represents the simultaneous scattering of the two new dipoles, a non-linear effect that tames the growth of the amplitude and drives the system towards saturation.

The interaction kernel,  $K_{\text{BK}}$ , describes the splitting of the parent dipole. For phenomenological accuracy, it is crucial to include running coupling corrections. Following the Balitsky prescription [33], the kernel is given by

$$K_{\text{BK}}(\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2) = \frac{N_c \alpha_s(\mathbf{r}_{01}^2)}{2\pi^2} \left[ \frac{\mathbf{r}_{01}^2}{\mathbf{r}_{02}^2 \mathbf{r}_{12}^2} + \frac{1}{\mathbf{r}_{02}^2} \left( \frac{\alpha_s(\mathbf{r}_{02}^2)}{\alpha_s(\mathbf{r}_{12}^2)} - 1 \right) + \frac{1}{\mathbf{r}_{12}^2} \left( \frac{\alpha_s(\mathbf{r}_{12}^2)}{\alpha_s(\mathbf{r}_{02}^2)} - 1 \right) \right]. \quad (2.9)$$

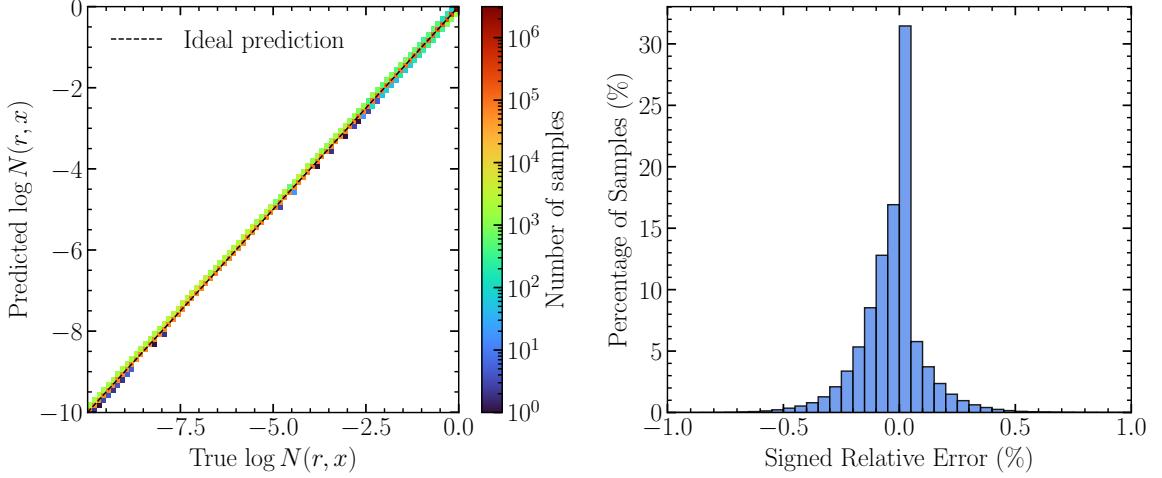
Here the position-space strong coupling constant,  $\alpha_s(r^2)$ , is taken at leading order with  $n_f = 3$  active quark flavors as [12]

$$\alpha_s(r^2) = \frac{12\pi}{(33 - 2n_f) \ln \left( \frac{4C^2}{r^2 \Lambda_{\text{QCD}}^2} \right)}. \quad (2.10)$$

Here,  $\Lambda_{\text{QCD}} = 0.241 \text{ GeV}$  and  $C^2$  is a parameter that sets the scale of the argument of the logarithm, which must be determined from fits to data [12, 17]. For large values of the dipole size  $r$ , we freeze  $\alpha_s$  to the value  $\alpha_{s,\text{max}} = 0.7$ .

## 3 Fitting procedure

The parameters of the initial dipole scattering amplitude are determined from deep-inelastic scattering data via a two-stage procedure. First, the BK evolution equation is solved for a large ensemble of initial conditions to generate a comprehensive library of theoretical predictions. Second, this library is used to train a neural network emulator that interpolates the solutions of the BK evolution across the parameter space. The resulting high-quality model allows for the rapid and efficient execution of a global fit to HERA data within the dipole picture.



**Figure 1:** Left: Two dimensional histogram comparing the exact and transformer emulated values of  $\ln N(r, x)$  for the BK dipole amplitude. The color scale indicates the density of samples, and the dashed line represents exact agreement. The mean and median relative errors on the validation set are 0.09% and 0.05%, respectively. Right: Distribution of the signed relative errors, zoomed to  $\pm 1\%$ . This highlights that the vast majority of predictions fall within a few per mille of the true values, demonstrating the emulator’s high accuracy.

### 3.1 Sampling and solving the BK equation

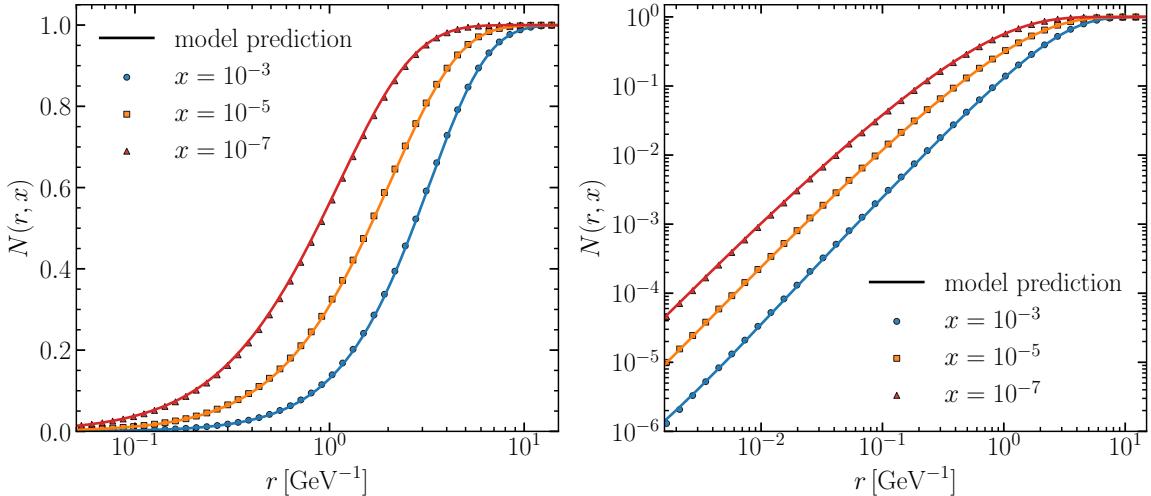
The BK equation (2.8) describes the evolution of the dipole amplitude  $N(r, x)$  with respect to the collision energy. The evolution is uniquely specified by an initial condition at  $x_0$ , and to assess the dependence on the initial rapidity scale we consider two different choices  $x_0 = 0.01$  and  $x_0 = 0.05$ . For the initial condition, we use a generalization of the MV model considered in Ref. [12]

$$N(r, x_0) = 1 - \exp \left[ - \left( \frac{r^2 Q_{s0}^2}{4} \right)^\gamma \ln \left( \frac{1}{\Lambda_{\text{QCD}} r} + e e_c \right) \right], \quad (3.1)$$

which is a function of the initial saturation scale  $Q_{s0}$ , the anomalous dimension  $\gamma$ , and an infrared regulator  $e_c$ . To efficiently map this parameter space, we use Latin hypercube sampling (LHS) to generate 10 000 parameter sets for  $(Q_{s0}, \gamma, e_c, C^2)$ . For each set, the running-coupling BK equation with the Balitsky kernel is solved using a numerical Julia implementation developed for this work. The solution is tabulated on a two-dimensional grid spanning transverse sizes  $10^{-6} \leq r/\text{GeV}^{-1} \leq 10^2$  and rapidities  $0 \leq \ln(x_0/x) \leq 16$ , which covers the relevant HERA kinematics [22]. This procedure yields a dataset of approximately  $4.7 \times 10^7$  values of  $N(r, x)$ , which serves as the training data for the neural network model.

### 3.2 Neural network emulation of the BK amplitude

Directly solving the BK equation for every trial set of parameters in a fit would be computationally prohibitive. To circumvent this we train a neural network to emulate the functional dependence of  $\ln N(r, x)$  on the variables  $(r, x)$  and the three parameters of the initial

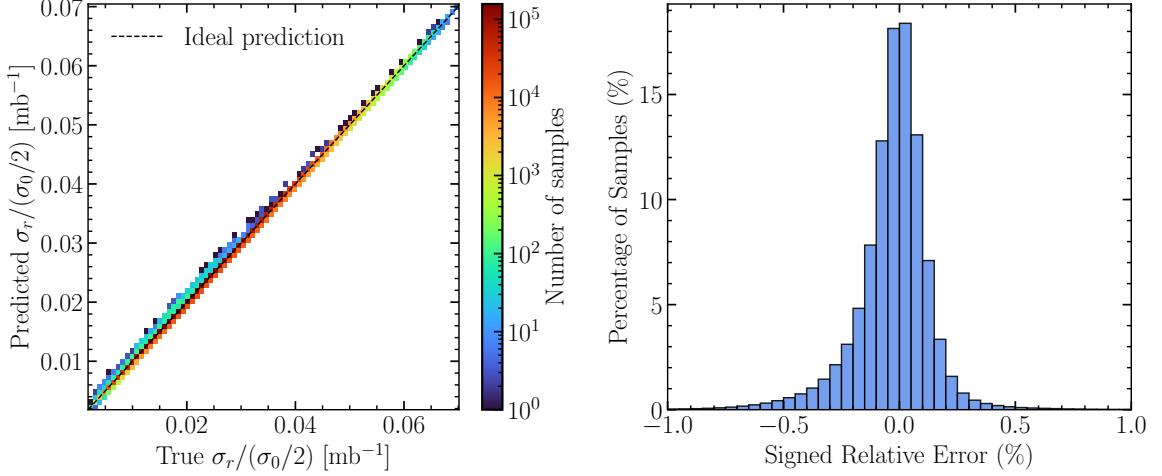


**Figure 2:** The emulated dipole amplitude  $N(r, x)$  (lines) is compared against the exact BK solution (points) for a representative out-of-training-sample parameter set at three different rapidities,  $x = 10^{-3}, 10^{-5}$ , and  $10^{-7}$ . The excellent agreement in both log-linear (left) and log-log (right) scales across the entire range of  $r$  demonstrates the model’s robust extrapolation capability. The chosen parameters are  $Q_{s0}^2 \simeq 0.07 \text{ GeV}^2$ ,  $\gamma = 1.01$ ,  $e_c = 24.68$ ,  $C^2 \simeq 4.65$ , and  $x_0 = 0.01$ , corresponding roughly to physical values found in previous fits [12, 17].

condition. Owing to its ability to capture correlations in an ordered sequence of input variables, we adopt a transformer architecture. The network takes as input the 6-dimensional vector  $(\log_{10} Q_{s0}, e_c, \gamma, \log_{10} C, \log_{10}(r/\text{GeV}^{-1}), x)$ . Each scalar is mapped by a shared linear projection to form a sequence of six tokens of dimension  $d_{\text{model}} = 128$ . A sinusoidal positional encoding is added to the sequence, which is then processed by a transformer encoder with four layers, eight attention heads, feed-forward width 512, dropout 0.1, and pre-normalization. The token representations are mean-pooled and passed to a two-layer multilayer perceptron (MLP) head [34] that outputs a raw logit function [35]. Applying a sigmoid to this logit yields the dipole amplitude  $N(r, x)$ , automatically enforcing the unitarity bound  $0 \leq N \leq 1$ . For diagnostics we also use  $\log N$  after the sigmoid.

For training we adopt a “log-MSE” loss: the network’s raw logit is passed through a sigmoid to yield  $N_{\text{pred}} \in (0, 1)$ , and we minimise the mean squared error between  $\log N_{\text{pred}}$  and  $\log N_{\text{true}}$ . This formulation with  $\log N$  simultaneously accommodates the dilute regime ( $N \ll 1$ ) and the saturation regime ( $N \approx 1$ ) without the need for a piece-wise or hybrid design.

Inputs are standardized using the STANDARDSCALER implementation from the Scikit-learn library [36]. The model is trained using the AdamW optimizer [37] with learning rate  $3 \times 10^{-4}$ , weight decay  $10^{-4}$ , cosine annealing schedule to  $3 \times 10^{-6}$ , batch size 4096, and early stopping (patience 20). Automatic mixed-precision training [38] and gradient clipping (max-norm 1.0) are applied. 90% of the generated data is used for training and the remainder for validation.



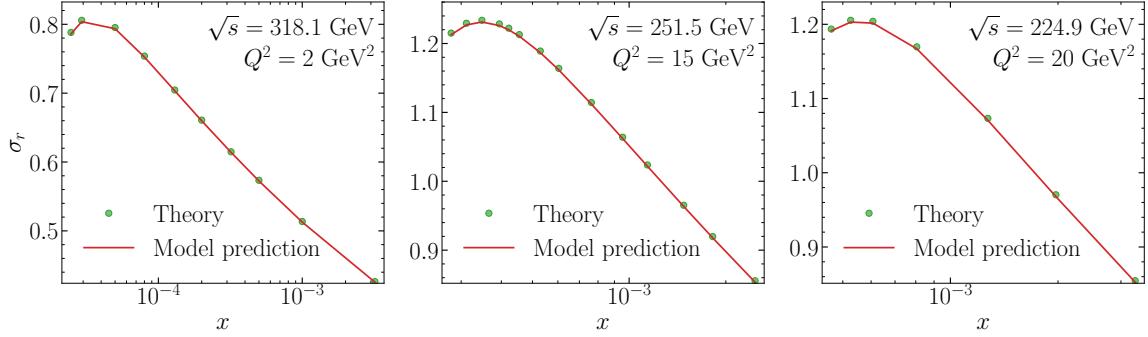
**Figure 3:** Left: Parity plot comparing the transformer emulator predictions for the DIS reduced cross section  $\sigma_r/(\sigma_0/2)$  (y-axis) against the exact dipole model values (x-axis). The dashed line indicates perfect agreement, with mean and median relative errors of 0.115% and 0.073%, respectively. Right: Distribution of the signed relative errors, shown within  $\pm 1\%$ , demonstrating that the overwhelming majority of predictions lie well below the per mille level. Together these demonstrate the high accuracy of the DIS surrogate model.

The trained emulator exhibits excellent performance, achieving mean and median relative errors of 0.09% and 0.05%, respectively, on the validation set. This high accuracy is visually demonstrated in Fig. 1. The left panel shows a strong one-to-one correlation between the emulated and exact values of  $\ln N(r, x)$  across many orders of magnitude. The distribution of relative errors, shown in the right panel, is sharply peaked at zero, confirming that the vast majority of predictions are accurate to within a few per mille.

Beyond its high accuracy on the validation set, it is important to demonstrate the emulator’s performance on representative test cases. In Fig. 2, we show a direct comparison of the emulated dipole amplitude with the exact BK solution for a parameter set not used in training. The excellent agreement across a wide range of  $r$  confirms that the emulator reproduces the BK evolution with high fidelity, providing a reliable surrogate for phenomenological applications.

#### 4 Application to DIS observables

With a robust emulator for the BK-evolved dipole amplitude established, we now proceed to its primary application: the calculation of DIS observables for a global analysis of HERA data. This requires convolving the emulated amplitude with the virtual photon wave functions to compute the reduced cross section,  $\sigma_r$ . While these integrals must be evaluated once to generate the training data for the emulator, repeating them for each of the 10 000 parameter sets across a large number of kinematic points during a global fit would be computationally prohibitive. The surrogate model circumvents this bottleneck by replacing the repeated numerical evaluations with fast predictions, thereby enabling



**Figure 4:** Comparison of the model prediction with the theory cross sections for  $e^+p$  DIS at three center-of-mass energies,  $\sqrt{s} = 318.1$ ,  $251.5$ , and  $224.9$  GeV. The green circles denote the theory values, while the red curves show the model prediction evaluated with a representative parameter set outside the training replicas. Each panel is drawn at fixed  $Q^2$  ( $2$ ,  $15$ ,  $20$   $\text{GeV}^2$  respectively).

efficient exploration of the parameter space. To overcome this final bottleneck, we construct a second, dedicated transformer based emulator. This network is trained to map directly from the six inputs ( $\log_{10} Q_{s0}$ ,  $\gamma$ ,  $e_c$ ,  $C$ ,  $\log_{10} x$ ,  $\log_{10} Q^2$ ) to the normalized reduced cross section  $\sigma_r / (\sigma_0/2)$ . Its architecture follows closely that used for  $N(r, x)$ , and it is trained on the  $\sim 4 \times 10^6$  values computed from the dipole amplitudes.

For the DIS cross section emulator, we employ a composite loss function that combines mean squared error (MSE) with a Smooth- $L_1$  term [39],

$$\mathcal{L}_{\text{DIS}} = 0.8 \text{ MSE}\left(\frac{\sigma_r^{\text{pred}}}{\sigma_0/2}, \frac{\sigma_r^{\text{true}}}{\sigma_0/2}\right) + 0.2 \text{ SmoothL1}\left(\frac{\sigma_r^{\text{pred}}}{\sigma_0/2}, \frac{\sigma_r^{\text{true}}}{\sigma_0/2}\right), \quad (4.1)$$

with the Smooth- $L_1$  evaluated at  $\beta = 0.1$ . While the MSE provides the primary measure of accuracy, we found in practice that adding a small Smooth- $L_1$  component helps stabilize training across the wide dynamic range of  $\sigma_r / (\sigma_0/2)$  by reducing the impact of occasional large residuals. The resulting surrogate model for the normalized DIS reduced cross section,  $\sigma_r / (\sigma_0/2)$ , achieves very high precision, with a mean relative error of  $0.115\%$  and a median error of  $0.073\%$ . Figure 3 illustrates this accuracy: the left panel shows a parity plot with near-perfect alignment along the diagonal, while the right panel presents the distribution of signed relative errors within  $\pm 1\%$ , with the bulk of predictions clustered well below the per-mille level.

As an intermediate validation step prior to the global fit, Fig. 4 compares the model prediction with the theory cross sections for  $e^+p$  DIS at three center-of-mass energies ( $\sqrt{s} = 318.1$ ,  $251.5$ ,  $224.9$  GeV). In each panel the green circles denote the theory values, while the thin red curve shows the prediction evaluated with a representative parameter set outside the training sample. Each panel is drawn at fixed  $Q^2$  values ( $2$ ,  $15$ ,  $20$   $\text{GeV}^2$ , respectively), with a logarithmic  $x$  axis and  $\sigma_r$  on the vertical axis. The excellent agreement demonstrates that the numerical accuracy of the method is well under control, providing a reliable basis for the subsequent phenomenological fits.

Parameter	Prior range	Fit for $x_0 = 0.01$		Fit for $x_0 = 0.05$	
		4-param	5-param	4-param	5-param
$Q_{s,0}^2 [\text{GeV}^2]$	[0.01, 0.11]	$0.063^{+0.001}_{-0.004}$	$0.068^{+0.024}_{-0.015}$	$0.025^{+0.0035}_{-0.0025}$	$0.045^{+0.0021}_{-0.0017}$
$e_c$	[0.5, 70.0]	$29.0^{+8.7}_{-3.1}$	$18.7^{+20.1}_{-9.6}$	$34.0^{+35.3}_{-3.2}$	$41.0^{+9.2}_{-2.6}$
$C^2$	[2.0, 20.0]	$4.45^{+0.99}_{-0.59}$	$4.82^{+1.56}_{-2.63}$	$17.4^{+2.5}_{-6.1}$	$10.2^{+1.6}_{-0.8}$
$\sigma_0/2 [\text{mb}]$	[12.0, 20.0]	$14.5^{+0.6}_{-0.4}$	$14.8^{+0.9}_{-2.4}$	$19.4^{+0.5}_{-2.0}$	$16.6^{+0.2}_{-0.4}$
$\gamma$	[0.9, 1.3]	1.00 (fixed)	$1.006^{+0.037}_{-0.019}$	1.00 (fixed)	$1.138^{+0.033}_{-0.012}$
$\chi^2/\text{dof}$	—	0.854	0.857	1.471	1.195

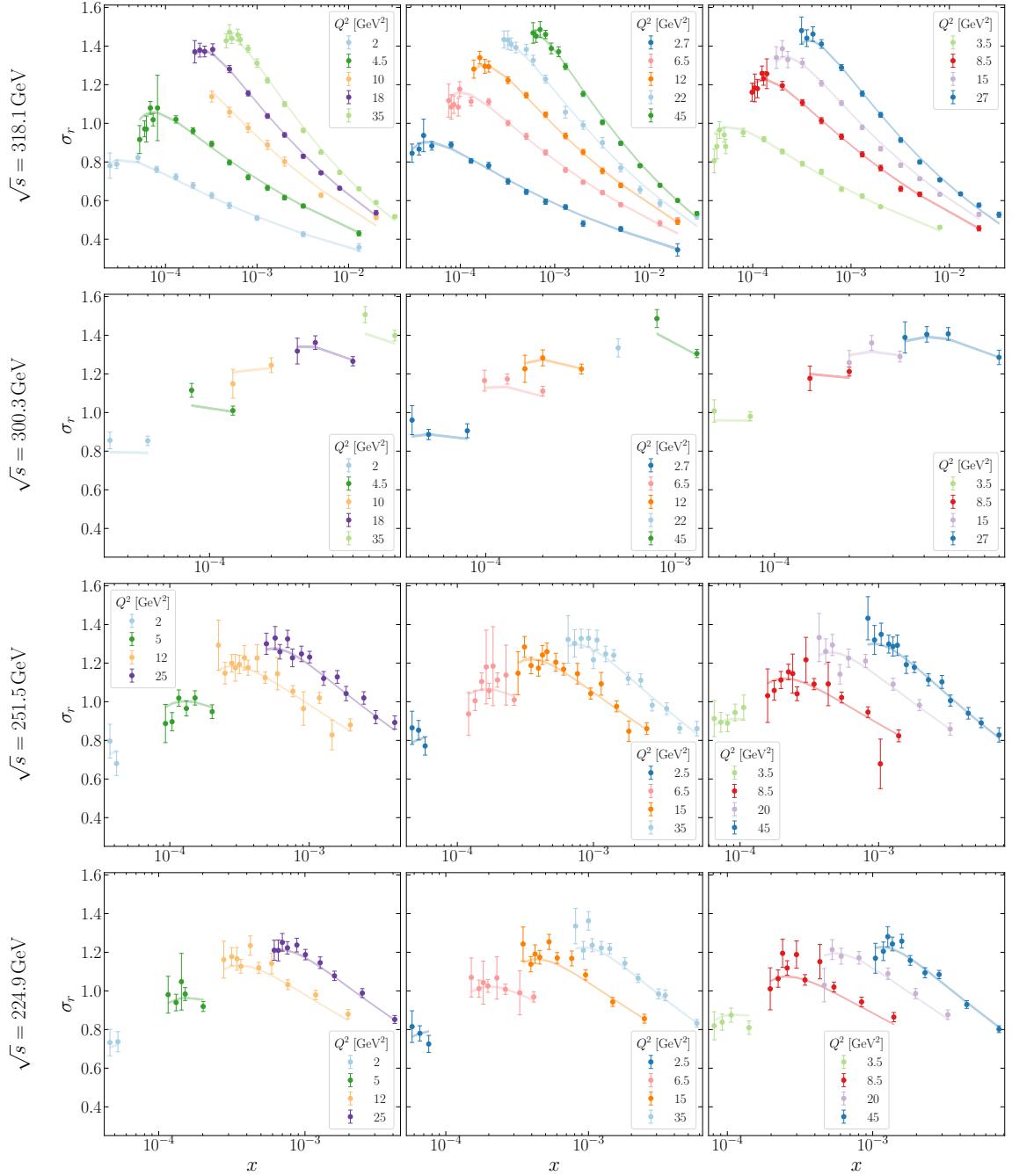
**Table 1:** Combined fit results for the 4-parameter ( $\gamma$  fixed at 1.0) and 5-parameter ( $\gamma$  free) models at two starting points  $x_0 = 0.01$  and  $x_0 = 0.05$ . Prior ranges reflect the updated bounds in the analysis. Entries show the central value with the 95% credible bounds.

This high-speed surrogate enables a full global fit to the combined HERA neutral current  $e^+p$  data [22]. The negligible evaluation cost of the emulator allows for the efficient scanning of millions of points in the parameter space. The best-fit parameters  $\boldsymbol{\theta} = (\sigma_0/2, Q_{s,0}^2, e_c, \gamma, C^2)$  are determined by minimizing the chi-squared function

$$\chi^2 = \sum_{i=1}^{N_{\text{data}}} \left( \frac{\sigma_{r,i}^{\text{data}} - \sigma_{r,i}^{\text{th}}(\boldsymbol{\theta})}{\delta\sigma_{r,i}^{\text{data}}} \right)^2, \quad (4.2)$$

where  $\sigma_{r,i}^{\text{th}}(\boldsymbol{\theta})$  is the model prediction and  $\delta\sigma_{r,i}^{\text{data}}$  denotes the total experimental uncertainty for point  $i$ . Statistical uncertainties are estimated using the replica method [40], in which independent fits are performed on replica data sets generated by adding Gaussian noise to the experimental measurements. This approach has been widely used in the extraction of transverse-momentum-dependent parton distribution functions [41–43]. In this work we use 200 replicas to obtain the uncertainty estimates. The minimization is performed using the `Minuit` algorithm from the `iminuit` package [44].

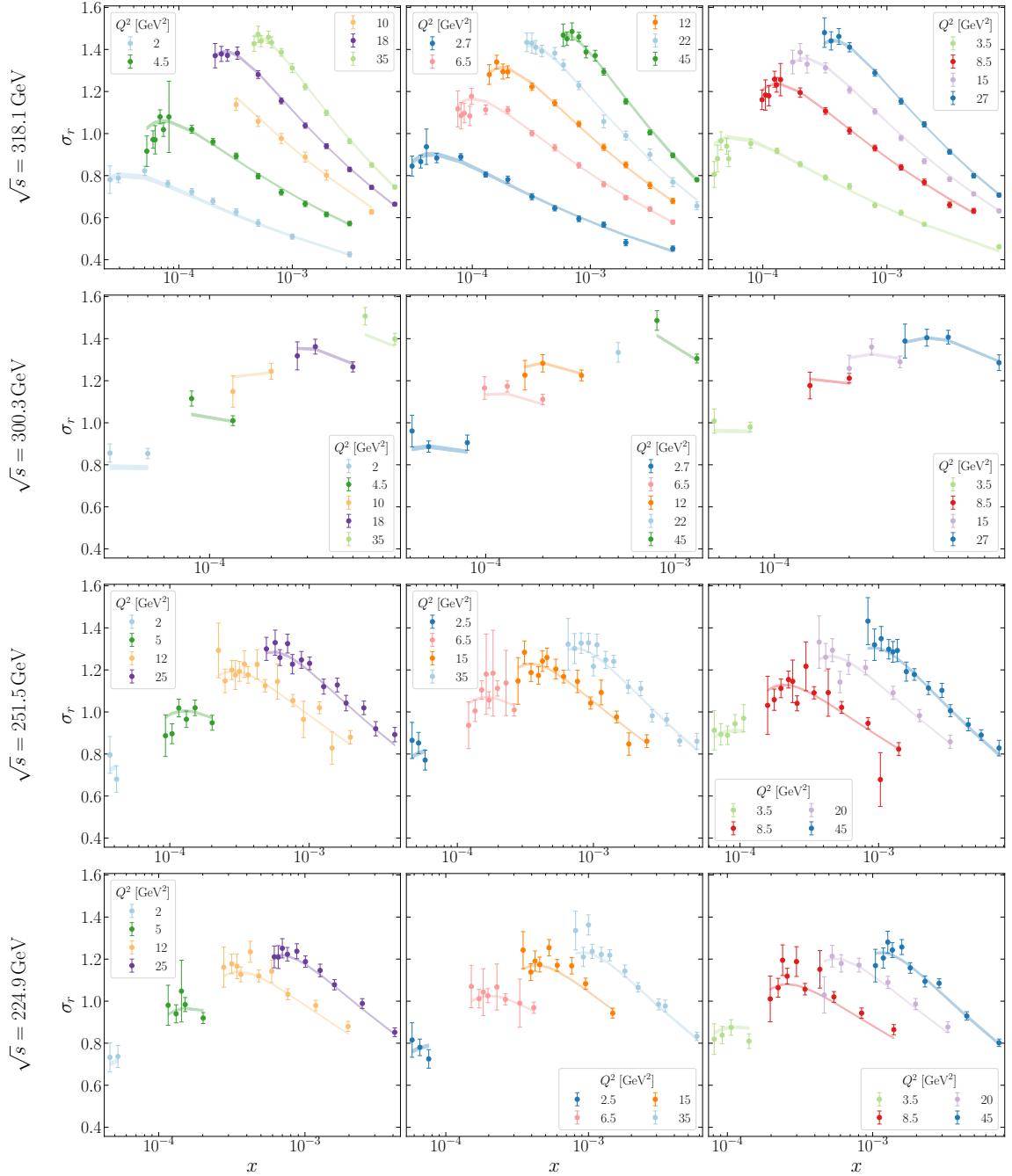
In our analysis we perform four distinct fits, corresponding to the two choices of the starting point  $x_0 = 0.01$  and  $x_0 = 0.05$ , and to configurations where the anomalous dimension  $\gamma$  is either fixed to 1.0 or treated as a free parameter with  $\gamma = 1$  corresponding to the standard MV model. Our motivation for considering  $x_0 = 0.05$ , in addition to the more standard choice  $x_0 = 0.01$ , is to assess the validity of the small- $x$  framework for more moderate values of  $x$ . The best fit parameters and corresponding  $\chi^2$  per degrees of freedom (dof) for all four fit configurations are summarized in Table 1. The overall fit quality is comparable to the recent Bayesian analysis [17]. For our fits with  $x_0 = 0.01$  we obtain  $\chi^2/\text{dof}$  values close to unity, in line with the results of Ref. [17]. When the starting point is shifted to  $x_0 = 0.05$ , the fit quality deteriorates somewhat, as reflected in the larger  $\chi^2/\text{dof}$  values. Comparing the parameters, we observe that the fitted anomalous dimension  $\gamma$  values remain close to 1.0, though the  $x_0 = 0.05$  fit prefers a slightly larger value. Differences also appear in the extracted  $\sigma_0/2$  and  $C^2$  parameters, reflecting the sensitivity



**Figure 5:** Reduced cross section  $\sigma_r(x, Q^2)$  compared with the combined HERA  $e^+p$  data at four center-of-mass energies,  $\sqrt{s} = 318.1, 300.3, 251.5$ , and  $224.9$  GeV. The curves show the result of the five-parameter fit (with  $\gamma$  free) performed with the evolution starting point  $x_0 = 0.05$ . The shaded bands represent the  $2\sigma$  uncertainty of the fitted prediction.

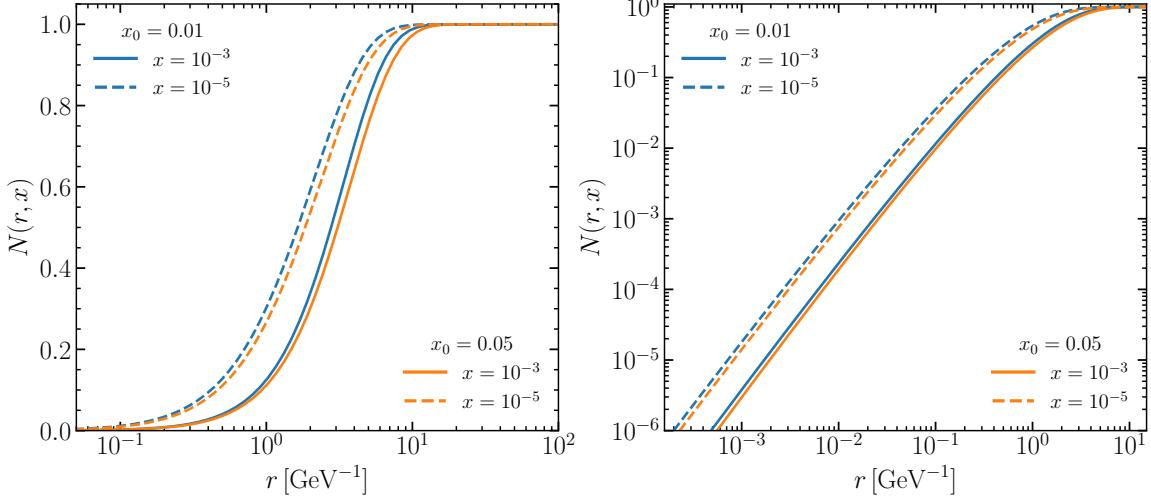
of the fits to the chosen starting point of the BK evolution. The smaller  $Q_{s,0}^2$  for  $x_0 = 0.05$  is expected, as the saturation scale increases with higher energy (i.e. lower  $x$ ).

Figures 5 and 6 present the phenomenological results of this analysis. They provide a



**Figure 6:** Same as Fig. 5, but with the evolution starting point  $x_0 = 0.01$ . The four panels correspond to  $\sqrt{s} = 318.1$ ,  $300.3$ ,  $251.5$ , and  $224.9$  GeV. The shaded bands again represent the  $2\sigma$  uncertainty of the fitted prediction.

direct comparison between our best-fit model and the combined HERA  $e^+p$  data for the reduced cross section  $\sigma_r$ , shown simultaneously across the four experimental center-of-mass energies  $\sqrt{s} = 318.1$ ,  $300.3$ ,  $251.5$ , and  $224.9$  GeV. The solid curves represent the results of our 5-parameter fit (with  $\gamma$  free), using the evolution starting point  $x_0 = 0.05$  (Fig. 5)



**Figure 7:** Comparison of the dipole amplitudes  $N(r, x)$  from the  $x_0 = 0.01$  (solid lines) and  $x_0 = 0.05$  (dashed lines) from the 5-parameter fits, evaluated at the same values of  $x$ . The left panel shows the results on a linear scale, while the right panel uses a logarithmic scale to emphasize the small- $r$  region. The two sets of curves become nearly identical after the appropriate rapidity shift, with only small residual differences visible around  $r \sim 1 \text{ GeV}^{-1}$  and at very small  $r$ . These remaining discrepancies could be further reduced once the normalization factor  $\sigma_0/2$  is included, but here we focus solely on the dipole amplitude itself.

and 0.01 (Fig. 6). The shaded bands indicate the  $2\sigma$  uncertainty envelope of the fitted prediction, demonstrating an excellent description of the data across the full kinematic range.

As a further cross-check, we also examine how the dipole amplitude depends on the choice of the initial rapidity scale. In Fig. 7 we compare the fitted dipole amplitudes  $N(r, x)$  obtained from the  $x_0 = 0.01$  and  $x_0 = 0.05$  fits, evaluated at the same values of  $x$ . Specifically, we compare the results at  $x = 10^{-3}$  and  $x = 10^{-5}$  for both initial conditions  $x_0 = 0.01$  and  $x_0 = 0.05$ . The left panel (linear scale) shows that the two curves nearly coincide, with only a small residual difference around the transition region  $r \sim 1 \text{ GeV}^{-1}$ . The right panel (log-log scale) demonstrates that in the small- $r$  region the two amplitudes are essentially identical, with deviations well below the percent level. This indicates that once the rapidity shift is properly taken into account, the BK evolution effectively eliminates the dependence on the arbitrary choice of  $x_0$ , leaving only minor residual differences around the transition region. The remaining small discrepancy is compensated by a larger value of  $\sigma_0/2$  for the  $x_0 = 0.05$  fit, such that the produced cross sections are very close to each other when away from the saturation region  $N \approx 1$ .

## 5 Summary

We have presented a new machine learning framework that, by employing transformer-based emulators, provides high-fidelity representations of both the BK-evolved dipole amplitude and the resulting DIS cross sections. This methodology dramatically reduces the computational cost of small- $x$  phenomenology, enabling parameter space exploration many orders of magnitude faster than direct numerical evolution. Our application of this framework to HERA data confirms that the BK equation, coupled with a three-parameter initial condition, provides an excellent description of inclusive DIS cross sections over a wide kinematic range.

The framework is readily extensible to several crucial areas. Future investigations will include the incorporation of the impact parameter [45–51], the implementation of next-to-leading order corrections to both the BK kernel [52, 53] and the photon wave functions [54–62], and the application to diffractive DIS measurements both at LO [63–66] and NLO [67–69]. A key advantage of our approach is the decoupling of the computationally intensive evolution from the fitting procedure, which permits the straightforward addition of new parameters or the coupling to other physical processes. This work thus provides a robust and flexible tool, paving the way for the next generation of precision global analyses in small- $x$  physics.

## Acknowledgment

We thank Bjoern Schenke for helpful discussions and Noah Moran for collaboration during the early stage of this project. Z.K. and J.P. are supported National Science Foundation under grant No. PHY-2515057. D.Y.S. is supported by the National Science Foundations of China under Grant No. 12275052, No. 12147101. This work is also supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, within the framework of the Saturated Glue (SURGE) Topical Theory Collaboration.

## References

- [1] A. Accardi et al., *Electron Ion Collider: The Next QCD Frontier: Understanding the glue that binds us all*, *Eur. Phys. J. A* **52** (2016), no. 9 268, [[arXiv:1212.1701](#)].
- [2] 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](#)].
- [3] D. P. Anderle et al., *Electron-ion collider in China*, *Front. Phys. (Beijing)* **16** (2021), no. 6 64701, [[arXiv:2102.09222](#)].
- [4] E. Iancu and R. Venugopalan, *The Color glass condensate and high-energy scattering in QCD*, pp. 249–3363. 3, 2003. [hep-ph/0303204](#).
- [5] F. Gelis, E. Iancu, J. Jalilian-Marian, and R. Venugopalan, *The Color Glass Condensate*, *Ann. Rev. Nucl. Part. Sci.* **60** (2010) 463–489, [[arXiv:1002.0333](#)].
- [6] I. Balitsky, *Operator expansion for high-energy scattering*, *Nucl. Phys. B* **463** (1996) 99–160, [[hep-ph/9509348](#)].

- [7] Y. V. Kovchegov, *Small  $x$   $F(2)$  structure function of a nucleus including multiple pomeron exchanges*, *Phys. Rev. D* **60** (1999) 034008, [[hep-ph/9901281](#)].
- [8] L. D. McLerran and R. Venugopalan, *Gluon distribution functions for very large nuclei at small transverse momentum*, *Phys. Rev. D* **49** (1994) 3352–3355, [[hep-ph/9311205](#)].
- [9] L. D. McLerran and R. Venugopalan, *Computing quark and gluon distribution functions for very large nuclei*, *Phys. Rev. D* **49** (1994) 2233–2241, [[hep-ph/9309289](#)].
- [10] L. D. McLerran and R. Venugopalan, *Green's functions in the color field of a large nucleus*, *Phys. Rev. D* **50** (1994) 2225–2233, [[hep-ph/9402335](#)].
- [11] 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** (2009) 034031, [[arXiv:0902.1112](#)].
- [12] T. Lappi and H. Mäntysaari, *Single inclusive particle production at high energy from HERA data to proton-nucleus collisions*, *Phys. Rev. D* **88** (2013) 114020, [[arXiv:1309.6963](#)].
- [13] 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** (2011) 1705, [[arXiv:1012.4408](#)].
- [14] B. Ducloué, E. Iancu, G. Soyez, and D. N. Triantafyllopoulos, *HERA data and collinearly-improved BK dynamics*, *Phys. Lett. B* **803** (2020) 135305, [[arXiv:1912.09196](#)].
- [15] G. Beuf, H. Hänninen, T. Lappi, and H. Mäntysaari, *Color Glass Condensate at next-to-leading order meets HERA data*, *Phys. Rev. D* **102** (2020) 074028, [[arXiv:2007.01645](#)].
- [16] A. Dumitru, H. Mäntysaari, and R. Paatelainen, *High-energy dipole scattering amplitude from evolution of low-energy proton light-cone wave functions*, *Phys. Rev. D* **107** (2023), no. 11 114024, [[arXiv:2303.16339](#)].
- [17] C. Casuga, M. Karhunen, and H. Mäntysaari, *Inferring the initial condition for the Balitsky-Kovchegov equation*, *Phys. Rev. D* **109** (2024), no. 5 054018, [[arXiv:2311.10491](#)].
- [18] H. Hänninen, A. Kykkänen, and H. Schlüter, *Reconstruction of the Dipole Amplitude in the Dipole Picture as a mathematical Inverse Problem*, [arXiv:2509.05005](#).
- [19] 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** (2023), no. 19 192301, [[arXiv:2211.03504](#)].
- [20] C. Casuga, H. Hänninen, and H. Mäntysaari, *Initial condition for the Balitsky-Kovchegov equation at next-to-leading order*, *Phys. Rev. D* **112** (2025), no. 3 034003, [[arXiv:2506.00487](#)].
- [21] **H1, ZEUS** Collaboration, F. D. Aaron et al., *Combined Measurement and QCD Analysis of the Inclusive  $e^- p$  Scattering Cross Sections at HERA*, *JHEP* **01** (2010) 109, [[arXiv:0911.0884](#)].
- [22] **H1, ZEUS** Collaboration, H. Abramowicz et al., *Combination of measurements of inclusive deep inelastic  $e^\pm p$  scattering cross sections and QCD analysis of HERA data*, *Eur. Phys. J. C* **75** (2015), no. 12 580, [[arXiv:1506.06042](#)].
- [23] L. N. Lipatov, *Reggeization of the Vector Meson and the Vacuum Singularity in Nonabelian Gauge Theories*, *Sov. J. Nucl. Phys.* **23** (1976) 338–345.

- [24] V. S. Fadin, E. A. Kuraev, and L. N. Lipatov, *On the Pomeranchuk Singularity in Asymptotically Free Theories*, *Phys. Lett. B* **60** (1975) 50–52.
- [25] E. A. Kuraev, L. N. Lipatov, and V. S. Fadin, *Multiregge processes in the Yang-Mills theory*, *Sov. Phys. JETP* **44** (1976), no. 3 443–451.
- [26] E. A. Kuraev, L. N. Lipatov, and V. S. Fadin, *The Pomeranchuk Singularity in Nonabelian Gauge Theories*, *Sov. Phys. JETP* **45** (1977) 199–204.
- [27] I. I. Balitsky and L. N. Lipatov, *The Pomeranchuk Singularity in Quantum Chromodynamics*, *Sov. J. Nucl. Phys.* **28** (1978) 822–829.
- [28] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and  $e+e-$  Annihilation by Perturbation Theory in Quantum Chromodynamics.*, *Sov. Phys. JETP* **46** (1977) 641–653.
- [29] V. N. Gribov and L. N. Lipatov, *Deep inelastic  $e p$  scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 438–450.
- [30] V. N. Gribov and L. N. Lipatov,  *$e+e-$  pair annihilation and deep inelastic  $e p$  scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 675–684.
- [31] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, *Nucl. Phys. B* **126** (1977) 298–318.
- [32] Y. V. Kovchegov and E. Levin, *Quantum Chromodynamics at High Energy*, vol. 33. Oxford University Press, 2013.
- [33] I. Balitsky, *Quark contribution to the small- $x$  evolution of color dipole*, *Phys. Rev. D* **75** (2007) 014001, [[hep-ph/0609105](#)].
- [34] M. Gardner and S. Dorling, *Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences*, *Atmospheric Environment* **32** (1998), no. 14 2627–2636.
- [35] H. Wei, R. Xie, H. Cheng, L. Feng, B. An, and Y. Li, *Mitigating neural network overconfidence with logit normalization*, in *Proceedings of the 39th International Conference on Machine Learning* (K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, eds.), vol. 162 of *Proceedings of Machine Learning Research*, pp. 23631–23644, PMLR, 17–23 Jul, 2022.
- [36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., *Scikit-learn: Machine learning in python*, *the Journal of machine Learning research* **12** (2011) 2825–2830.
- [37] I. Loshchilov and F. Hutter, *Decoupled weight decay regularization*, in *International Conference on Learning Representations*, 2017.
- [38] P. Micikevicius, S. Narang, J. Alben, G. F. Diamos, E. Elsen, D. García, B. Ginsburg, M. Houston, O. Kuchaiev, G. Venkatesh, and H. Wu, *Mixed precision training*, *ArXiv abs/1710.03740* (2017).
- [39] J. Terven, D.-M. Cordova-Esparza, J.-A. Romero-González, A. Ramírez-Pedraza, and E. A. Chávez-Urbiola, *A comprehensive survey of loss functions and metrics in deep learning*, *Artificial Intelligence Review* **58** (2025), no. 7 195.
- [40] **NNPDF** Collaboration, R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, A. Piccione, J. Rojo, and M. Ubiali, *A Determination of parton distributions with faithful*

*uncertainty estimation*, *Nucl. Phys. B* **809** (2009) 1–63, [[arXiv:0808.1231](#)]. [Erratum: *Nucl.Phys.B* 816, 293 (2009)].

- [41] V. Moos, I. Scimemi, A. Vladimirov, and P. Zurita, *Determination of unpolarized TMD distributions from the fit of Drell-Yan and SIDIS data at  $N^4LL$* , [\[arXiv:2503.11201\]](#).
- [42] **MAP (Multi-dimensional Analyses of Partonic distributions)** Collaboration, A. Bacchetta, V. Bertone, C. Bissolotti, G. Bozzi, M. Cerutti, F. Delcarro, M. Radici, L. Rossi, and A. Signori, *Flavor dependence of unpolarized quark transverse momentum distributions from a global fit*, *JHEP* **08** (2024) 232, [\[arXiv:2405.13833\]](#).
- [43] M. G. Echevarria, Z.-B. Kang, and J. Terry, *Global analysis of the Sivers functions at NLO+NNLL in QCD*, *JHEP* **01** (2021) 126, [\[arXiv:2009.10710\]](#).
- [44] H. Dembinski and P. O. et al., *scikit-hep/iminuit*, .
- [45] 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** (2025), no. 5 054033, [\[arXiv:2411.13533\]](#).
- [46] J. Cepila, J. G. Contreras, and M. Matas, *Collinearly improved kernel suppresses Coulomb tails in the impact-parameter dependent Balitsky-Kovchegov evolution*, *Phys. Rev. D* **99** (2019), no. 5 051502, [\[arXiv:1812.02548\]](#).
- [47] J. Cepila, J. G. Contreras, and M. Matas, *Predictions for nuclear structure functions from the impact-parameter dependent Balitsky-Kovchegov equation*, *Phys. Rev. C* **102** (2020), no. 4 044318, [\[arXiv:2002.11056\]](#).
- [48] J. Cepila, J. G. Contreras, and M. Vaculciak, *Solutions to the Balitsky-Kovchegov equation including the dipole orientation*, *Phys. Lett. B* **848** (2024) 138360, [\[arXiv:2309.02910\]](#).
- [49] 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** (2025), no. 9 096015, [\[arXiv:2412.08571\]](#).
- [50] J. Cepila, M. Matas, and M. Vaculciak, *Probing nuclear structure with the Balitsky-Kovchegov equation in full impact-parameter dependence*, [\[arXiv:2509.02115\]](#).
- [51] 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** (2019), no. 5 054015, [\[arXiv:1907.12123\]](#).
- [52] I. Balitsky and G. A. Chirilli, *Next-to-leading order evolution of color dipoles*, *Phys. Rev. D* **77** (2008) 014019, [\[arXiv:0710.4330\]](#).
- [53] I. Balitsky and G. A. Chirilli, *NLO evolution of color dipoles in  $N=4$  SYM*, *Nucl. Phys. B* **822** (2009) 45–87, [\[arXiv:0903.5326\]](#).
- [54] I. Balitsky and G. A. Chirilli, *Photon impact factor in the next-to-leading order*, *Phys. Rev. D* **83** (2011) 031502, [\[arXiv:1009.4729\]](#).
- [55] 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** (2013), no. 1 014013, [\[arXiv:1207.3844\]](#).
- [56] G. Beuf, *NLO corrections for the dipole factorization of DIS structure functions at low  $x$* , *Phys. Rev. D* **85** (2012) 034039, [\[arXiv:1112.4501\]](#).
- [57] 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** (2016), no. 5 054016, [\[arXiv:1606.00777\]](#).

- [58] G. Beuf, *Dipole factorization for DIS at NLO: Combining the  $q\bar{q}$  and  $q\bar{q}g$  contributions*, *Phys. Rev. D* **96** (2017), no. 7 074033, [[arXiv:1708.06557](#)].
- [59] H. Hänninen, T. Lappi, and R. Paatelainen, *One-loop corrections to light cone wave functions: the dipole picture DIS cross section*, *Annals Phys.* **393** (2018) 358–412, [[arXiv:1711.08207](#)].
- [60] G. Beuf, T. Lappi, and R. Paatelainen, *Massive quarks in NLO dipole factorization for DIS: Longitudinal photon*, *Phys. Rev. D* **104** (2021), no. 5 056032, [[arXiv:2103.14549](#)].
- [61] G. Beuf, T. Lappi, and R. Paatelainen, *Massive Quarks at One Loop in the Dipole Picture of Deep Inelastic Scattering*, *Phys. Rev. Lett.* **129** (2022), no. 7 072001, [[arXiv:2112.03158](#)].
- [62] G. Beuf, T. Lappi, and R. Paatelainen, *Massive quarks in NLO dipole factorization for DIS: Transverse photon*, *Phys. Rev. D* **106** (2022), no. 3 034013, [[arXiv:2204.02486](#)].
- [63] H. Kowalski, T. Lappi, C. Marquet, and R. Venugopalan, *Nuclear enhancement and suppression of diffractive structure functions at high energies*, *Phys. Rev. C* **78** (2008) 045201, [[arXiv:0805.4071](#)].
- [64] S. Munier and A. Shoshi, *Diffractive photon dissociation in the saturation regime from the Good and Walker picture*, *Phys. Rev. D* **69** (2004) 074022, [[hep-ph/0312022](#)].
- [65] C. Marquet, *A Unified description of diffractive deep inelastic scattering with saturation*, *Phys. Rev. D* **76** (2007) 094017, [[arXiv:0706.2682](#)].
- [66] T. Lappi, A. D. Le, and H. Mäntysaari, *Rapidity gap distribution of diffractive small-xp events at HERA and at the EIC*, *Phys. Rev. D* **108** (2023), no. 11 114023, [[arXiv:2307.16486](#)].
- [67] G. Beuf, H. Hänninen, T. Lappi, Y. Mulian, and H. Mäntysaari, *Diffractive deep inelastic scattering at NLO in the dipole picture: The  $q\bar{q}^-g$  contribution*, *Phys. Rev. D* **106** (2022), no. 9 094014, [[arXiv:2206.13161](#)].
- [68] G. Beuf, T. Lappi, H. Mäntysaari, R. Paatelainen, and J. Penttala, *Diffractive deep inelastic scattering at NLO in the dipole picture*, *JHEP* **05** (2024) 024, [[arXiv:2401.17251](#)].
- [69] A. Kaushik, H. Mäntysaari, and J. Penttala, *Diffractive deep inelastic scattering in the dipole picture: the  $q\bar{q}g$  contribution in exact kinematics*, [arXiv:2510.24171](#).