Bayesian approach to determine the initial condition for the Balitsky-Kovchegov equation

Carlisle Casuga,[a,b,*] **Mikko Karhunen**[a,b] **and Heikki Mäntysaari**[a,b]

[a]*University of Jyväskylä, P.O. Box 35, 40014 University of Jyväskylä, Finland*

[b]*Helsinki Institute of Physics, P.O. Box 64, 00014 University of Helsinki, Finland*

E-mail: carlisle.doc.casuga@jyu.fi, mikko.a.karhunen@student.jyu.fi,
heikki.mantysaari@jyu.fi

We present posterior distributions of parameters that characterize the nonperturbative initial input for the Balitsky-Kovchegov evolution equation. The BK equation evolves an initial dipole-target scattering amplitude at moderate x_{Bj} values toward smaller values. We use Bayesian inference to constrain the model parameters against the precise combined reduced cross section data from HERA, and obtain probability distributions for the free parameters. Our results show the data's preference for the anomalous dimension to be $\gamma \sim 1$. The distributions provide a rigorous method for propagating uncertainties of the BK initial condition for other CGC calculations. We demonstrate this for inclusive quark production in proton-proton collisions and nuclear modification factor for the total deep inelastic scattering cross section to be measured at the EIC.

---

*Speaker

1. Introduction

Results from the future Electron-Ion Collider (EIC) [1] must be met with high precision theoretical predictions in order to probe saturation and proton structure at small-x. The color glass condensate (CGC) effective field theory [2] provides a convenient framework to describe scattering in the high energy regime. Along with the next-to-leading corrections for the CGC cross sections, global analysis of the nonperturbative input to evolution equations like the Balitsky-Kovchegov (BK) equation [3, 4] is necessary to move towards higher theoretical precision. The BK initial condition has been fitted to HERA data before in Refs. [5–8] at the leading order and Refs. [9, 10] at the next-to-leading order. This study [11] also aims to constrain the BK initial condition. Contrary to previous fits, our analysis provides uncertainties for the parameter estimates for the first time. This input is necessary for statistically rigorous propagation of uncertainties when calculating observables in CGC.

In the dipole picture of deep inelastic scattering (DIS), the virtual photon fluctuates into a quark-antiquark pair and interacts with the proton target modeled as a CGC field. The lifetime of the dipole is much larger than the time scale of its interaction with the proton. One can then factorize the transverse (T) and longitudinal (L) photon-proton cross section as

$$\sigma_{T,L}^{\gamma^* p} = \sigma_0 \sum_f \int d^2\mathbf{r} \int \frac{dz}{4\pi} |\psi^{\gamma^* \to q\bar{q}}(\mathbf{r}, Q^2, z)|^2 N(\mathbf{r}, x_{\text{Bj}}). \tag{1}$$

Here $\psi^{\gamma^* \to q\bar{q}}(\mathbf{r}, Q^2, z)$ describes the $\gamma^* \to q\bar{q}$ splitting at photon virtuality Q^2 and the longitudinal momentum fraction carried by the quark, z. The amplitude, $N(\mathbf{r}, x_{\text{Bj}})$, describes the dipole-target scattering at a particular dipole size, $\mathbf{r}$, and Bjorken-x, x_{Bj}. The BK equation evolves a nonperturbative initial amplitude towards small x_{Bj} values or higher rapidities, $Y = 1/\ln x_{\text{Bj}}$. For the initial condition, we use a McLerran-Venugopalan model [12] inspired parametrization as in Ref. [6],

$$N(\mathbf{r}, x = x_0) = 1 - \exp\left[-\frac{\left(\mathbf{r}^2 Q_{\text{s},0}^2\right)^\gamma}{4} \ln\left(\frac{1}{|\mathbf{r}|\Lambda_{\text{QCD}}} + e_c \cdot e\right)\right]. \tag{2}$$

The parameters describing the initial condition are the following: the parameter that controls $Q_s, Q_{\text{s},0}^2$, the anomalous dimension, γ, and infrared regulator, e_c. We introduce the transverse proton size, $\sigma_0/2$, another free parameter to replace impact parameter dependence such that $\int d^2\mathbf{b} \to \sigma_0/2$. The running of the strong coupling in the transverse coordinate space is parametrized as $\alpha_s \sim (\ln(C^2))^{-1}$ where the C^2 is a parameter that controls the speed of evolution of the BK kernel. Our study focuses on finding the probability distribution of these parameters by constraining against the most recent combined inclusive deep inelastic ep scattering data from HERA [13] in the region 2.0 GeV$^2 < Q^2 <$ 50.0 GeV^2. Finally, our analysis neglects heavy quarks and is limited to the leading order (impact factor and BK equation).

2. Bayesian Inference

We find posterior distributions for the parameters using Bayesian inference setup. In this work, we use a similar procedure as in other works in the field used i.e. in extracting properties

PoS(DIS2024)075

of the quark-gluon plasma [14, 15] and event-by-event fluctuating proton geometry [16]. The workflow involves a combination of a gaussian process emulator (GPE) and a Monte Carlo Markov Chain (MCMC) sampler. Because the MCMC sampler requires a computationally efficient model surrogate, this study uses a GPE [17] in place of the model. The emulator only needs training from model calculations across points in the parameter space to learn the parameter dependence of the model. It represents each predictive point as a gaussian process, characterized by a mean and standard deviation value, by fitting a covariance kernel with hyperparameters over the training points.

The MCMC importance sampler is an algorithm that employs a random walk across the parameter space finding areas where the model and data values are closest. These are areas of high posterior. Bayes' theorem defines the posterior function as

$$P(\mathbf{y}(\boldsymbol{\theta})|\mathbf{y}_{\text{exp}}) \propto P(\mathbf{y}_{\text{exp}}|\mathbf{y}(\boldsymbol{\theta}))P(\boldsymbol{\theta}), \tag{3}$$

where $P(\mathbf{y}_{\text{exp}}|\mathbf{y}(\boldsymbol{\theta}))$ is the likelihood function and $P(\boldsymbol{\theta})$ is the prior at a certain parameter point $\boldsymbol{\theta}$. The likelihood is a function of the difference between the model calculation at certain $\boldsymbol{\theta}$ and the experimental value, taking into account experimental and model uncertainties. Meanwhile, the prior, as used in this study, is a flat distribution that spans the parameter space that encodes *prior* knowledge about possible values or range of the parameters. In the study, the MCMC sampling is done with `emcee` [18] that have multiple random walkers, whose positions are dependent on each other, exploring the parameter space. The acceptance probability of a step of a walker is the ratio of the posterior function between the proposed and the current step. This way the walkers will eventually converge to areas of high posterior and their chain will form samples of the posterior distribution.

3. Results

The five-dimensional posterior distribution for $\boldsymbol{\theta} = (Q_{\text{s},0}^2, \gamma, e_c, C^2, \sigma_0/2)$ obtained through Bayesian analysis are the blue plots shown in Figure 1. Moreover, the red plots present another setup where we fix $\gamma = 1$. The diagonal histograms represent the probability distributions for each parameter. The distributions for most likely values of the parameters have values close to those obtained in previous leading order dipole fits [5–8]. We also get a very good agreement with the HERA data, obtaining $\chi^2/\text{dof} = 1.02$ averaged over many samples for both the 4- and 5-parameter setups. Correlated systematic uncertainties from the experimental data are included in this analysis, the effect of which are wider posterior distributions but unchanged correlations between parameters.

We find well-constrained parameters except e_c. We expected that the dipole amplitude is not very sensitive to e_c especially for small dipoles. In the setup where γ is a free parameter, the posterior distribution is wider and the anomalous dimension covers values of $\gamma = 0.95...1.05$. This is a favorable result as values $\gamma \geq 1$ produce negative 2DFT values hence negative cross sections [19] for inclusive quark production.

The off-diagonal plots inform the correlations between each pair of parameters, for example, the negative correlation between $Q_{\text{s},0}^2$ and $\sigma_0/2$. In the region where the contribution from small dipoles dominate, the dipole-target amplitude is $N(r) \sim (Q_{\text{s},0}^2/Q^2)^\gamma$ where $\sigma^{\gamma^* p} \sim \frac{\sigma_0}{2}N(r)$. This also explains the correlation between γ and $\sigma_0/2$.

3
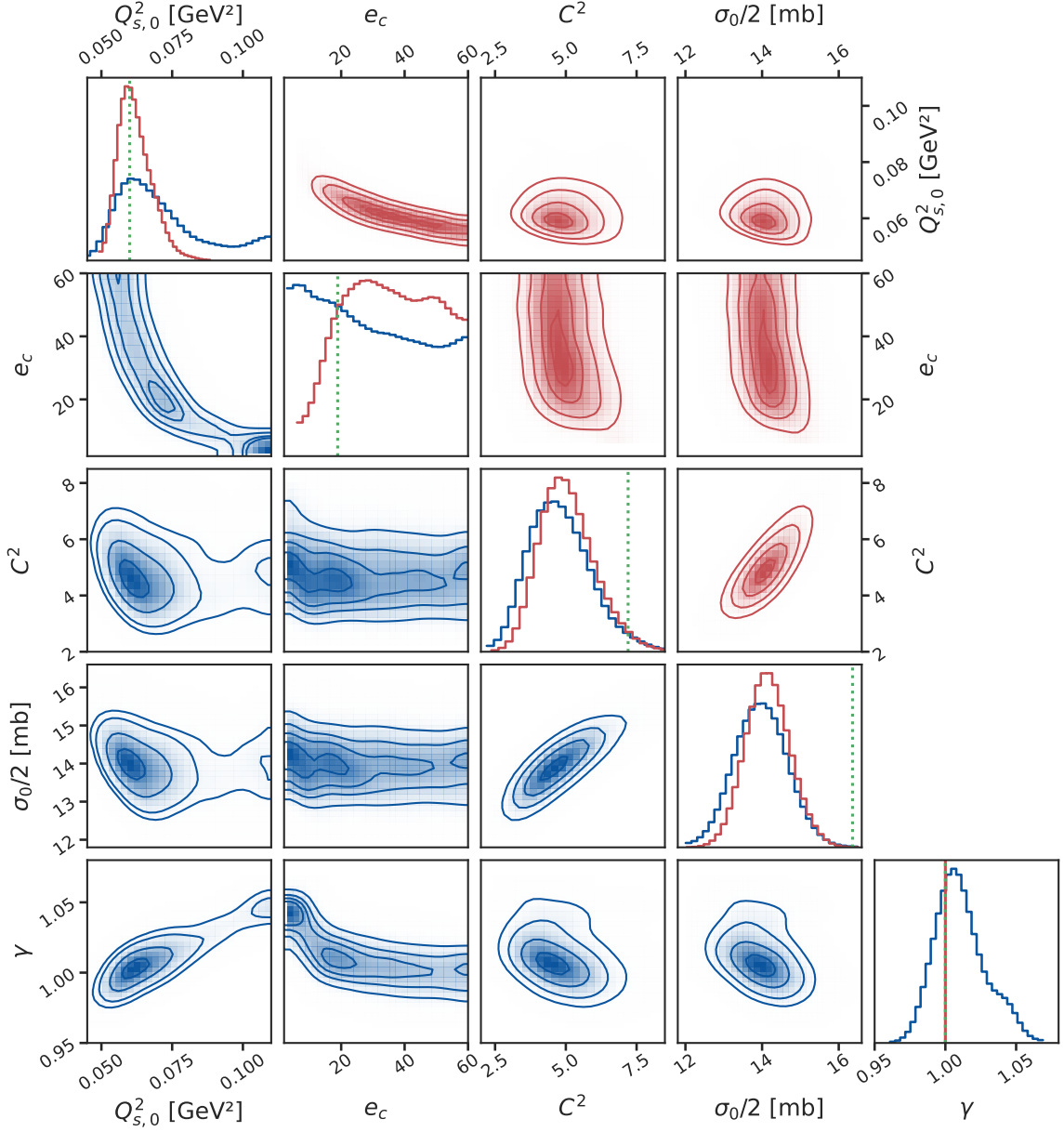
Figure 1: Posterior distributions for both the 4-parameter (red) and 5-parameter (blue) case. The dotted green line shows best fit values from MV^e parametrization of Ref. [6]

The posterior distributions provide a streamlined method of propagating uncertainties to observables that can be calculated in the CGC framework. We demonstrate explicit examples by calculating single inclusive forward quark production in proton-proton collisions from samples of the posterior. It is proportional to the 2-dimensional Fourier transform of the initial dipole-proton amplitude

$$\tilde{S}_P(\mathbf{k}) = \int d^2\mathbf{r} \, e^{-i\mathbf{k}\cdot\mathbf{r}} \left[1 - N(\mathbf{r})\right], \tag{4}$$

shown in Figure 2a. The results illuminate the importance of the $\gamma \leq 1$ constraint as there are
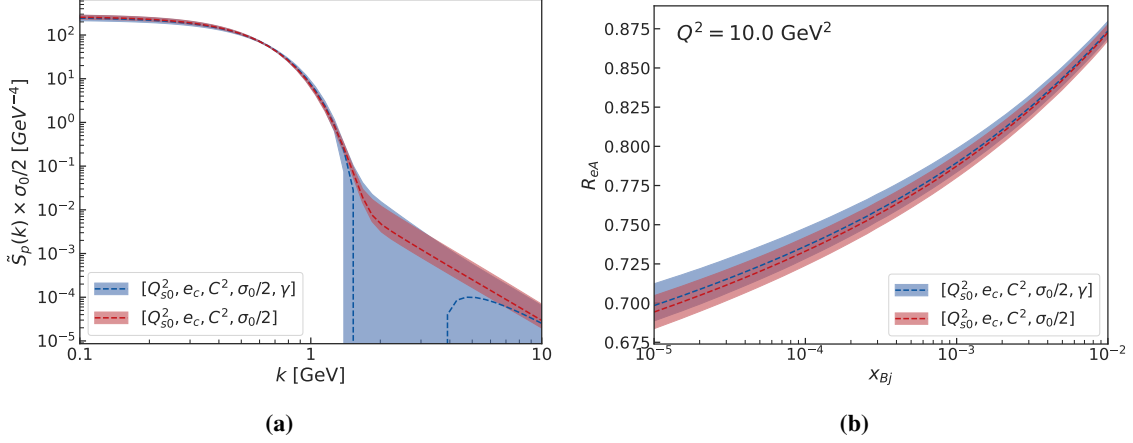
Figure 2: Applications of posterior distributions: (a) 2DFT of the dipole-proton amplitude scaled by $\sigma_0/2$ at initial x_{Bj} in the 4-parameter (red) and 5-parameter (blue) cases. (b) Nuclear modification factor, R_{eA} for the F_2 of as a function of x_{Bj} at fixed $Q^2 = 10.0$ GeV^2 and $A = 197$.

negative values for the 2DFT of the dipole amplitude due to $\gamma \geq 1$ values in the posterior for the 5-parameter setup.

Figure 2b shows the nuclear modification ratio, $R_{eA} = F_{2,A}/(A F_{2,p})$, for the structure function, F_2. We calculate it for an electron-Au collision for which the R_{eA} will be measured at the EIC. The initial condition for the dipole-nucleus scattering amplitude at a fixed impact parameter is related to the transverse thickness function, $T_A(\mathbf{b})$, which is the unintegrated Woods-Saxon potential modeling the distribution of nucleons in the nucleus as in Ref. [6]. Unlike the case for the inclusive quark production cross section where the 2σ uncertainties could reach $\sim 150\%$ the mean value, the uncertainties effectively cancel for the modification factor.

4. Summary & Outlook

We have presented results from the Bayesian inference of the nonperturbative initial condition for the BK evolution equation. Our fits agree very well with the constraining HERA data. We also successfully account for correlated systematic uncertainties from the experimental data. Contrary to previous fits of the initial condition, our results provide uncertainties to parameter estimates enabling a rigorous method to propagate uncertainties for CGC calculations with examples we calculated in the paper.

The Bayesian setup allows easy extension to next-to-leading order analysis and towards constraining the initial condition for the NLO BK equation. For a global analysis approach, we will also include heavy quark data as an additional constraint to the parameters in the future.

Acknowledgments

References

[1] R. Abdul Khalek et al., *Science Requirements and Detector Concepts for the Electron-Ion Collider: EIC Yellow Report*, *Nucl. Phys. A* **1026** (2022) 122447 [`2103.05419`].

[2] F. Gelis, E. Iancu, J. Jalilian-Marian and R. Venugopalan, *The Color Glass Condensate*, *Ann. Rev. Nucl. Part. Sci.* **60** (2010) 463 [`1002.0333`].

[3] 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`].

[4] I. Balitsky, *Operator expansion for high-energy scattering*, *Nucl. Phys. B* **463** (1996) 99 [`hep-ph/9509348`].

[5] 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 [`0902.1112`].

[6] 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 [`1309.6963`].

[7] 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 [`1012.4408`].

[8] B. Ducloué, E. Iancu, G. Soyez and D.N. Triantafyllopoulos, *HERA data and collinearly-improved BK dynamics*, *Phys. Lett. B* **803** (2020) 135305 [`1912.09196`].

[9] 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 [`2007.01645`].

[10] 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) 192301 [`2211.03504`].

[11] C. Casuga, M. Karhunen and H. Mäntysaari, *Inferring the initial condition for the balitsky-kovchegov equation*, *Physical Review D* **109** (2024) [`2311.10491`].

[12] L.D. McLerran and R. Venugopalan, *Computing quark and gluon distribution functions for very large nuclei*, *Phys. Rev. D* **49** (1994) 2233 [`hep-ph/9309289`].

[13] H1, ZEUS collaboration, *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) 580 [1506.06042].

[14] J.E. Bernhard, J.S. Moreland, S.A. Bass, J. Liu and U. Heinz, *Applying Bayesian parameter estimation to relativistic heavy-ion collisions: simultaneous characterization of the initial state and quark-gluon plasma medium*, *Phys. Rev. C* **94** (2016) 024907 [1605.03954].

[15] J.E. Parkkila, A. Onnerstad, S.F. Taghavi, C. Mordasini, A. Bilandzic, M. Virta et al., *New constraints for QCD matter from improved Bayesian parameter estimation in heavy-ion collisions at LHC*, *Phys. Lett. B* **835** (2022) 137485 [2111.08145].

[16] H. Mäntysaari, B. Schenke, C. Shen and W. Zhao, *Bayesian inference of the fluctuating proton shape*, *Phys. Lett. B* **833** (2022) 137348 [2202.01998].

[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel et al., *Scikit-learn: Machine learning in Python*, *Journal of Machine Learning Research* **12** (2011) 2825 [1201.0490].

[18] D. Foreman-Mackey, D.W. Hogg, D. Lang and J. Goodman, *emcee: The mcmc hammer*, *PASP* **125** (2013) 306 [1202.3665].

[19] B.G. Giraud and R. Peschanski, *Fourier-positivity constraints on QCD dipole models*, *Phys. Lett. B* **760** (2016) 26 [1604.01932].