UC Berkeley
UC Berkeley Previously Published Works

Title

New metric improving Bayesian calibration of a multistage approach studying hadron and inclusive jet suppression

Permalink

https://escholarship.org/uc/item/9ss4b144

Journal

Physical Review C, 109(6)

ISSN

2469-9985

Authors

Fan, W

Vujanovic, G

Bass, SA

et al.

Publication Date

2024-06-01

DOI

10.1103/physrevc.109.064903

Copyright Information

Peer reviewed

New metric improving Bayesian calibration of a multistage approach studying hadron and inclusive jet suppression

W. Fan,[1,*] G. Vujanovic,[2,3,†] S. A. Bass,[1] A. Angerami,[4] R. Arora,[5] S. Cao,[6,2] Y. Chen,[7,8] T. Dai,[1] L. Du,[9] R. Ehlers,[10,11] H. Elfner,[12,13,14] R. J. Fries,[15,16] C. Gale,[9] Y. He,[17,18] M. Heffernan,[9] U. Heinz,[19] B. V. Jacak,[20,21] P. M. Jacobs,[20,21] S. Jeon,[9] Y. Ji,[22] L. Kasper,[23] M. Kordell, II,[15,16] A. Kumar,[9,2] J. Latessa,[24] Y.-J. Lee,[7,8] R. Lemmon,[25] D. Liyanage,[19] A. Lopez,[26] M. Luzum,[26] A. Majumder,[2] S. Mak,[22] A. Mankolli,[23] C. Martin,[10] H. Mehryar,[24] T. Mengel,[10] J. Mulligan,[20,21] C. Nattrass,[10] J. Norman,[27] J.-F. Paquet,[1] C. Parker,[15,16] J. H. Putschke,[2] G. Roland,[7,8] B. Schenke,[28] L. Schwiebert,[24] A. Sengupta,[15,16] C. Shen,[2,29] C. Sirimanna,[2] D. Soeder,[1] R. A. Soltz,[2,4] I. Soudi,[2] M. Strickland,[30] Y. Tachibana,[31,2] J. Velkovska,[23] X.-N. Wang,[32,20,21] and W. Zhao[2]

(JETSCAPE Collaboration)

[1]*Department of Physics, Duke University, Durham, North Carolina 27708, USA*
[2]*Department of Physics and Astronomy, Wayne State University, Detroit, Michigan 48201, USA*
[3]*Department of Physics, University of Regina, Regina, SK S4S 0A2, Canada*
[4]*Lawrence Livermore National Laboratory, Livermore, California 94550, USA*
[5]*Research Computing Group, University Technology Solutions, The University of Texas at San Antonio, San Antonio, Texas 78249, USA*
[6]*Institute of Frontier and Interdisciplinary Science, Shandong University, Qingdao, Shandong 266237, China*
[7]*Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[8]*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[9]*Department of Physics, McGill University, Montréal QC H3A 2T8, Canada*
[10]*Department of Physics and Astronomy, University of Tennessee, Knoxville, Tennessee 37996, USA*
[11]*Physics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, USA*
[12]*GSI Helmholtzzentrum für Schwerionenforschung, 64291 Darmstadt, Germany*
[13]*Institute for Theoretical Physics, Goethe University, 60438 Frankfurt am Main, Germany*
[14]*Frankfurt Institute for Advanced Studies, 60438 Frankfurt am Main, Germany*
[15]*Cyclotron Institute, Texas A&M University, College Station, Texas 77843, USA*
[16]*Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77843, USA*
[17]*Guangdong Provincial Key Laboratory of Nuclear Science, Institute of Quantum Matter, South China Normal University, Guangzhou 510006, China*
[18]*Guangdong-Hong Kong Joint Laboratory of Quantum Matter, Southern Nuclear Science Computing Center, South China Normal University, Guangzhou 510006, China*
[19]*Department of Physics, The Ohio State University, Columbus, Ohio 43210, USA*
[20]*Department of Physics, University of California, Berkeley, California 94270, USA*
[21]*Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94270, USA*
[22]*Department of Statistical Science, Duke University, Durham, North Carolina 27708, USA*
[23]*Department of Physics and Astronomy, Vanderbilt University, Nashville, Tennessee 37235, USA*
[24]*Department of Computer Science, Wayne State University, Detroit, Michigan 48202, USA*
[25]*Daresbury Laboratory, Daresbury, Warrington, Cheshire WA4 4AD, United Kingdom*
[26]*Instituto de Fìsica, Universidade de São Paulo, C.P. 66318, 05315-970 São Paulo, SP, Brazil*
[27]*Oliver Lodge Laboratory, University of Liverpool, Liverpool L69 7ZE, United Kingdom*
[28]*Physics Department, Brookhaven National Laboratory, Upton, New York 11973, USA*
[29]*RIKEN BNL Research Center, Brookhaven National Laboratory, Upton, New York 11973, USA*
[30]*Department of Physics, Kent State University, Kent, Ohio 44242, USA*
[31]*Akita International University, Yuwa, Akita-City 010-1292, Japan*
[32]*Key Laboratory of Quark and Lepton Physics (MOE) and Institute of Particle Physics, Central China Normal University, Wuhan 430079, China*

---

*Corresponding author: wenkai.fan@duke.edu

†Corresponding author: gojko.vujanovic@uregina.ca

We study parton energy-momentum exchange with the quark gluon plasma (QGP) within a multistage approach composed of in-medium Dokshitzer-Gribov-Lipatov-Altarelli-Parisi evolution at high virtuality, and (linearized) Boltzmann transport formalism at lower virtuality. This multistage simulation is then calibrated in comparison with high-p_T charged hadrons, D mesons, and the inclusive jet nuclear modification factors, using Bayesian model-to-data comparison, to extract the virtuality-dependent transverse momentum broadening transport coefficient $\hat{q}$. To facilitate this undertaking, we develop a quantitative metric for validating the Bayesian workflow, which is used to analyze the sensitivity of various model parameters to individual observables. The usefulness of this new metric in improving Bayesian model emulation is shown to be highly beneficial for future such analyses.

I. INTRODUCTION

Colliding QCD bound states at relativistic energies can lead to the excitation of its fundamental degrees of freedom known as partons. Some of these are highly energetic and generate a spray of particles known as a jet. While proton-proton collisions allow the study of the fragmentation of partons and their subsequent decay into hadronic bound states in the vacuum [1–4], the showering of jets in high-energy heavy-ion collisions inherently includes interactions with the hot and dense nuclear medium known as the quark gluon plasma (QGP) [5–19]. The modification of jets in nucleus-nucleus (*A-A*) collisions compared with jets in proton-proton (*p-p*) collisions is referred to as jet quenching. As jets are generated early in heavy-ion collisions, their partonic content samples the properties of the QGP throughout its evolution and can be described by perturbative QCD, which provides us with an established approach to study them (see Refs. [20–23] and references therein). Given that parton lifetime (or virtuality) plays an important role in how partons interact with the QGP, and no single Monte Carlo approach for parton energy loss has been devised that describes all virtualities at once, a multiscale approach is the preferred option, as explored herein.

The interaction between an energetic parton and the QGP medium is divided into two regimes determined by the virtuality ($t = E^2 - |\mathbf{p}|^2 = Q^2$) of the parton. Parton evolution in the high-virtuality ($t^2 \gg \hat{q}E$) regime is described by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution modified to include nuclear medium effects [24–26] based on the higher-twist formalism [19,27,28]. The jet-medium interaction is encapsulated in various transport coefficients governing the exchange of energy-momentum between jet partons and those in the QGP: The transverse momentum diffusion of jet partons in the QGP is encoded in $\hat{q}$ [29,30], while longitudinal transfers (not used herein) are contained within $\hat{e}$ and $\hat{e}_2$ [31,32].

With every split in the DGLAP or vacuum like stage the virtuality undergoes a reduction. Once the virtuality reaches $t_s \sim \sqrt{\hat{q}E}$, the switching virtuality (t_s) between the DGLAP and transport stages, multiple scatterings from the medium maintain the virtuality at the scale $\sqrt{\hat{q}E}$. In this effort, t_s is treated as a free parameter, tuned using Bayesian methods. Below t_s, the virtuality scale is considered close to that of the medium in our simulations, and thus rate equations [15,33,34] become an apt description of parton evolution in the QGP. Finally, once partons reach low energies and (and low

virtualities), hadronization occurs via PYTHIA's string fragmentation present within the JETSCAPE framework [35,36]. The high- and low-virtuality parton energy-loss regimes are incorporated inside the JETSCAPE framework, which provides a model-agnostic communication layer among jet energy-loss models, allowing for a multistage event generator to be created, as is the case in this study. Beyond jet-medium interactions, the model-agnostic nature of the JETSCAPE framework has been used to study and interpret simulations of the nuclear bulk medium itself [37,38], while the framework also provides a dynamical communication layer between simulation of the nuclear medium and the energy-loss calculations. Finally, the JETSCAPE framework has developed a set of Bayesian tools to constrain jet-quenching calculations in heavy-ion collisions [39].

This study focuses on improving Bayesian tools to constrain the nuclear modification factor of inclusive jets, light hadrons and D mesons (a discussion on particulars of heavy flavor production can be found in Ref. [40]). The combination of highly dimensional parameter spaces explored in high-energy heavy-ion collisions simulations (see, e.g., Refs. [37,39,41–45]) together with the high computational requirements to generate realistic simulations, necessitates the use of model emulators to accelerate the Markov Chain Monte Carlo computations employed when obtaining the posterior parameter distributions in Bayesian analysis. Given that the presence of emulators is currently unavoidable inside large-scale Bayesian analysis, a new measure quantifying the performance of an emulator is needed.

For reliable predictions with quantified uncertainty, we introduce herein a new measure for quantifying the performance of Gaussian process (GP) emulators to approximate full model calculations. This novel measure is based on the Kullback-Leibler divergence within closure tests. The novel measure proposed is inspired from the work done in devising scoring rules [46–49] in Bayesian statistics. To fully appreciate the usefulness of this new measure, a Bayesian inference (calibration) of model parameters is performed where a selection of hadron and inclusive jet suppression observables are used, focusing solely on central (0%–10%) nuclear collisions at the CERN Large Hadron Collider (LHC). As our calculation currently yields a large statistical uncertainty for heavy flavor observables, validation of the Bayesian inference workflow is especially important, and presents a good scenario to test our new performance metric of GP emulators. We also test its sensitivity to various model parameters and to different

observables, thus highlighting its usefulness for future Bayesian analysis. Finally, we present Bayesian constraints on model parameters explored in our limited Bayesian analysis, and provide comparisons with experimental data.

This work is organized as follows: Section II presents details regarding the multistage energy-loss calculation herein as well as provide details about the hydrodynamical simulation of the QGP with which partons will interact. Section III presents our Bayesian setup, with the new measure quantifying the performance for GP emulation being presented in Sec. III D. The best performing GP emulator is then used within a small-scale Bayesian calibration. Section IV is reserved for concluding remarks and present an outlook of how our current Bayesian analysis can be improved in the future.

II. SIMULATION SETUP

In the following sections we describe our simulation of parton evolution in the QGP. Section II A describes the models used to explain the interaction of jet partons with the QGP, while Sec. II B provides details of how QGP is evolved.

A. Parton interactions with the quark gluon plasma

After initial parton production in PYTHIA, the evolution of high-energy and high-virtuality partons is calculated in MATTER (Modular All Twist Transverse-scattering Elastic-drag and Radiation) [50,51], which uses the higher twist formalism [19,27,28] to explain parton interactions within the QGP. A virtuality ordered shower is thus developed for massless and massive [52,53] partons. Once hard partons in the shower reach a low virtuality regime, further evolution proceeds via the linear Boltzmann transport (LBT) model [54]. The LBT interactions between the hard partons and the QGP are preserving parton virtuality while modifying their energy, and three-momentum direction. Thus,

MATTER evolves partons with virtuality $t > t_s$—t_s being the switching virtuality—while LBT simulates those with $t \leqslant t_s$. The connection time between the PYTHIA shower and the energy loss models is chosen to be 0.6 fm/c but the dependence of the nuclear modification factor on this quantity is found to be weak [55,56].

The reason for the weak dependence on the connection time stems from the fact that medium-induced modifications to the jet shower evolution are suppressed for highly virtual partons, thus making the shower more vacuum-like [55]. Furthermore, as t_s is being varied, so is the balance between the high virtuality portion of the evolution in MATTER versus the low virtuality portion in LBT. Our Bayesian exploration constrains t_s in a way that reproduces the observed R_{AA}, while simultaneously probing the virtuality dependence of $\hat{q}$ in MATTER.

Following the evolution in LBT, the JETSCAPE framework determines whether partons undergo further splittings in MATTER (i.e., for parton leaving the QGP with enough virtuality to continue showering in the vacuum) or whether they need to hadronize (hadronization is handled via fragmentation in PYTHIA) using the colorless string hadronization routine [4].

1. The MATTER simulation

Parton decays in MATTER are calculated using the Sudakov form factor. The probability for *no* decay for a parton is given by

$$\Delta(t, t_{\min}) = \exp\left[-\int_{t_{\min}}^{t} dt' \int_{z_{\min}}^{z_{\max}} dz \frac{dN^{\text{tot}}}{dzdt'}\right] \frac{dN^{\text{tot}}}{dzdt'}$$

$$= \frac{dN^{\text{vac}}}{dzdt'} + \frac{dN^{\text{med}}}{dzdt'}, \qquad (1)$$

where $dN^{\text{vac}}/dzdt' + dN^{\text{med}}/dzdt'$ includes all possible decay channels of a given parton. For instance, according to soft collinear effective theory (SCET) [52], the decay of heavy quark $Q \rightarrow Q + g$ gives

$$\frac{dN^{\text{vac}}}{dzdt} + \frac{dN^{\text{med}}}{dzdt} = \frac{\alpha_s(t)}{2\pi} \frac{P_{g \leftarrow Q}(z)}{t} \left\{ 1 + \int_0^{\tau_Q^+} d\tau^+ \frac{2 - 2\cos\left(\frac{\tau^+}{\tau_Q^+}\right)}{z(1-z)t(1+\chi)^2} \left[\left(\frac{1+z}{2}\right) - \chi + \left(\frac{1+z}{2}\right)\chi^2 \right] \hat{q} \right\}. \qquad (2)$$

In Eq. (2), z labels the momentum fraction of the daughter heavy quark, M is the mass of the heavy quark, $\chi = (1 - z)^2 M^2/l_\perp^2$, with $l_\perp^2$ being the relative transverse momentum square between the outgoing daughter partons, determined via $z(1-z)t = l_\perp^2(1+\chi)$, while t is the virtuality of the heavy quark and $P_{g \leftarrow Q}(z) = C_F\left(\frac{1+z^2}{1-z}\right)$ is the splitting function and $C_F = 4/3$. The light flavor result [19,27,28,50], is recovered in the limit $M \rightarrow 0$. The integral over light-cone time τ^+ in Eq. (2) assumes the medium is in its rest frame, with the upper bound $\tau_Q^+ = 2q^+/t$ being given by the ratio of forward light-cone momentum $q^+ = (q^0 + \mathbf{q} \cdot \hat{n})/\sqrt{2}$ (with $\hat{n} = \mathbf{q}/|\mathbf{q}|$), and the virtuality t.

As the $g \rightarrow Q + \bar{Q}$ has not yet been calculated using the SCET approach devised in Ref. [52], this phenomenological

study estimates the gluon splitting into heavy quarks using the light flavor formula [50], and reduces the kinematic range using [53]

$$z_{\min} = \frac{t_0 + M^2}{t} + O\left(\left(\frac{t_0 + M^2}{t}\right)^2\right),$$

$$z_{\max} = 1 - \frac{t_0 + M^2}{t} + O\left(\left(\frac{t_0 + M^2}{t}\right)^2\right), \qquad (3)$$

assuming $M^2/t \ll 1$, $t_0/t \ll 1$, and $t_0 = 1$ GeV^2. Imposing $z_{\max} > z_{\min}$ as well as $t > t_{\min}$, requires that $t_{\min} = 2(M^2 + t_0)$. The determination of t and z proceeds in the same way as for $Q \rightarrow Q + g$ (more details are in Ref. [53]).

The transverse momentum broadening $\hat{q}[T(\tau^+)]$ acquired by the quark as it traverses the QGP is the only quantity that depends on τ^+ through the temperature T. From the hard thermal loop (HTL) approximation [57], $\hat{q}$ is

$$\hat{q}^{\text{HTL}} = C_a \frac{42\zeta(3)}{\pi} \alpha_s^2 T^3 \ln\left(\frac{cET}{4m_D^2}\right), \tag{4}$$

where $\zeta(3) \approx 1.20205$ is Apéry's constant, $C_a = N_c = 3$ the number of colors, while the Debye mass is $m_D^2 = 6\pi\alpha_s T^2$, and $c \approx 5.7$ [58]. The studies [53,55] showed that a constant effective α_s^{eff} can be improved by allowing the coupling to run with the scale $\mu^2 = 2ET$ via

$$\hat{q}^{\text{HTL}} = C_a \frac{42\zeta(3)}{\pi} \alpha_s(\mu^2)\alpha_s^{\text{eff}} T^3 \ln\left(\frac{cET}{4m_D^2}\right), \tag{5}$$

where

$$m_D^2 = \frac{4\pi\alpha_s^{\text{eff}} T^2}{3}\left(N_c + \frac{N_f}{2}\right) \stackrel{N_f=3}{=} 6\pi\alpha_s^{\text{eff}} T^2,$$

$$\alpha_s(\mu^2) = \begin{cases} \alpha_s^{\text{eff}}, & \mu^2 < \mu_0^2 \\ \frac{4\pi}{11 - 2N_f/3}\frac{1}{\ln\frac{\mu^2}{\Lambda^2}}, & \mu^2 > \mu_0^2, \end{cases} \tag{6}$$

$$p_1^\mu \partial_\mu f_1(x_1, p_1) = \mathcal{C}_{\text{el}}[f_1] + \mathcal{C}_{\text{inel}}[f_1],$$

$$\mathcal{C}_{\text{el}}[f_1] = \frac{d_2}{2}\int dP_2 \int dP_3 \int dP_3 (2\pi)^4 \delta^{(4)}(p_1 + p_2 - p_3 - p_4)|\mathcal{M}_{1,2\to3,4}|^2 \lambda_2(s, t, u)$$

$$\times \{f_3(\mathbf{p}_3)f_4(\mathbf{p}_4)[1 \pm f_1(\mathbf{p}_1)][1 \pm f_2(\mathbf{p}_2)] - f_1(\mathbf{p}_1)f_2(\mathbf{p}_2)[1 \pm f_3(\mathbf{p}_3)][1 \pm f_4(\mathbf{p}_4)]\}, \tag{8}$$

where d_2 is the spin-color degeneracy of parton "2," $\int dP_i \equiv \int \frac{d^3 p_i}{(2\pi)^3 2p_i^0}$ with $i = 2, 3, 4$, while $\lambda_2(s, t, u) = \theta(s - 2m_D^2)\theta(s + t - m_D^2)\theta(-t - m_D^2)$. The same $2 \to 2$ scattering rates are also used in MATTER.

The medium-induced gluon radiation describing $2 \to 3$ processes uses the same higher twist formulation as that employed in Eq. (2) of the MATTER simulation. The latter has an average number of gluons emitted from a hard quark (between time t and $t + \Delta t$):

$$\bar{N}^{\text{med}}(t \to t + \Delta t) \approx \Delta t \int dz dk_\perp^2 \frac{dN^{\text{med}}}{dz dk_\perp^2 dt},$$

$$\frac{dN^{\text{med}}}{dz dk_\perp^2 dt} = \frac{2\alpha_s P(z)}{\pi k_\perp^4}\hat{q}\left(\frac{k_\perp^2}{k_\perp^2 + z^2 M^2}\right)^4 \sin^2\left(\frac{t - t_i}{2\tau_f}\right). \tag{9}$$

A Poisson probability distribution is employed to sample independent successive emissions, with the probability of emitting n gluons being

$$\mathcal{P}(n) = \frac{(\bar{N}^{\text{med}})^n}{n!}e^{-\bar{N}^{\text{med}}}, \tag{10}$$

with Λ being chosen such that $\alpha_s(\mu_0^2) = \alpha_s^{\text{eff}}$ at $\mu_0^2 = 1$ GeV^2 [59]. The effective parametrization of the t-dependent $\hat{q}$ is [53,55]

$$\frac{\hat{q}(t)}{\hat{q}^{\text{HTL}}} = H(t) = \frac{c_0}{1 + c_1 \ln^2(t) + c_2 \ln^4(t)}, \tag{7}$$

where $\hat{q}^{\text{HTL}}$ is given in Eq. (5), c_1 as well as c_2 are tunable parameters, and $c_0 = 1 + c_1 \ln^2(t_s) + c_2 \ln^4(t_s)$ is an overall normalization ensuring $\hat{q}(t)/\hat{q}^{\text{HTL}} \in [0, 1]$ for $t > t_s$. Note that currently the virtuality dependence of $\hat{q}$ is assumed to be the same regardless of the mass of the quark [53]. Finally, MATTER also includes elastic $2 \to 2$ scatterings using leading-order perturbative QCD matrix elements as explored in detail below.

2. The linearized Boltzmann transport simulation

Once a parton enters the linearized Boltzmann transport (LBT) at $t < t_s$, its virtuality remains unchanged (see, e.g., Ref. [21] and references therein). The LBT relies on solving the Boltzmann equation taking into account $2 \to 2$ and $2 \to 3$ processes. The $2 \to 2$ scattering processes consist of leading-order perturbative QCD matrix elements. The evolution of the momentum and position distribution of a hard quark Q with momentum p_1 is given by

while the probability of a total inelastic process is $\mathcal{P}_{\text{inel.}} = 1 - e^{-\bar{N}^{\text{med}}}$. The procedure to determine whether (and how many) elastic vs inelastic scatterings inside the QGP have occurred is explored in detail in Ref. [55]. The only undetermined coefficient in LBT is the strong coupling α_s, which has a fixed component $\alpha_s^{(\text{eff})} = 0.3$, and a running component $\alpha_s(\mu^2)$ [53].

B. Evolution of the QCD medium

The evolution of the QCD medium used herein is performed using a boost-invariant $(2 + 1)$-dimensional [$(2 + 1)$D] model which involves three stages: a prehydrodynamic, hydrodynamic, and a hadronic transport stage [41,60–62]. The prehydrodynamic stage is based on the T$_R$ENTo initial condition for Pb-Pb collisions [63], which is followed by a collisionless Boltzmann evolution for a proper time of $\tau_{FS} = 1.2$ fm/c. Free-streaming generates a nontrivial initial profile used inside a $(2 + 1)$D hydrodynamical simulation. 400 T$_R$ENTo initial Pb-Pb configurations were generated within the 0%–10% centrality class at $\sqrt{s_{NN}} = 5.02$ TeV. The centrality selection is done using soft charged hadron multiplicity, as outlined in Refs. [41,64] (see also Ref. [65]). The most central collision was chosen to maximize the potential constraint that could be imposed $\hat{q}$. This requires maximizing

the space-time size of the QGP medium, and minimizing the effects of hadronic energy loss, which this study does not model. The relevant parameters used for simulating the evolution of the QCD medium are extracted from a Bayesian model-to-data comparison, explained in Refs. [41,64]. The event-by-event setup of the soft medium has been found to be important for the proper description of jet energy loss due to the added fluctuations in the medium [66]. The hydrodynamical simulation [60,67] is stopped once all fluid cells reach below $T_c = 154$ MeV [68], at which point all fields are converted into particles using the Cooper-Frye prescription [64,69,70], following which the ultrarelativistic molecular dynamics (UrQMD) [61,62] hadronic (Boltzmann) transport simulation is used.

III. BAYESIAN INFERENCE

There are four parameters in the aforementioned multistage energy-loss approach: the effective coupling constant α_s^{eff}, the switching virtuality $t_s = Q_s^2$, and (c_1, c_2) that control the virtuality dependence of $\hat{q}(t)$. Previous studies [53,55] have briefly explored the effects of these parameters on the charged hadron, D meson, and inclusive jet R_{AA}. However, the full correlation between the parameters and the observables, as well as the full capability of this multistage approach to describe experimental data, remain to be quantified. Bayesian inference can help answer these questions.

The Bayes' theorem states that the posterior distribution of the parameter set $\mathbf{x}$, given the experimental observation $\mathbf{y}_{\text{exp}}$, is proportional to the product of the prior distribution $q(\mathbf{x})$ and the likelihood function $\mathcal{L}(\mathbf{y}_{\text{exp}}|\mathbf{x})$:

$$p(\mathbf{x}|\mathbf{y}) \propto \mathcal{L}(\mathbf{y}_{\text{exp}}|\mathbf{x})q(\mathbf{x}). \quad (11)$$

The prior $q(\mathbf{x})$, as the name suggests, represents our prior knowledge of the parameter values. The likelihood function $\mathcal{L}(\mathbf{y}_{\text{exp}}|\mathbf{x})$ is the probability of observing $\mathbf{y}_{\text{exp}}$ given a specific parameter set $\mathbf{x}$:

$$\mathcal{L}(\mathbf{y}|\mathbf{x}) = \frac{\exp\left\{-\frac{1}{2}[\mathbf{f}(\mathbf{x}) - \mathbf{y}_{\text{exp}}]^\mathsf{T}\Sigma^{-1}[\mathbf{f}(\mathbf{x}) - \mathbf{y}_{\text{exp}}]\right\}}{\sqrt{(2\pi)^m \det\Sigma}}, \quad (12)$$

where m is the dimension of $\mathbf{y}_{\text{exp}}$, $\Sigma = \Sigma_M + \Sigma_{\text{exp}}$ is the uncertainty covariance matrix, which takes into account both model and experimental uncertainties, and $f(\mathbf{x})$ is the model calculation given the parameters $\mathbf{x}$. In the case of model uncertainties Σ_M, solely statistical are accounted for herein.

If $f(\mathbf{x})$ is known for an arbitrary $\mathbf{x}$, then one can perform a Markov chain Monte Carlo (MCMC) random walk through the parameter space to extract the posterior parameter distribution. However, each point in the parameter space requires at least $O(10^4)$ CPU hours to compute, meaning it is computationally prohibitive to walk in this parameter space by performing a full model simulation at each step. A surrogate model that can mimic the actual model with a reasonable computational cost is needed. The Gaussian process (GP) emulator is chosen as a fast surrogate model yielding both mean and covariance information. The surrogate model is trained on the set of precomputed $(\mathbf{x}, f(\mathbf{x}))$ pairs called the training data stored as $(X_{\text{train}}, \mathbf{y}_{\text{train}})$. X_{train} has dimensions $m \times k$, where m is the number of training data and k is the dimension

of the parameter set. $\mathbf{y}_{\text{train}}$ is a $m \times 1$ vector, since at each training point, just one dimension of the model output $f(\mathbf{x})$ is emulated. A GP essentially interpolates between all training data. Mathematically, one assumes that all desired outputs $\mathbf{y}$ to be predicted at inputs X, along with the known outputs $\mathbf{y}_{\text{train}}$ at the training points X_{train}, follow a multivariate normal distribution:

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{y}_{\text{train}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \mu_{\text{train}} \end{pmatrix}, \begin{pmatrix} K(X, X) & K(X, X_{\text{train}}) \\ K(X_{\text{train}}, X) & K(X_{\text{train}}, X_{\text{train}}) \end{pmatrix} \right),$$

where K denotes the covariance matrix. The distribution of $\mathbf{y}$ is then given by

$$\mathbf{y} \sim \mathcal{N}(K(X, X_{\text{train}})K^{-1}(X_{\text{train}}, X_{\text{train}})\mathbf{y}_{\text{train}}, K(X, X)$$
$$- K(X, X_{\text{train}})K^{-1}(X_{\text{train}}, X_{\text{train}})K(X_{\text{train}}, X)). \quad (13)$$

Each element in the covariance matrix K is calculated with the kernel function $k(\mathbf{x}, \mathbf{x}')$ that characterizes the correlation between two points in the parameter space [37]. The kernel encodes the prior Bayesian belief of the function being mimicked. Common kernel choices include [71]

(1) the radial basis function (RBF) kernel $k(r) = \sigma^2 \exp(-r^2/2l^2)$ (equivalent to the Matérn kernel with $\nu \to \infty$),
(2) the Matérn ($\nu = 5/2$) kernel $k(r) = \sigma^2(1 + \sqrt{5}r/l + 5r^2/3l^2)\exp(-\sqrt{5}r/l)$,
(3) the Matérn ($\nu = 3/2$) kernel $k(r) = \sigma^2(1 + \sqrt{3}r/l)\exp(-\sqrt{3}r/l)$,
(4) the white noise kernel $k(r) = \sigma^2\delta(r)$,

where $r = |x - x'|$, while σ and l are hyperparameters that are assigned a possible window and then optimized to maximize the likelihood of fit of the Gaussian process to the training data.

Since a GP emulator maps onto a one-dimensional space, in principle one would need dim($\mathbf{y}$) of GP emulators for all the data points. However, as the simulation results are in fact correlated (e.g., the measured or calculated R_{AA} points at different p_T are positively correlated with each other), dimensional reduction of the data is possible using principal component analysis (PCA). The principal component decomposition allows us to select the number (N_{PC}) of principal components (PCs)—a subset of vectors in dim($\mathbf{y}$)-dimensional space—that emulate the majority of the variance in training data. The additional uncertainty introduced by GP emulation and truncating uncertainty induced via PCA selection are accounted for in the covariance matrix Σ.

The last ingredient needed for an efficient Bayesian inference workflow is to devise the optimal distribution of training data points. Latin hypercube sampling of the parameter space is used to optimally distribute training points. One runs the full simulations at the hypercube-sampled design points, selects the first few principal components in PCA (N_{PC} containing the best optimal number of PCs), and trains N_{PC} GP emulators using the appropriate kernel functions discussed above. To predict the model output at a new point in the parameter space, one runs the N_{PC} GP emulators at this new point and then inverse-transforms the outputs from the principal component space onto the original observable (output) space. Besides

TABLE I. Prior parameters ranges in our Bayesian calibration.

Parameter	α_s^{eff}	Q_s	c_1	c_2
Range	0.1–0.5	1.5–4	1–10	50–300

generating the training data, the two key factors that affect model emulation process is the choice of the kernel in GP, and the number of principal components N_{PC} in PCA. One of the main results of this work—presented in Sec. III D—is to devise a quantitative measure of the performance of the GP emulator, allowing to choose the best emulator for a Bayesian model-to-data comparison. Finally, using the optimal emulator, the MCMC random walk in the parameter space is performed to extract the posterior distribution of the parameters within a limited Bayesian model-to-data comparison.

A. Calibration setup

The prior range for the model parameters are considered to be uniform distributions and are listed in Table I.

Those ranges are selected based on previous exploration of these parameters in Refs. [55,65]. Due to constrained computation budget of this work, 50 design points are drawn by latin hypercube sampling (see Fig. 1). For each design point, roughly 400 000 events are generated and distributed evenly among 400 fluid simulations. Sizable statistical fluctuations are observed especially for charged hadron and D meson R_{AA}, which will impact our calibration. The validity of the Bayesian analysis against model calculation fluctuations will be verified in Sec. III D, with further details in Appendix.

Given the computational resources available, this work focuses on Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV and 0%–10% centrality. A previous study [72] shows that calibrating to different collision energies independently versus at the same time may yield slightly different posteriors. Equivalently, one may need to use different values for the same parameter in different collision systems [73]. We leave such a exploration to a future study.

As for the experimental observables that will be calibrated to, we choose the nuclear modification factor R_{AA} for charged hadrons, D mesons, and inclusive jets. R_{AA}^X is defined as

$$R_{AA}^X = \frac{\frac{d\sigma_{AA}^X}{dp_T}}{\frac{d\sigma_{pp}^X}{dp_T}} = \frac{\sum_\ell \frac{dN_{AA,\ell}^X}{dp_T} \hat{\sigma}_\ell(\hat{p}_T)}{\sum_\ell \frac{dN_{pp,\ell}^X}{dp_T} \hat{\sigma}_\ell(\hat{p}_T)}, \quad (14)$$

where dN_{AA}^X/dp_T and dN_{pp}^X/dp_T are the multiplicity in the experimentally given p_T bin of the quantity X, which is either charged hadrons, D mesons, or jets originating from A-A and p-p collisions, respectively. The spectrum $dN_{AA,\ell}^X/dp_T$ is calculated utilizing the multistage model presented in Sec. II A. The total cross section for producing a hard scattering process is broken down in several segments ℓ of exchanged transverse momentum $\hat{p}_T$ contributing to the hard scattering at the level $\hat{\sigma}_\ell$. Each hard scattering event is sampled by PYTHIA. Many samplings of $\hat{\sigma}_\ell$, spanning a large kinematic range of the collision, are combined to produce $d\sigma^X/dp_T$. To avoid
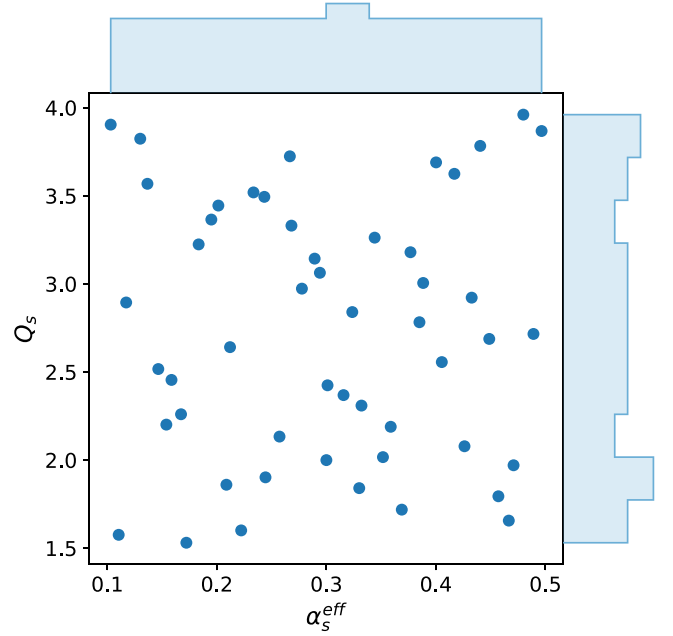

FIG. 1. Distribution of input parameter α_s^{eff} and Q_s from all 50 sampled design points within the prior range.

complications from hard-thermal hadronic recombinations and nonperturbative effects, the charged hadron and D meson are sampled solely for $p_T \geqslant 7$ GeV.

An important step before training GP emulators is to verify that the dynamic range of model calculations spans that of the experimental observations. Our model calculations depicted in Fig. 2 cover the measured range in R_{AA}^X, with each unique color in each plot corresponding to calculation from a single design point.

At first glance, it may appear that the statistical fluctuations are significant, especially for D meson R_{AA}, which could affect the validity of the Bayesian inference.[1] However, we have performed extensive validations of the GP emulator, in Sec. III B, GP emulator closure tests in Sec. III C, as well as stability tests of the posterior distribution in Appendix to addresses these concerns.

B. Emulator validation

The GP emulator's performance validation is presented in Fig. 3, where a direct comparison between emulator predictions and model calculations at five random design points can be seen. Here the RBF with a white noise kernel is used along with five PCs. The emulator predictions fit the model calculations well, and seem to cut off some statistical fluctuations,

---

[1]While generating the training data under a fixed computation budget, we were pursuing a balance between the number of design points to cover the parameter space and the number of events for each design point. This balance required a reduction in statistics for the D meson R_{AA} at every design point, to ensure sufficient number of design points are available to cover the parameter space.
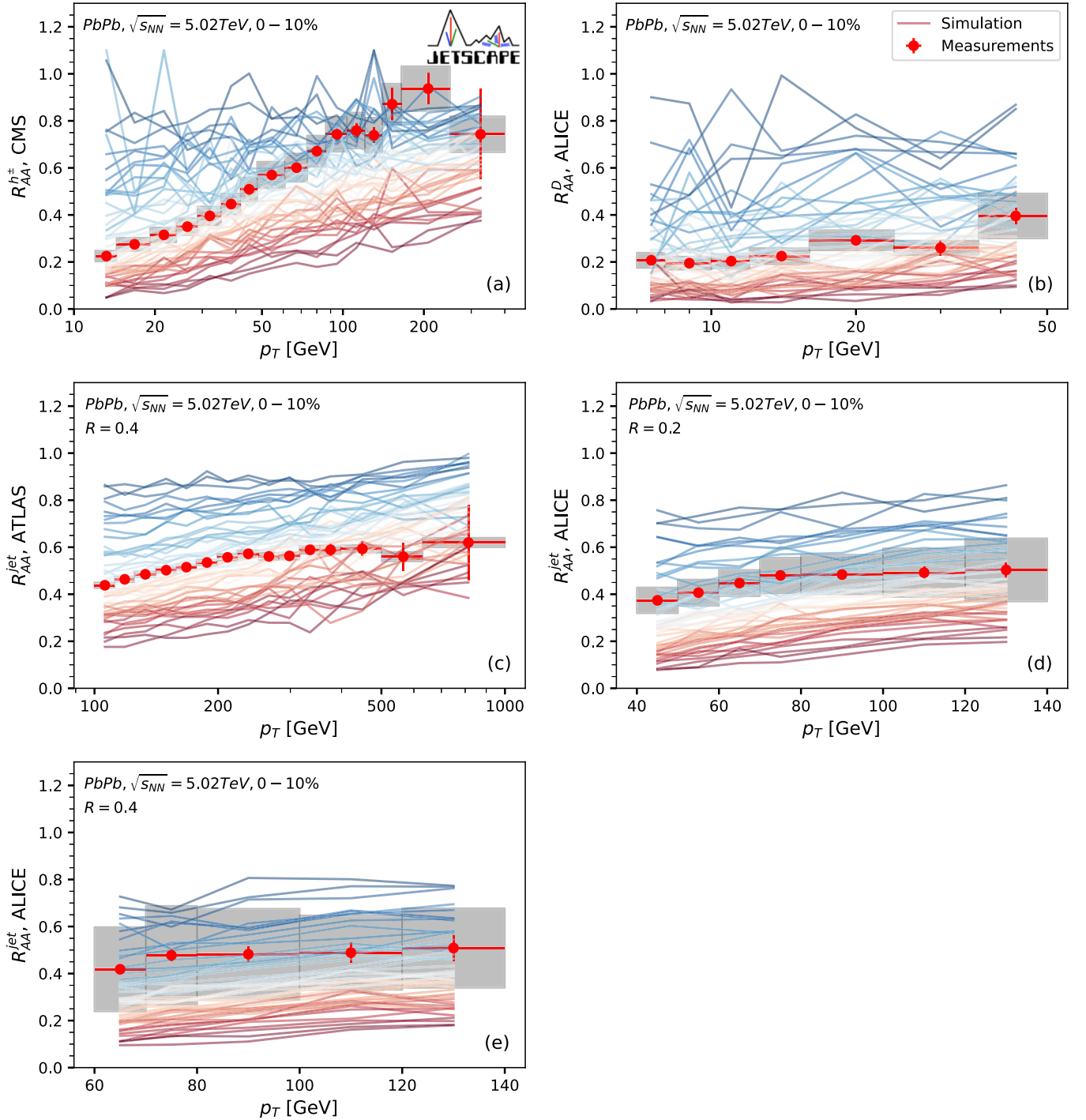
FIG. 2. A comparison between model calculation using parameters from all the design points and experimental data is shown, focusing on LHC data from Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV and 0%–10% centrality. Leading hadrons theory to data comparisons are presented in panels (a) and (b). Specifically, panel (a) shows charged hadron R_{AA} theoretical calculations against experimental data from CMS [74], while panel (b) focuses on D-meson R_{AA} calculations against measurements from ALICE [75]. Panels (c) through (e) are reserved inclusive jet R_{AA} comparisons, where jets are reconstructed using the anti-k_{T} algorithm. In panel (c), theoretical calculations are contrasted against data from ATLAS [76], while panels (d) and (e) center on describing the jet radius dependence using data from ALICE [77] at $R = 0.2$ and $R = 0.4$, respectively. Each unique color in each plot corresponds to calculation from a single design point.
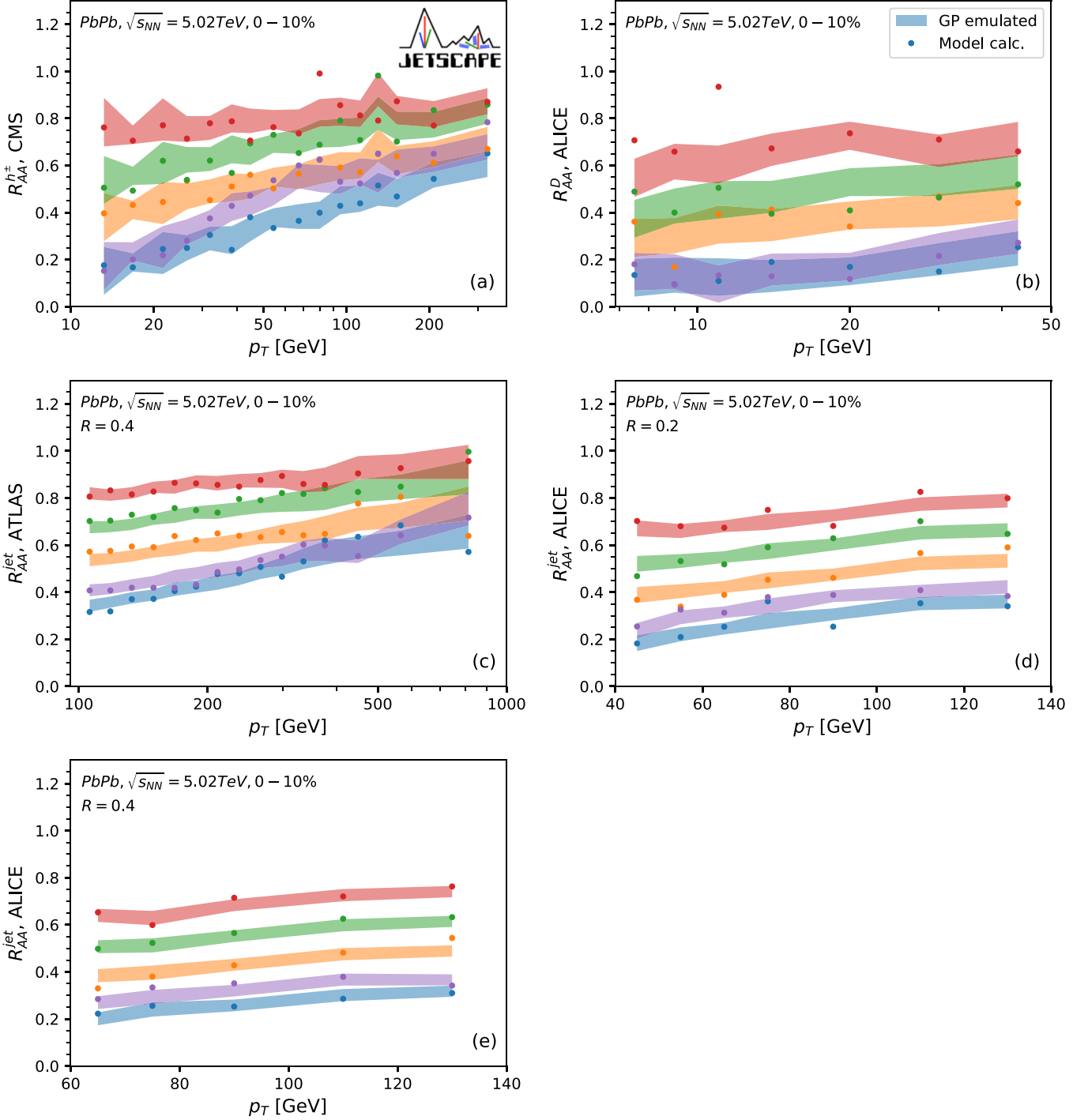
FIG. 3. Comparison between emulator predictions and model calculations at five random design points selected from the sample as the one depicted in Fig. 1. The combination of radial basis function with a white noise kernel is used along with five principal components (PCs). The panels follow the same categories as in Fig. 2, while the colored bands correspond to the region covered by one standard deviation.

stemming form using a subset of all PCs. The full performance of the emulator is plotted in Fig. 4, where the emulator response versus model calculation is plotted for all design points. The emulator seems to perform the best at predicting inclusive jet R_{AA} reconstricted using the anti-k_{T} algorithm [see Figs. 4(c)–4(e)], followed by charged hadron R_{AA} in Fig. 4(a), and finally the D meson R_{AA} in Fig. 4(d). Figure 4(f) depicts

the histograms of the relative difference between model calculations and emulator predictions for each observable. The distributions can all be fitted by a Gaussian centered near the origin, implying that the emulator induces little systematic bias in predicting the model. More quantitatively, the standard deviation is around 6%–9% when predicting inclusive jet R_{AA}, 12% when predicting charged hadron R_{AA}, and 30% when
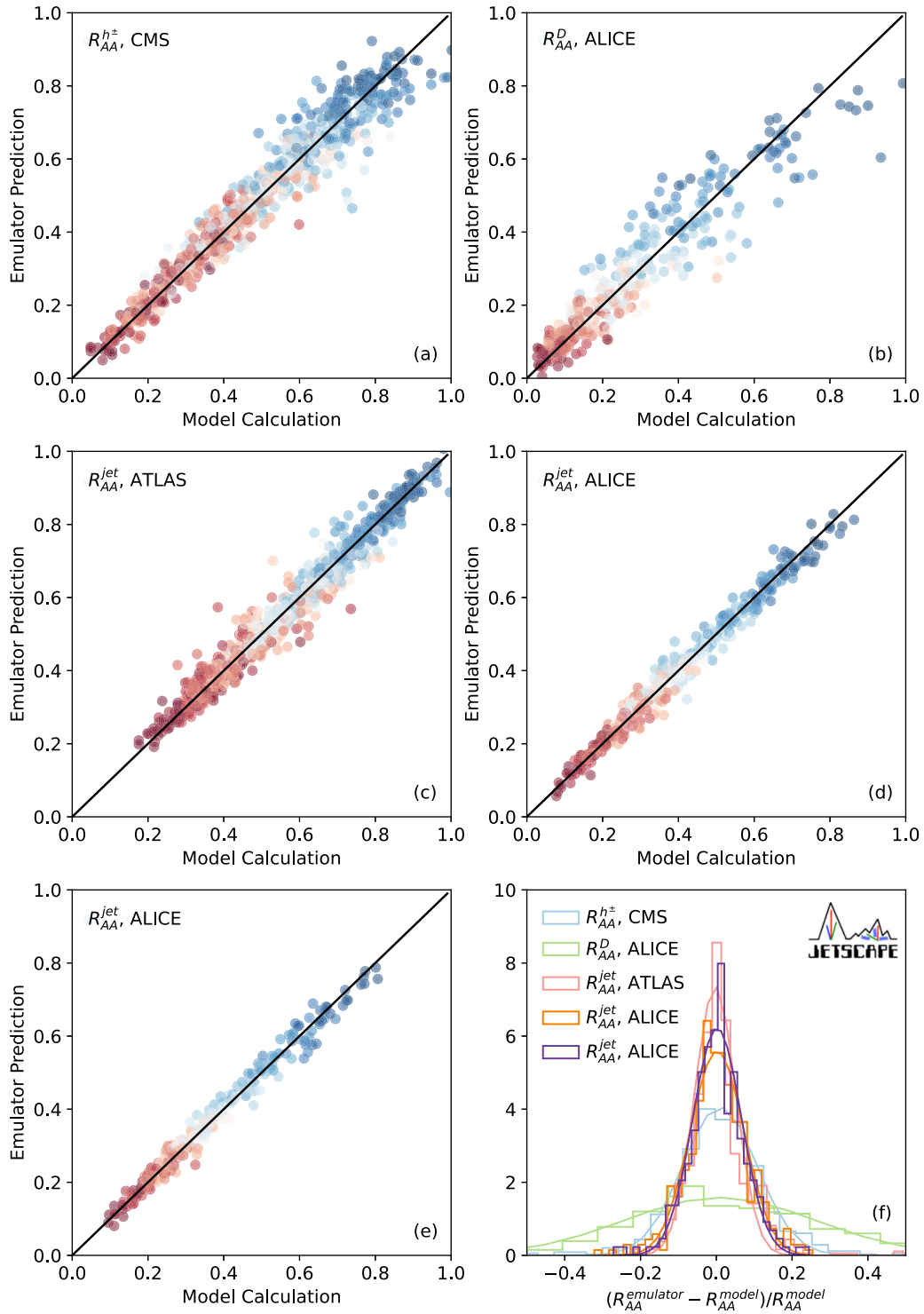
FIG. 4. A detailed comparison between emulator predictions and model calculations at all design points. The first five plots are scatter plots that display the model calculation and emulator prediction for each observable. An ideal emulator gives the same output as the model, thus lying along the back solid line in panels (a) through (e). Each color in panels (a)–(e) are calculated using the same design point as in Fig. 2. Panel (f) shows histograms of the relative difference between model calculation and emulator prediction for different observables. The combination of RBF and white noise kernel, along with five PCs, are used throughout.
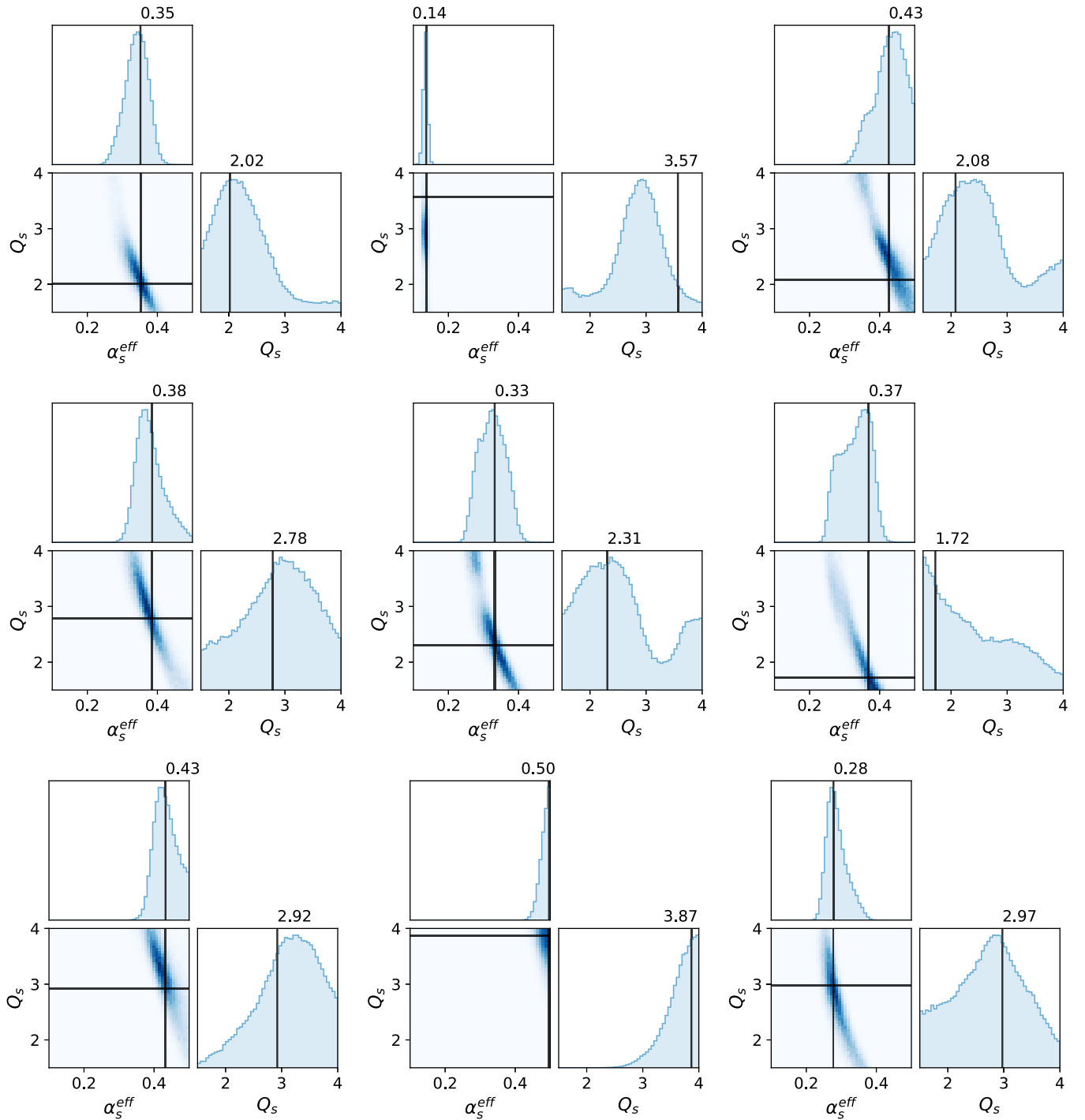
FIG. 5. Closure test results for α_s^{eff} and Q_s at nine random design points. The black lines represent the truth values. The RBF and white noise kernel is used along with five PCs.

predicting D meson R_{AA}.[2] Note that when calculating the emulator prediction at one design point, the GP emulation

training dataset will exclude the data from that specific design point. Thus, different emulators are trained for each design point and the emulator does not know the truth values when making predictions, as required for cross validation [78] via closure tests.

---

[2]As a point of comparison, if all the design points are used to train a single emulator, the uncertainty of the relative difference will be slightly reduced (around 5%–7% when predicting inclusive jet R_{AA}, 10% when predicting charged hadron R_{AA}, and 25% when predicting D meson R_{AA}). However, we did not perform closure cross

---

validation on emulators using all design points, given the computational resources available for this work.

C. Closure test

Being able to predict the training data does not guarantee that our emulator can constrain the model parameters well. If the data are not sensitive to some parameters, those parameters may end up with a wide posterior. Furthermore, if there are degeneracies in the model, i.e., multiple combinations of model parameters can describe the same set of data, the posterior distributions becomes multimodal. These scenarios can be checked for by performing a cross-validation closure test, whereby one design point is taken out from the training process and treated as the truth. The emulator is trained without the truth point and then the posterior distributions of the parameters are drawn. Since the truth values for the parameters is known in this case, one can make a comparison between the posterior parameter distributions and the truth. If there are infinitely many design points and zero statistical fluctuation at each design point, one expects a very narrow peak in the posterior distribution in a closure test near the truth value (or several peaks in the case of degeneracy).

The results for the closure test at nine random design points is shown in Fig. 5. For each panel in this figure, a single design point is removed when training GP emulator, i.e., the remaining 49 points are used for training. The trained GP emulator is then employed as a surrogate within the MCMC random walk to obtain the posterior parameter distribution shown. Notice that the posterior distribution are often peaked near the truth values for α_s^{eff} and Q_s. This procedure is repeated eight additional times, each time another random training point is removed, thus producing the remaining eight panels in Fig. 5. In that figure, the RBF kernel augmented with a white noise kernel was used to produce results, for reasons that are explored in Sec. III D 2.[3] Repeating this procedure for each parameter combination of the simulated model, as was done herein, ensures that the emulator does not exhibit undesired behavior within the spanned parameter space.

D. Selection of optimal emulator settings for Bayesian inference

In the previous section, while the closure tests at nine random design points seem to perform well, it is difficult quantify emulator performance by merely looking at posterior distributions, let alone compare the performance between different kernel functions and N_{PC}. A performance measure of the GPE is thus devised leveraging the notion of scoring rules [46–49] in Bayesian statistics. Scoring rules are traditionally used for evaluating the accuracy of probabilistic predictive models [48]. In this study such a rule is used for a novel purpose of fitting emulator parameters for Bayesian analysis. The proposed measure, derived below, quantifies the amount of information loss induced by GPE modeling via an information-theoretic approach.

The main contribution to GPE information loss stems from the second moment of the posterior distribution (as will be shown below). We define a quantity Δ_d which measures the
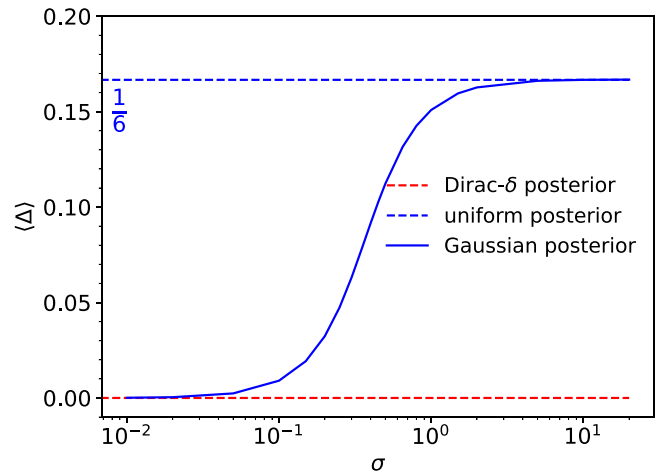
FIG. 6. $\langle\Delta\rangle$ calculated with Gaussian posterior centered at the truth versus the standard deviation σ of the Gaussian posterior $p(x)$. The analytical result using a uniform posterior and Dirac-δ distribution posterior are shown with dashed lines.

(second moment) deviation of the posterior distribution away from the truth value $x_{\text{truth}}^{(d)}$ of the parameter x, while using the dth design point as the truth:

$$\Delta_d = \int \left(\frac{x - x_{\text{truth}}^{(d)}}{x_{\text{max}} - x_{\text{min}}} \right)^2 p_d(x)dx, \qquad (15)$$

where $|x_{\text{max}} - x_{\text{min}}|$ is the allowed range of a parameter to be constrained, and $p_d(x)$ is the marginalized posterior distribution obtained from the case where dth design point is excluded. This is a new quantitative measure of the emulator's performance at recovering the truth from the mock data.

Using Δ_d defined as a closure test for one design point, averaging over all design points allows us to obtain an overall performance of the emulator:

$$\langle\Delta\rangle = \frac{1}{N_d} \sum_{d=1}^{N_d} \Delta_d. \qquad (16)$$

In our study, the number of design points is $N_d = 50$.

To understand how $\langle\Delta\rangle$ measures the deviation of the posterior distribution, the following example is illustrative. Suppose there are infinitely many design points uniformly distributed among the prior range. One calculates the values of $\langle\Delta\rangle$ with a uniform posterior $p(x)$ or a Gaussian posterior distribution centered at the truth with the variance being a free parameter to get a hint of the magnitude of $\langle\Delta\rangle$. For the case of a uniform $p(x)$ posterior, $\langle\Delta\rangle = 1/6$ is immediately obtained. If one assumes a Gaussian posterior distribution centered at the truth, Fig. 6 shows how $\langle\Delta\rangle$ changes as a function of σ: the standard deviation. $\langle\Delta\rangle$ approaches the value calculated with a uniform posterior when $\sigma \to \infty$, and goes to 0 as σ decreases. Thus, in closure tests, it is desirable for $\langle\Delta\rangle$ to be as close to zero as possible[4] while extracting the parameter

---

[3]Note, however, that it is difficult to compare the posterior distributions with different shapes generated using different kernels and N_{PC}.

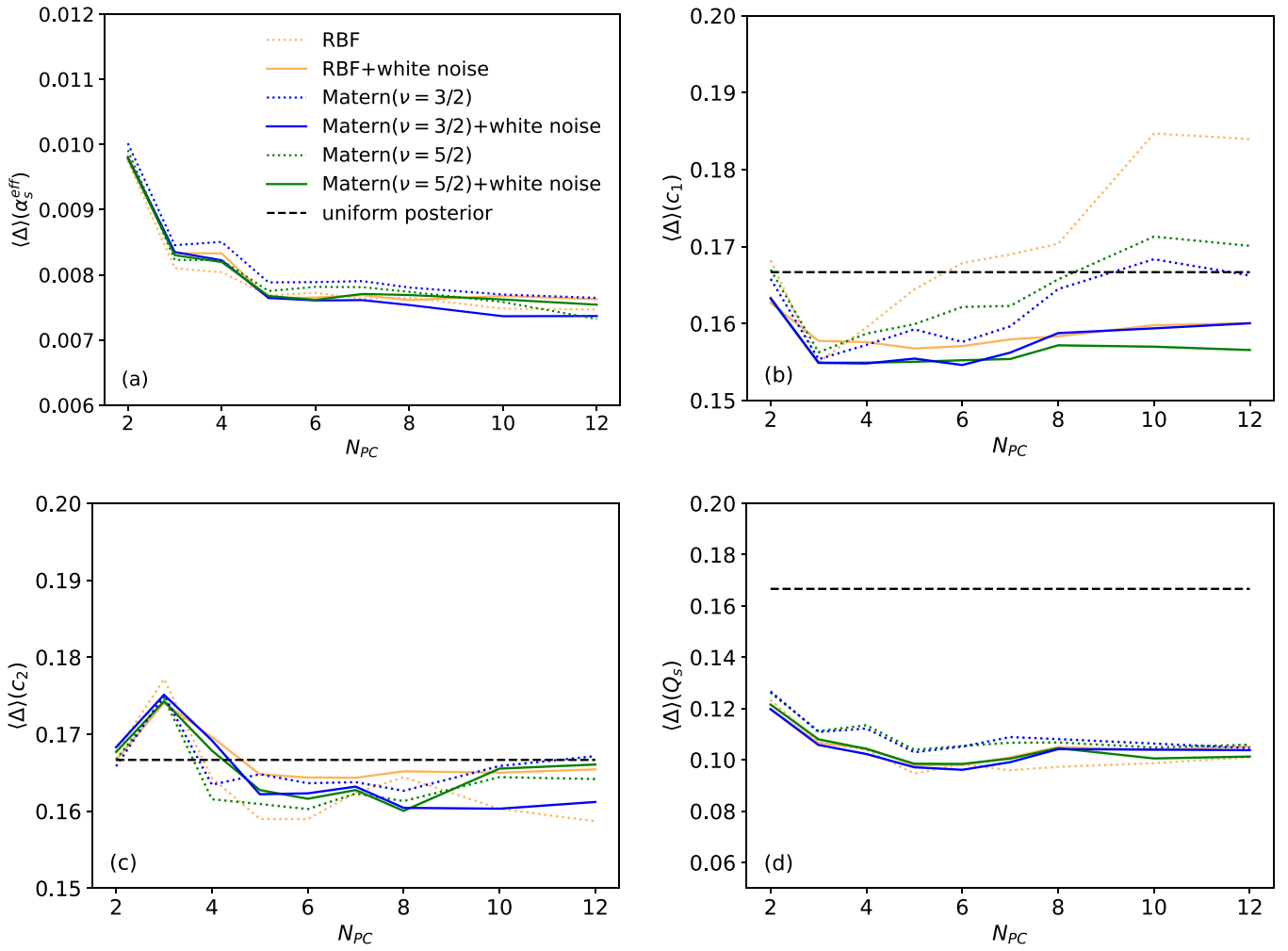[4]Note that, when $\langle\Delta\rangle = 0$, the GP emulator is indistinguishable from the full model.

FIG. 7. Comparison of $\langle\Delta\rangle$ using different kernels and various number of principal components. The panels explore the sensitivity of $\langle\Delta\rangle$ to (a) α_s^{eff}, (b) c_1, (c) c_2, and (d) Q_s.
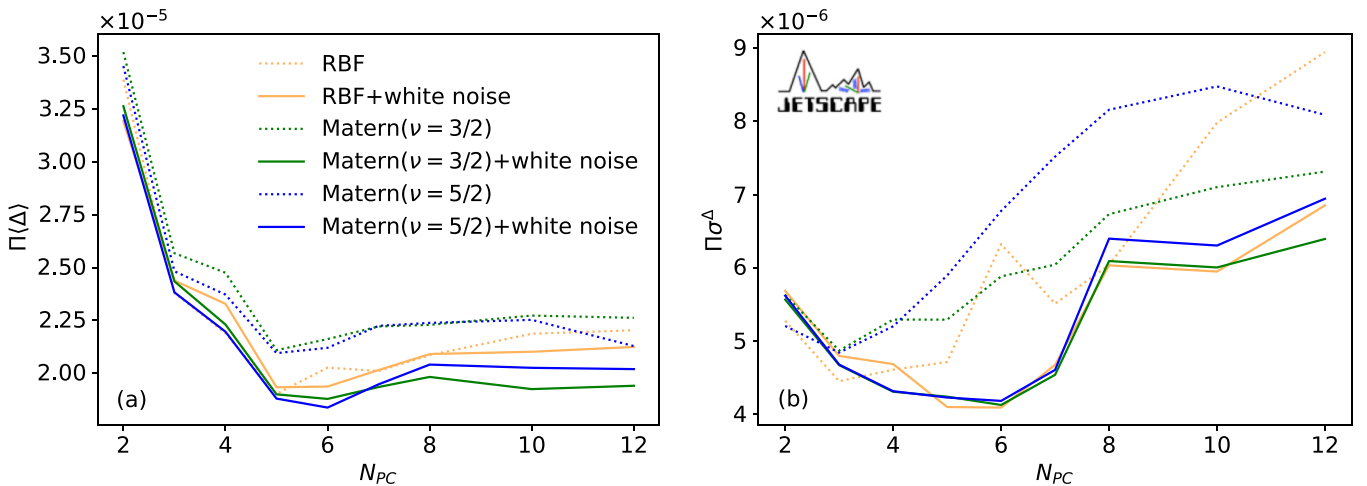

FIG. 8. Comparison of (a) $\Pi\langle\Delta\rangle$ and (b) $\Pi\langle\Delta\rangle$ for different parameters with different kernels, training data selection, and number of principal components.
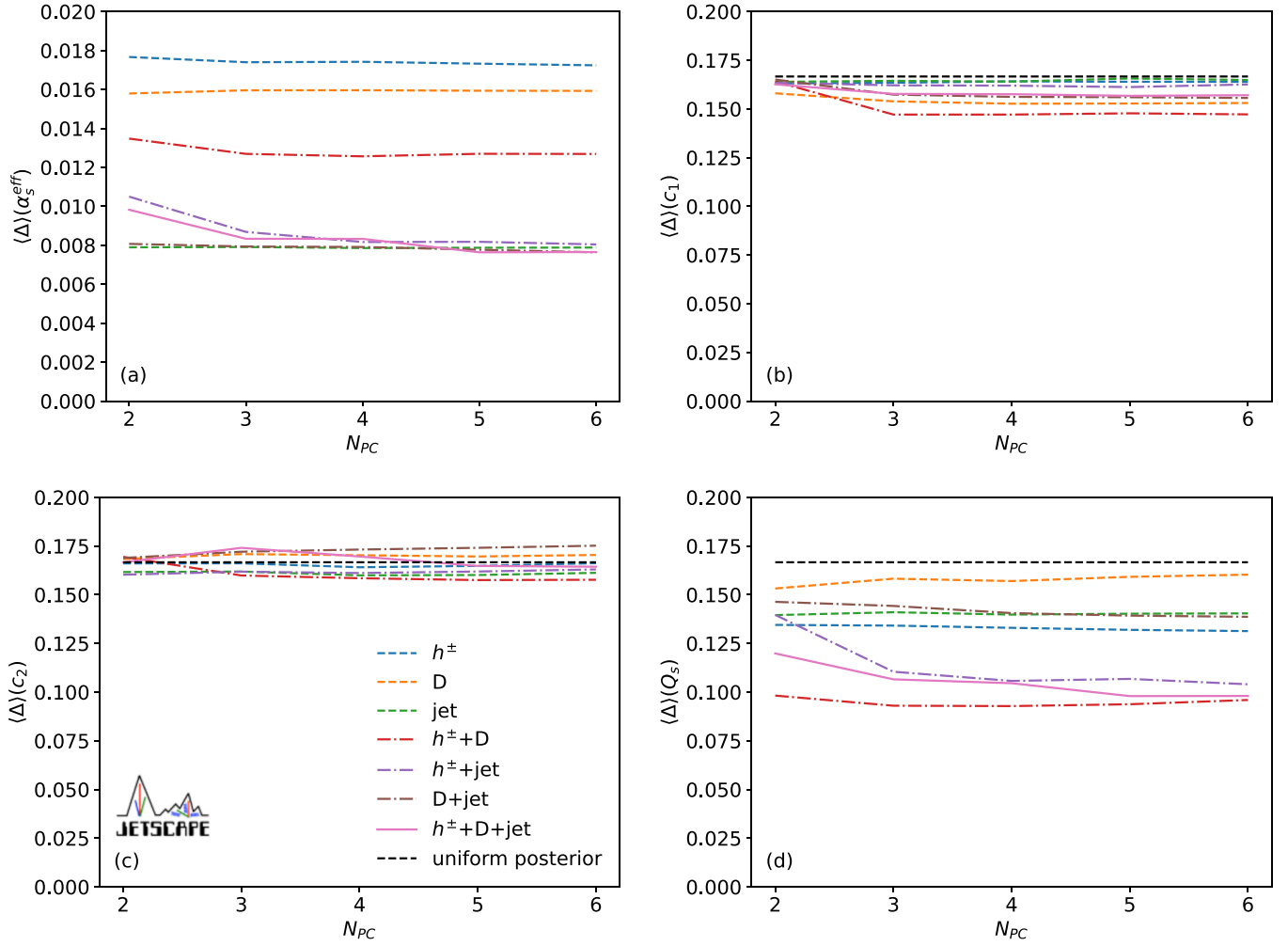
FIG. 9. The change of sensitivity of $\langle\Delta\rangle$ as different types of observables are considered. Panel (a) explores the change in sensitivity of $\langle\Delta\rangle(\alpha_s^{eff})$, panels (b) and (c) focus on $\langle\Delta\rangle(c_1)$, and $\langle\Delta\rangle(c_2)$, respectively, while panel (d) investigates the change in sensitivity for $\langle\Delta\rangle(Q_s)$ as various observables are incorporated.

values in the posterior because that brings the sensitivity of the GP emulator to model-parameters closer to the sensitivity of the model itself to those parameters.

We also recommend that, in practice, the averaging over Δ_d (yielding $\langle\Delta\rangle$) should be done using closure tests for cross validation [78], rather than separating the simulation dataset into a training and testing set for the GP emulator. The first approach via cross-validated closure tests ensures there is less variability in $\langle\Delta\rangle$ (and thus more stable fits for emulator parameters), since each data point is used for both training and testing (in different folds). The latter approach, which performs a single training-testing split of the simulated data, introduces greater variability in $\langle\Delta\rangle$ (especially for the 50 design points used herein), which in turn induces greater instability in model fitting. The latter can be of course be overcome by increasing the number of design points, but this is prohibitively expensive for our study.

1. Connection to the Kullback-Leibler divergence

The Kullback-Leibler divergence [79] is defined between two probability distributions, the posterior $p(x)$ and the prior $q(x)$, as

$$D_{KL}(p|q) = \int p(x) \ln\left(\frac{p(x)}{q(x)}\right) dx. \qquad (17)$$

This provides an information-theoretic measure of how one probability distribution p differs from a second probability distribution q when the latter is used as a reference. In particular, for Bayesian analysis, this can be interpreted as a measure of information change when updating the prior distribution to a posterior distribution [80]. We restrict ourselves to a one-dimensional system to ensure a clarity of our argument, with the generalization to multiple dimensions being straightforward.

D_{KL} is ill-defined (i.e., diverges) when the prior distribution $q(x)$ is a Dirac-δ function centered at x_0, the so-called
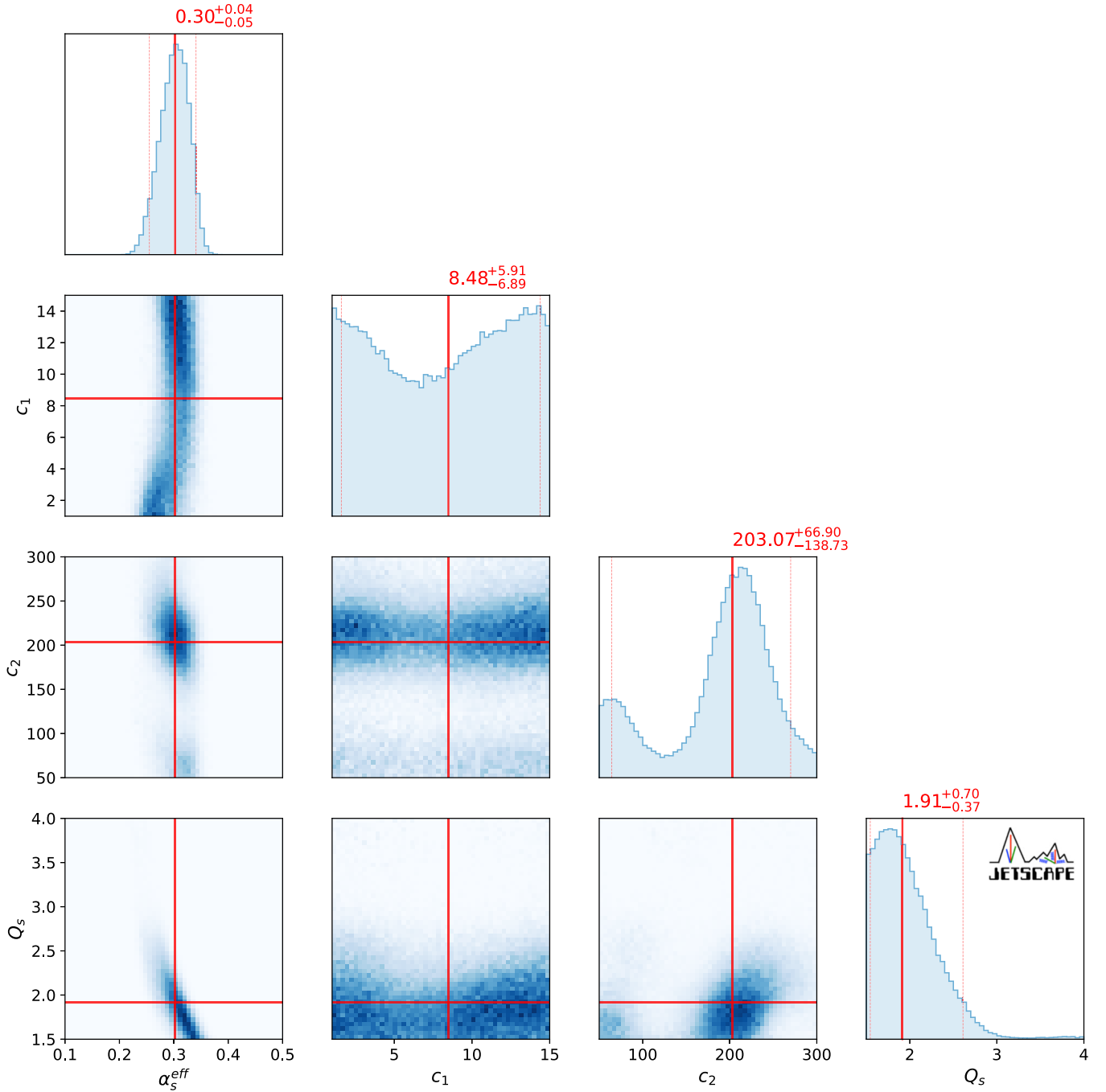
FIG. 10. The posterior distribution of model parameters. The emulator is using five PC and the RBF and white noise kernel combination and trained from 50 design points.

"true" value. This can be seen using the limit representation $\delta(x - x_0) = \lim_{\sigma_q \to 0^+} [\frac{1}{\sqrt{\pi}\sigma_q} e^{-(x-x_0)^2/\sigma_q^2}]$ as

$$D_{KL}(p|q) = \int dx\, p(x) \lim_{\sigma_q \to 0^+} \left[\ln \left(\frac{p(x)}{\frac{1}{\sqrt{\pi}\sigma_q} e^{-(x-x_0)^2/\sigma_q^2}} \right) \right]$$

$$= \lim_{\sigma_q \to 0^+} \left[\sigma_q^{-2} \int p(x)(x-x_0)^2 dx + O(\ln \sigma_q) \right].$$

$$(18)$$

However, noticing that the divergence is caused by the prior $q(x)$, with the posterior $p(x)$ being σ_q independent, the non-singular part in Eq. (18) is

$$\hat{\Delta}(x_0) = \int p(x)(x-x_0)^2 dx, \qquad (19)$$

which bares resemblance to Δ_d. Indeed, the shape of $\Delta_d(x_{\text{truth}})$ in Eq. (15) is the same as $\hat{\Delta}(x_0)$ in Eq. (19), as required, the only difference being an x-independent overall normalization.
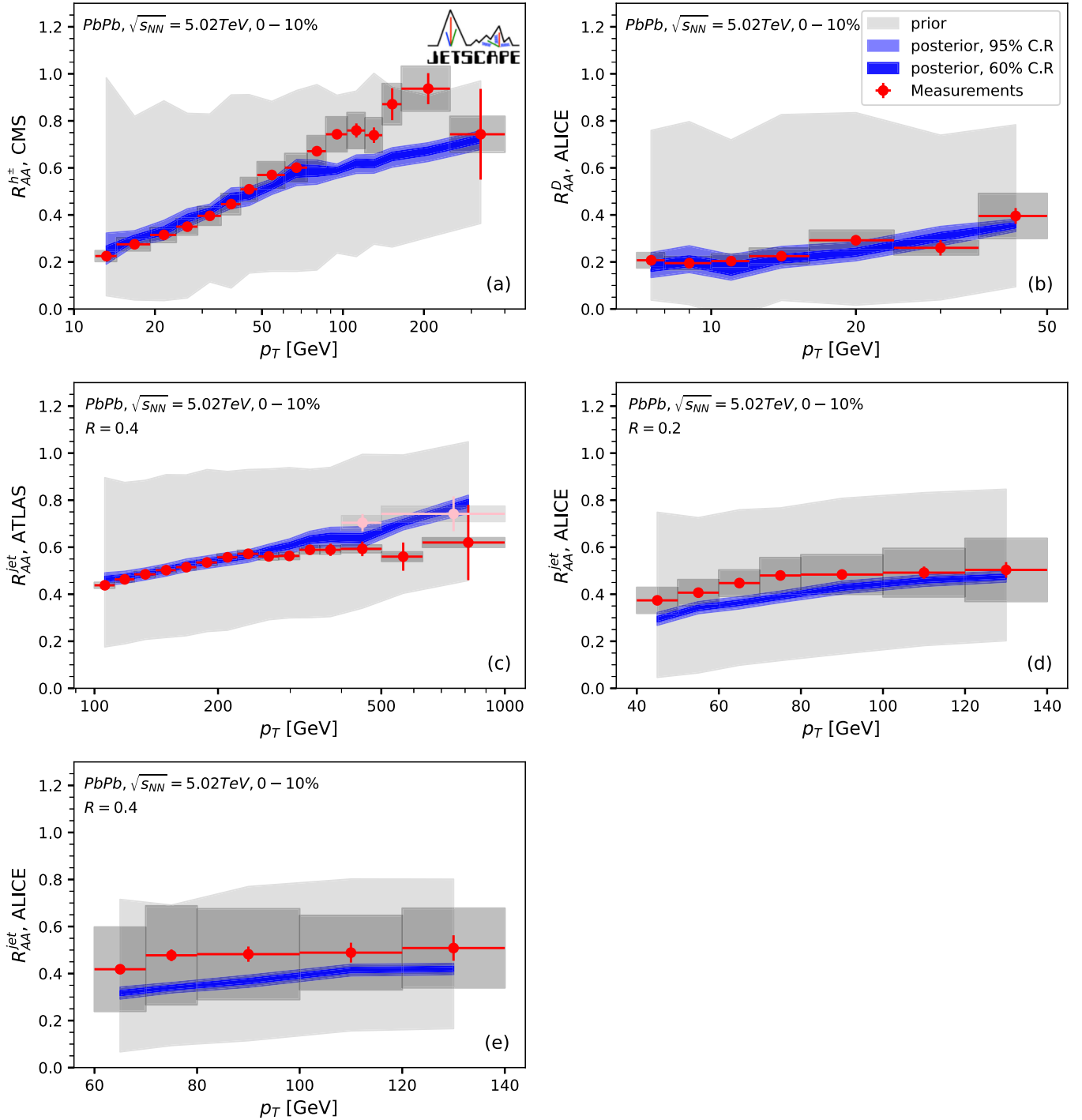
FIG. 11. Comparison between the posterior distribution of the observables and experiment data. Note that the pink (light gray) colored CMS data [81] in panel (c) are not used in training the GP emulator nor are they used in the subsequent Bayesian model-to-data comparison.

The above derivation indicates that Δ_d is a measure of the finite part in D_{KL} when the prior is a Dirac-δ distribution represented as the limit of a Gaussian distribution. D_{KL} was not used before as a metric in closure test because of this divergence, which we have regulated. This connection is another motivation for using Δ_d and $\langle \Delta \rangle$ in closure tests. For a comparison with other metrics, see Ref. [65].

Notice that the interpretation of the Kullback-Leibler divergence is different in closure tests compared with model-to-data Bayesian comparisons. In closure tests, the starting point is one where the information is maximized, i.e., the exact parameters are known *a priori*. However, since that parameters set is not used in the GP emulator, there is information loss which is captured by D_{KL}. In this case, a small D_{KL} is desirable.

Conversely, when using the trained GP emulators in Bayesian model-to-data comparisons, we are using a uniform prior distribution $q(\mathbf{x})$ to extract the posterior distribution via Bayesian inference. In that case, the posterior has gained information relative to the prior and a larger D_{KL} means parameters are better constrained.

2. Comparison between different kernels

Six types of kernels are compared in this section: the RBF kernel, the Matérn ($\nu = 5/2$), the Matérn ($\nu = 3/2$), and linear combinations of these three kernels with the white noise kernel. The results are shown in Fig. 7. From the various panels therein, one can see that $\langle\Delta\rangle(\alpha_s^{\mathrm{eff}})$ has the smallest values, and thus model dependence on α_s^{eff} is best captured by the emulator. The second best emulated parameter is Q_s as can be seen by $\langle\Delta\rangle(Q_s)$ in Fig. 7(d). Finally, Figs. 7(b) and 7(c) show that the emulator struggles to capture model sensitivity to c_1 and c_2, as is seen in $\langle\Delta\rangle(c_1)$ and $\langle\Delta\rangle(c_2)$.[5] This means the emulators are having trouble recovering these two parameters given the current level of uncertainties. To pick the optimal settings for the emulator, the product of all the $\langle\Delta\rangle$ is investigated. That is,

$$\Pi\langle\Delta\rangle = \langle\Delta\rangle(\alpha_s^{\mathrm{eff}})\langle\Delta\rangle(Q_s)\langle\Delta\rangle(c_1)\langle\Delta\rangle(c_2), \quad (20)$$

which is computed for different emulators in Fig. 8. As can be seen therein, a minimum in $\Pi\langle\Delta\rangle$ curve exists and gives the optimal choice of N_{PC}. Also note that the white noise kernel improves the overall performance, thus highlighting the importance of including the white noise kernel in the GP emulator.

One can also look at the variance of Δ, i.e., σ^Δ, which is calculated as

$$\sigma^\Delta = \frac{1}{N_d}\sum_{d=1}^{N_d=50}(\Delta_d - \langle\Delta\rangle)^2. \quad (21)$$

Notice that σ^Δ is not an absolute indication of the emulator's performance because it is possible to still have large mean and small variance (or small mean and large variance) at the same time (for a comparison with other metrics, see Ref. [65], Sec. 6.4.1). Nevertheless, a large σ^Δ indicates that the emulator is not performing well on at least some of the design points. The product of all the σ^Δ, which is

$$\Pi\sigma^\Delta = \sigma^\Delta(\alpha_s^{\mathrm{eff}})\sigma^\Delta(Q_s)\sigma^\Delta(c_1)\sigma^\Delta(c_2), \quad (22)$$

is shown in Fig. 8(b). Interestingly, the smallest values are also achieved at around five or six principal components, and with the kernels that include white noise. The decision to use $N_{\mathrm{PC}} = 5$ is based on the results presented in Fig. 8. In this work, we use the (RBF + white noise) kernel with five principal components for its overall performance and consistency with previous studies.

The method discussed in this section is quite general and can easily be applied to other Bayesian analysis projects utilizing the Gaussian Process emulator as a surrogate model.

---

[5]Note $\Delta = 1/6$ is indicated by the blue (gray) dashed line, which is the analytical result assuming a uniform posterior.
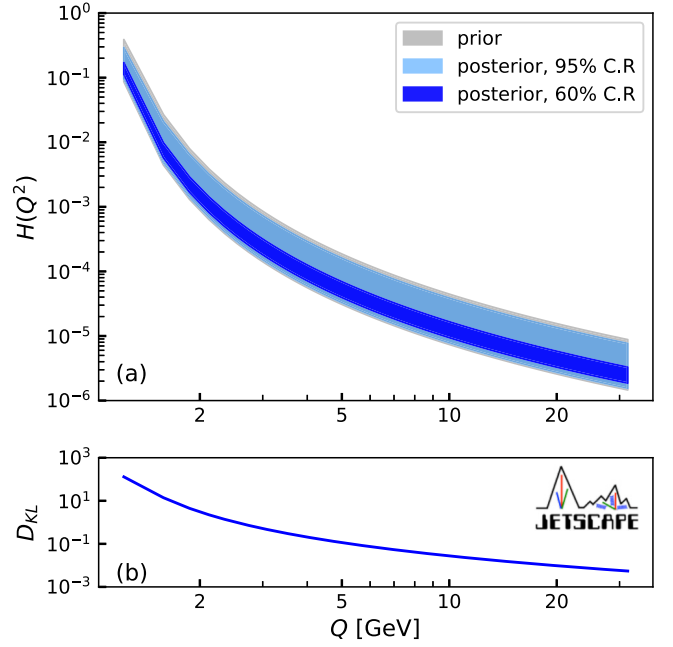

FIG. 12. Panel (a) depicts the prior (light gray), 95% credible region of the posterior (light blue), and 60% credible region of the posterior (deep blue) of $H(t) = H(Q^2)$ defined in Eq. (7). Panel (b) displays the corresponding information gain using the Kullback-Leibler divergence D_{KL}.

Key ingredients for training the emulator, including choice of the kernel and the number of principal components, can all be determined by computing $(\langle\Delta\rangle, \Pi\langle\Delta\rangle)$ as well as $(\sigma^\Delta, \Pi\sigma^\Delta)$. The level of constraint on each parameter, can also be reflected by $\langle\Delta\rangle$. Note, however, that model-related (i.e., theoretical) systematic uncertainties have not yet been taken into account. One should also keep in mind that this is a measure of the average performance over the entire parameter space being searched. Currently, the average is the best estimate we can get since the truth (optimal) parameters are not known *a priori*.

3. Observable sensitivity analysis

$\langle\Delta\rangle$ can also be used to measure the sensitivity of model parameters to different observables. There are three categories of observables in our calibration: charged hadrons, D mesons, and inclusive jet R_{AA}. The $\langle\Delta\rangle$ when calibrating to different combinations of these three types of observables are shown in Fig. 9. We find that inclusive jet R_{AA} observables are most sensitive to $\langle\Delta\rangle(\alpha_s^{\mathrm{eff}})$, while charged hadron R_{AA} governs the size of $\langle\Delta\rangle(Q_s)$. Of course, a combination of two or more of these observables improves $\langle\Delta\rangle$. No significant constraint can be derived on average for c_1 and c_2. The fact that Δ can be used to identify which observable(s) contribute the most to emulator prediction is a key feature guiding future emulator performance improvement. Together with the fact that Δ can also quantify the performance of GP emulation kernels as discussed in Sec. III D 2, makes it an invaluable quantity to compute when training emulators.
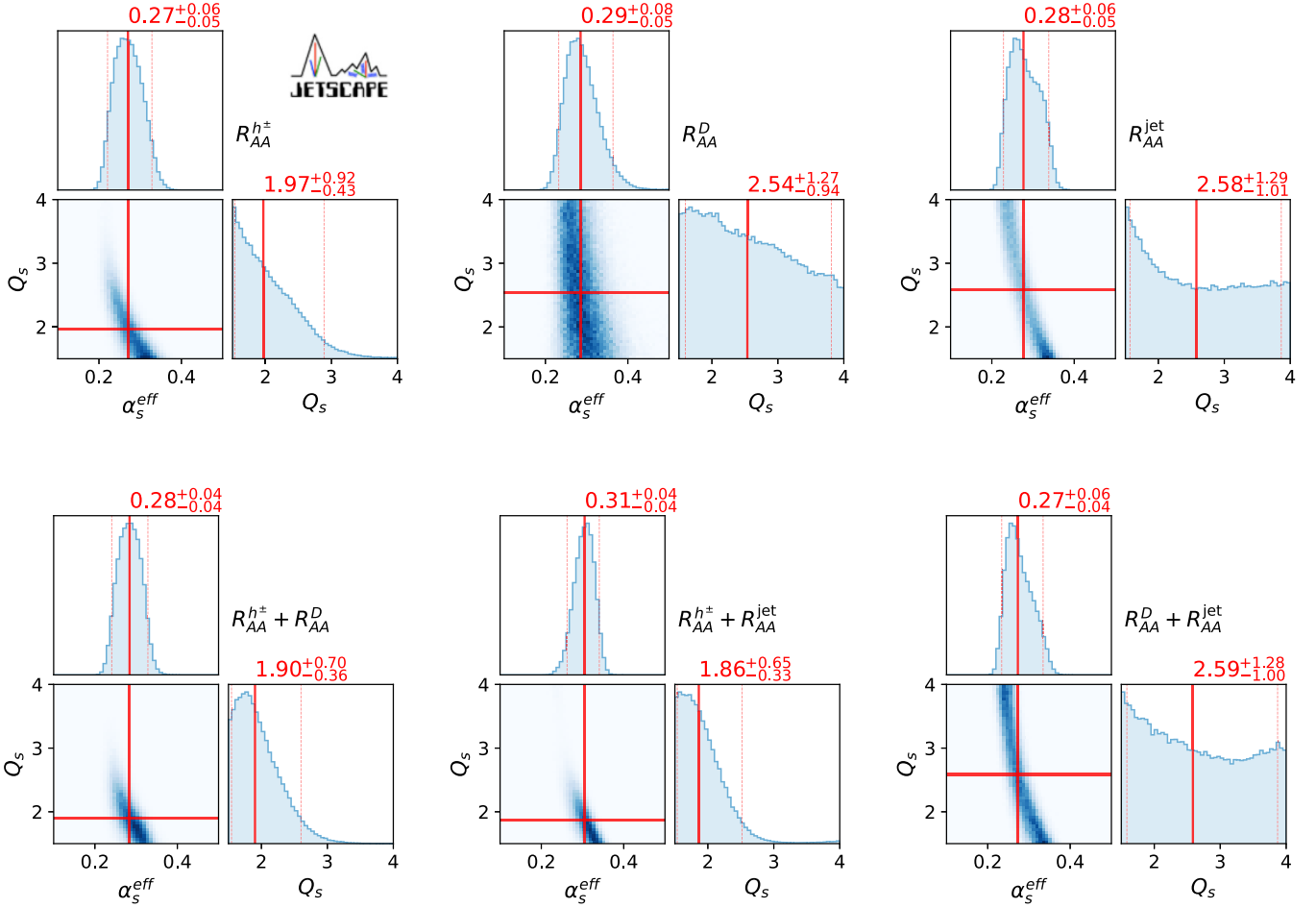
FIG. 13. Posterior distribution of α_s^{eff} and Q_s, calibrated using different observables individually as well as pairwise.

E. Bayesian inference results

The posterior distribution of the parameters calibrating all five experimental datasets is shown in Fig. 10. The posterior distribution of the observables compared with data are shown in Fig. 11. Relative to the parameters used in Refs. [53,55], the posterior distribution in Fig. 10 suggests similar value for α_s^{eff} and a slightly smaller Q_s. Indeed, Refs. [53,55] show that smaller values of Q_s shifts the charged hadron and D meson R_{AA} upwards, a little closer to data compared with results found in Refs. [53,55], compensating somewhat for the mismatch in charged hadron R_{AA} at high p_T seen in our earlier study.

While the case of a virtuality independent $\hat{q}$ has been phenomenologically ruled out in Refs. [53,55], Fig. 10 shows a weak constraint on c_1 and a slight preference for large c_2 values. An intuitive explanation for why c_1 and c_2 are hard to constrain is that the energy loss in the MATTER regime goes to zero when c_1 and c_2 go to infinity. Furthermore, MATTER's energy loss is quickly reduced when c_1 and c_2 values are non-negligible as is the case herein, and was shown in Refs. [53,55] using much higher statistics.[6] In the future,

a measurement with smaller uncertainties at high p_T can be beneficial in helping better constrain c_1 and c_2, together with higher statistics of theoretical calculations to improve D-meson R_{AA} predictions.

Figure 12 shows the constraint on $H(Q^2)$ resulting from this Bayesian analysis. Figure 12(a) shows the prior, 95% credible region, and 60% credible region of $H(t) = H(Q^2)$ as a function of Q. The Kullback-Leibler divergence (defined in Sec. III D 1) depicted in Fig. 12(b) monotonically decreases with Q. This is likely due to the fact that the second-order term and the fourth-order term in $H(Q^2)$ are comparable in magnitude in this region. Thus, the *joint distribution* of c_1 and c_2 at low Q is constrained, rather than their individual distributions. While tighter constraints on parameters involved in describing $H(Q^2)$ should be sought after in the future, Fig. 12(b) confirms that there is significant information gained about $H(Q^2)$ at (mostly) lower values of Q^2. A different parametrization choice for $H(Q^2)$ may yield better constraints on its parameter values, a fact actively being explored within the JETSCAPE Collaboration [82].

F. Sensitivity to different observables

We analyze in Fig. 13 how different observables (and observable combinations) affect the constraints presented in

---

[6] $\langle\Delta\rangle(c_1)$ and $\langle\Delta\rangle(c_2)$ also reflect the small sensitivity of (c_1, c_2) to observables.

Fig. 10. Since the constraint for c_1 and c_2 are weak, only the posterior distribution of α_s^{eff} and Q_s are shown in Fig. 13. One can see that charged-hadron data are the main reason why large Q_s is disfavored. Indeed, looking at the bottom panels in Fig. 13 one sees that any combination including charged hadron R_{AA} provides a stronger sensitivity to Q_s, compared with using jet and D mesons, thus furthering our understanding of the results in Fig. 10. Looking at the off-diagonal joint distribution plots in Fig. 13 also reveals that α_s^{eff} and Q_s are anticorrelated; an observation that may be useful in future Bayesian analysis.

IV. CONCLUSION

In this work, we performed Bayesian inference for a multistage parton energy-loss approach in heavy-ion collisions. The model calculation is calibrated to charged hadron, D meson, and inclusive jet R_{AA} measurements. The challenge, however, is the sizable uncertainties in both experimental measurements and model calculation (emulation). We have validated our Bayesian workflow in both the forward direction, i.e., mapping from model parameters to observables, and the inverse direction, in the context of closure tests. Specifically, we have proposed the $\langle \Delta \rangle$ metric in cross-validation closure tests which we connected to the Kullback-Leibler divergence. With this metric, the optimal settings for emulating the model calculation was determined. The $\langle \Delta \rangle$ metric can also give guidance about the sensitivity of each parameter to different observables. Its usefulness for future Bayesian inference studies is what makes $\langle \Delta \rangle$ an important quantity to calculate and is the main result of our work.

The results of our Bayesian analysis finds optimal values for α_s^{eff} and Q_s to be similar to what have been used in previous studies [53,55,65], although now these values are on a much firmer footing. The constraints for c_1 and c_2 are much weaker due to their small sensitivity to the current observables. The small sensitivity is verified by both the $\langle \Delta \rangle$ metric and arguments about the influence of c_1 and c_2 on parton energy loss [65]. Another outcome of our Bayesian analysis is the identification that α_s^{eff} is most sensitive to inclusive jet R_{AA}, while Q_s is most sensitive to charged hadron R_{AA}.

While there are sizable experiment uncertainties in the data against which our model was calibrated, there are also sizable theoretical uncertainties owing, in part, to the computation resources available for simulation. The latter restricted us to a subset of experimental data included in our Bayesian study. To improve theoretical uncertainties, one could adopt a more complex parametrization for $\hat{q}$ to account for the potential difference between data and theory at high p_T. One could also sample c_1 and c_2 on a logarithmic scale to improve the design point density at large c_1 and c_2. Our calibration should also include more observables, different collision energies and systems, as well as centralities. The JETSCAPE Collaboration has an ongoing Bayesian analysis that will address most of these points [82]. Furthermore, we expect that improved experimental uncertainties, such as releasing the full covariance matrix, would benefit future Bayesian studies. Thus, a more detailed discussion of the uncertainties for both experiment (e.g., off-diagonal experimental systematic uncertainty) and

theory (systematic model uncertainty), should be included in a future study.

APPENDIX: STABILITY OF THE POSTERIOR TO FLUCTUATIONS

The closure test offers an excellent check that the emulator can reproduce mock data (model calculation) at many design points. However, the relation between the posterior distribution and the uncertainty level of the training data is still not explored. It is difficult to reduce the fluctuations of the training data because it requires running more events. The
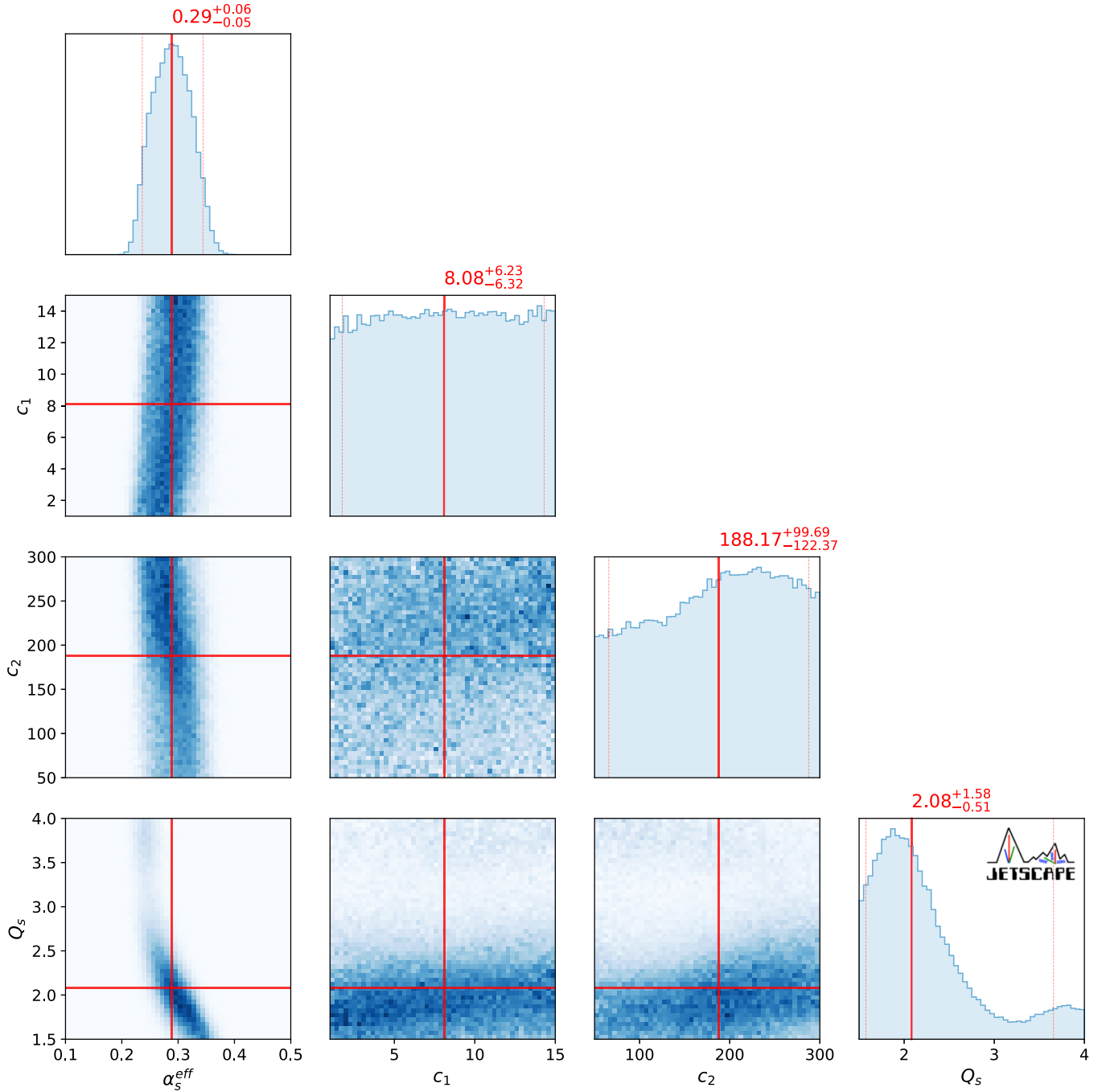
FIG. 14. The posterior distribution of the model parameters. Only 1/3 of events are used for each design point.

other direction is easier to explore and can be investigated in two ways.

First, we can reduce the statistics when generating the training data, which we have done by only using 1/3 of events at each design point. Thus, the effects of statistical fluctuations for each observable are scaled up and can be readily appreciated. Figure 14 shows the posterior distribution of the parameters. Compared with what is shown in Fig. 10, the posterior distribution is not wildly altered in Fig. 14, except for weaker constraint on c_1 and c_2, which remain the roughly the same. The effects of increasing the statistics by

a factor of three significantly improve the constraint on the α_s^{eff} and Q_s parameters, while the effects on (c_1, c_2) are more modest.

In Fig. 15, we study how improved statics affect $\langle \Delta \rangle$. A marked improvement is seen on $\langle \Delta \rangle(\alpha_s^{\mathrm{eff}})$ and $\langle \Delta \rangle(Q_s)$, while $\langle \Delta \rangle$ for c_1 and c_2 is modest, hovering around $1/6$, showing that constraints on c_1 and c_2 almost do not change when tripling the statistics. More statistics in D meson R_{AA} calculations, although a different parametrization for $\hat{q}(t)$ may be needed and is being explored in Ref. [82]. Of course, reduced experimental uncertainties will also improve this
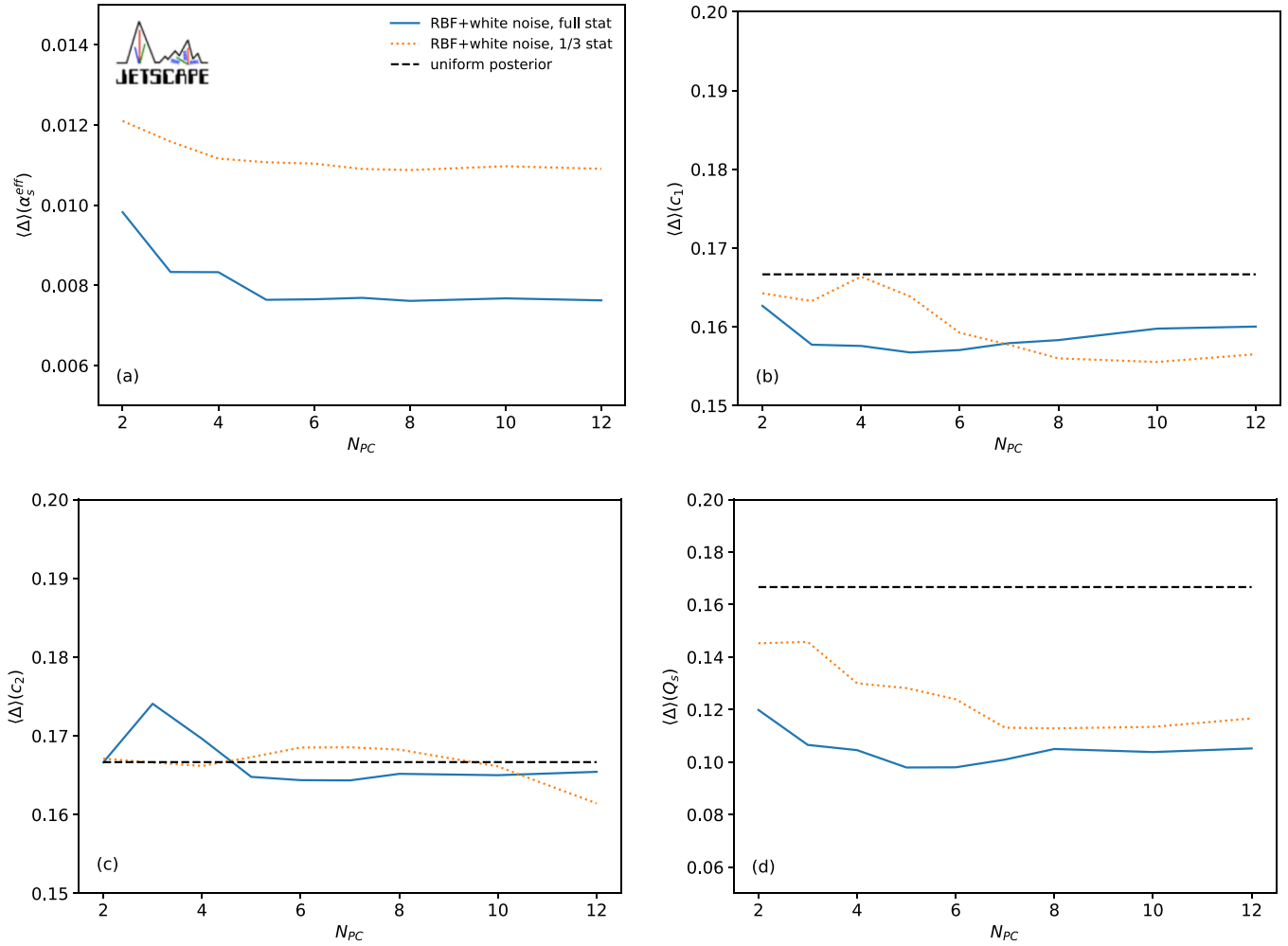
FIG. 15. Comparison of $\langle\Delta\rangle$ for different parameters with the same settings for the emulator but different statistics for the training data.

situation. Finally, the last element we considered is adding more Gaussian noise to all the model calculations for each design point. That is, every observable gets an equal amount of additional statistical fluctuation. In Fig. 16, we can see the updated posterior distribution when an additional noise with 0.05 standard deviation is introduced. Similar posterior distributions of the parameters are seen compared with Figs. 10 and 14.
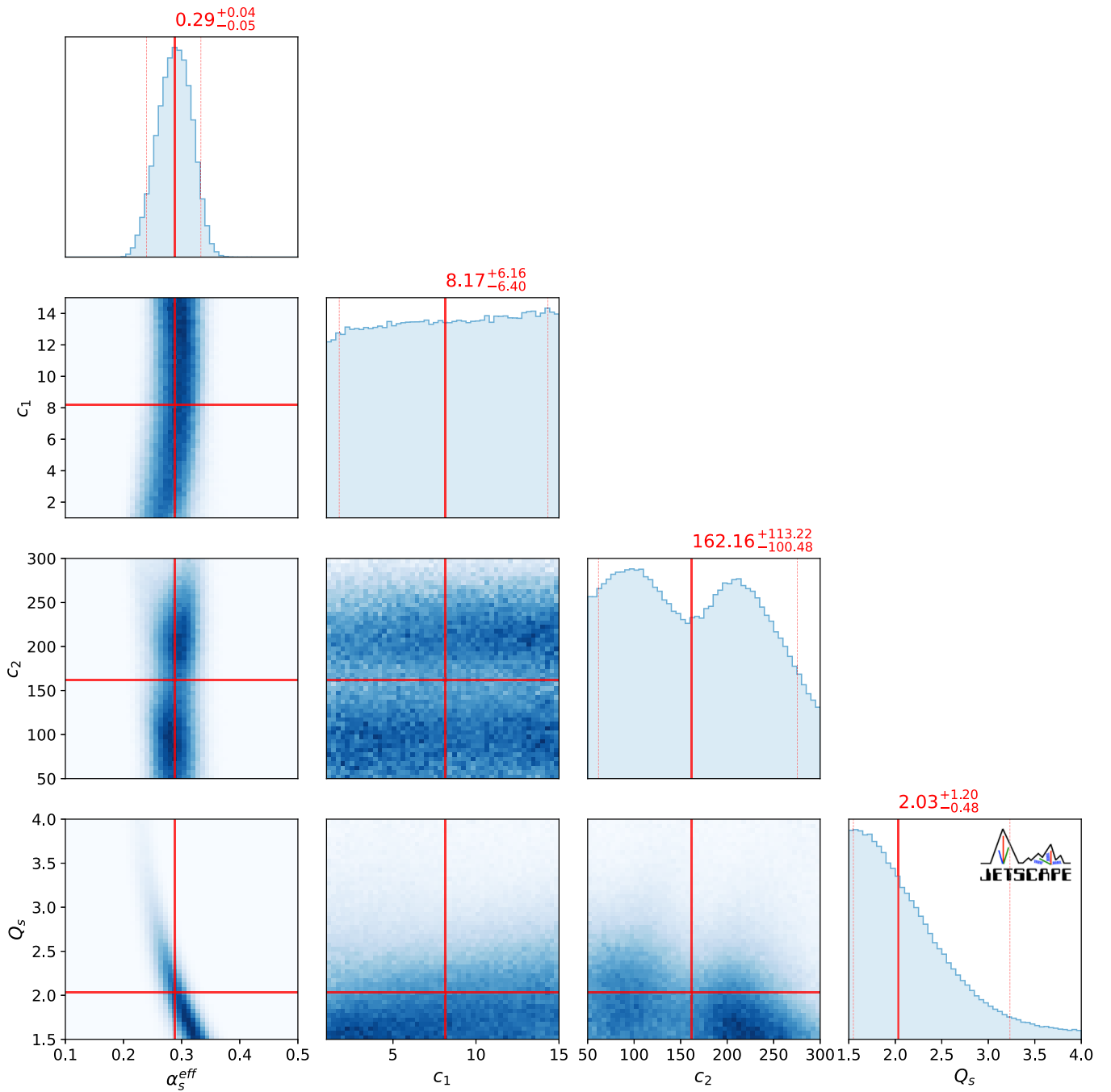
FIG. 16. The posterior distribution of the model parameters with added Gaussian noise to all training data. The Gaussian noise has a mean $\mu = 0$ and standard deviation $\sigma = 0.05$.

[1] R. D. Field, Applications of Perturbative QCD, in *Frontiers in Physics* (Redwood, California: Addison-Wesley, 1989), Vol. 77.

[2] P. Skands, S. Carrazza, and J. Rojo, Tuning PYTHIA 8.1: The Monash 2013 tune, Eur. Phys. J. C **74**, 3024 (2014).

[3] First tuning of HERWIG/JIMMY to ATLAS data, ATL-PHYS-PUB-2010-014 (2010).

[4] A. Kumar *et al.* (The JETSCAPE Collaboration), Jetscape framework: $p + p$ results, Phys. Rev. C **102**, 054906 (2020).

[5] J. D. Bjorken, Energy loss of energetic partons in quark-gluon plasma: possible extinction of high p(t) jets in hadron-hadron collisions, FERMILAB-PUB-82-059-THY (1982).

[6] D. A. Appell, Jets as a probe of quark-gluon plasmas, Phys. Rev. D: Part. Fields **33**, 717 (1986).

[7] J. P. Blaizot and L. D. McLerran, Jets in expanding quark-gluon plasmas, Phys. Rev. D: Part. Fields **34**, 2739 (1986).

[8] X.-N. Wang and M. Gyulassy, Gluon shadowing and jet quenching in $A + A$ collisions at $\sqrt{s} = 200A$ GeV, Phys. Rev. Lett. **68**, 1480 (1992).

[9] M. Gyulassy and X.-n. Wang, Multiple collisions and induced gluon bremsstrahlung in QCD, Nucl. Phys. B **420**, 583 (1994).

[10] R. Baier, Y. L. Dokshitzer, A. H. Mueller, S. Peigne, and D. Schiff, Radiative energy loss and $p_\perp$ broadening of high-energy partons in nuclei, Nucl. Phys. B **484**, 265 (1997).

[11] B. G. Zakharov, Radiative energy loss of high-energy quarks in finite size nuclear matter and quark-gluon plasma, JETP Lett. **65**, 615 (1997).

[12] U. A. Wiedemann, Gluon radiation off hard quarks in a nuclear environment: opacity expansion, Nucl. Phys. B **588**, 303 (2000).

[13] M. Gyulassy, P. Lévai, and I. Vitev, Jet quenching in thin quark gluon plasmas. I: formalism, Nucl. Phys. B **571**, 197 (2000).

[14] X. Guo and X.-N. Wang, Multiple scattering, parton energy loss and modified fragmentation functions in deeply inelastic eA scattering, Phys. Rev. Lett. **85**, 3591 (2000).

[15] S. Jeon and G. D. Moore, Energy loss of leading partons in a thermal QCD medium, Phys. Rev. C **71**, 034901 (2005).

[16] P. B. Arnold, G. D. Moore, and L. G. Yaffe, Photon and gluon emission in relativistic plasmas, J. High Energy Phys. 06 (2002) 030.

[17] M. Djordjevic and M. Gyulassy, Heavy quark radiative energy loss in QCD matter, Nucl. Phys. A **733**, 265 (2004).

[18] M. Djordjevic and U. W. Heinz, Radiative energy loss in a finite dynamical QCD medium, Phys. Rev. Lett. **101**, 022302 (2008).

[19] A. Majumder, Hard collinear gluon radiation and multiple scattering in a medium, Phys. Rev. D **85**, 014023 (2012).

[20] A. Majumder and M. Van Leeuwen, The theory and phenomenology of perturbative QCD based jet quenching, Prog. Part. Nucl. Phys. **66**, 41 (2011).

[21] S. Cao and X.-N. Wang, Jet quenching and medium response in high-energy heavy-ion collisions: a review, Rep. Prog. Phys. **84**, 024301 (2021).

[22] M. Connors, C. Nattrass, R. Reed, and S. Salur, Jet measurements in heavy ion physics, Rev. Mod. Phys. **90**, 025005 (2018).

[23] A. Majumder, A comparative study of jet-quenching schemes, J. Phys. G **34**, S377 (2007).

[24] A. Majumder and C. Shen, Suppression of the high p_T charged hadron R_{AA} at the LHC, Phys. Rev. Lett. **109**, 202301 (2012).

[25] A. Majumder, The in-medium scale evolution in jet modification, arXiv:0901.4516.

[26] W. T. Deng and X.-N. Wang, Multiple parton scattering in nuclei: Modified DGLAP evolution for fragmentation functions, Phys. Rev. C **81**, 024902 (2010).

[27] X.-N. Wang and X. f. Guo, Multiple parton scattering in nuclei: parton energy loss, Nucl. Phys. A **696**, 788 (2001).

[28] G.-Y. Qin and A. Majumder, Perturbative QCD description of heavy and light flavor jet quenching, Phys. Rev. Lett. **105**, 262301 (2010).

[29] R. Baier, Jet quenching, Nucl. Phys. A **715**, 209c (2003).

[30] A. Kumar, A. Majumder, and J. H. Weber, Jet transport coefficient $\hat{q}$ in lattice QCD, Phys. Rev. D **106**, 034505 (2022).

[31] A. Peshier, The QCD collisional energy loss revised, Phys. Rev. Lett. **97**, 212301 (2006).

[32] A. Majumder, Elastic energy loss and longitudinal straggling of a hard jet, Phys. Rev. C **80**, 031902(R) (2009).

[33] S. Turbide, C. Gale, S. Jeon, and G. D. Moore, Energy loss of leading hadrons and direct photon production in evolving quark-gluon plasma, Phys. Rev. C **72**, 014906 (2005).

[34] G.-Y. Qin, J. Ruppert, C. Gale, S. Jeon, and G. D. Moore, Jet energy loss, photon production, and photon-hadron correlations at RHIC, Phys. Rev. C **80**, 054909 (2009).

[35] J. H. Putschke *et al.*, The JETSCAPE framework, arXiv:1903.07706.

[36] https://github.com/JETSCAPE/JETSCAPE.

[37] D. Everett *et al.* (JETSCAPE Collaboration), Multisystem Bayesian constraints on the transport coefficients of QCD matter, Phys. Rev. C **103**, 054904 (2021).

[38] D. Everett *et al.* (JETSCAPE Collaboration), Phenomenological constraints on the transport properties of QCD matter with data-driven model averaging, Phys. Rev. Lett. **126**, 242301 (2021).

[39] S. Cao *et al.* (JETSCAPE Collaboration), Determining the jet transport coefficient $\hat{q}$ from inclusive hadron suppression measurements using Bayesian parameter estimation, Phys. Rev. C **104**, 024905 (2021).

[40] A. Andronic, F. Arleo, R. Arnaldi, A. Beraudo, E. Bruna, D. Caffarri, Z. C. del Valle, J. G. Contreras, T. Dahms, A. Dainese *et al.*, Heavy-flavour and quarkonium production in the LHC era: from proton–proton to heavy-ion collisions, Eur. Phys. J. C **76**, 107 (2016).

[41] J. E. Bernhard, J. S. Moreland, and S. A. Bass, Bayesian estimation of the specific shear and bulk viscosity of quark–gluon plasma, Nat. Phys. **15**, 1113 (2019).

[42] G. Nijs, W. van der Schee, U. Gürsoy, and R. Snellings, Bayesian analysis of heavy ion collisions with the heavy ion computational framework trajectum, Phys. Rev. C **103**, 054909 (2021).

[43] M. R. Heffernan, C. Gale, S. Jeon, and J.-F. Paquet, Bayesian quantification of strongly-interacting matter with color glass condensate initial conditions, arXiv:2302.09478.

[44] D. Liyanage, O. Sürer, M. Plumlee, S. M. Wild, and U. Heinz, Bayesian calibration of viscous anisotropic hydrodynamic simulations of heavy-ion collisions, Phys. Rev. C **108**, 054905 (2023).

[45] F.-L. Liu, X.-Y. Wu, S. Cao, G.-Y. Qin, and X.-N. Wang, Constraining the equation of state with heavy quarks in the quasi-particle model of QCD matter, Phys. Lett. B **848**, 138355 (2024).

[46] J. E. Matheson and R. L. Winkler, Scoring rules for continuous probability distributions, Manag. Sci. **22**, 1087 (1976).

[47] R. L. Winkler, Evaluating probabilities: Asymmetric scoring rules, Manag. Sci. **40**, 1395 (1994).

[48] T. Gneiting and A. E. Raftery, Strictly proper scoring rules, prediction, and estimation, J. Am. Stat. Assoc. **102**, 359 (2007).

[49] A. P. Dawid and M. Musio, Theory and applications of proper scoring rules, METRON **72**, 169 (2014).

[50] A. Majumder, Incorporating space-time within medium-modified jet event generators, Phys. Rev. C **88**, 014909 (2013).

[51] S. Cao and A. Majumder, Nuclear modification of leading hadrons and jets within a virtuality ordered parton shower, Phys. Rev. C **101**, 024903 (2020).

[52] R. Abir and A. Majumder, Drag-induced radiative energy loss from semihard heavy quarks, Phys. Rev. C **94**, 054902 (2016).

[53] W. Fan *et al.* (JETSCAPE Collaboration), Multiscale evolution of charmed particles in a nuclear medium, Phys. Rev. C **107**, 054901 (2023).

[54] T. Luo, S. Cao, Y. He, and X.-N. Wang, Multiple jets and γ-jet correlation in high-energy heavy-ion collisions, Phys. Lett. B **782**, 707 (2018).

[55] A. Kumar *et al.* (JETSCAPE Collaboration), Inclusive jet and hadron suppression in a multistage approach, Phys. Rev. C **107**, 034911 (2023).

[56] Y. Tachibana *et al.* (JETSCAPE Collaboration), Hard jet substructure in a multi-stage approach, arXiv:2301.02485 [hep-ph].

[57] Y. He, T. Luo, X.-N. Wang, and Y. Zhu, Linear Boltzmann transport for jet propagation in the quark-gluon plasma: Elastic processes and medium recoil, Phys. Rev. C **91**, 054908 (2015); **97**, 019902(E) (2018).

[58] S. Caron-Huot, Transverse diffusion ("$\hat{q}$") at weak coupling, Nucl. Phys. A **820**, 115c (2009).

[59] A. Kumar, A. Majumder, and C. Shen, Energy and scale dependence of $\hat{q}$ and the "jet puzzle", Phys. Rev. C **101**, 034908 (2020).

[60] C. Shen, Z. Qiu, H. Song, J. Bernhard, S. Bass, and U. Heinz, The iEBE-VISHNU code package for relativistic heavy-ion collisions, Comput. Phys. Commun. **199**, 61 (2016).

[61] S. A. Bass, M. Belkacem, M. Bleicher, M. Brandstetter, L. Bravina, C. Ernst, L. Gerland, M. Hofmann, S. Hofmann, J. Konopka *et al.*, Microscopic models for ultrarelativistic heavy ion collisions, Prog. Part. Nucl. Phys. **41**, 255 (1998).

[62] M. Bleicher, E. Zabrodin, C. Spieles, S. A. Bass, C. Ernst, S. Soff, L. Bravina, M. Belkacem, H. Weber, H. Stöcke *et al.*, Relativistic hadron-hadron collisions in the ultra-relativistic quantum molecular dynamics model, J. Phys. G **25**, 1859 (1999).

[63] J. S. Moreland, J. E. Bernhard, and S. A. Bass, Alternative ansatz to wounded nucleon and binary collision scaling in high-energy nuclear collisions, Phys. Rev. C **92**, 011901(R) (2015).

[64] J. E. Bernhard, Bayesian parameter estimation for relativistic heavy-ion collisions, Ph.D. thesis, Duke University, 2018; arXiv:1804.06469.

[65] W. Fan, Multi stage heavy quark transport in ultra-relativistic heavy-ion collisions, Ph.D. thesis, Duke University, 2022; arXiv:2207.12452.

[66] J. Noronha-Hostler, B. Betz, J. Noronha, and M. Gyulassy, Event-by-event hydrodynamics + jet energy loss: A solution to the $R_{AA} \otimes v_2$ puzzle, Phys. Rev. Lett. **116**, 252301 (2016).

[67] H. Song and U. W. Heinz, Causal viscous hydrodynamics in 2 + 1 dimensions for relativistic heavy-ion collisions, Phys. Rev. C **77**, 064901 (2008).

[68] A. Bazavov *et al.* (HotQCD Collaboration), Equation of state in (2 + 1)-flavor QCD, Phys. Rev. D **90**, 094503 (2014).

[69] M. McNelis, D. Everett, and U. Heinz, Particlization in fluid dynamical simulations of heavy-ion collisions: The iS3D module, Comput. Phys. Commun. **258**, 107604 (2021).

[70] P. Huovinen and H. Petersen, Particlization in hybrid models, Eur. Phys. J. A **48**, 171 (2012).

[71] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, 2006), Vol. 2.

[72] Y. Xu, J. E. Bernhard, S. A. Bass, M. Nahrgang, and S. Cao, Data-driven analysis for the temperature and momentum dependence of the heavy-quark diffusion coefficient in relativistic heavy-ion collisions, Phys. Rev. C **97**, 014907 (2018).

[73] W. Ke and X.-N. Wang, QGP modification to single inclusive jets in a calibrated transport model, J. High Energy Phys. 05 (2021) 041.

[74] V. Khachatryan *et al.* (The CMS Collaboration), Charged-particle nuclear modification factors in PbPb and pPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, J. High Energy Phys. 04 (2017) 039.

[75] S. Acharya *et al.* (The ALICE Collaboration), Measurement of D^0, D^+, D^{*+} and D_s^+ production in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, J. High Energy Phys. 10 (2018) 174.

[76] M. Aaboud *et al.* (The ATLAS Collaboration), Measurement of the nuclear modification factor for inclusive jets in Pb+Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV with the ATLAS detector, Phys. Lett. B **790**, 108 (2019).

[77] S. Acharya *et al.* (ALICE Collaboration), Measurements of inclusive jet spectra in pp and central Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, Phys. Rev. C **101**, 034911 (2020).

[78] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning* (Springer, New York, NY, 2009), Vol. 2.

[79] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley-Interscience, Hoboken, NJ, 2006).

[80] S. Oladyshkin and W. Nowak, The connection between Bayesian inference and information theory for model selection, information gain and experimental design, Entropy **21**, 1081 (2019).

[81] A. M. Sirunyan *et al.* (The CMS Collaboration), First measurement of large area jet transverse momentum spectra in heavy-ion collisions, J. High Energy Phys. 05 (2021) 284.

[82] R. Ehlers (JETSCAPE Collaboration), Bayesian analysis of QGP jet transport using multi-scale modeling applied to inclusive hadron and reconstructed jet data, Acta Phys. Pol. B Proc. Suppl. **16**, 62 (2023).

[83] R. Pordes, D. Petravick, B. Kramer, D. Olson, M. Livny, A. Roy, P. Avery, K. Blackburn, T. Wenaus, F. Würthwein *et al.*, The open science grid, J. Phys.: Conf. Ser. **78**, 012057 (2007).

[84] I. Sfiligoi, D. C. Bradley, B. Holzman, P. Mhashilkar, S. Padhi, and F. Wurthwein, The pilot way to grid resources using glidein-WMS, WRI World Congress **2**, 428 (2009).