Machine-learning prediction for quasiparton distribution function matrix elements

Rui Zhang,[1,*] Zhouyou Fan,[1] Ruizi Li,[1] Huey-Wen Lin,[1,2,†] and Boram Yoon[3]

[1]*Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA*
[2]*Department of Computational Mathematics, Science & Engineering, Michigan State University, East Lansing, Michigan 48824, USA*
[3]*Computer, Computational, and Statistical Sciences CCS-7, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

(Received 23 November 2019; accepted 3 February 2020; published 24 February 2020)

There have been rapid developments in the direct calculation in lattice QCD (LQCD) of the Bjorken-x dependence of hadron structure through large-momentum effective theory (LaMET). LaMET overcomes the previous limitation of LQCD to moments (that is, integrals over Bjorken x) of hadron structure, allowing LQCD to directly provide the kinematic regions where the experimental values are least known. LaMET requires large-momentum hadron states to minimize its systematics and allow us to reach small-x reliably. This means that very fine lattice spacing to minimize lattice artifacts at order $(P_z a)^n$ will become crucial for next-generation LaMET-like structure calculations. Furthermore, such calculations require operators with long Wilson-link displacements, especially in finer lattice units, increasing the communication costs relative to that of the propagator inversion. In this work, we explore whether machine-learning algorithms can make predictions of correlators to reduce the computational cost of these LQCD calculations. We consider two algorithms, gradient-boosting decision tree and linear models, applied to LaMET data, the matrix elements needed to determine the kaon and η_s unpolarized parton distribution functions (PDFs), meson distribution amplitude (DA), and the nucleon gluon PDF. We find that both algorithms can reliably predict the target observables with different prediction accuracy and systematic errors. The predictions from smaller displacement z to larger ones work better than those for momentum p due to the higher correlation among the data.

DOI: 10.1103/PhysRevD.101.034516

I. INTRODUCTION

In the early days, probing hadron structure with lattice QCD (LQCD) was limited to only the first few moments, due to complications arising from the breaking of rotational symmetry by the discretized Euclidean spacetime. The nonzero lattice spacing breaks the symmetry group of Euclidean spacetime from $O(4)$ to the discrete hypercubic subgroup $H(4)$. Due to the reduced symmetry, the required operators are more complicated and often either suffer from divergences or mix with other operators under renormalization. This is treatable but complicated. As a result, even with increasing computational resources becoming available to the lattice-QCD community, LQCD hadronic structure calculations were limited to the lowest few moments (see Refs. [1,2] and references within for more details). Although modeling the x dependence to reproduce the calculated lattice moments to gain information on the x dependence [3] was attempted, this will only give the combinations of the difference between quark and antiquark contributions rather than individual (anti)quark contributions. Experiments such as E665 at FNAL can probe nucleon sea flavor asymmetry, meaning that lattice QCD would be excluded if it could only apply traditional moment calculations. Similarly, STAR at RHIC is probing the polarized (anti)quark structure of nucleon. The future electron-ion collider will further study sea structure. Facing these challenges, LQCD required a new computationally friendly approach to extend its applicability to calculations of PDFs and catch up with ongoing experimental efforts.

Large-momentum effective theory (LaMET) [4] is one of the most widely adopted new methods for calculating the full x dependence of hadron structure. In the LaMET framework, we take an operator containing an integral of gluonic field strength along a line and boost the nucleon momentum toward the speed of light, tilting the spacelike line segment toward the light-cone direction. The time-independent, nonlocal (in space) correlators at finite P_z can

*zhangr60@msu.edu
†hwlin@pa.msu.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP[3].

be directly evaluated on the lattice. For example, the quark unpolarized distribution of a hadron can be calculated via

$$q_{\rm lat}(x,\mu,P_z)=\int\frac{dz}{4\pi}e^{izk}\times\langle\vec{P}|\bar{\psi}(z)\Gamma\left(\prod_n U_z(n\hat{z})\right)\psi(0)|\vec{P}\rangle, \tag{1}$$

where U_z is a discrete gauge link in the z direction, $\Gamma=\gamma_t$, $x=k/P_z$, μ is the renormalization scale and $\vec{P}$ is the momentum of the hadron, taken such that $P_z\to\infty$. The $q_{\rm lat}(x,\mu,P_z)$, often called the "quasi-PDF" [5], is related to the light-cone PDF through a factorization theorem, where the former can be factorized into a perturbative matching coefficient and the latter, up to power corrections suppressed by the nucleon momentum. This factorization theorem is founded in LaMET [4,6–9], where the matching coefficient can be calculated exactly in perturbation theory. Lattice-QCD results using LaMET already include the isovector quark PDF of the nucleon [10–14], the pion generalized parton distribution [15], the meson distribution amplitudes (DAs) [16,17] and the nonperturbative renormalization in the regularization-independent momentum subtraction scheme [18,19]. Certain technical issues regarding the nonperturbative renormalization were raised and addressed in Refs. [14,18–23]. The finite-volume effect in nucleon quasi-PDF was studied in [24].

Even with these promising results published and efforts ongoing, much work remains to be done. For example, most work so far has been limited to a single ensemble; more detailed studies incorporating the systematic errors from lattice artifacts, such as finite-volume and lattice spacing, is necessary to reach precision LQCD PDFs. Larger boost momentum in the hadron is important to suppress finite-momentum corrections, as well as getting the antiquark distribution and small-x quark distribution corrections. Ensembles with smaller lattice spacing ($a^{-1}>4$ GeV) will become the crucial factors in the next generation of LaMET calculations. To reach larger boost momentum P_z, a smaller lattice spacing is needed to control the $(P_z a)^n$ lattice artifacts, similar to how heavy-quark studies must control the heavy-quark mass artifacts at order $(m_q a)^n$. Likely, more than $O(100,000)$ calculations will be needed to get a good signal-to-noise for the three-point correlators and allow us to disentangle the excited states from the ground state. A larger number of degrees of freedom will be necessary to keep the finite-volume systematic within the consensus optimal region, $M_\pi L\approx 4$. More communication costs will be incurred transporting the Wilson link from one side of the lattice to the other, which can easily become a dominating cost for the calculation. Although optimizing communication efficiency may address the latter problem, we are hoping to find a method that will work for both the large-momentum and Wilson-link displacement issues that are characteristic of LaMET and its similar approaches.

Recently, authors of Ref. [25] introduced a machine-learning (ML) approach predicting observables by taking advantage of the correlations between lattice-QCD observables. Two types of data with high-statistics measurements, $O(100,000)$, were used in the studies: nucleon isovector charges and the CP-violating phase induced by the quark chromoelectric dipole moment interactions. The authors found a reduction in the computational cost by 7%–35%, depending on observable, showing very promising potential for lattice-QCD applications. In this work, we are interested in finding out how well the ML approach would work with the difficult computational situation in LaMET-type observables and whether computational cost can be saved for future large lattices (say, $64^3\times 192$ and above). Although this paper focuses on the discussion with quasi-PDF, what we learn here also applies to the pseudo-PDF[1] [27–30] correlators since the building blocks of matrix elements are the same.

The structure of this paper is as follows: In Sec. II we briefly describe the two ML algorithms used in this work. Section III demonstrates the application of both algorithms to LaMET-type observables, including the correlators from the meson distribution amplitude, kaon and η_s parton distribution functions and nucleon gluon parton distribution function. We compare results of the ML predictions. We summarize the conclusions and future prospects of this work in Sec. IV.

II. MACHINE-LEARNING ALGORITHM

ML works by optimizing a prediction model mapping between input and output data, creating a function approximating the relationship between them, inferred from data. The model is built from a set of data whose label (output) is known, and it is applied to make predictions of the labels for a new set of data whose label is unknown, assuming that there exists a consistent mapping function between the input and output data. In this study, we use regression algorithms, a class of ML approaches, to make quantitative predictions of lattice-QCD measurements. Specifically, the supervised ML regression algorithms we use are the simple linear regression and the gradient boosting tree (GBT) algorithms [31] implemented in the PYTHON scikit-learn package [32]. Although all the data we use in this test have labels, we divide them into the labeled and unlabeled sets, by hiding the label for the unlabeled dataset. Then, the labeled dataset is used for training and bias-correction procedure, while the unlabeled dataset is used for the test of the trained regression algorithm.

Gradient boosting is one of the techniques creating a strong model from an ensemble of weak prediction models

[1]Another paper [26] applied the neural network (NN) algorithm to the inversion problem to reconstruct PDF from pseudo-PDF matrix elements, though the model was trained and tested on mock datasets instead of real lattice data.

[33,34]. For GBT, the shallow decision trees are used as the weak learners (nested mappings as elements of a more complicated function approximation) in series and building the active model:

$$f_k(x) = \sum_{i=1}^{k} r_i h_i(x), \qquad f(x) = f_{N_{\text{est}}}(x) \tag{2}$$

where N_{est} is the number of estimators, r_i is the learning rate, and $h_i(x)$ is the function used in the base decision tree to minimize the loss function L:

$$h_i(x) = \arg\min_h \sum_j L(y_j, f_{i-1}(x_j) + h(x_j)) \tag{3}$$

where the subscript j iterates over the training-data samples. In this work, the loss function is chosen to be the mean squared error, and the depth of the decision tree is fixed at 3. To optimize the ML predictions, we must choose model parameters in Eq. (2) within the proper range. Two parameters are tuned explicitly in this process: the learning rate r, and the number of estimators N_{est}.

The prediction accuracy of GBT is compared with those of the linear regression model

$$f^{\text{lin}}(\vec{x}) = \theta_0 + \vec{\theta} \cdot \vec{x} \tag{4}$$

for the same set of data. Quantitatively, the quality of the prediction accuracy of the regression models is represented by the fit variance F_v defined as

$$F_v = 1 - (\langle (C_{\text{ul}} - C_{\text{pred}})^2 \rangle - \langle C_{\text{ul}} - C_{\text{pred}} \rangle^2)/\sigma^2, \tag{5}$$

where C_{ul} and C_{pred} are the observed and predicted measurements on unlabeled dataset, respectively, and σ^2 is the variance of the observed measurements. The higher value of F_v indicates the better fit quality, and the maximum value of F_v is 1, which shows a perfect prediction, $C_{\text{pred}} = C_{\text{ul}}$. In practical calculations, the F_v can be calculated on the bias-correction dataset of the labeled dataset, which is described below. However, we use the unlabeled dataset for the calculation of the F_v, because C_{ul} are available in this test study.

Prediction from a ML algorithm may have bias due to prediction error. We follow the bias-correction strategy introduced in Ref. [25] to remove the bias in our estimate and define the bias-corrected prediction as

$$\langle C_{\text{pred,BC}} \rangle = \langle C_{\text{pred}} \rangle_{\text{ul}} + \langle C_{\text{BC}} - C_{\text{pred}} \rangle_{\text{BC}}, \tag{6}$$

where the brackets with subscripts "ul" and "BC" denote averages over the unlabeled and bias-correction datasets, respectively. After bias correction, the expectation value of the prediction becomes the same as the expectation value of the ground truth, and its statistical error includes the systematic error due to inaccurate predictions. After the bias correction, therefore, our main concern is reducing the statistical error of the final estimate.

TABLE I. The convention for the subscripts we use in this work.

Subscript	Convention
in	input to the model
pred	prediction of the model
pred, BC	bias-corrected prediction
tr	labeled training data
BC	labeled bias-correction data
lb	all labeled data
ul	unlabeled data

We normalize the labeled data so that the standard deviation of each input measurement becomes 1 before we pass it to the ML algorithms. Each subset of the data (training, bias-correction, and unlabeled datasets) described in Ref. [25] are chosen such that the configurations are evenly distributed. The convention of notations throughout this work is given in Table I.

The errors of the predictions are estimated using the bootstrap method. We randomly pick the bootstrap samples for labeled and unlabeled datasets, and partition the labeled one into training and BC datasets. We train the model and estimate the bias correction on each bootstrap sample of labeled data. We make prediction on the corresponding sample of unlabeled data and calculate the average of the results for unlabeled data. The error is then estimated over all bootstrap samples.

III. APPLICATION TO LATTICE QUASI-PDF MATRIX ELEMENTS

A. Predictions of meson quasi-DA measurements

Meson DAs ϕ_M are important universal quantities appearing in many factorization theorems, which allow for the description of exclusive processes at large-momentum transfers $Q^2 \gg \Lambda^2_{\text{QCD}}$ [35,36]. Such quantities can be calculated using LaMET [4,8] by calculating the time-independent spatial correlators (the quasi-DA) on the lattice, followed by a matching procedure with corrections suppressed by the hadron momentum. The light-cone meson DA

$$\phi_M(x,\mu) = \frac{i}{f_M} \int \frac{d\xi}{2\pi} e^{i(x-1)\xi n \cdot P} \times \langle M(P)|\bar{\psi}_1(0) n \cdot \gamma \gamma_5 U(0,\xi n) \psi_2(\xi n)|0\rangle \tag{7}$$

can be extracted from the quasi-DA

$$\tilde{\phi}_M(x,\mu_R,P_z) = \frac{i}{f_M} \int \frac{dz}{2\pi} e^{i(x-1)zP_z} \times \langle M(P)|\bar{\psi}_1(0)\gamma_z\gamma_5 \prod_{x=0}^{z-1} U_z(x,t)\psi_2(z)|0\rangle \tag{8}$$

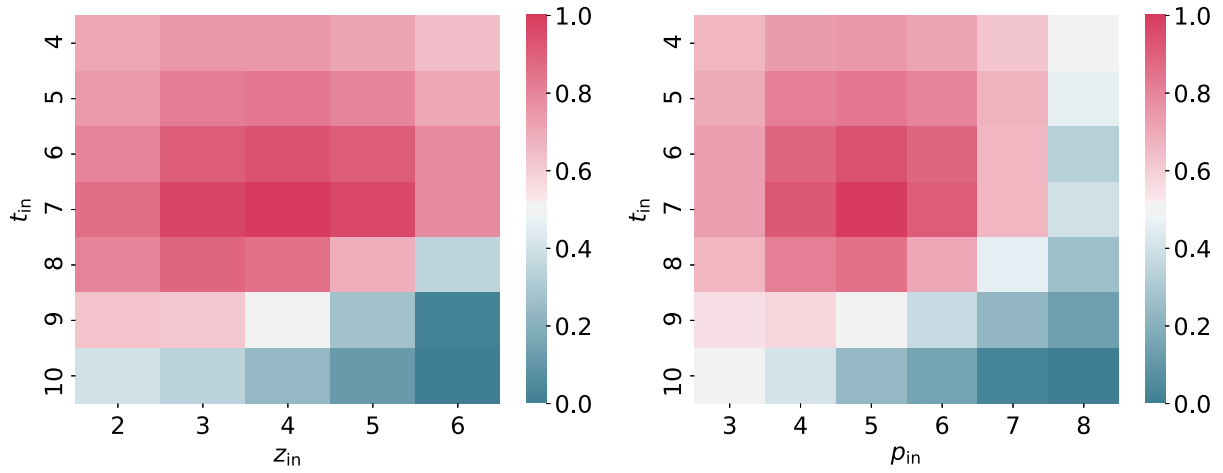

FIG. 1. Correlations between target η_s DA C_{2pt} data at $z_{pred} = 4$, $p_{pred} = 5$, $t_{pred} = 7$ with input data at a different link length (momentum) and time slice for z prediction (left) and p prediction (right). The correlation decays quickly, especially at larger t.

through the matching [37]

$$\tilde{\phi}_M(x,\mu_R,P_z) = \int dy Z_\phi(x,y,\mu,\mu_R,P_z)\,\phi_M(y,\mu) + \mathcal{O}\left(\frac{\Lambda_{QCD}}{P_z^2},\frac{m_M^2}{P_z^2}\right) \quad (9)$$

according to LaMET. The quasi-DA can be obtained by computing the following correlators for K^- and η_s, as presented in the Refs. [16,17]:

$$C_{2pt}(z,P,t) = \langle 0|\int d^3y e^{i\vec{P}\cdot\vec{y}}\bar{\psi}_1(\vec{y},t)\gamma_z\gamma_5\prod_{x=0}^{z-1}U_z(y+x\hat{z},t) \times \psi_2(\vec{y}+z\hat{z},t)\bar{\psi}_2(0,0)\gamma_5\psi_1(0,0)|0\rangle \quad (10)$$

where $\{\psi_1,\psi_2\}$ are $\{u,s\}$ for K^- and $\{s,s\}$ for η_s, $U(\vec{x},\vec{x}+z)$ is the Wilson line connecting lattice site $\vec{x}$ to $\vec{x}+z\hat{z}$.

We perform a calculation using gauge ensembles with clover valence fermions on a $48^3\times144$ lattice with $2+1+1$ flavors (degenerate up and down, strange, and charm degrees of freedom) of highly improved staggered quarks (HISQ) [38] generated by the MILC Collaboration [39]. The lattice spacing $a \approx 0.06$ fm, and $m_\pi^{sea} = 310$ MeV. Hypercubic (HYP) smearing [40] is applied to the configurations. The bare quark masses and clover parameters are tuned to recover the lowest pion mass of the staggered quarks in the sea. Correlators are calculated from momentum-smearing sources [41] using 20 source locations on each of the 95 configurations (1900 measurements in total).

We make two predictions using the ML algorithm. One is to predict the correlators at larger link length z_{pred} from the correlators at $z_{in} < z_{pred}$. The other is to predict the correlators of larger momentum p_{pred} from the correlators of $p_{in} < p_{pred}$.

To determine what input data to use for these predictions, we first check the correlations among datasets with different momenta, link lengths and time slices. The results are shown in Fig. 1. Here, we set the target data to be the 2-point quasi-DA correlators at $p_{pred} = 5$, $z_{pred} = 4$ with input data $p_{in} = 4$, $z_{in} = 4$ for p prediction and $p_{in} = 5$, $z_{in} < 4$ for z prediction. We select the time slice $t_{pred} = 7$ to check the correlations.

Despite the larger error, larger time slices have a weaker correlation with the target data. This suggests that we should use input data close to the time slice of the target data. On the other hand, we should be able to extend the range of momentum or links of the input.

In the training process, we tried different parameters for learning rate in $\{0.5, 0.2, 0.1, 0.02, 0.01, 0.005, 0.002\}$ and the number of estimators in $\{100, 150, 200, 250, 300\}$. The corresponding fit variance are plotted in a heat map with range [0,1], as shown in Fig. 2. Considering the fit quality for both p predictions and z predictions, we selected parameters $r = 0.1$, $N_{est} = 150$ as having highest fit quality in both cases; these will be used for further meson-DA predictions.

The datasets were evenly distributed into three parts: training data, bias-correction data, and unlabeled test data. In practice, we want to minimize the labeled data size without sacrificing much prediction quality. We varied the amount of training data and bias-correction data from 300 to 500, while keeping the number of unlabeled test data $N_{ul} = 900$ fixed, to look for a best trade-off between reduced data size and prediction quality. The results are shown in Fig. 3. When correlation is obvious, small number of training and bias-correction datasets provides precise estimate that is very close to the true observations for the unlabeled dataset. When correlation is vague, the prediction becomes more precise as one increases the size of the training or the bias-correction datasets. Based on the plot, we picked $N_{tr} = 400$, $N_{BC} = 500$ for further estimations.

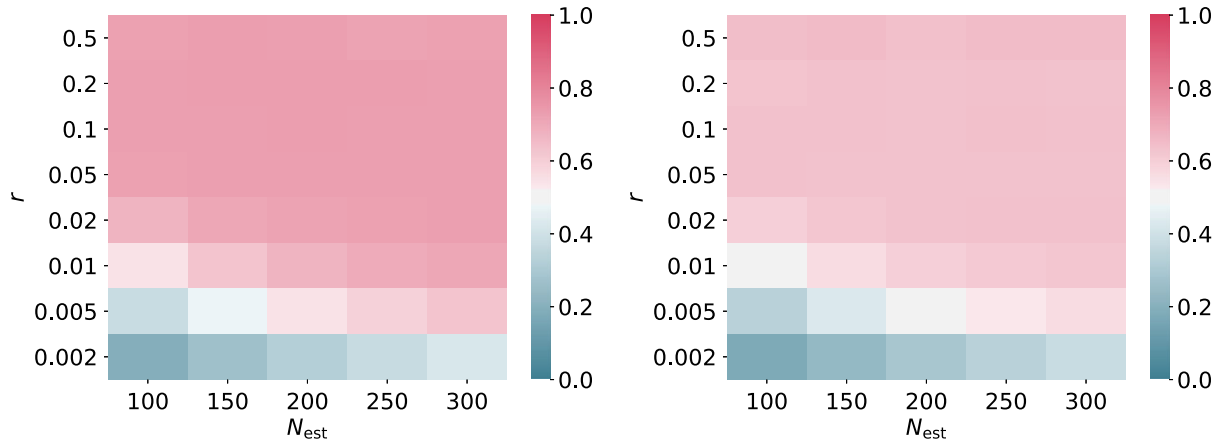

FIG. 2. Fit variance F_v of the unlabeled η_s DA data for the $p_{\text{pred}} = 5$, $z_{\text{pred}} = 4$ prediction at $t_{\text{pred}} = 4$ from $z_{\text{in}} = 3$ (left) or $p_{\text{in}} \in [3, 4]$ (right). $N_{\text{tr}} = 400$, $N_{\text{ul}} = 1000$. It is clear that more estimators are needed for smaller learning rate. Increasing N_{est} without worsening the prediction indicates that the model is robust to overfitting.

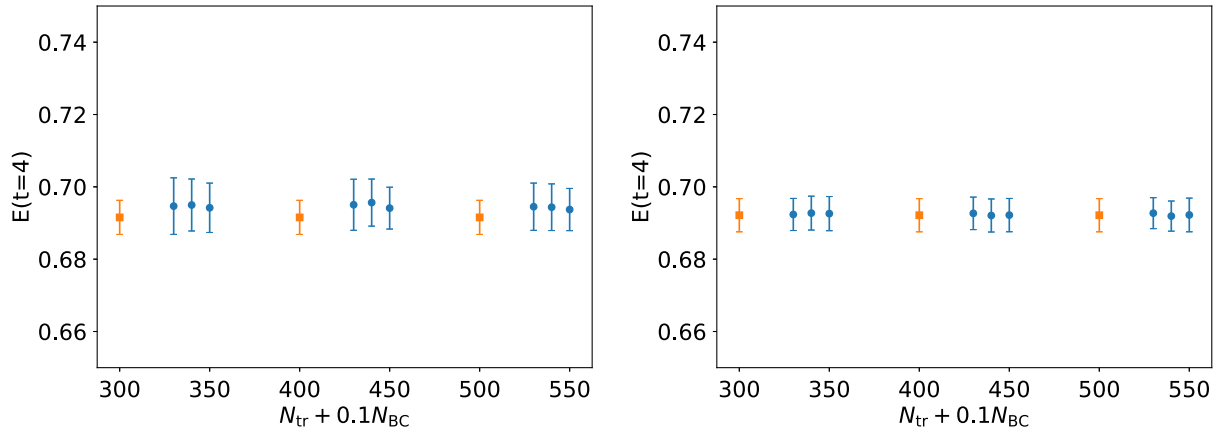

FIG. 3. The observed and z-predicted η_s DA effective mass of $p_{\text{pred}} = 5$, $z_{\text{pred}} = 4$ at $t_{\text{pred}} = 4$ with input $p_{\text{pred}} = 5$, $z_{\text{pred}} \in [0, 3]$, $t_{\text{in}} \in [3, 5]$ for different choices of training-data counts and bias-correction data counts. The left (right) plot is the prediction of GBT (linear) model. The horizontal axis is $N_{\text{tr}} + 0.1N_{\text{BC}}$, with $N_{\text{ul}} = 900$ fixed. The GBT parameters are $N_{\text{est}} = 150$, $r = 0.1$. The blue points are predictions with bias correction for the unlabeled test data, and the brown points are observations for unlabeled test data.

To further check the consistency of our predictions with the observations, we calculate the effective mass from $C_{2\text{pt}}$ and compare the results. The effective mass is defined as

$$E(t) = \ln \frac{C_{2\text{pt}}(t)}{C_{2\text{pt}}(t+1)}. \tag{11}$$

Then, we compared different input data to be used for z prediction in Table II. The bias correction makes the prediction noisier by converting the systematic error into statistical error, which improves the accuracy of the prediction for most cases.

For small datasets, such as what we have for the quasi-DA data, it can be difficult for the GBT model to extract the

TABLE II. Effective mass calculated from the prediction of η_s DA $C_{2\text{pt}}$ at $p_{\text{pred}} = 5$, $z_{\text{pred}} = 4$, $t_{\text{pred}} = 7$ with different models and different inputs. Models are trained with $N_{\text{tr}} = 400$, $N_{\text{BC}} = 500$, $N_{\text{ul}} = 900$, $N_{\text{est}} = 150$, $r = 0.1$. The linear model is more accurate than GBT. Both models have better performance for z prediction than p prediction.

Type	Input	Method	E_{tr}	E_{pred}	$E_{\text{pred,BC}}$	E_{ul}	F_v
p-pred	$p_{\text{in}} \in [3, 4]$, $z_{\text{in}} = 4$, $t_{\text{in}} = 7$	GBT	0.679(11)	0.684(13)	0.683(14)	0.6923(80)	0.50(13)
		linear	0.679(11)	0.6960(86)	0.6961(91)	0.6920(74)	0.911(43)
z-pred	$p_{\text{in}} = 5$, $z_{\text{in}} \in [0, 3]$, $t_{\text{in}} = 7$	GBT	0.679(11)	0.694(13)	0.692(12)	0.6923(80)	0.62(14)
		linear	0.679(11)	0.6913(76)	0.6912(75)	0.6920(74)	0.99935(40)

TABLE III. Effective mass calculated from the prediction of η_s DA C_{2pt} at $p_{pred} = 5$, $z_{pred} = 4$, $t_{pred} = 10$ with different models and different input time slices. Models are trained with $N_{tr} = 400$, $N_{BC} = 500$, $N_{ul} = 900$, $N_{est} = 150$, $r = 0.1$. The linear model has better performance on correlated cleaner data but fails when more uncorrelated noisy data input are included. The GBT is more stable and less sensitive to these inputs.

Type	Input	Method	E_{tr}	E_{pred}	$E_{pred,BC}$	E_{ul}	F_v
p-pred	$p_{in} \in [3,4]$, $z_{in} = 4$, $t_{in} = 10$	GBT	0.686(43)	0.678(45)	0.683(40)	0.675(26)	0.36(19)
		linear	0.683(37)	0.692(39)	0.695(39)	0.676(27)	0.72(13)
p-pred	$p_{in} \in [3,4]$, $z_{in} = 4$, $t_{in} \in [7,13]$	GBT	0.686(43)	0.677(51)	0.676(43)	0.675(26)	0.25(27)
		linear	0.683(37)	0.675(88)	0.677(77)	0.676(27)	−0.13(85)

nonlinear pattern of the training dataset. As a consequence, the fit quality of the GBT model for the test data is poor. Instead, the simpler linear regression shows better performance. Sometimes, however, with input data when the dataset is noisy (e.g., larger-t data), the linear regression fails with poor prediction quality, as shown in Table III, while GBT was able to capture the correlation and make predictions. Using cleaner and more correlated data like the closest time slice, momentum and link can significantly improve the fit quality for linear regression.

After determining the parameters, we run the ML program and show the effective mass of our predictions along with the observed datasets for both p_{pred} and z_{pred} predictions in Fig. 4. The linear model works well for z prediction, but the GBT model and p predictions still need to be improved.

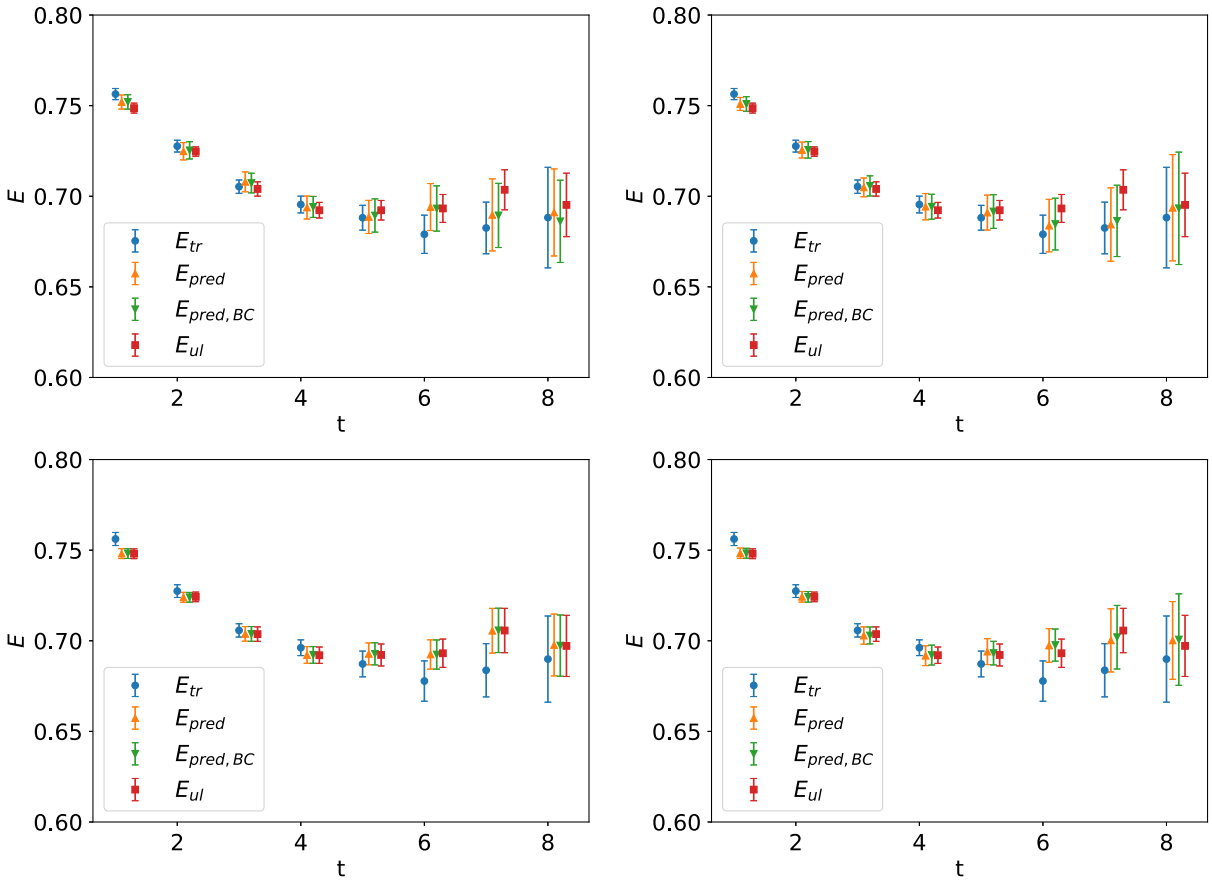

FIG. 4. The observed/predicted η_s DA effective mass at $p_{pred} = 5$, $z_{pred} = 4$ from $p_{in} = 5$, $z_{in} \in [0,3]$ (left) and $p_{in} \in [3,4]$, $z_{in} = 4$ (right). The top (bottom) plots are obtained by using GBT (linear) model with $N_{tr} = 400$, $N_{BC} = 500$, $N_{ul} = 900$. The GBT parameters are $N_{est} = 150$, $r = 0.1$. The linear model shows better consistency with the unlabeled data, while at some time slices the GBT model fails to give a good prediction.

B. Predictions of kaon quasi-PDFs

As Nambu-Goldstone bosons associated with dynamical chiral SU(3) symmetry breaking, the pion and kaon serve as a fundamental test ground for our understanding of QCD theory at the hadronic scale. The *ab initio* calculation of hadron PDFs from lattice QCD provides theoretical background for particle-discovery experiments and Standard-Model (SM) tests at colliders [42]. After decades of theoretical and experimental efforts, the precision required in PDFs for more stringent tests of the SM has increased significantly. In Ref. [43], we presented the first direct lattice calculation of the valence-quark distribution in the pion, using the MILC HISQ coarse ensemble with $M_\pi \approx 330$ MeV. Since the computational cost of quasi-PDF measurements on an ensemble at lighter pion mass or reduced lattice spacing would increase significantly, in this work we investigate a ML algorithm to reduce the computational cost.

We test on the meson unpolarized quasi-PDF measurements on the lattice:

$$C_{3\mathrm{pt}}(z,t) = \langle 0| \int d^3y e^{-iy\cdot P} M_{\mathrm{ps}}(\vec{y}, t_{\mathrm{sep}}) \bar{s}(z,t)\gamma_4 \times \prod_{x=0}^{z-1} U_z(x,t) s(0,t) \bar{M}_{\mathrm{ps}}(\vec{0},0)|0\rangle, \quad (12)$$

$$C_{2\mathrm{pt}}(t_{\mathrm{sep}}) = \langle 0| \int d^3y e^{-iy\cdot P} M_{\mathrm{ps}}(\vec{y}, t_{\mathrm{sep}}) \bar{M}_{\mathrm{ps}}(\vec{0},0)|0\rangle, \quad (13)$$

where $C_{3\mathrm{pt}}$ is the three-point correlator, $C_{2\mathrm{pt}}$ is the two-point correlator, $M_{\mathrm{ps}} = \bar{q}\gamma_5 q$ is the pseudoscalar meson operator, z is the length of the Wilson link, $U_\mu(x,t)$ is the gauge link, and γ_i are Dirac spinor matrices. For this study we use Wilson clover valence quarks on a MILC HISQ ensemble. The lattice spacing is $a \approx 0.12$ fm, the lattice volume $V = 40^3 \times 64$, and the pion mass $M_\pi^{\mathrm{sea}} \approx 220$ MeV. The valence-quark masses are tuned to match the valence pion to the sea pion mass. We adopt Gaussian momentum smearing [41] to generate quark sources, to enhance the ground-state signal at nonzero momentum near 1.55 GeV. The Gaussian smearing width is chosen to be 3, with 50 iterations, and the momentum parameter $k = 4.82$. Measurements are done on 495 configurations, using 4 quark-source locations per configuration, making 1960 measurements in total. Measurements are averaged over these quark sources before being passed to the ML algorithm, as this has shown to provide predictions with smaller statistical errors. The ratio of the three-point correlator ($C_{3\mathrm{pt}}$) to the two-point correlator ($C_{2\mathrm{pt}}$) is a useful way to extract the matrix elements:

$$R(t) = C_{3\mathrm{pt}}(t)/C_{2\mathrm{pt}}(t_{\mathrm{sep}}) \quad (14)$$

where $R(t)$ is the ratio at the operator insertion time t, and t_{sep} is the meson source and sink temporal separation.

1. Kaon quasi-PDF results

For the kaon quasi-PDF, the meson operator is $K = \bar{u}\gamma_5 s$. We first check the correlation for three-point correlators with insertion operator γ_4. Generally, the correlations are better than for the DA case. The correlations between different time are shown in Fig. 5. The correlation is insensitive to insertion time, but sensitive to the difference between two-point time slice and three-point source-sink time separation. Because the correlators are similar for different insertion times, we can use all the insertion time slices as input in the same procedure. An anomaly is observed in the momentum correlation in Fig. 6, which may be due to the different number of measurements for $p \in \{3,4\}$ and $p \in \{5,6\}$, since we had an extra run for $p \in \{5,6\}$ with different source locations. The link

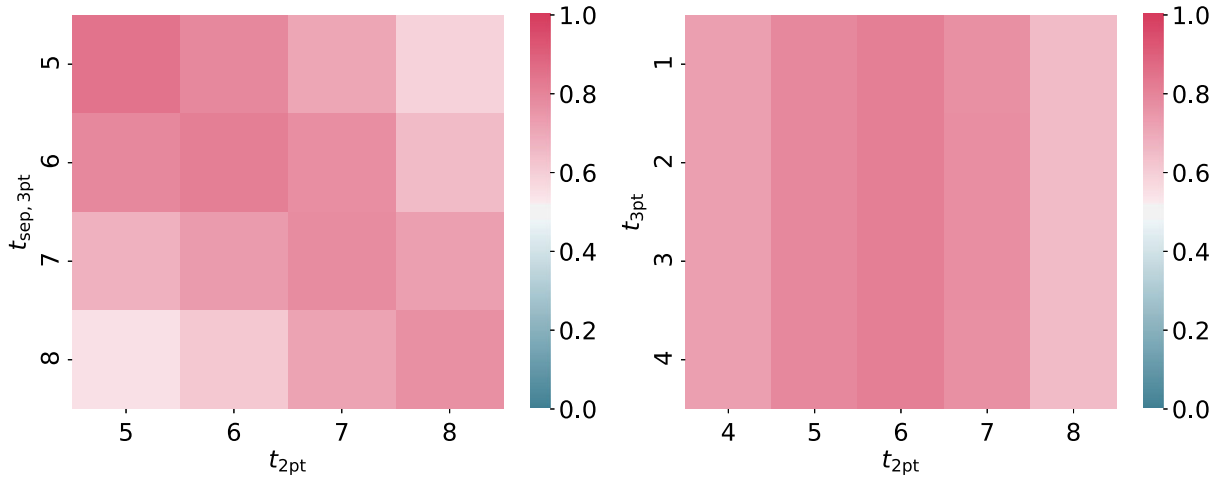

FIG. 5. Correlations between $C_{2\mathrm{pt}}$ and $C_{3\mathrm{pt}}$ of kaon quasi-PDF at different time separations (left, with insertion time $t_{3\mathrm{pt}} = t_{\mathrm{sep}}/2$) and at different insertion times (right, with $t_{\mathrm{sep}} = 6$). The correlation is insensitive to the insertion time.

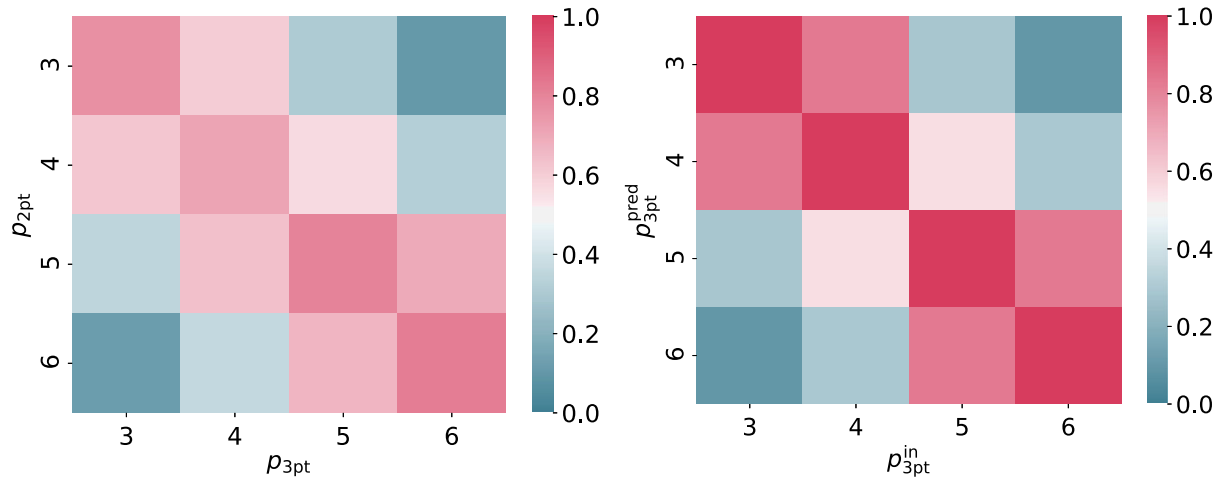

FIG. 6. Correlations between the kaon quasi-PDF three-point correlators and two-point correlators (left) or three-point correlators (right) at different momenta. The three-point correlation seem to be clustered; $p \in \{3,4\}$ and $p \in \{5,6\}$ are correlated separately. Thus, the prediction of $p_{\text{pred}} = 5$ from smaller momentum has bad quality.

correlation is then displayed in Fig. 7. The correlation decays slowly in the z direction, suggesting that we may use more data at different links as inputs.

Again, we compare the parameters used for the GBT model. Figure 8 shows the fit-variance estimate F_v from both the z prediction (with $p_{\text{in}} = p_{\text{pred}}$ and $z_{\text{in}} < z_{\text{pred}}$), and the p prediction (with $p_{\text{in}} < p_{\text{pred}}$ and $z_{\text{in}} = z_{\text{pred}}$) using the GBT model trained on 400 measurements. The horizontal axis shows the number of estimators N_{est}, and the vertical axis shows the learning rate r. The target measurement is at $p_{\text{pred}} = 4$, $z_{\text{pred}} = 4$, $t_{\text{sep}} = 5$, and $t = 2$. For each prediction we used both $C_{3\text{pt}}$ and the $C_{2\text{pt}}$. Thus, in either case a set of fit parameters can be chosen as, e.g., $N_{\text{est}} = 150$, $r = 0.1$. As expected, with reduced learning rate, one needs

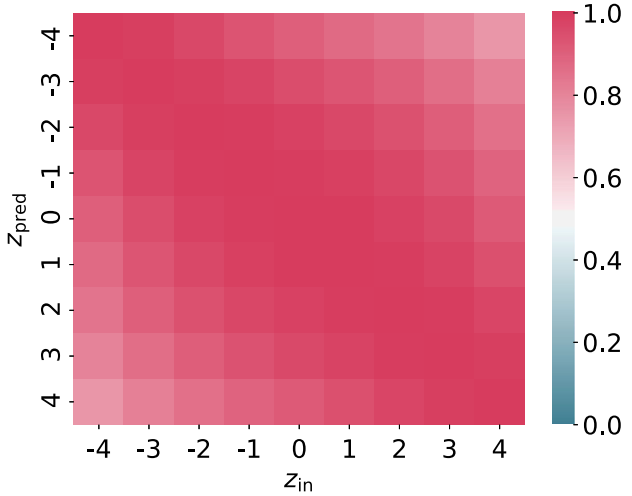

FIG. 7. Correlations between the kaon quasi-PDF three-point correlators with $p_{\text{in}} = p_{\text{pred}} = 3$ at different link lengths. Shorter link lengths have better correlation, and the correlation decay in the z direction is slow.

more estimators to achieve a similar fit variance. With fixed learning rate, the fit variance becomes stable when we keep increasing N_{est}, indicating that the model is robust to overfitting.

Figure 9 compares the final predictions among various training and bias-correction measurements: N_{tr} and N_{BC} are selected from $\{80, 160, 240, 320, 400\}$, and the number of unlabeled measurements is fixed to $N_{\text{ul}} = 1180$. The fit parameters are adopted as above. We observe a reduced error size of final predictions with increased N_{tr} and N_{BC}.

Using p prediction on kaon quasi-PDFs can reduce the computational cost, because calculating $C_{3\text{pt}}$ at different momenta requires the calculation of different propagators from different sequential sources. The effective computational savings of the ML calculation can be derived by considering the number of propagators needed to achieve the same precision as in a calculation without ML. In our case, to use $p_{\text{in}} = 3$ to predict $p_{\text{pred}} = 4$, we need to calculate N_{in} propagators at $p_{\text{in}} = 3$ and $N_{\text{BC}} + N_{\text{tr}}$ propagators at $p_{\text{pred}} = 4$ for the ML setup. Then, we can use the model to obtain the N_{ul} predictions at $p_{\text{pred}} = 4$. This amount of data is equivalent to a non-ML calculation with N_{in} propagators at $p_{\text{in}} = 3$ and $N_{\text{ul}} \times \sigma^2(R_{\text{ul}})/\sigma^2(R_{\text{comb}})$ propagators at $p_{\text{pred}} = 4$. The cost with ML can be quantified by

$$\text{Cost} = \frac{N_{\text{in}} + N_{\text{BC}} + N_{\text{tr}}}{N_{\text{in}} + N_{\text{ul}} \left\langle \frac{\sigma^2(R_{\text{ul}})}{\sigma^2(R_{\text{comb}})} \right\rangle_t} \tag{15}$$

where N_{BC}, N_{tr} and N_{ul} are the numbers of propagator calculations (which represent the computational cost) needed to obtain the corresponding datasets (bias-correction, training and unlabeled), and N_{in} is that of the input data. The ratio $\sigma^2(R_{\text{ul}})/\sigma^2(R_{\text{comb}})$ is the scaling factor of the effective number of measurements we can obtain by employing the

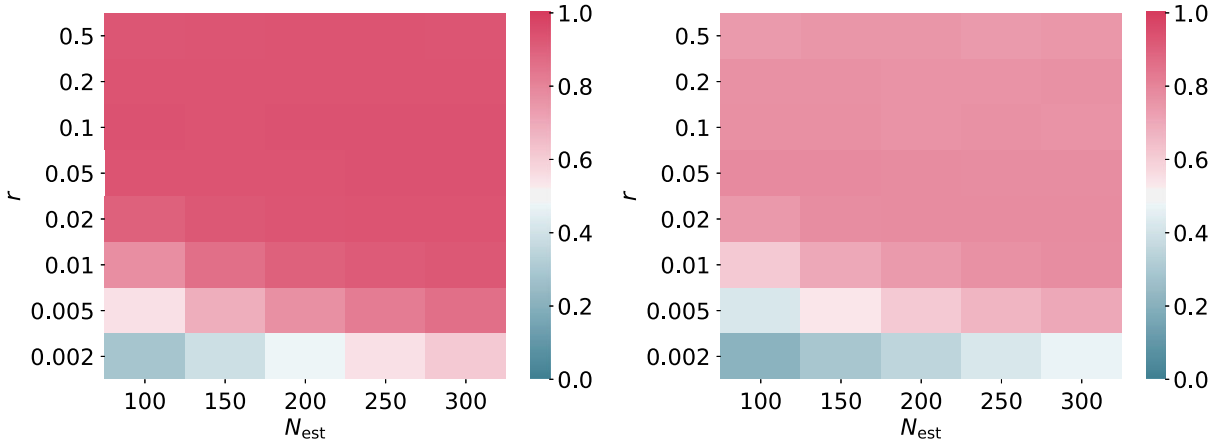

FIG. 8. Estimates of the fit variance $F_v(t=2)$ as a function of learning rate r and number of estimators N_{est} from the kaon quasi-PDF measurements at $p_{\text{pred}}=4$, $z_{\text{pred}}=4$, $t_{\text{sep}}=5$. $N_{\text{tr}}=400$ and $N_{\text{ul}}=1180$ are used. The left (right) plot shows the results from the z prediction (p prediction). The z prediction has a much better fit variance because of the good correlations between close links.

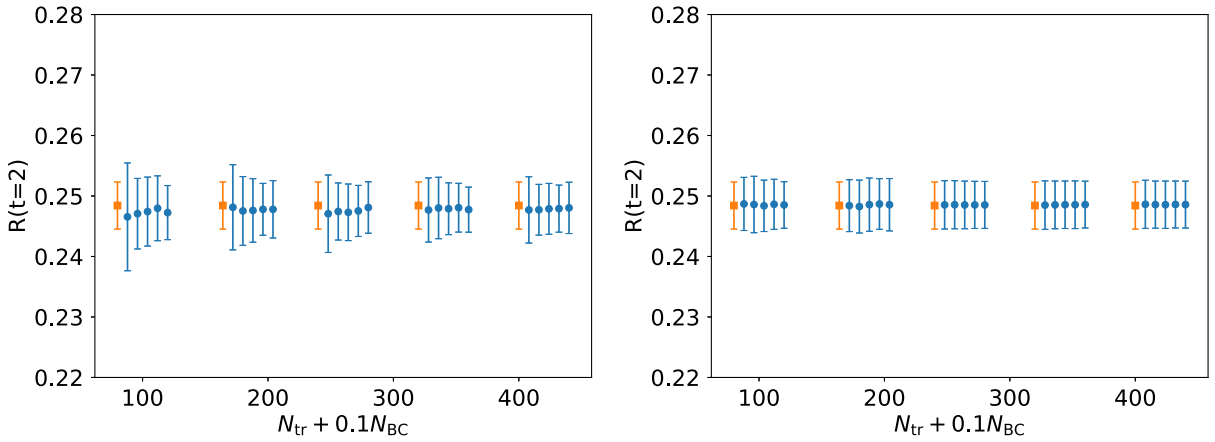

FIG. 9. Observations and predictions of the ratio $R(t=2)$ of kaon quasi-PDF correlators at $p_{\text{pred}}=4$, $z_{\text{pred}}=4$, $t_{\text{sep}}=5$ from input data at $p_{\text{in}}=4$, $z_{\text{in}}\in[0,3]$, $t_{\text{sep}}=5$. The left and right sides show the results from using the GBT and linear models, respectively. We use $N_{\text{est}}=150$, $r=0.1$ for the GBT model. The horizontal axis shows $N_{\text{tr}}+0.1N_{\text{BC}}$, and the number of unlabeled measurements is fixed to 1180. Points in blue are for predictions with bias correction, and orange for observations.

ML prediction, accounting for the increase of statistical error due to prediction error. We assume that the errors of observables scale as $1/\sqrt{N}$ as the number of measurements increases. For the cost estimate, we use an average value of the ratios over different insertion time slices. We calculate the $R_{\text{comb}}=C_{3\text{pt}}^{\text{comb}}/C_{2\text{pt}}$ here from each bootstrap sample by taking the weighted average of the measurements on labeled data and BC predictions on unlabeled data in each sample:

$$C_{3\text{pt}}^{\text{comb}}=\frac{\bar{C}_{3\text{pt}}^{\text{pred,BC}}/\sigma^2(C_{3\text{pt}}^{\text{pred,BC}})+\bar{C}_{3\text{pt}}^{\text{lb}}/\sigma^2(C_{3\text{pt}}^{\text{lb}})}{1/\sigma^2(C_{3\text{pt}}^{\text{pred,BC}})+1/\sigma^2(C_{3\text{pt}}^{\text{lb}})} \tag{16}$$

while the error $\sigma(R_{\text{comb}})$ is estimated from all bootstrap results. A smaller cost indicates higher prediction efficiency, so we vary N_{tr} and N_{BC} to find the optimal cost reduction, as shown in Fig. 10. By choosing optimal N_{tr} and N_{BC}, we can obtain about 20% reduction in computational cost.

Figure 11 shows this set of fitted results from both the z prediction and p prediction at $N_{\text{tr}}=240$, $N_{\text{BC}}=240$, while Table IV compares several sets of p and z predictions and observations. The last column of the table shows the fit quality.

We compare the predicted ratios for these models in Fig. 11. The z predictions are consistent with unlabeled data for both models, but the p predictions still need to be improved.

2. η_s quasi-PDF results

For the η_s quasi-PDF data, the meson operator is $\eta_s=\bar{s}\gamma_5 s$. The η_s data have better signals, and the correlations among η_s data show the same patterns as those of the kaon. Therefore, we select the same parameters for the model

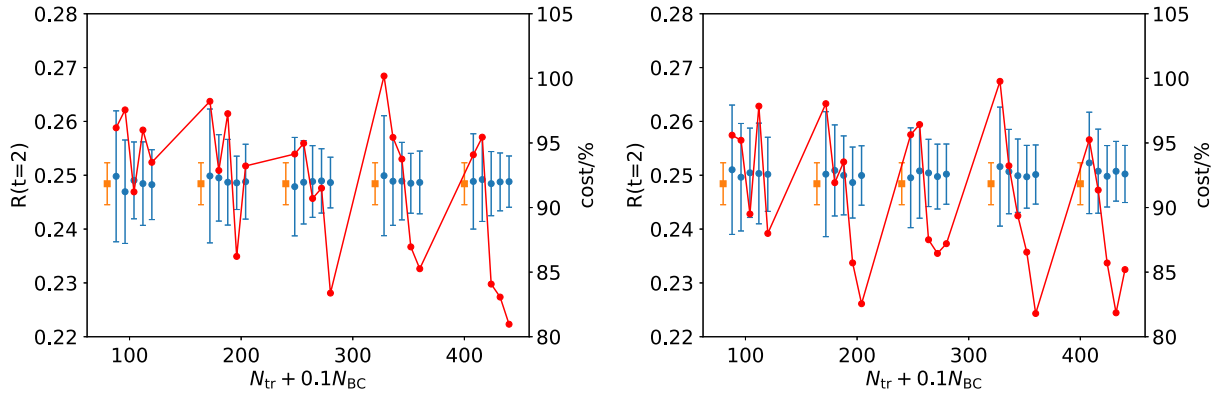

FIG. 10. Observations and predictions of the ratio $R(t = 2)$ of kaon quasi-PDF correlators at $p_{pred} = 4$, $z_{pred} = 4$, $t_{sep} = 5$ from input data at $p_{in} = 3$, $z_{in} = 4$, $t_{sep} = 5$. The red line shows the effective cost averaged on $R(t \in [1, 4])$. The left and right sides show the results from using the GBT and linear models, respectively. We use $N_{est} = 150$, $r = 0.1$ for the GBT model. The horizontal axis shows $N_{tr} + 0.1N_{BC}$, and the number of unlabeled measurements is fixed to 1180. Points in blue are for predictions with bias correction, and orange for observations.

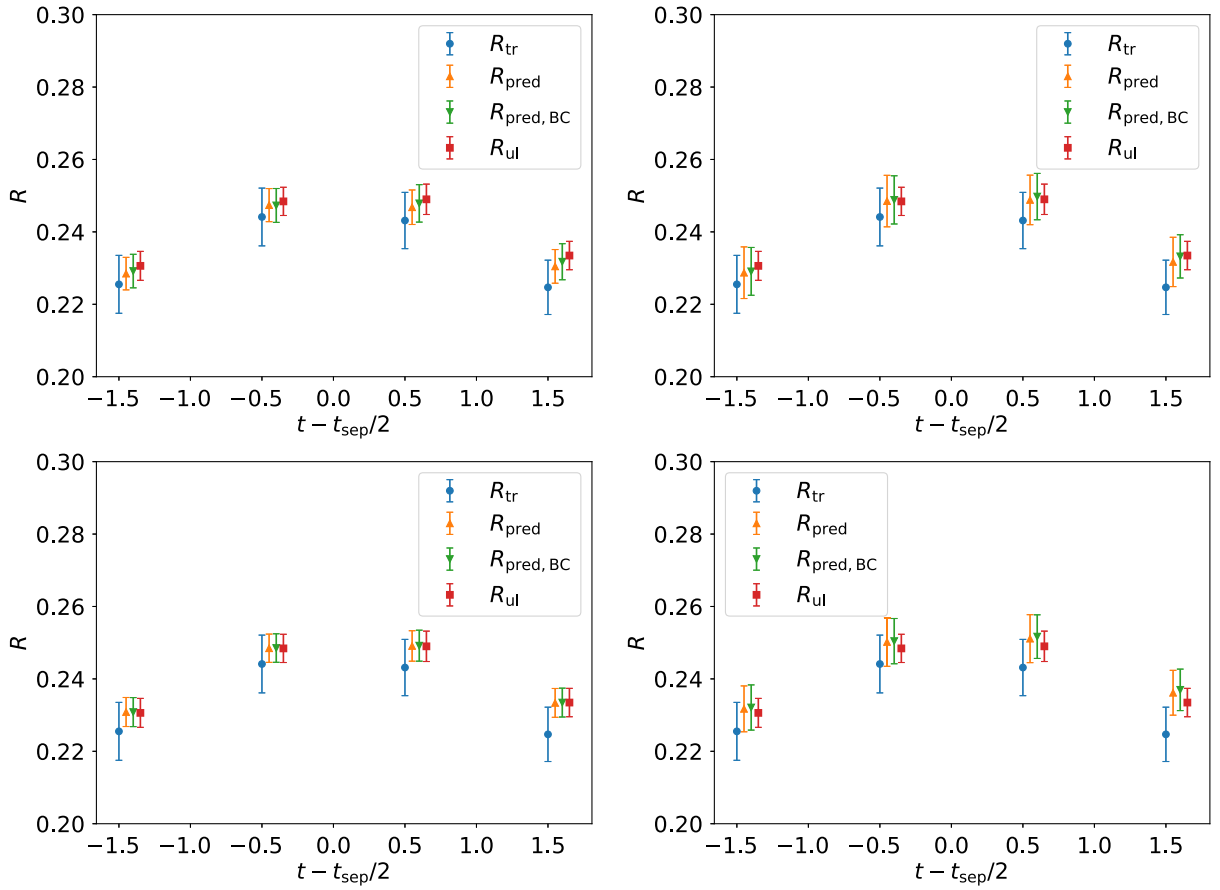

FIG. 11. The ratio $R(t)$ of the kaon quasi-PDF correlators at $z_{pred} = 4$, $p_{pred} = 4$ from direct measurements and the predictions of the three models. The top (bottom) row is GBT (linear) model with $N_{est} = 150$, $r = 0.1$. $N_{tr} = N_{BC} = 240$ and $N_{ul} = 1180$ are used. The left column uses $z_{in} \in [0, 3]$, $p_{in} = 4$ as inputs, while the right column uses $z_{in} = 4$, $p_{in} = 3$ as inputs. z predictions are better than p predictions.

TABLE IV. Observations and predictions of the ratio $R(t=2)$ of the kaon quasi-PDF correlators at $p_{\text{pred}}=4$, $z_{\text{pred}}=4$, $t_{\text{sep}}=5$ from different models and inputs. We use $N_{\text{est}}=150$, $r=0.1$ for the GBT model. The models are trained on 240 measurements with 240 bias-correction measurements, and then tested on 1180 unlabeled measurements. The linear model shows a better fit variance than GBT.

Type	Input	Method	R_{tr}	R_{pred}	$R_{\text{pred,BC}}$	R_{comb}	R_{ul}	F_v
p-pred	$p_{\text{in}}=3$, $z_{\text{in}}=4$	GBT	0.2441(70)	0.2430(60)	0.2439(56)	0.2435(51)	0.2471(35)	0.692(41)
		linear	0.2441(70)	0.2479(63)	0.2480(58)	0.2472(54)	0.2471(35)	0.772(29)
z-pred	$p_{\text{in}}=4$, $z_{\text{in}}\in[0,3]$	GBT	0.2441(70)	0.2458(40)	0.2455(41)	0.2456(32)	0.2471(35)	0.890(26)
		linear	0.2441(70)	0.2470(36)	0.2473(36)	0.2466(32)	0.2471(35)	0.998(1)

training, $N_{\text{est}}=150$, $r=0.1$, $N_{\text{tr}}=N_{\text{BC}}=240$, $N_{\text{ul}}=1180$. By comparing Fig. 12 and Fig. 8, we can see that the fit quality is slightly improved by the cleaner dataset. We infer that with more labeled kaon quasi-PDF data available for model training, the kaon model will show better performance as well. The predictions compared with observations are shown in Fig. 14. Both z predictions and p predictions are more precise compared to the kaon case. Figure 13 shows the cost on different $N_{\text{tr}}/N_{\text{BC}}$ set, the linear model shows a better optimal reduction than the kaon case. Overall, the cost reductions are 12%–18% at optimal choices of the sizes of the training and bias-correction datasets.

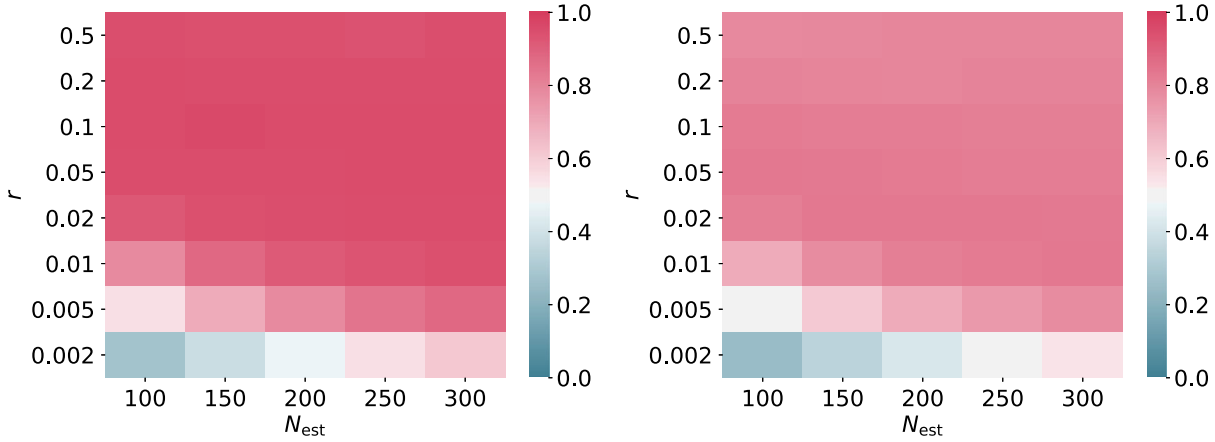

FIG. 12. Estimates of the fit variance F_v as a function of learning rate r and number of estimators N_{est} from the η_s quasi-PDF measurements at $p_{\text{pred}}=4$, $z_{\text{pred}}=4$, $t_{\text{sep}}=5$, and $t_{\text{pred}}=2$. $N_{\text{tr}}=400$ and $N_{\text{ul}}=1180$ are used. The left (right) side shows the results from the z prediction (p prediction). The performance is better than the model of kaon data.

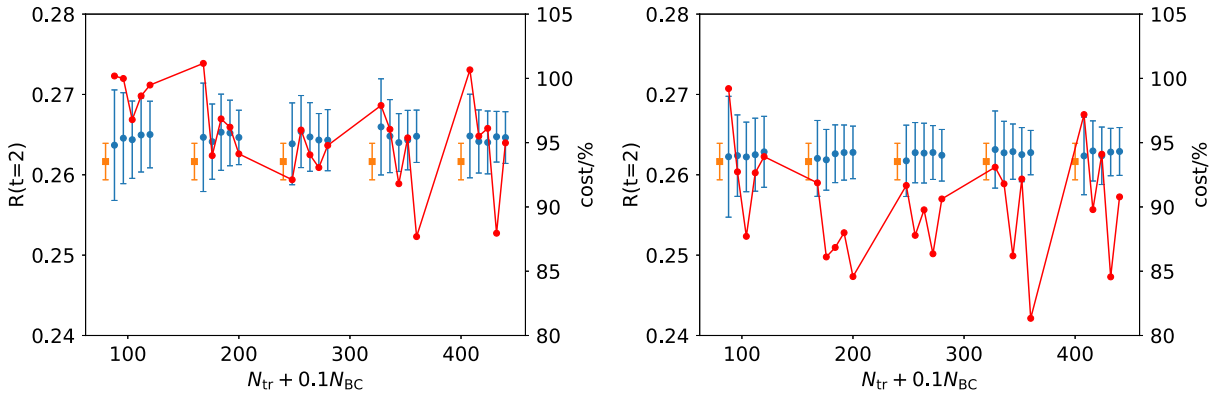

FIG. 13. Observations and predictions of the ratio $R(t=2)$ of η_s quasi-PDF correlators at $p_{\text{pred}}=4$, $z_{\text{pred}}=4$, $t_{\text{sep}}=5$ from input data at $p_{\text{in}}=3$, $z_{\text{in}}=4$, $t_{\text{sep}}=5$. Red line is the effective cost averaged on $R(t\in[1,4])$. The left and right sides show the results from using the GBT and linear models, respectively. We use $N_{\text{est}}=150$, $r=0.1$ for the GBT model. The horizontal axis shows $N_{\text{tr}}+0.1N_{\text{BC}}$, and the number of unlabeled measurements is fixed to 1180. Points in blue are for predictions with bias correction, and orange for observations.

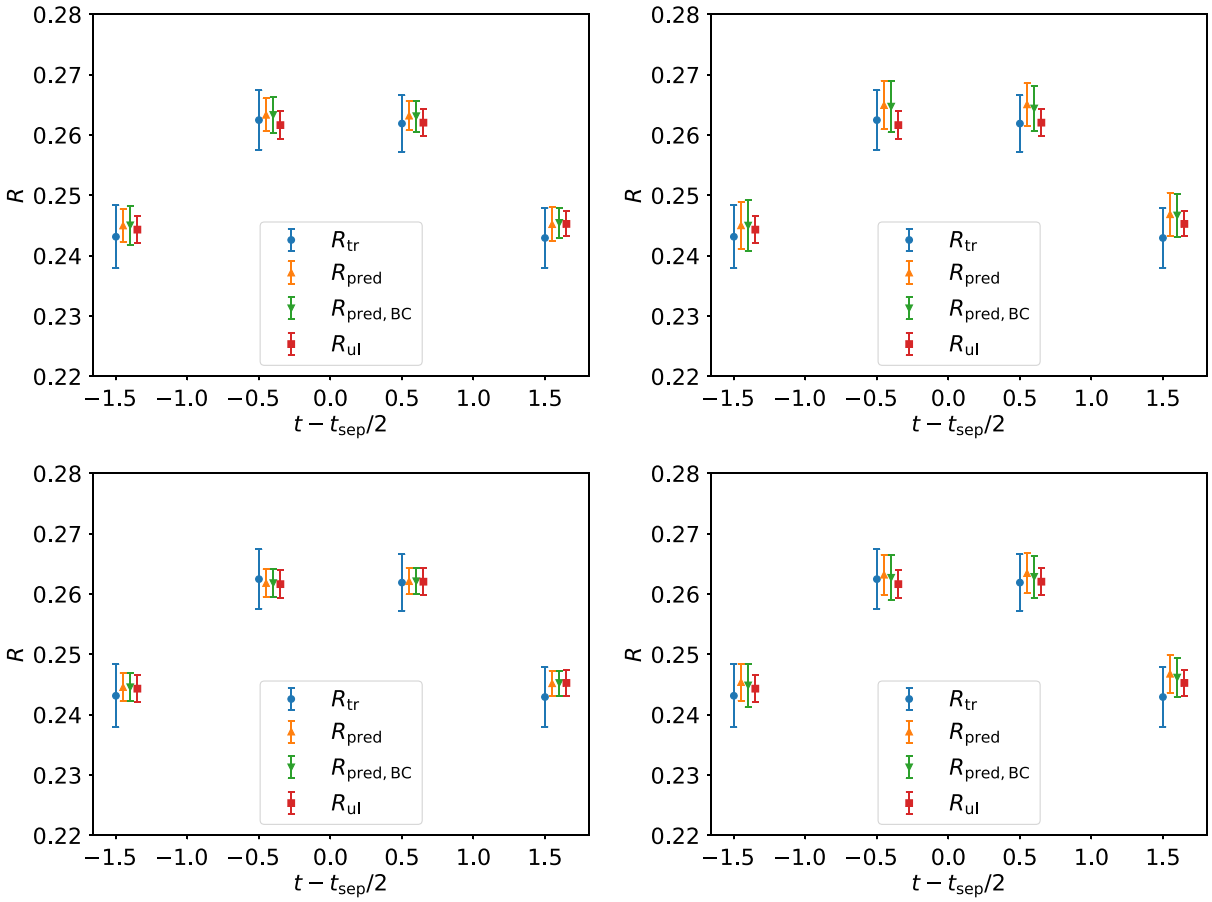

FIG. 14. The ratio $R(t)$ of η_s quasi-PDF correlators at $z_{\text{pred}} = 4$, $p_{\text{pred}} = 4$ from direct measurements and the predictions of the two models. The top (bottom) row is the GBT (linear) model. The left column uses $z_{\text{in}} \in [0, 3]$, $p_{\text{in}} = 4$, the right column uses $z_{\text{in}} = 4$, $p_{\text{in}} = 3$. The model performs better on these cleaner datasets. $N_{\text{tr}} = N_{\text{BC}} = 240$, $N_{\text{ul}} = 1180$ and $N_{\text{est}} = 150$, $r = 0.1$ are used in the model training.

C. Gluon quasi-PDF matrix elements

The gluon PDF contributes at next-to-leading order to deep inelastic scattering (DIS) cross sections, and it enters at leading order in jet production. Global fits have combined the data from both DIS and jet-production cross sections, and constraints on the gluon PDF from the experimental side are improving. However, on the theoretical side the gluon PDF is poorly known. PDF cannot be calculated using perturbative QCD. Recently, it has been found that they can be calculated directly in lattice QCD using large-momentum effective field theory. The gluon unpolarized quasi-PDF matrix elements are computed on the lattice using

$$C_{3\text{pt}}(z; t_{\text{sep}}, t) = \langle 0|\Gamma \int d^3y e^{-iy\cdot P} \chi(\vec{y}, t_{\text{sep}}) F_{\mu t}(z, t) \times \left[\prod_{x=0}^{z-1} U(x, t)\right] F^{z\mu}(0, t) \chi(\vec{0}, 0)|0\rangle, \quad (17)$$

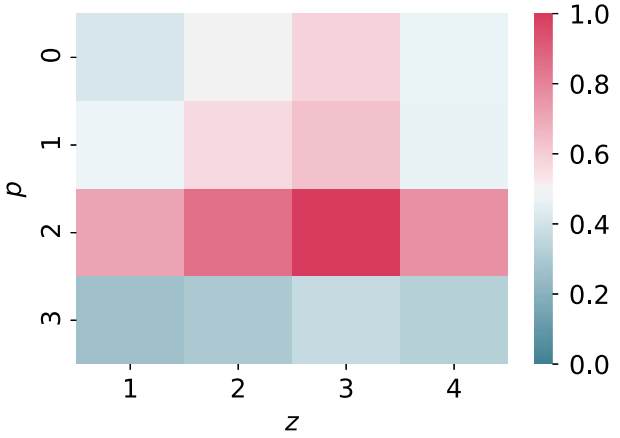

FIG. 15. Correlation coefficient between the three-point correlation function at $p_{\text{pred}} = 2$, $z_{\text{pred}} = 3$ and at various choices of p and z calculated using the overlap valence fermions. Different z at the same p cases show higher correlation than different p at the same z cases.

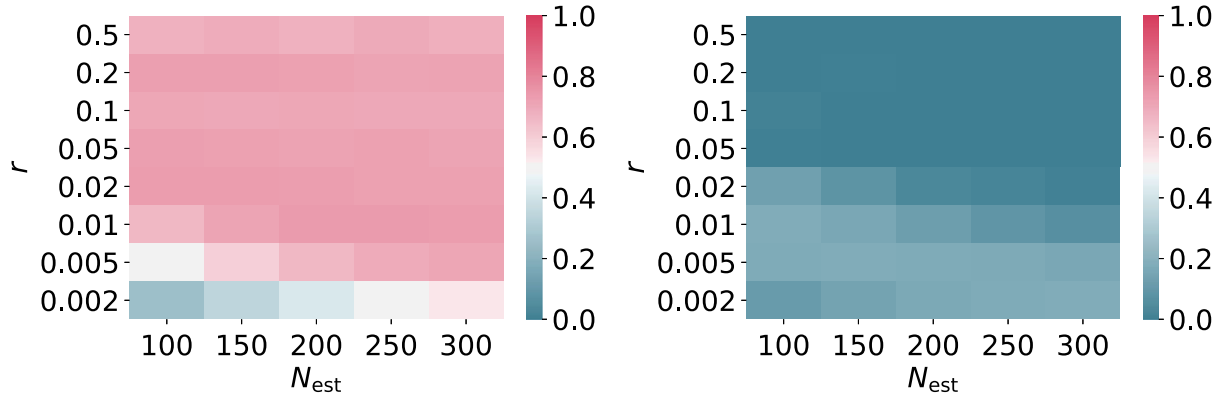

FIG. 16. Gluon-correlator ratio fit variance for the z prediction (left) and p prediction (right) for the overlap valence fermions at $p_{\text{pred}} = 5$, $z_{\text{pred}} = 3$, $t_{\text{sep}} = 8$ at $t = 4$ from $p_{\text{in}} = 5$, $z_{\text{in}} = 2$, $t_{\text{sep}} = 8$ and $p_{\text{in}} = 4$, $z_{\text{in}} = 3$, $t_{\text{sep}} = 8$ using $N_{\text{tr}} = 61440$, $N_{\text{BC}} = 61440$, and $N_{\text{ul}} = 81920$. Fit variance is closely related to the correlations between the input data and unlabeled data. z prediction works much better than p prediction.

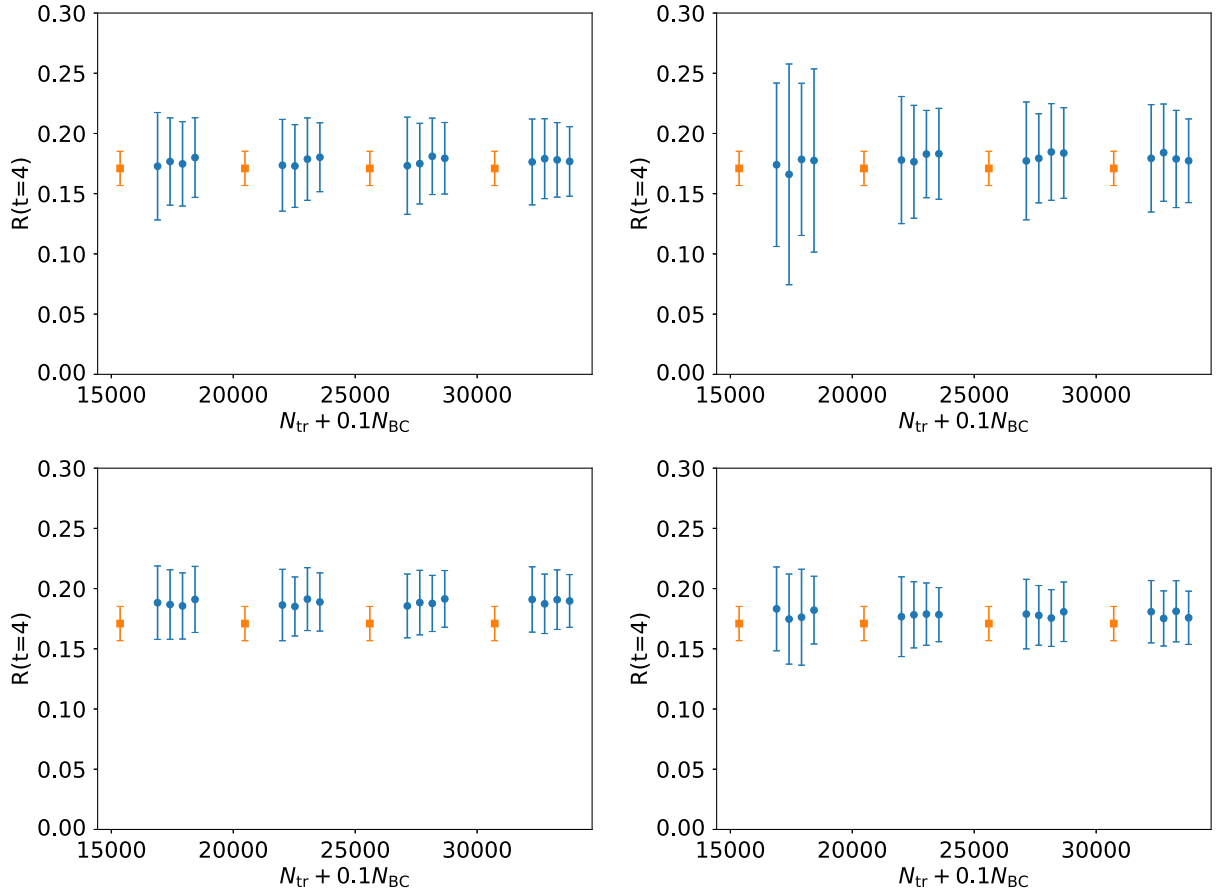

FIG. 17. The GBT (left) and linear-regressor (right) results. The observed p- and z-predicted gluon-correlator ratios are for the overlap valence fermions at $p_{\text{pred}} = 2$, $z_{\text{pred}} = 3$ at $t_{\text{sep}} = 8$ by using $r = 0.02$ and $N_{\text{est}} = 150$ for different counts of training data and bias-correction data. The horizontal axis is $N_{\text{tr}} + 0.1N_{\text{BC}}$, with $N_{\text{ul}} = 143360$ fixed. The blue points are predictions with bias correction for the unlabeled test data, and the orange points are observations for unlabeled test data.

$$C_{2\mathrm{pt}}(z;t_{\mathrm{sep}}) = \langle 0|\Gamma \int d^3y e^{-iy\cdot P}\chi(\vec{y},t_{\mathrm{sep}})\chi(\vec{0},0)|0\rangle, \quad (18)$$

where $C_{3\mathrm{pt}}$ is the three-point correlator, $C_{2\mathrm{pt}}$ is the two-point correlator, $\mathcal{O}(z,t)$ is the gluon operator, $\chi = \epsilon^{abc}[u^{aT}(x)i\gamma_4\gamma_2\gamma_5 d^b(x)]u^c(x)$ is the nucleon interpolation field, $\{a,b,c\}$ are color indices, $\Gamma = \frac{1}{2}(1+\gamma_4)$, and the field tensor $F_{\mu\nu}$ is defined by

$$F_{\mu\nu} = \frac{i}{8a^2 g}(\mathcal{P}_{[\mu,\nu]} + \mathcal{P}_{[\nu,-\mu]} + \mathcal{P}_{[-\mu,-\nu]} + \mathcal{P}_{[-\nu,\mu]}), \quad (19)$$

where the plaquette $\mathcal{P}_{\mu,\nu} = U_\mu(x)U_\nu(x+a\hat{\mu})U_\mu^\dagger(x+a\hat{\nu})U_\nu^\dagger(x)$ and $\mathcal{P}_{[\mu,\nu]} = \mathcal{P}_{\mu,\nu} - \mathcal{P}_{\nu,\mu}$. To improve the signal, we studied 1, 3, 5, 10 steps of HYP smearing [40] on the gluon momentum faction $\langle x\rangle_g$ in Eq. (3) of Ref. [44]. After applying the renormalization to the bare matrix elements, the results from different numbers of HYP-smearing steps are consistent with each other and with phenomenology results 0.42(2) within the uncertainties [44] with the exception of the 10-step. Therefore, we apply 5 steps of HYP smearing to the gluon quasi-PDF operators in this work. The ratio R of the three-point correlator to the two-point correlator follows the same definition as in Eq. (14).

We use valence overlap fermions on RBC gauge configurations [45] with $2+1$ flavors of domain-wall fermions (DWF), lattice volume $L^3 \times T = 24^3 \times 64$, lattice spacing $a = 0.1105(3)$ fm, and pion mass $m_\pi^{\mathrm{sea}} = 330$ MeV. We also compute clover valence quarks on the MILC $N_f = 2+1+1$ HISQ configurations [46] with $L^3 \times T = 32^3 \times 96$, $a = 0.0888(8)$ fm, and $m_\pi^{\mathrm{sea}} = 313$ MeV. For the nucleon two-point function, considering all time slices and independent smeared point sources, the number of measurements for the two-point functions is $200 \times 128 \times 8 = 204,800$ on the RBC-24I lattices and $300 \times 16 \times 6 = 28,800$ on the MILC-a09m310 lattices.

1. Predictions of the gluon correlators with the overlap valence fermions

To make z/p predictions on correlators based on smaller z/p values, we should first check the correlations among correlators with different momenta and link lengths.

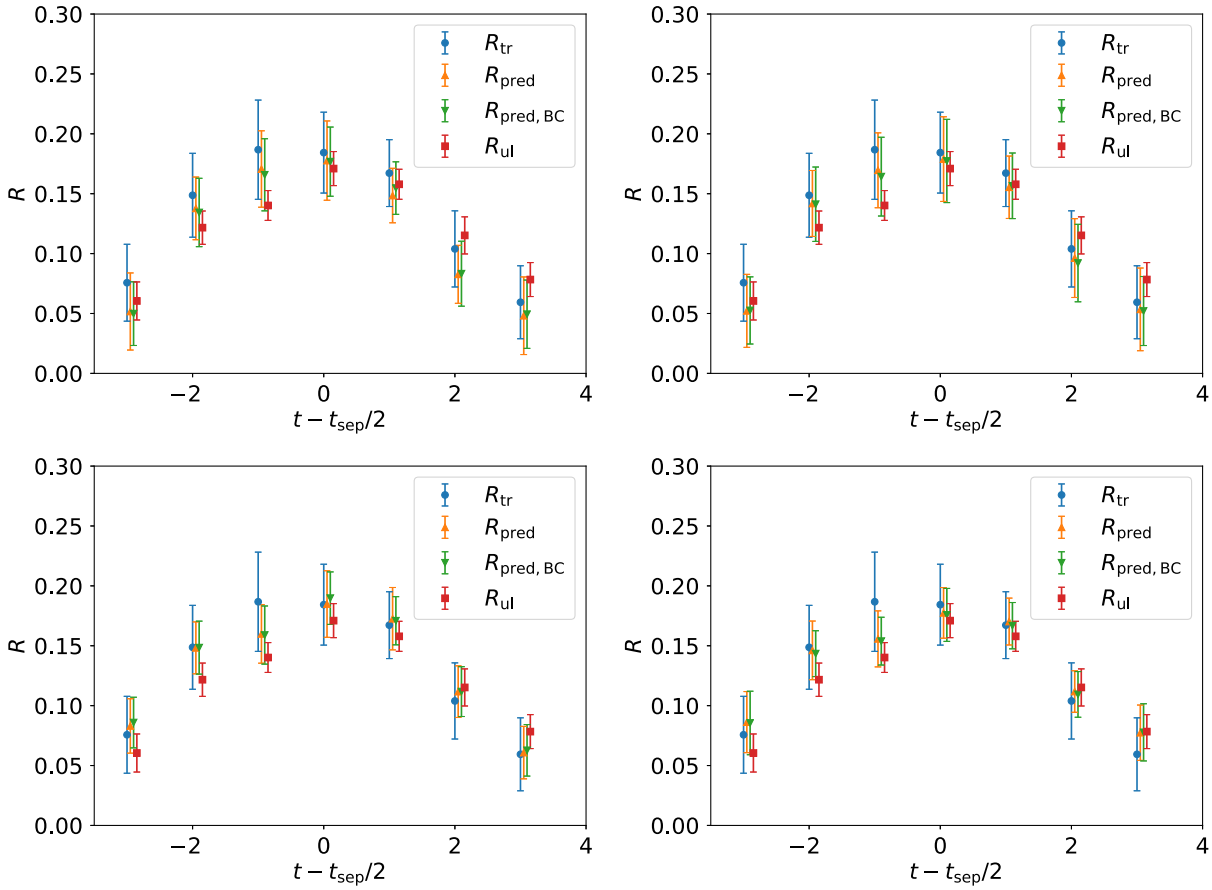

FIG. 18. The observed/predicted gluon-correlator $C_{3\mathrm{pt}}$ and $C_{2\mathrm{pt}}$ ratio of the overlap valence fermions lattice ensemble at $p_{\mathrm{pred}} = 2$, $z_{\mathrm{pred}} = 3$ from $p_{\mathrm{in}} = 1$, $z_{\mathrm{in}} = 3$ (upper) and $p_{\mathrm{pred}} = 2$, $z_{\mathrm{pred}} = 3$ from $p_{\mathrm{in}} = 2$, $z_{\mathrm{in}} = 2$ (lower) by using $N_{\mathrm{tr}} = 30720$, $N_{\mathrm{BC}} = 30720$, $N_{\mathrm{ul}} = 1433600$, $r = 0.02$, and $N_{\mathrm{est}} = 150$. The GBT and linear-regressor results are shown on the left and right, respectively. The predictions with bias correction do not improve much over the raw predictions.

TABLE V. Observations and predictions of gluon-correlator ratios for the overlap valence fermions observations and predictions at $p_{\mathrm{pred}} = 2$, $z_{\mathrm{pred}} = 3$, $t_{\mathrm{sep}} = 8$, $t = 4$ by using $N_{\mathrm{tr}} = 30720$, $N_{\mathrm{BC}} = 30720$, $N_{\mathrm{ul}} = 1433600$, $r = 0.02$, and $N_{\mathrm{est}} = 150$. For the z predictions, the linear model shows a better fit variance than GBT. The p predictions are bad for both models, because the correlations are poor, as shown in Fig. 18.

Type	Input	Method	R_{tr}	R_{pred}	$R_{\mathrm{pred,BC}}$	R_{ul}	F_v
p-pred	$p_{\mathrm{in}} = 1$, $z_{\mathrm{in}} = 3$	GBT	0.184(34)	0.178(33)	0.177(29)	0.171(14)	0.07(18)
		linear	0.184(34)	0.179(35)	0.177(35)	0.171(14)	−0.05(38)
z-pred	$p_{\mathrm{in}} = 2$, $z_{\mathrm{in}} = 2$	GBT	0.184(34)	0.185(28)	0.189(22)	0.171(14)	0.53(12)
		linear	0.184(34)	0.177(21)	0.176(22)	0.171(14)	0.665(79)

In Fig. 15, we show the correlations between the three-point correlation function at $p_{\mathrm{pred}} = 2$, $z_{\mathrm{pred}} = 3$ and the same three-point correlation functions at various choices of momenta $p_{\mathrm{in}} = \{0, 1, 3\}$ and link lengths $z_{\mathrm{in}} = \{1, 2, 3, 4\}$. The source-sink time separation is fixed to $t_{\mathrm{sep}} = 8$. We notice that the correlations between different momenta are weaker than the correlations between different link lengths in this case, which will result in a relatively low p-prediction fit variance as shown in Fig. 16.

The fit variances F_v from the p prediction and z prediction are shown in Fig. 16 with different learning rates in $\{0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5\}$ and different numbers of estimators in $\{100, 150, 200, 250, 300\}$. The target measurement is with $p_{\mathrm{in}} = [0, 1]$, $p_{\mathrm{pred}} = 2$, $z_{\mathrm{in}} = z_{\mathrm{pred}} = 3$, $t_{\mathrm{sep}} = 8$, and $t = 4$. We used both $C_{3\mathrm{pt}}$ and $C_{2\mathrm{pt}}$ for prediction. Thus, considering the F_v for z/p prediction shown in Fig. 16, we choose $r = 0.02$, $N_{\mathrm{est}} = 150$ as the parameter set we will use in further work.

For p prediction, we varied the number of training data and bias-correction data from 15360 to 30720, while keeping the number of unlabeled test data $N_{\mathrm{ul}} = 143360$ fixed, to compare their performance. The results are shown in Fig. 17. We will use $N_{\mathrm{tr}} = 30720$, $N_{\mathrm{BC}} = 30720$ in the following p/z prediction.

With the ML model parameters and the dataset we obtained from the overlap-fermion ensembles, we show the result of our prediction along with the observed datasets for both p_{pred} and z_{pred} predictions in Fig. 18. In the prediction, we can use any $p_{\mathrm{in}} < p_{\mathrm{pred}}$ or $z_{\mathrm{in}} < z_{\mathrm{pred}}$ for prediction. In Table V, two-point and three-point correlator data at $p_{\mathrm{in}} = 1$, $z_{\mathrm{in}} = 3$, $t_{\mathrm{sep}} = 8$ or $p_{\mathrm{in}} = 2$, $z_{\mathrm{in}} = 2$, $t_{\mathrm{sep}} = 8$ are used for predicting the $p_{\mathrm{pred}} = 2$, $z_{\mathrm{pred}} = 3$, $t_{\mathrm{sep}} = 8$ ratio. The data for insertion time $t = 4$ are shown. From the table we can see that the p predictions are bad for both models, because the correlations are weak, as shown in Fig. 18. The z predictions are better than p predictions, and the linear model performs better than GBT.

2. Predictions of the gluon correlators for clover valence fermions

We repeat the procedure we established from the overlap valence fermions for the clover fermions, checking the correlations among correlators with different momenta and link lengths. In Fig. 19, we show the correlations between the three-point correlation functions at $p_{\mathrm{pred}} = 5$, $z_{\mathrm{pred}} = 3$ at various values of $p_{\mathrm{in}} = \{0, 2, 4\}$, $z_{\mathrm{in}} = \{0, 1, 2, 3\}$. The source-sink time separation is fixed $t_{\mathrm{sep}} = 8$. The correlations between different momenta are much stronger than in the overlap case, which leads to a much higher p-prediction fit variance, as shown in Fig. 20. The reason that the correlations of clover fermion case are stronger than overlap-fermion case is the construction of the sources of proton correlator are different in two cases. In overlap fermion, we use grid spatial source which needs gauge averaging to get consistent correlators that are due to weak correlation properties. While the clover fermion does not have this kind of problem because of using one spatial location per time source.

We use the same fit-variance F_v estimation as in the overlap case. The target measurement is $p_{\mathrm{in}} = 4$, $p_{\mathrm{pred}} = 5$, $z_{\mathrm{in}} = 2$, $z_{\mathrm{pred}} = 3$, $t_{\mathrm{sep}} = 8$, and $t = 4$. We obtain $r = 0.2$, $N_{\mathrm{est}} = 200$ as the parameters we will use in the following process from Fig. 20. These two figures indicate stronger correlations between input and target data are needed to obtain good results for the fit variance.

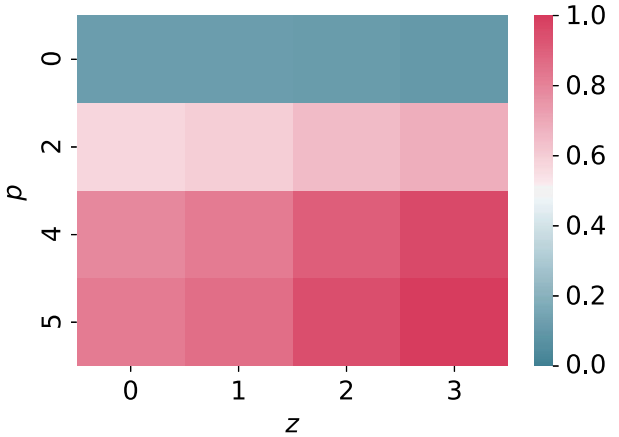

FIG. 19. Correlation coefficients between the three-point correlation functions at $p_{\mathrm{pred}} = 5$, $z_{\mathrm{pred}} = 3$ at various values of $p_{\mathrm{in}} = \{0, 2, 4\}$, $z_{\mathrm{in}} = \{0, 1, 2, 3\}$ calculated using the clover valence fermions. Different z at the same p cases show higher correlation than different p at the same z cases.

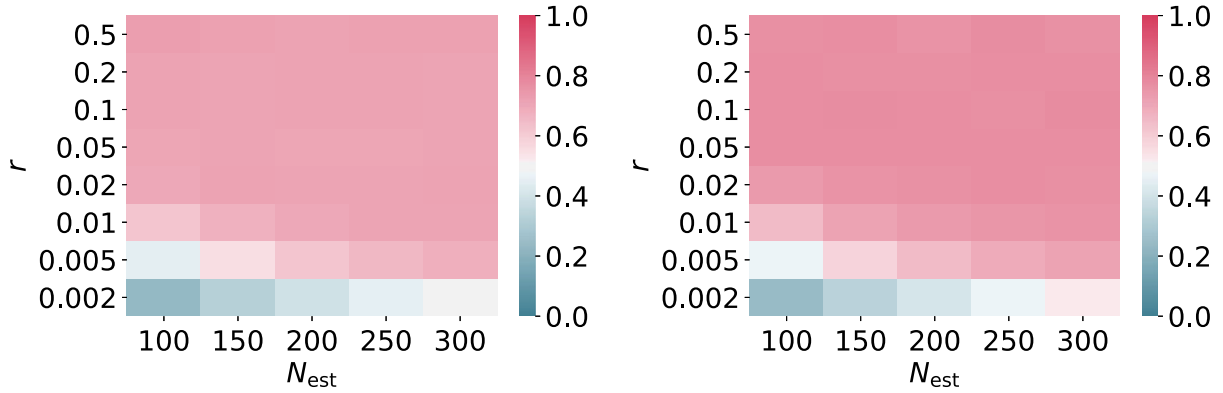

FIG. 20. Gluon-correlator ratio fit variance for the z prediction (left) and p prediction (right) for the clover valence fermions at $p_{\text{pred}} = 5$, $z_{\text{pred}} = 3$, $t_{\text{sep}} = 8$ at $t = 4$ from $p_{\text{in}} = 5$, $z_{\text{in}} = 2$, $t_{\text{sep}} = 8$ and $p_{\text{in}} = 4$, $z_{\text{in}} = 3$, $t_{\text{sep}} = 8$ using $N_{\text{tr}} = 2880$, $N_{\text{BC}} = 2880$, and $N_{\text{ul}} = 23040$. With a stronger correlation between input and target data, smaller learning rate and number of estimators are needed to have good fit-variance score.

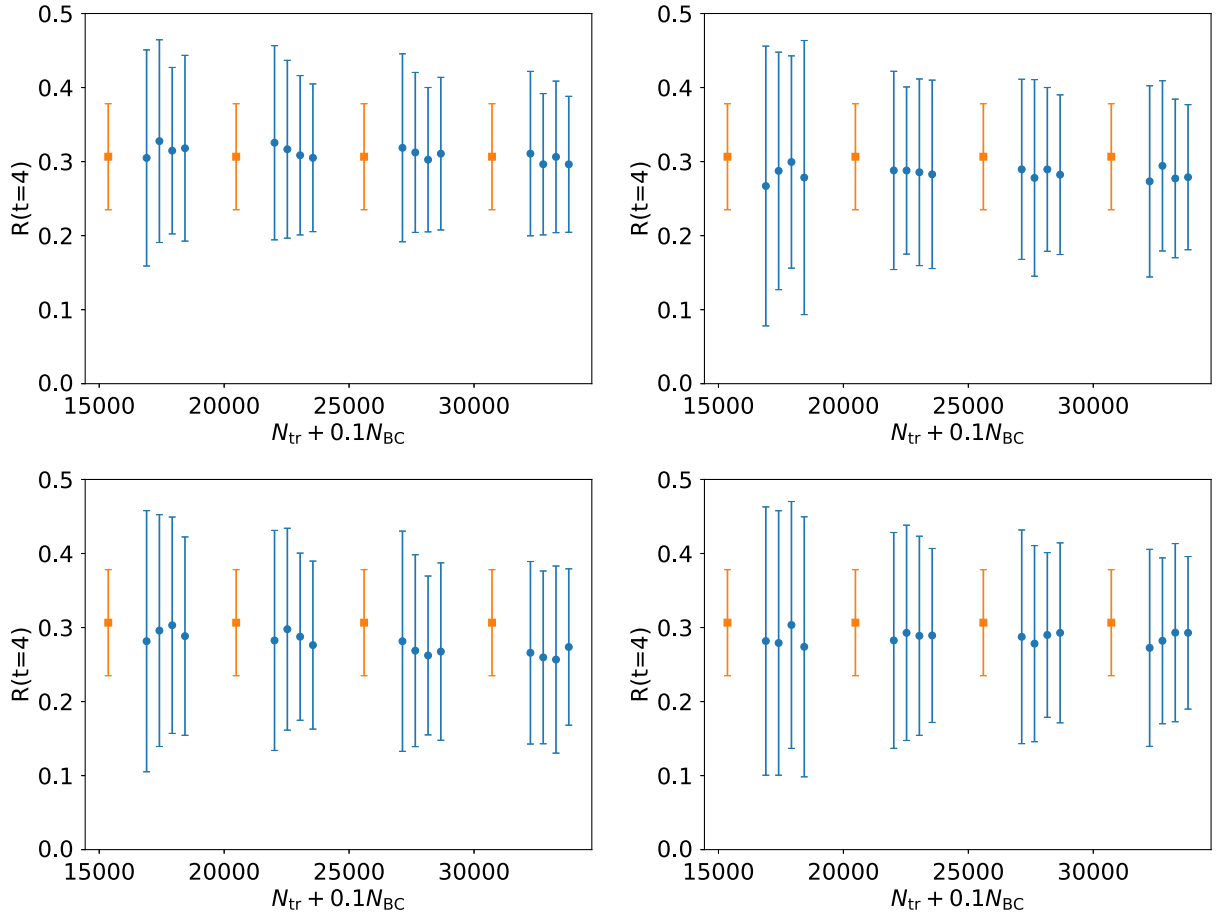

FIG. 21. The observed/predicted gluon-correlator $C_{3\text{pt}}$ and $C_{2\text{pt}}$ ratio of the clover valence fermions $p_{\text{pred}} = 5$, $z_{\text{pred}} = 3$ at $t_{\text{sep}} = 8$ by using $r = 0.2$ and $N_{\text{est}} = 200$ for different counts of training data and bias-correction data. The horizontal axis is $N_{\text{tr}} + 0.1N_{\text{BC}}$, with $N_{\text{ul}} = 23040$ fixed. The GBT and linear-regressor results are shown on the left and right, respectively. The blue points are predictions with bias correction for the unlabeled test data, and the orange points are observations for unlabeled test data.

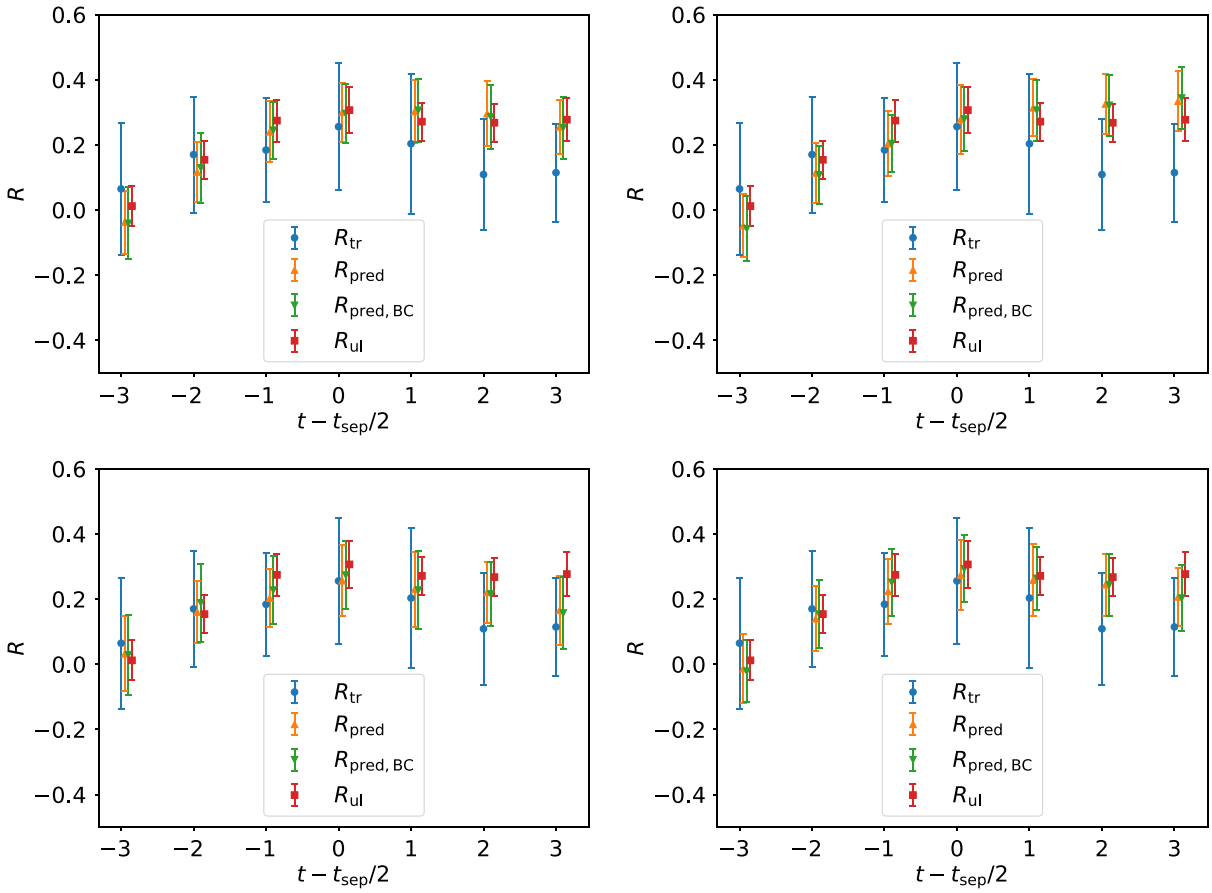

FIG. 22. The observed/predicted gluon-correlator ratios of the clover valence fermions at $p_{\text{pred}} = 5$, $z_{\text{pred}} = 3$ from $p_{\text{in}} = 4$, $z_{\text{in}} = 3$ (upper) and $p_{\text{pred}} = 5$, $z_{\text{pred}} = 3$ from $p_{\text{in}} = 5$, $z_{\text{in}} = 2$ (lower) by using $N_{\text{tr}} = 2880$, $N_{\text{BC}} = 2880$, $N_{\text{ul}} = 230400$, $r = 0.2$, and $N_{\text{est}} = 200$. The GBT and linear-regressor results are shown in the left and right columns, respectively. The predictions with bias correction do not much improve the raw predictions.

Again, to compare their performance we varied the number of training data and bias-correction data from 1440 to 2880, while keeping the number of unlabeled test data $N_{\text{ul}} = 23040$ fixed. The observed, p- and z-predicted gluon correlator $C_{3\text{pt}}$ and $C_{2\text{pt}}$ ratio of the clover valence fermions $p_{\text{pred}} = 5$, $z_{\text{pred}} = 3$ at $t_{\text{sep}} = 8$ are shown in Fig. 21. Comparing with these results, we will use $N_{\text{tr}} = 2880$, $N_{\text{BC}} = 2880$ in the following p and z predictions.

The observed/predicted gluon-correlator ratios of the clover valence fermions of the GBT and linear-regressor model at $p_{\text{pred}} = 5$, $z_{\text{pred}} = 3$ are shown in Fig. 22. The linear model gives a slightly better results. In Table VI, two-point and three-point correlator data at $p_{\text{in}} = 4$, $z_{\text{in}} = 2$, $t_{\text{sep}} = 8$ are used for predicting $p_{\text{pred}} = 2$, $z_{\text{pred}} = 3$, $t_{\text{sep}} = 8$ correlator. The data at insertion time $t = 4$ are shown. Compared with the overlap-fermion result in Table V, the fit

TABLE VI. Observations and predictions of gluon-correlator ratio for the clover valence fermions and predictions at $p_{\text{pred}} = 5$, $z_{\text{pred}} = 3$, $t_{\text{sep}} = 8$, $t = 4$ using $N_{\text{tr}} = 2880$, $N_{\text{BC}} = 2880$, $N_{\text{ul}} = 23040$, $r = 0.2$, and $N_{\text{est}} = 200$. The linear model shows a better fit variance than GBT.

Type	Input	Method	R_{tr}	R_{pred}	$R_{\text{pred,BC}}$	R_{ul}	F_v
p-pred	$p_{\text{in}} = 4$, $z_{\text{in}} = 3$	GBT	0.26(19)	0.300(91)	0.296(92)	0.307(72)	0.733(62)
		linear	0.26(19)	0.28(11)	0.279(98)	0.307(72)	0.845(60)
z-pred	$p_{\text{in}} = 5$, $z_{\text{in}} = 2$	GBT	0.26(19)	0.26(11)	0.27(11)	0.307(72)	0.704(62)
		linear	0.26(19)	0.27(11)	0.29(10)	0.307(72)	0.819(51)

TABLE VII. Summary of fit variance F_v for all the cases we investigate. The larger value of F_v indicates the better predictions, and a perfect prediction yields $F_v = 1.0$. In general, z predictions work better than p predictions, and linear model shows better performance than GBT on our dataset.

Type	Method	η_sDA	Kaon PDF	Overlap Gluon PDF	Clover Gluon PDF
z-pred	GBT	0.62(14)	0.890(26)	0.53(12)	0.704(62)
	linear	0.99935(40)	0.998(1)	0.665(79)	0.819(51)
p-pred	GBT	0.50(13)	0.692(14)	0.07(18)	0.733(62)
	linear	0.911(43)	0.772(29)	−0.05(38)	0.845(60)

variance is much higher, due to the input data having stronger correlations with the target data.

IV. SUMMARY

In this article, we applied the ML technique to quasi-DA and quasi-PDF correlators. Using both GBT model and linear model, we tried to predict the $C_{2\mathrm{pt}}$ for meson quasi-DAs and the $C_{3\mathrm{pt}}$ for meson and gluon quasi-PDFs at larger momenta and link lengths, which are noisier and need more computational resources. By predicting from the computationally less expensive data, we are able to reduce the computational cost. Systematic uncertainties from the ML prediction errors are converted to the statistical uncertainties by using the bias-correction procedure. With the full bootstrap resampling, we effectively estimated and compared the errors of different model predictions.

Table VII summarizes the best fit variances F_v of all predictions we investigated. It is observed that for meson datasets, the data from different links are more correlated than those of different momenta. Consequently, the z predictions for both models work much better than p predictions. The ML approach on the z prediction of quasi-DAs and meson quasi-PDFs is very precise, while the p predictions and the predictions for gluon quasi-PDFs show relatively worse precision. By comparing two ML regression models, we find that the linear model is preferred on cleaner datasets when the correlations between input data and target data are good enough, such as the z prediction of meson-DAs and meson PDFs. On the other hand, the GBT model is more robust to noisy and less-obviously correlated inputs. For the p prediction of meson quasi-PDFs, both models are able to give a computational cost reduction of 16%.

ACKNOWLEDGMENTS

We thank the MILC Collaboration and RBC Collaboration for sharing the lattices used to perform this study. The LQCD calculations were performed using the Chroma software suite [47]. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 through Energy Research Computing Allocation Process (ERCAP); facilities of the USQCD Collaboration, which are funded by the Office of Science of the U.S. Department of Energy, and supported in part by Michigan State University through computational resources provided by the Institute for Cyber-Enabled Research. R. L., Z. F., H. L. and R. Z. are supported by the U.S. National Science Foundation under Grant No. PHY 1653405 "CAREER: Constraining Parton Distribution Functions for New-Physics Searches." B. Y. is supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics under Contract No. 89233218CNA000001 and by the Los Alamos National Laboratory (LANL) LDRD program. B. Y. also acknowledges support from the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research and Office of Nuclear Physics, Scientific Discovery through Advanced Computing (SciDAC) program.

[1] S. Aoki *et al.* (Flavour Lattice Averaging Group), arXiv: 1902.08191.
[2] H.-W. Lin *et al.*, Prog. Part. Nucl. Phys. **100**, 107 (2018).
[3] W. Detmold, W. Melnitchouk, and A. W. Thomas, Eur. Phys. J. Direct **3**, 13 (2001).
[4] X. Ji, Phys. Rev. Lett. **110**, 262002 (2013).
[5] W. Detmold, R. G. Edwards, J. J. Dudek, M. Engelhardt, H.-W. Lin, S. Meinel, K. Orginos, and P. Shanahan (USQCD Collaboration), Eur. Phys. J. A **55**, 193 (2019).
[6] X. Ji, J.-H. Zhang, and Y. Zhao, Phys. Rev. Lett. **111**, 112002 (2013).
[7] Y. Hatta, X. Ji, and Y. Zhao, Phys. Rev. D **89**, 085030 (2014).
[8] X. Ji, Sci. China Phys. Mech. Astron. **57**, 1407 (2014).

[9] X. Ji, J.-H. Zhang, and Y. Zhao, Phys. Lett. B **743**, 180 (2015).
[10] H.-W. Lin, J.-W. Chen, S. D. Cohen, and X. Ji, Phys. Rev. D **91**, 054510 (2015).
[11] C. Alexandrou, K. Cichy, V. Drach, E. Garcia-Ramos, K. Hadjiyiannakou, K. Jansen, F. Steffens, and C. Wiese, Phys. Rev. D **92**, 014502 (2015).
[12] J.-W. Chen, S. D. Cohen, X. Ji, H.-W. Lin, and J.-H. Zhang, Nucl. Phys. **B911**, 246 (2016).
[13] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, F. Steffens, and C. Wiese, Phys. Rev. D **96**, 014513 (2017).
[14] H.-W. Lin, J.-W. Chen, T. Ishikawa, and J.-H. Zhang (LP3 Collaboration), Phys. Rev. D **98**, 054504 (2018).
[15] J.-W. Chen, H.-W. Lin, and J.-H. Zhang, arXiv:1904.12376.
[16] J.-H. Zhang, J.-W. Chen, X. Ji, L. Jin, and H.-W. Lin, Phys. Rev. D **95**, 094514 (2017).
[17] J.-W. Chen, L. Jin, H.-W. Lin, A. Schäfer, P. Sun, Y.-B. Yang, J.-H. Zhang, R. Zhang, and Y. Zhao, Nucl. Phys. **B939**, 429 (2019)
[18] J.-W. Chen, T. Ishikawa, L. Jin, H.-W. Lin, Y.-B. Yang, J.-H. Zhang, and Y. Zhao, Phys. Rev. D **97**, 014505 (2018).
[19] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, H. Panagopoulos, and F. Steffens, Nucl. Phys. **B923**, 394 (2017).
[20] M. Constantinou and H. Panagopoulos, Phys. Rev. D **96**, 054506 (2017).
[21] J. Green, K. Jansen, and F. Steffens, Phys. Rev. Lett. **121**, 022004 (2018).
[22] J.-W. Chen, T. Ishikawa, L. Jin, H.-W. Lin, Y.-B. Yang, J.-H. Zhang, and Y. Zhao, Chin. Phys. C **43**, 103101 (2019).
[23] J.-W. Chen, T. Ishikawa, L. Jin, H.-W. Lin, A. Schäfer, Y.-B. Yang, J.-H. Zhang, and Y. Zhao, Sci. China Phys. Mech. Astron. **62**, 991021 (2019).
[24] H.-W. Lin and R. Zhang, Phys. Rev. D **100**, 074502 (2019).
[25] B. Yoon, T. Bhattacharya, and R. Gupta, Phys. Rev. D **100**, 014504 (2019).
[26] J. Karpie, K. Orginos, A. Rothkopf, and S. Zafeiropoulos, J. High Energy Phys. 04 (2019) 057.
[27] K. Orginos, A. Radyushkin, J. Karpie, and S. Zafeiropoulos, Phys. Rev. D **96**, 094503 (2017).
[28] A. V. Radyushkin, Phys. Rev. D **96**, 034025 (2017).
[29] B. Joó, J. Karpie, K. Orginos, A. V. Radyushkin, D. G. Richards, R. S. Sufian, and S. Zafeiropoulos, Phys. Rev. D **100**, 114512 (2019).
[30] B. Joó, J. Karpie, K. Orginos, A. Radyushkin, D. Richards, and S. Zafeiropoulos, J. High Energy Phys. 12 (2019) 081.
[31] A. Natekin and A. Knoll, Front. Neurorobotics **7**, 21 (2013).
[32] 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, J. Mach. Learn. Res. **12**, 2825 (2011).
[33] J. H. Friedman, Ann. Stat. **29**, 1189 (2001).
[34] J. H. Friedman, Computational Statistics and Data Analysis **38**, 367 (2002).
[35] M. Beneke, G. Buchalla, M. Neubert, and C. T. Sachrajda, Phys. Rev. Lett. **83**, 1914 (1999).
[36] M. Beneke, G. Buchalla, M. Neubert, and C. T. Sachrajda, Nucl. Phys. **B606**, 245 (2001).
[37] X. Ji, A. Schäfer, X. Xiong, and J.-H. Zhang, Phys. Rev. D **92**, 014039 (2015).
[38] E. Follana, Q. Mason, C. Davies, K. Hornbostel, G. P. Lepage, J. Shigemitsu, H. Trottier, and K. Wong (HPQCD and UKQCD Collaboration), Phys. Rev. D **75**, 054502 (2007).
[39] A. Bazavov *et al.* (MILC Collaboration), Phys. Rev. D **87**, 054505 (2013).
[40] A. Hasenfratz and F. Knechtli, Phys. Rev. D **64**, 034504 (2001).
[41] G. S. Bali, B. Lang, B. U. Musch, and A. Schäfer, Phys. Rev. D **93**, 094515 (2016).
[42] A. C. Aguilar *et al.*, Eur. Phys. J. A **55**, 190 (2019).
[43] J.-W. Chen, L. Jin, H.-W. Lin, Y.-S. Liu, A. Schäfer, Y.-B. Yang, J.-H. Zhang, and Y. Zhao, Phys. Rev. D **100**, 034505 (2019).
[44] Z.-Y. Fan, Y.-B. Yang, A. Anthony, H.-W. Lin, and K.-F. Liu, Phys. Rev. Lett. **121**, 242001 (2018).
[45] T. Blum, P. Boyle, N. Christ, J. Frison, N. Garron, R. Hudspith, T. Izubuchi, T. Janowski, C. Jung, A. Jüttner *et al.*, Phys. Rev. D **93**, 074505 (2016).
[46] A. Bazavov, C. Bernard, J. Komijani, C. DeTar, L. Levkova, W. Freeman, S. Gottlieb, R. Zhou, U. Heller, J. Hetrick *et al.*, Phys. Rev. D **87**, 054505 (2013).
[47] R. G. Edwards, B. Joo (SciDAC, LHPC, and UKQCD Collaborations), Nucl. Phys. B, Proc. Suppl. **140**, 832 (2005).