

Bayesian Parameter Estimation of Relativistic Heavy Ion Collisions Simulation with Viscous Anisotropic Hydrodynamics Modeling

Dananjaya Liyanage,¹ Özge Sürer,² Matthew Plumlee,^{2,3} and Ulrich Heinz¹

¹*Department of Physics, The Ohio State University, Columbus, OH 43210, USA*

²*Northwestern-Argonne Institute of Science and Engineering (NAISE), Northwestern University, Evanston, IL 60208, USA*

³*Industrial Engineering and Management Sciences Department, Northwestern University, Evanston, IL 60208, USA*

State-of-the-art hydrodynamic simulation models for relativistic heavy ion collisions are only applicable after $O(1)$ fm/c after the collision due to large pressure gradients that are present at early time. As a solution, pre-hydrodynamic stage which models the early stage evolution as a conformal, weakly interacting gas is used before the hydrodynamic stage. The transition from pre-hydrodynamic to the hydrodynamic stage is discontinuous and introduce a considerable theoretical ambiguity to the model. Recently a novel hydrodynamic model, Viscous Anisotropic Hydrodynamic (VAH) which can handle large pressure anisotropies has been introduced as an alternative solution. VAH is applicable at very early times of the collision and it smoothly matches to conventional second-order viscous hydrodynamics at late times. In this work we present a Bayesian parameter estimation study for VAH model using the experimental data for Pb-Pb collisions at LHC ($\sqrt{s_{NN}} = 2.76$ TeV). We find that the VAH model can fit the experimental data very well and we present the most probable values for the model parameters as constrained by the experimental data.

I. INTRODUCTION

The atomic nucleus is made up of fundamental particles called quarks and gluons. The study of quarks poses a unique challenge because it can not be found as an isolated particle in nature. This fundamental property of the quarks is called Quark Confinement. Thus, scientists have to resort to other methods like studying systems that consist of multiple quarks [1, 2]. Quark Gluon Plasma (QGP) is one such exotic phase of matter which filled the universe right after the big bang before it cooled down to produce the hadronic matter that we observe today [3].

The largest experimental facility in the world, located in Europe which is called the large hadron collider (LHC), collide nuclei in relativistic high speeds to create QGP in laboratories [4]. At these high energies all of the baryonic nuclear matter of nuclei pass through each other without stopping and only a fraction of the total energy is deposited in the mid region of the collision [5]. The matter created right after the collision is highly out of equilibrium and has only weak interactions. Then it expands and cools down to produce slightly out of equilibrium quark gluon plasma. The QGP further expands and reduce its temperature until it reaches the freeze out temperature where the QGP gets converted into hadrons. Then these hadrons fly out towards the detectors while decaying and interacting with other hadrons.

The QGP produced in the relativistic heavy ion collisions has extremely short lifetime ($\sim 10^{-23}$ s) and size ($\sim 10^{-14}$ m). Due to this the QGP can only be studied from the final hadrons that it emits at the freeze out. A central PbPb collision produces on average 30,000 particles at the LHC and inferring the properties of the QGP from the detected particles and the information about the four momenta is formidable challenge which require advanced simulation and statistical techniques [6].

There have been many large scale simulation studies carried out recently to infer the properties of QGP using Bayesian statistical techniques [7–17]. One of the major limitations of

these current simulations for relativistic heavy ion collisions is the breaking down of the hydrodynamic stage of the simulation due to large pressure gradients at early stage of the collision. To circumvent this issue, all of the previous work used a pre-hydrodynamic stage which evolve the QGP first as a conformal, weakly interacting gas up-to a time $O(1)$ fm/c where the pressure gradients becomes weaker and second order viscous hydrodynamic codes become applicable. This approach introduces a considerable theoretical ambiguity to the model due to discontinuity of the Equation of state in the (conformal) pre-hydrodynamic stage and (non-conformal) hydrodynamic stage. This discontinuity can give rise to many unphysical effects including a large bulk viscosity values in the hydrodynamic stage [18].

In this work we study a novel simulation model for relativistic heavy-ion collisions without aforementioned discontinuity. We use the JETSCAPE simulation framework [17] for relativistic heavy ion collisions but with one important modification. We replace the second order viscous hydrodynamic stage with a novel hydrodynamic module called viscous anisotropic hydrodynamics [19] and remove the conformal pre-hydrodynamic stage. Viscous anisotropic hydrodynamic can handle large pressure gradients compared to traditional hydrodynamic modules and thus allows us to start the hydrodynamic stage at a very early time (0.5 fm/c) and then it matches smoothly to the second order viscous hydrodynamic stage at a later time.

II. OVERVIEW OF THE HYBRID MODEL FOR RELATIVISTIC HEAVY-ION COLLISIONS

The dynamical evolution of relativistic heavy ion collisions involve physics ranging from small to large length scales and demands a multistage modeling approach with different modules simulating different stages of the evolution. We discuss each stage of the evolution of relativistic heavy ion collisions and how we simulate them below.

A. Initial conditions

Our simulation for relativistic heavy ion collision starts right after the two incoming nuclei collide. In the lab frame the nuclei travel at a speed closer to light before the collision and due to the Lorentz length contraction they are spread out on the transverse plane with a small width along the collision axis. Due to gluon saturation at these high energies the nuclei consist of mostly gluons and when the pancake shaped nuclei collide other most of the Byronic matter pass through each other without stopping. The interactions of gluons between the two nuclei create deposit large amount of energy in the mid-rapidity region. Our starting point is the phenomenological model T_{RENT}O that stochastically simulate the shape and size of the energy deposition at mid-rapidity right after the baryon matter in the nuclei traverse each other. Here we summarize how T_{RENT}O models the initial energy density for relativistic heavy ion collisions [20, 21].

First, the distribution of a single nucleon inside the nucleus is assumed to be a three-dimensional Gaussian with a nucleon width parameter w .

$$\rho(\mathbf{x}_\perp) = \int_{-\infty}^{\infty} \frac{dz}{(2\pi w^2)^{3/2}} \exp\left(-\frac{\mathbf{x}_\perp^2 + z^2}{2w^2}\right). \quad (1)$$

Then the position of nucleons inside each nucleus is sampled using a Wood-Saxon distribution for both the nuclei. An impact parameter is then sampled from a uniform distribution assuming the collisions are minimum-biased. Next the collision probability is sampled for each pairwise collision of the two nuclei by taking into account the proton-proton collision cross section at the specified center of mass energy. The nucleons that do not undergo any collisions are then thrown away. The remaining nucleons in each nucleus are labeled as participants and their distribution functions are integrated along the longitudinal direction to get the participant density function. The fluctuated thickness function for A then reads

$$T_A(\mathbf{x}_\perp) = \sum_{i=1}^{N_{part}} \int_{-\infty}^{\infty} dz \gamma_i \rho(\mathbf{x} - \mathbf{x}_i). \quad (2)$$

Here γ_i are independent random weights sampled from a gamma distribution with unit mean and a standard deviation of σ_k . These additional fluctuations capture the large multiplicity fluctuations observed in minimum-biased proton-proton collisions. x_i are participant nucleon positions in A. $T_B(\mathbf{x}_\perp)$ follows analogously.

The participant density function has the information about nucleon density at each point on the transverse plane to the collision axis. Next, the following phenomenological model is derived with the T_{RENT}O ansatz that the energy deposition at mid-rapidity after a time τ_0 is the harmonic mean of the two nuclei thickness functions multiplied by a constant N .

$$T_R(\mathbf{x}_\perp; p) = \left(\frac{T_A^p(\mathbf{x}_\perp) + T_B^p(\mathbf{x}_\perp)}{2} \right)^{1/p} \quad (3)$$

$$\epsilon(\mathbf{x}_\perp) = \frac{1}{\tau_0} \frac{dE}{d\eta d^2\mathbf{x}_\perp} = \frac{1}{\tau_0} N T_R(\mathbf{x}_\perp; p) \quad (4)$$

In this work we have assumed the longitudinal boost invariance for the collision system following Ref. [22] and start the viscous anisotropic hydrodynamics at $\tau_0 = 0.05 \text{ fm}/c$. We use this initial condition model to simulate Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76 \text{ TeV}$. The Wood-Saxon parameters used for Pb are $R = 6.62 \text{ fm}$ and $\alpha = 0.546 \text{ fm}$. The nucleon width (w), the Normalization (N), harmonic mean parameter (p) and standard deviation of the gamma distribution (σ_k) are considered to be parameters of interests that we would find from experimental data. In addition we also consider d_{min} parameter that fix the minimum distance between nucleons inside the nucleus as a free parameter that we calibrate later on using the experimental data.

B. Viscous Anisotropic Hydrodynamics

Hydrodynamic is an effective theory that can accurately model macroscopic properties of many dynamical systems found in nature, ranging from small scale systems like cold atomic gases to large scale galaxy formations. Hydrodynamic modeling has also been extensively used in relativistic heavy ion collisions for more than a decade [23]. It is the workhorse in many of the simulations of dynamical evolution of QGP [10, 15, 22].

Taking the macroscopic degrees of freedom to be the energy density ϵ and u_μ to be the fluid four velocity (i.e. the four-velocity of the local rest frame (LRF) relative to the global frame) energy momentum tensor for ideal hydrodynamics (zeroth order in gradients of the macroscopic degrees of freedom) found to be,

$$T^{\mu\nu} = \epsilon u^\mu u^\nu - p \Delta^{\mu\nu}. \quad (5)$$

Here $\Delta^{\mu\nu} = g^{\mu\nu} - u^\mu u^\nu$ is the space like projector and the metric, $g^{\mu\nu}$ signature is taken to be "mostly minus" (+,-,-,-). p is the equilibrium pressure in the LRF which can be found as a function of the energy density, using the equation of state for the underlying microscopic system. In ideal hydrodynamics it is assumed that the equilibrium pressure is isotropic and no spatial direction is special in the energy momentum tensor.

This assumption can be relaxed by considering higher order gradient corrections to the energy momentum tensor. The second order viscous hydrodynamics consider up to two orders of gradients of the macroscopic degrees of freedom (ϵ, u^μ) to construct the energy momentum tensor as shown below.

$$T^{\mu\nu} = \epsilon u^\mu u^\nu - (p + \Pi) \Delta^{\mu\nu} + \pi^{\mu\nu} \quad (6)$$

The bulk stress Π and shear stress tensor $\pi^{\mu\nu}$ are defined using the first and second order gradients. In the LRF the pressure is no longer isotropic and the shear stress tensor determines the size of the pressure anisotropy. The second order hydrodynamic relies on the assumption that these pressure

anisotropies introduced by the higher order gradient corrections are small compared to the energy density and equilibrium pressure. Thus second order viscous anisotropic hydrodynamics is not applicable for situations with large pressure anisotropies similar to what we find at the early stage of relativistic heavy ion collisions.

The momentum distribution of the QGP is highly anisotropic right after the relativistic heavy ion collision [24]. Initially, the medium has a large expansion rate along the longitudinal beam direction and can be approximated by boost invariant longitudinal velocity profile (Bjorken flow [25]). The transverse expansion rate of the medium is initially small and gradually builds up due to transverse pressure gradients. This introduce large pressure anisotropy between transverse and longitudinal directions to the beam.

State of the art hybrid cascade simulation models for relativistic heavy ion collisions circumvent this issue by introducing a free streaming stage to evolve the QGP before feeding it to the hydrodynamic simulation. The free streaming model assume the QGP to be consist of massless particles that do not interact and expand isotropically with the speed of light. In this work, we discard this overly simplified picture about the QGP matter in the state of the art simulation models for relativistic heavy ion collisions by using viscous anisotropic hydrodynamics.

The formulation of viscous anisotropic hydrodynamic starts with further decomposing the spatial projector, $\Delta^{\mu\nu}$ along the beam direction (z^μ) and perpendicular to the beam direction ($\Xi^{\mu\nu}$) as $\Delta^{\mu\nu} = \Xi^{\mu\nu} - z^\mu z^\nu$ in the LRF [26]. The Energy momentum tensor for viscous anisotropic hydrodynamic is found to be as following,

$$T^{\mu\nu} = \epsilon u^\mu u^\nu + P_L z^\mu z^\nu - P_\perp \Xi^{\mu\nu} + 2W_{\perp z}^{(\mu} z^{\nu)} + \pi_\perp^{\mu\nu} \quad (7)$$

The evolution equations for the energy density and the flow velocity is obtained from the conservation laws for energy and momentum.

$$\partial_\mu T^{\mu\nu} = 0 \quad (8)$$

The equilibrium pressure that appear in the equations is found using the lattice QCD calculation provided by the HotQCD collaboration [27].The dynamical evolution for the source terms in the conservation equations, $P_L, P_\perp, W_{\perp z}^{(\mu} z^{\nu)}, \pi_\perp^{\mu\nu}$ are derived assuming a weakly-coupled dilute fluid whose microscopic physics can be described by the relativistic Boltzmann-Vlasov equation with a medium-dependent mass in Ref. [28]. There the bulk relaxation time and shear relaxation time are parameterized as follow.

$$\tau_\pi = \frac{\eta}{s\beta_\pi} \quad \tau_\Pi = \frac{\zeta}{s\beta_\Pi} \quad (9)$$

β_π, β_Π and anisotropic transport coefficients are computed with the quasiparticle kinetic theory model discussed in

Ref. [28]. β_π and β_Π are the temperature dependent isotropic thermodynamic integrals that are defined in Eq.82 in Ref. [28].

The temperature dependent specific shear viscosity, $\frac{\eta}{s}(T)$ and specific bulk viscosity, $\frac{\zeta}{s}(T)$ is parameterized following the Ref. [22].

$$\frac{\eta}{s}(T) = \max \left[\frac{\eta}{s} \Big|_{\text{lin}}(T), 0 \right], \quad (10)$$

with

$$\begin{aligned} \frac{\eta}{s} \Big|_{\text{lin}}(T) &= a_{\text{low}}(T-T_\eta) \Theta(T_\eta-T) + (\eta/s)_{\text{kink}} \\ &\quad + a_{\text{high}}(T-T_\eta) \Theta(T-T_\eta). \end{aligned} \quad (11)$$

$$\begin{aligned} \frac{\zeta}{s}(T) &= \frac{(\zeta/s)_{\text{max}} \Lambda^2}{\Lambda^2 + (T - T_\zeta)^2}, \\ \Lambda &= w_\zeta [1 + \lambda_\zeta \text{sign}(T - T_\zeta)]. \end{aligned} \quad (12)$$

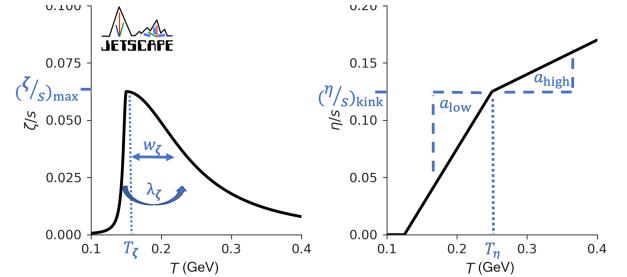


FIG. 1. We use the same pasteurization for temperature dependence of specific shear and bulk viscosities as shown in the above figure from Ref. [22]. (Dan: Do we need this figure?)

Apart from these eight parameters related to the viscosity there is a one parameter that we will infer from the experimental data later in this work. R is the initial longitudinal to transverse pressure ratio that can be tuned at the initialization of viscous anisotropic hydrodynamic stage.

$$\begin{aligned} P_L &= \frac{3RP_{eq}}{2+R} \\ P_\perp &= \frac{3P_{eq}}{2+R} \end{aligned} \quad (13)$$

where $R \in (0.3, 1)$. By using this parameterization we also assume that the initial bulk pressure is $\Pi = \frac{2P_\perp + P_L}{3} + P_{eq} = 0$. Furthermore we assume initial fluid velocity is static ($u = (1, 0, 0, 0)$) and the residual shear stresses are $W_{\perp z}^{(\mu} z^{\nu)} = \pi_\perp^{\mu\nu} = 0$ to initialize the energy momentum tensor using the Eq. 7 and the TRENTo initial energy profile.

C. Particilization

Thus far we have discussed how the QGP is modeled as an anisotropically expanding liquid starting with a phenomenological initial energy deposition model right after the collisions. As the QGP fluid expands and cools down it reaches a critical temperature where the fluid cells get converted into particles. We keep this temperature as a free parameter in our model and call it the switching temperature T_{sw} . We use experimental data to infer the value of the switching temperature later in this work. The freezeout fluid cells that reach the switching temperature are found using freezeout finder at every two time steps of the VAH evolution. The switching hypersurface that has the space time coordinate and the volume element $d^3\sigma_\mu$ of all fluid cells that reach this constant switching temperature is found using the code CORNELIUS [29]. The hypersurface is passed to iS3Dhadron sampler [30, 31]. There the Cooper-Frye [32] prescription for particilization is used to convert all the energy and momentum of the fluid into different hadron types and momenta on the switching hypersurface.

The Lorentz-invariant particle momentum is given by,

$$P^0 \frac{dN_i}{d^3P} = \frac{g_i}{(2\pi)^3} \int_{\Sigma} d^3\sigma_\mu P^\mu f_i(X; P). \quad (14)$$

The four-momentum is P and the phase distribution function for the particle species i is given by $f_i(X, P)$. The switching hypersurface is Σ with a normal vector to the surface σ_μ .

For a locally equilibrated hadron resonance gas this distribution function takes the form

$$f_{eq,i}(x, p) = \frac{g_i}{\exp[\frac{p \cdot u(x)}{T(x)} - \alpha_i(x)] + \Theta_i}. \quad (15)$$

Where $g_i = 2s_i + 1$ are the spin degeneracy for species i . $u^\mu(x)$ is the fluid velocity, $T(x)$ is the temperature, $\alpha_B(x) = \mu_B(x)/T(x)$ is the baryon chemical potential-to-temperature ratio and $\Theta_i \in [-1, 1]$ accounts for the quantum statistics of fermions and bosons, respectively.

At the switching hypersurface, Σ , the QGP liquid still has large dissipation effects present and can not be modeled as a locally equilibrated hadron resonance gas. Therefore one can assume different forms for the out of local equilibrium distribution function. A study about the theoretical uncertainty introduced by these different assumptions and its importance in Bayesian parameter inference can be found in Ref. [14]. In this work we only use one of the particilization model due to limitations of computational resources and leave accessing the theoretical uncertainty due to other particilization models for future work. We choose a particilization model that assume out of local-equilibrium distribution function as a momentum-anisotropic leading order term and a residual correction, $f_i(x, p) = f_{a,i}(x, p) + \delta f_i(x, p)$.¹ $\delta f_i(x, p)$ can be derived using a linearization approach such as Chapman-Enskog

expansion combined with the relaxation time approximation (RTA) of the Boltzmann equation [33]. Unfortunately such linearization approaches are known to turn the total distribution functions negative at high momenta when large dissipative effects are present [34]. The root cause of such issues lie in the truncation of the underlying expansions to the first order in the dissipative flows. One way to circumvent such negative probability arising in the distribution function is to modify it to include the viscous corrections as arguments inside the exponential. Such approaches are called "modified equilibrium" prescriptions for the viscous corrections [34, 35].

In this work we use such "modified equilibrium" distribution function that is called Pratt-Torrieri-McNelis-Anisotropic (PTMA) distribution.

$$f_{a,i}^{\text{PTMA}} = \frac{\mathcal{Z} g_i}{\exp[\frac{\sqrt{P'^2 + m_i^2}}{\Lambda}] + \Theta_i} \quad (16)$$

Complete derivation of the PTMA distribution function and full details can be found in Ref. [33].

D. Hadronic Afterburner

A hadronic afterburner allow sampled hadrons from the switching hypersurface to kinetically decay, scatter and form resonances until they completely decouple from each other and dynamically reach chemical and kinetic freezeout [36–43]. In this work we use the Boltzmann transport code SMASH [44].

To get sufficient particle statistics with minimal computational cost, we oversample the switching hypersurface similarly to previous work using hybrid cascade models for relativistic heavy ion collisions [7–17]. We oversample each hypersurface a maximum number of 1000 or until we have 10^5 number of hadrons in the ensemble. Finally, the experimental observables that are measured by ALICE detector are calculated using the final-state particles and their momentum given by the SMASH output [45–48].

III. INTRODUCTION TO BAYESIAN PARAMETER ESTIMATION

We represent the simulation model with a mathematical function $\mathbf{y}_{\text{sim}}(\cdot)$ that takes values of parameters $\mathbf{x} = (x_1, \dots, x_q) \in \mathcal{X}$ and returns output $\mathbf{y}_{\text{sim}}(\mathbf{x}) \in \mathbb{R}^d$. A Bayesian parameter estimation process uses observations from the experiments, denoted by vector $\mathbf{y}_{\text{exp}} = (y_{\text{exp},1}, \dots, y_{\text{exp},d})$ to constrain the uncertainty on the parameters \mathbf{x} . In this paper, we study the Bayesian calibration of relativistic heavy ion collisions to align the simulation outputs

¹ Here, we prefer to use the anisotropic distribution function to be consistent

with the VAH formulation and specifically how the evolution of the source terms are modeled using anisotropic distribution function. VAH in general can also be used with any other type of particilization methods.

$\mathbf{y}_{\text{sim}}(\mathbf{x})$ with the experimental data \mathbf{y}_{exp} . This is done using a statistical model of the form

$$\mathbf{y}_{\text{exp}} = \mathbf{y}_{\text{sim}}(\mathbf{x}) + \boldsymbol{\epsilon}, \quad (17)$$

where $\boldsymbol{\epsilon} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$ denotes the residual error following a multivariate normal distribution (MVN) with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. In this work, we consider $d = 98$ experimental observables. The experimental observables also have experimental uncertainties associated with them. Our simulations are also stochastic and thus the simulation outputs also have uncertainties. Hence, finding the model parameter values that would fit the experimental data (i.e., solving the inverse problem) taking into account all of the uncertainties requires an advanced probabilistic framework.

The popular frequentist interpretation of probability is the number of occurrences of a certain outcome divided by the number of time an experiment is repeated when experiment is done large number of times. Bayesian interpretation of probability provides a simpler and more intuitive definition which does not involve repeated experiments. It defines probability as degree of belief about an hypothesis considering all available information [49]. Under this interpretation of the probability, Bayes' rule becomes a powerful inference tool, and it can be written as

$$\mathcal{P}(\mathbf{x}|\mathbf{y}_{\text{exp}}) = \frac{\mathcal{P}(\mathbf{y}_{\text{exp}}|\mathbf{x})\mathcal{P}(\mathbf{x})}{\mathcal{P}(\mathbf{y}_{\text{exp}})}. \quad (18)$$

The term $\mathcal{P}(\mathbf{x}|\mathbf{y}_{\text{exp}})$ on the left hand side of Eq. (18) is called the posterior (short for “the posterior probability density”), which is the probability of the model parameter \mathbf{x} to take a particular value given the experimental data \mathbf{y}_{exp} . $\mathcal{P}(\mathbf{x}|\mathbf{y}_{\text{exp}})$ is the main quantity of interest in Bayesian parameter inference. Here, $\mathcal{P}(\mathbf{x})$ represents the prior probability for the parameters to take value \mathbf{x} given the information from previous experiments and other theoretical bounds. $\mathcal{P}(\mathbf{y}_{\text{exp}}|\mathbf{x})$ is the likelihood function, describing the probability that model output with a given set of model parameters \mathbf{x} agrees with the experimental data \mathbf{y}_{exp} . Since $\boldsymbol{\epsilon} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$ in Eq. (17), we can write the likelihood $\mathcal{P}(\mathbf{y}_{\text{exp}}|\mathbf{x})$

$$\frac{1}{\sqrt{|2\pi\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2}(\mathbf{y}_{\text{sim}}(\mathbf{x}) - \mathbf{y}_{\text{exp}})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_{\text{sim}}(\mathbf{x}) - \mathbf{y}_{\text{exp}})\right], \quad (19)$$

where $d \times d$ matrix $\boldsymbol{\Sigma}$ represents the total uncertainty, obtained by adding the experimental and simulation uncertainties: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\text{exp}} + \boldsymbol{\Sigma}_{\text{sim}}(\mathbf{x})$.

The functional form of the posterior is generally not known analytically, in particular not for heavy ion collisions. To find the most likely range for the parameters \mathbf{x} and quantify their uncertainty requires numerical techniques for finely sampling from the posterior. This is typically achieved by using Markov Chain Monte Carlo (MCMC) techniques.² To analyze the

posterior, we need to produce thousands of samples via the MCMC techniques and the number of samples needed scales with the dimensionality of parameters and observables. To produce each sample, MCMC techniques have to evaluate the right hand side of the Eq. (18) multiple times for different values of \mathbf{x} . When the simulation model is computationally expensive evaluating the likelihood $\mathcal{P}(\mathbf{y}_{\text{exp}}|\mathbf{x})$ becomes computationally expensive. Often MCMC methods require millions of evaluations of the likelihood and thus it can make Bayesian parameter estimation computationally infeasible for expensive simulation models, i.e. relativistic heavy ion collision simulations. To overcome this computational barrier, we build emulators which are computationally cheap surrogates for the complex simulations as explained in details below [50].

IV. EMULATORS FOR RELATIVISTIC HEAVY-ION COLLISIONS

Relativistic heavy ion collision simulations are computationally expensive. For each set of model parameter values, the simulation has to run multiple times with randomly fluctuating initial conditions to produce measurements that are comparable with experimental measurements. For the JETSCAPE framework the full-model simulations for 2000 fluctuating initial conditions take on average $\mathcal{O}(1000)$ CPU hours [17]. A majority (80%) of the CPU time is spent on the hadron transport stage after partcileization; the remaining CPU time (20%) is mostly utilized by the hydrodynamic QGP evolution code. Bayesian parameter estimations require millions of evaluations of the simulation and it renders the inference for heavy-ion collision model computationally infeasible.

As a solution, we build computationally cheap surrogates which are called Gaussian Process (GP) emulators for the expensive simulation. GP emulators have to be trained on a sparse set of simulation data before they can be used to predict the simulation outputs. This step of getting the training data for the emulators is the most computationally demanding step in Bayesian parameter estimation. In this work, we employ several novel methods to significantly reduce cut the computational cost: a) we use a novel Minimum Energy Design (MED) for selecting the model parameter values to run the full simulation compared to the previously used method of Latin Hypercube Design (LHD) [51, 52]; (b) instead of getting all the stochastic simulation data at once, we follow a sequential process and get batches of simulation data with increasing accuracy; (c) we test several GP emulation methods for relativistic heavy-ion collisions and pick the best emulation method after a thorough validation. All of these approaches are being explained in details in the following.

A. Gaussian Process emulation

In this section, we provide an overview for GP emulators when the simulation model has one-dimensional output (e.g., $d = 1$). Later, we provide details in Section IV C to extend the GP emulators when $d > 1$. Let $f(\mathbf{x})$ denote the scalar simulation output at parameter point \mathbf{x} . A Gaussian

² These techniques require only relative probabilities, so the normalization $\mathcal{P}(\mathbf{y}_{\text{exp}})$ in the denominator on the right of Eq. (18) (which is independent of the parameters to be inferred) does not need to be calculated.

process is a stochastic process $\{f(\mathbf{x}) \in \mathbb{R} : \mathbf{x} \in \mathcal{X}\}$, for which any finite collection of points $f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)$ have a joint Gaussian distribution. Here we are building a GP with n training points. A GP is fully characterized by a mean function $\mu(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$ and a covariance function $k(\mathbf{x}, \mathbf{x}') = \text{Cov}[f(\mathbf{x}), f(\mathbf{x}')]$. This will be denoted as

$$f(\cdot) \sim \text{GP}(\mu(\cdot), k(\cdot, \cdot)).$$

The mean function $\mu(\mathbf{x})$ denotes the mean of the process while the covariance function controls the smoothness of its sample paths. In practice, the mean function $\mu(\cdot)$ prior to conditioning is typically set to be a constant μ . In our work, we have set it to zero. There are several popular choices for the covariance function $k(\cdot, \cdot)$, including Gaussian³, Matérn, and cubic covariances [53].

The details of GP fitting procedure is described in App. A. Once the GP is fitted, the goal is to predict the simulation output at an unseen point \mathbf{x}_{new} as follows. Let $\mathbf{K} = [k(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1}^n$ be the covariance matrix for the training data, and let n -dimensional vector \mathbf{f} be $\mathbf{f} = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)]$. Conditioning on the data \mathbf{f} (and assuming fixed parameters for the co-variance $k(\mathbf{x}, \mathbf{x}')$), the posterior distribution of f at a new point on the parameter space \mathbf{x}_{new} can be shown to be [54]

$$[f(\mathbf{x}_{\text{new}})|\mathbf{f}] \sim N(\mu^*(\mathbf{x}_{\text{new}}), \sigma^{2*}(\mathbf{x}_{\text{new}})), \quad (20)$$

where the posterior mean and variance are given by

$$\begin{aligned} \mu^*(\mathbf{x}_{\text{new}}) &= \mu + \mathbf{k}_{\text{new}}^\top (\mathbf{K} + \gamma^2 \mathbf{I}_n)^{-1} (\mathbf{f} - \mu \mathbf{1}_n) \\ \sigma^{2*}(\mathbf{x}_{\text{new}}) &= k(\mathbf{x}_{\text{new}}, \mathbf{x}_{\text{new}}) - \mathbf{k}_{\text{new}}^\top (\mathbf{K} + \gamma^2 \mathbf{I}_n)^{-1} \mathbf{k}_{\text{new}}. \end{aligned} \quad (21)$$

Here, $\mathbf{k}_{\text{new}} = [k(\mathbf{x}_{\text{new}}, \mathbf{x}_i)]_{i=1}^n$ is the covariance vector between the n existing design points of full-model runs and a new, interpolated point in the parameter space. Eqs. (20, 21) provide the basis for emulator modeling: the posterior mean $\mu^*(\mathbf{x}_{\text{new}})$ serves as the emulator model prediction at a new point \mathbf{x}_{new} , and the posterior variance $\sigma^{2*}(\mathbf{x}_{\text{new}})$ yields a quantification of emulator model uncertainty. A key appeal of GP emulators is that both their prediction and uncertainty can be efficiently computed via such closed-form expressions.

B. Generating the simulation data for the emulators

The most computationally expensive step in any Bayesian parameter estimation for relativistic heavy-ion physics is generating the simulation data necessary to train and test reliable, accurate emulators. All previous studies have used a Min-Max Latin Hypercube design (LHD) to get the design points that are needed to generate the simulation data for the emulators

[7–17]. LHD fills the multidimensional model parameter input space while making sure that each parameter is uniformly distributed with in its full range [52]. The emulators that are being built using LHD has the same level of accuracy everywhere in the parameter space.

Building emulators for Bayesian parameter estimation is a unique application of emulators. The emulators that are build for this purpose need only to accurately predict the simulation output for high posterior regions in the model parameter space. Naturally this leads to a sampling method where some regions in the input parameter space have more design points than the other and thus deviating largely from a LHD. This targeted sampling approach saves computational cost by requiring smaller number of high precision simulation data for Bayesian parameter estimation while sacrificing the emulation accuracy in the uninteresting regions in the model parameter space. This idea of sequential sampling of only interesting regions in the parameter space when building emulators is known as active learning in statistics literature (See Chapter 6 in [55] for a detailed review). It has been used in many applications including batch-sequential design for delta smelt conservation [56] and sequential design for deep GP emulators [57].

In this work, we take an active learning approach to generate the simulation data for the first time in Bayesian parameter estimation study of relativistic heavy ion collisions. Our simulation is stochastic and to calculate observables from our simulation we run the simulation multiple times with the same model parameter values but with different initial conditions. Each simulation we run with a different initial condition is called an event. When the number of events that we have for a design point increases, the accuracy of the final simulation output also increases. The design with higher number of events are computationally very expensive and we call them the high precision designs.

First, we use low accuracy design points to estimate an intermediate posterior for an unseen point. Informed by the initial design, we decide on which region in the parameter space we need to put more weight when sampling and use MED to do sequential intelligent sampling. MED selects the points that minimize the total potential energy given by [58], and the fast procedure for generating MED samples are provided in [51]. Suppose that we start with n_0 samples and their corresponding simulation outputs. Based on MED selection criterion, the $(j+1)$ th point is selected

$$\mathbf{x}_{j+1} = \arg \max_{\mathbf{x} \in \mathcal{L}} \min_{i=1:j} \mathcal{P}^{1/2q}(\mathbf{y}_{\text{exp}}|\mathbf{x}) \mathcal{P}^{1/2q}(\mathbf{y}_{\text{exp}}|\mathbf{x}_i) d(\mathbf{x}, \mathbf{x}_i) \quad (22)$$

where \mathcal{L} represents the candidate list of parameters, and $d(\mathbf{x}, \mathbf{x}_i)$ represents the Euclidean distance between \mathbf{x} and \mathbf{x}_i . In our experiments, we generate a list \mathcal{L} of parameters using LHD. In Eq. (22), the likelihood $\mathcal{P}(\mathbf{y}_{\text{exp}}|\mathbf{x})$ cannot be evaluated for an unseen point \mathbf{x} since the simulation output $\mathbf{y}_{\text{sim}}(\mathbf{x})$ has not obtained yet. However, since we can build an emulator using n_0 samples, we can estimate the likelihood for any unseen point \mathbf{x} . The following are the five batches of design points we ran the simulation in the chronological order.

- a) We first start with a Latin Hypercube design with 200

³ In the statistical literature the Gaussian function is often called a “squared-exponential”, indicated here by the superscript SE.

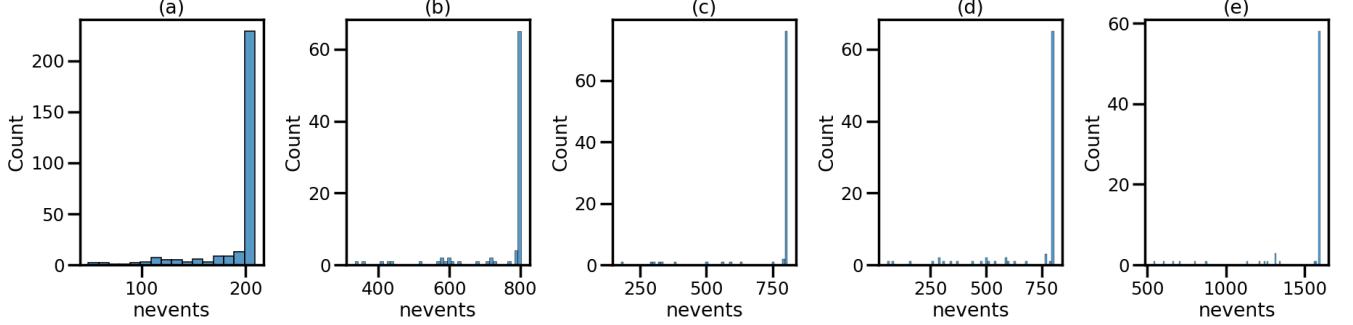


FIG. 2. Histogram of events per design in each design batch. (a) 300 design points generated using sliced LHD in batches of 50 with 200 events per each design point (b) 90 designs with 800 events per designs from LHD, (c) First batch of 90 design points from MED with 800 events per design, (d) Second batch of 90 design points from MED with 800 events per design, (e) Final batch of 70 design points from MED with 1600 events per design.

events per each design point. In total we have 300 design points generated from a Latin Hypercube design. We use this as our starting point and build emulators and then calculate a preliminary posterior.

- b) We have a 90 design points generated from a Latin Hypercube design with 800 events per design. These simulations are used as test data to check emulation accuracy.
- c) The first batch of 90 design points with 800 events per design from the Minimum Energy design.
- d) The second batch of 90 design points with 800 events per design from the Minimum Energy Design.
- e) Final batch of 70 design points with 1600 events per design from Minimum Energy Design.

The Fig. 2 show the distribution of events per design point in each of the batches from (a)-(e). When training and testing emulators we first discard the designs which have more than 5% failure rate to ensure that the simulation data is free of large errors.

C. Prepossessing and dimensional reduction of the simulation data

The Gaussian Process emulators that we use in this work can only predict a single output. Given that we have 98 observables as the simulation outputs, we would require 98 individual Gaussian Processes in total which is computationally expensive. To avoid this we first perform dimensional for the 98 observables into handful of dimensions using Principal Component Analysis (PCA).

Before we use PCA first we standardize all of our simulation training data (Combine all the batches together except (b)) to

have a zero mean and a unit variance as in the following.

$$\tilde{Y}_j^l = \frac{Y_j^l - \mu^l}{\sigma^l}, \quad (23)$$

$$\mu^l = \sum_i \frac{Y_i^l}{N_{\text{train}}}, \quad (\sigma^l)^2 = \sum_i \frac{(Y_i^l - \mu^l)^2}{N_{\text{train}}}.$$

Y_i^l is the l^{th} observable from the design i , and i is summed over all designs in the training data sets. \tilde{Y}_j^l is the standardized j^{th} design point output. We have 98 observables as outputs from the simulation and hence $l = 1, 2, \dots, 98$. We use this standardized simulation data set to perform PCA as follows [17].

Let \mathbf{Y} be a $n \times m$ matrix with elements $Y_{i,j} = \tilde{Y}_i^j$ where n is the number of observables from the simulation and m is the number of design points in the training data. Principal component analysis finds a linear transformation matrix \mathbf{A} with dimension of $m \times q$ which projects \mathbf{Y} to a subspace \mathbf{Z} by $\mathbf{Z} = \mathbf{YA}$. \mathbf{Z} is $q \times m$ dimensional and to go back to the original observable space again the transformation $\tilde{\mathbf{Y}} = \mathbf{ZA}^\top$ can be used. The q ($\leq n$) dimension is chosen by considering the difference between $\tilde{\mathbf{Y}}$ and \mathbf{Y} . More specifically, we calculate $\sigma_p = \sum_{i=1}^n \sum_{j=1}^m (\mathbf{Y}_{i,j} - \tilde{\mathbf{Y}}_{i,j})^2 / nm$ for $p = 1, 2, \dots, n$ and define cumulative explained variance β_p for each p value as below.

$$\beta_q = \frac{\sum_{i=1}^p \sigma_i}{\sum_{j=1}^m \sigma_j} \quad (24)$$

β_p for each value of $p = (1, 2, \dots, 98)$ quantifies the amount of information we keep when reducing the dimension of the original observable space. $p = 98$ will surely leads to the maximum value possible for $\beta_p = 1$ because there is no information loss of if we do not reduce the dimensionality. In this work we find that for $p = 12$ the $\beta_{12} = 0.98$ and thus we will reduce the dimensionality of the observables from 98 dimension to 12 and only make GPs to interpolate and predict each of these principal components. The explained variance for the first 12 principal components are shown in Fig. 3.

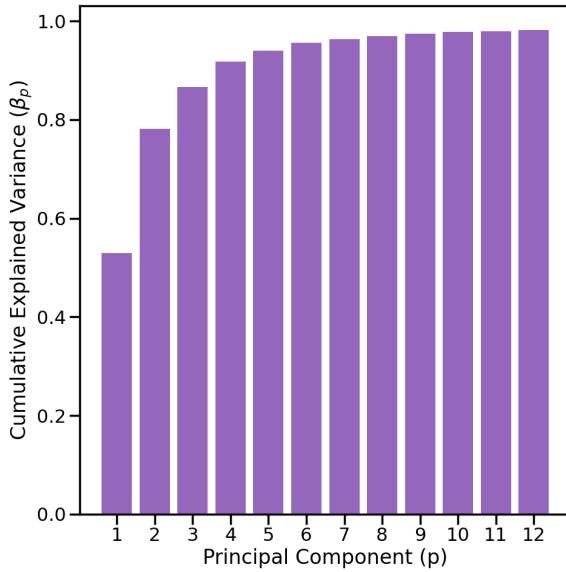


FIG. 3. The cumulative explained variance (β_q) when increasing the number of principal components (q)

D. Principal Component Stochastic Kriging (PCSK) emulation method

Ozge, Matt would you be able to fill in this section? Relativistic heavy ion collision simulations are stochastic simulations. Thus the simulation output provides a mean and a variance associated with it. The accuracy of our simulation data varies for different design batches that we have (a-e). Treating all the simulation data batches in the same footing and building emulators thus can introduce significant errors. The PCSK emulators provide a solution to this issue by incorporating the variance associated with simulation prediction when building emulators.

E. Principal Component Gaussian Process Regression (PCGPR) emulation method

PCGPR is the standard emulation method that has been used in all the previous Bayesian parameter studies in the relativistic heavy ion collisions [7–17]. PCGPR method emulates each of the principal components using independent Gaussian Processes. In this study, we employ the anisotropic Gaussian covariance function, widely used for computer experiment emulators [54]:

$$k^{\text{SE}}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp \left[- \sum_{j=1}^q \frac{(x_j - x'_j)^2}{2l_j^2} \right]. \quad (25)$$

Here $\sigma^2 > 0$ is a variance parameter controlling the variation of the process around its mean, while the parameters $l_j > 0$ ($j = 1, 2, \dots, q$) are characteristic length-scales for each of the model q model inputs. Larger l_j induce stronger

correlations between nearby points, resulting in smoother sample paths, whereas smaller l_j result in more wiggly sample paths.

F. Principal Component Gaussian Process Regression with Grouping (PCGPRG) emulation method

PCGPRG is an extension of the previous emulation method PCGPR but with one modification. Here first we separate our observables into two groups knowing that some of the observables are more prone to noise and would behave very differently compared to the others. In this study we separate the flow observable ($v_2^{(\text{ch})}\{2\}, v_3^{(\text{ch})}\{2\}, v_4^{(\text{ch})}\{2\}$) from the rest due to the fact that the flow observables require much more events to be accurately calculated and thus more noisy. Then we do the PCGPR emulation for each sub group and construct the emulators. Here we also keep the total number of principal components to 12 for fair comparison with other methods. PCGPR emulators for the flow observables produce 8 principal components which has a cumulative explained variance score of 0.965. PCGPR emulators for the collection of all other observables produce 4 principal components which has a cumulative explained variance score of 0.98. This difference in the number of principal components needed confirms the assumption that the flow observable behave differently to the rest of the observables. The flow observables are prone to more noise and that increases the number of principal components needed to capture the variance of the data.

G. Tests for Emulation Accuracy

In this section we test the accuracy of the emulators built using each of the above three methods. We use (a),(c),(d),(e) data sets to train the emulators and then use the data set (b) as a test set. The emulator accuracy is measured using Coefficient of determination (R^2 score) for j th observable from our simulation is defined as follows.

$$R_j^2 = 1 - \frac{\sum_{i=1}^m (y_{i,j}^{\text{test}} - y_{i,j}^{\text{emulation}})^2}{\sum_{i=1}^m (y_{i,j}^{\text{test}} - \mu_j)^2} \quad (26)$$

$$\mu_j = \frac{\sum_{i=1}^m y_{i,j}^{\text{test}}}{m}$$

Here $i = (1, 2, \dots, m)$ is the i^{th} design point from the test set and j stands for j^{th} observable out of the 98 observables that we have. μ_j is the mean of all the test simulation data for j th observable. The maximum value R^2 score can take is 1 which happen when all of the emulation predictions are identical to the simulation output. This would happen only if we have perfect emulators. R^2 score can get arbitrary negative values depending on how much the emulator predictions deviate from the simulation outputs for the testing data.

The R^2 scores calculated in this work for the three emulation methods (a) PCSK, (b) PCGPR, (c) PCGPRG are shown in Fig. 4. Each color corresponds to a different observable type

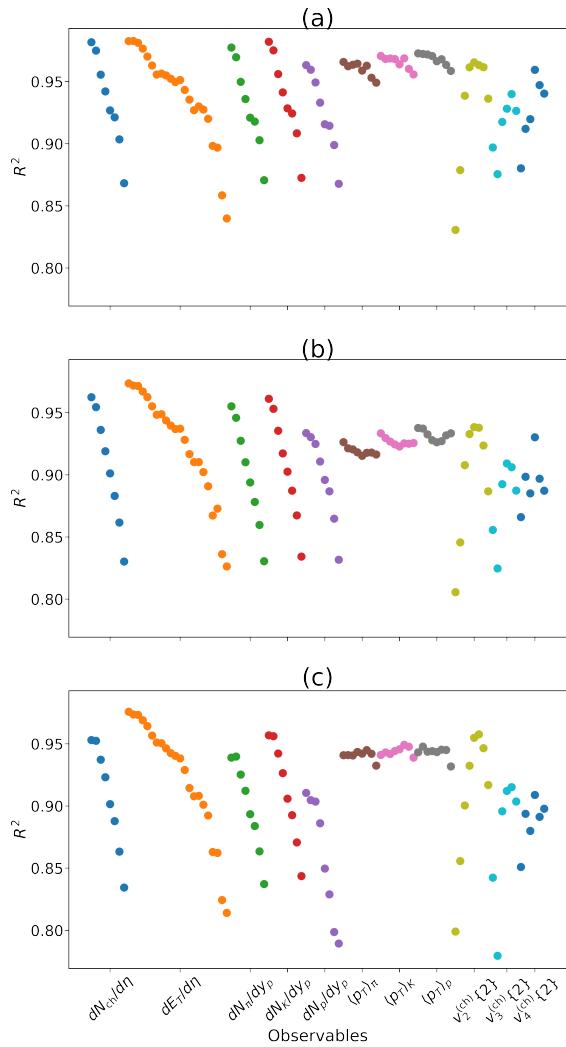


FIG. 4. R^2 scores calculated using the test simulations and emulator predictions for three different emulation methods. (a) PCSK, (b) PCGPR, (c) PCGPRG.

and each dot corresponds to increasing centrality in a particular observable type. In total there are 98 dots corresponding to the 98 observables that we consider in this work.

From Fig. 4 we conclude that PCSK emulators have the best overall performance compared to the other two methods. There is a general tendency for the R^2 scores to reduce as we go from PCSK emulators (a) to PCGPR (b) and they further reduce in PCGPRG (c) emulators. Thus for the analysis in the rest of the paper we have chosen the PCSK emulators.

What about calibration curve plots and emulation vs simulation plots? We have too many plots and can not include all.

V. CLOSURE TESTS

Following Ref. [17] we perform closure tests before the final Bayesian parameter estimation as a check to confirm that

our Bayesian parameter estimation setup is accurate. We randomly pick nine design points from our most accurate simulation data set (e) which has 1600 events per design. We train our emulators without including these nine design points in our training data. Then the simulation outputs of these nine design points are taken as pseudo-experimental data and we perform the Bayesian parameter inference to find the most probable values of the model parameters that can reproduce the pseudo-experimental data. Then we compare the inferred model parameter values with the known truth values that generated these pseudo-experimental data and validate our Bayesian parameter estimation framework.

Fig. 5 shows the closure test results for the temperature dependent specific bulk and specific shear viscosity values. From the figure it is clear that the true model parameter value (the dotted line) has been well captured by the inferred posteriors for these model parameters. We have also checked the agreement of the posterior for other parameters in the model and we find them to be accurately inferred.

VI. BAYESIAN PARAMETER ESTIMATION WITH LHC MEASUREMENTS

With the Bayesian parameter inference framework and trained and validated emulators we are now ready to perform the final Bayesian parameter inference using the Experimental data for heavy ion collisions. In this work we use the LHC experimental data for Pb-Pb collisions at LHC with center of mass energy of $\sqrt{s_{NN}} = 2.76$ TeV. We use the following measurements from ALICE collaboration similarly to the recent JETSCAPE analysis [17].⁴

- the charged particle multiplicity $dN_{ch}/d\eta$ [45] for bins in 0–70% centrality;
- the transverse energy $dE_T/d\eta$ [46] for bins in 0–70% centrality;
- the multiplicity dN/dy and mean transverse momenta $\langle p_T \rangle$ of pions, kaons and protons [47] for bins in 0–70% centrality;
- the two-particle cumulant harmonic flows $v_n\{2\}$ for $n = 2, 3, 4$, for bins in 0–70% centrality for $n = 2$, and for bins in 0–50% centrality for $n = 3$ and 4 [48];

In total we have 98 observables as the experimental data for the Bayesian parameter inference. The 15 model parameters that we infer and the priors we use for each of the parameters are given in Table I. For all the model parameters we use uniform priors that have the same ranges as in Ref. [17]. The parameter R in Eq. 13 which determines the initial pressure

⁴ We omit the fluctuation in the mean transverse momentum $\delta p_T/p_T$ [59] for bins 0–70% centrality in our analysis. We found excluding this observable increase the validation score R^2 for all observables. We suspect this is due to the current number of events per design being insufficient to calculate mean transverse momentum accurately.

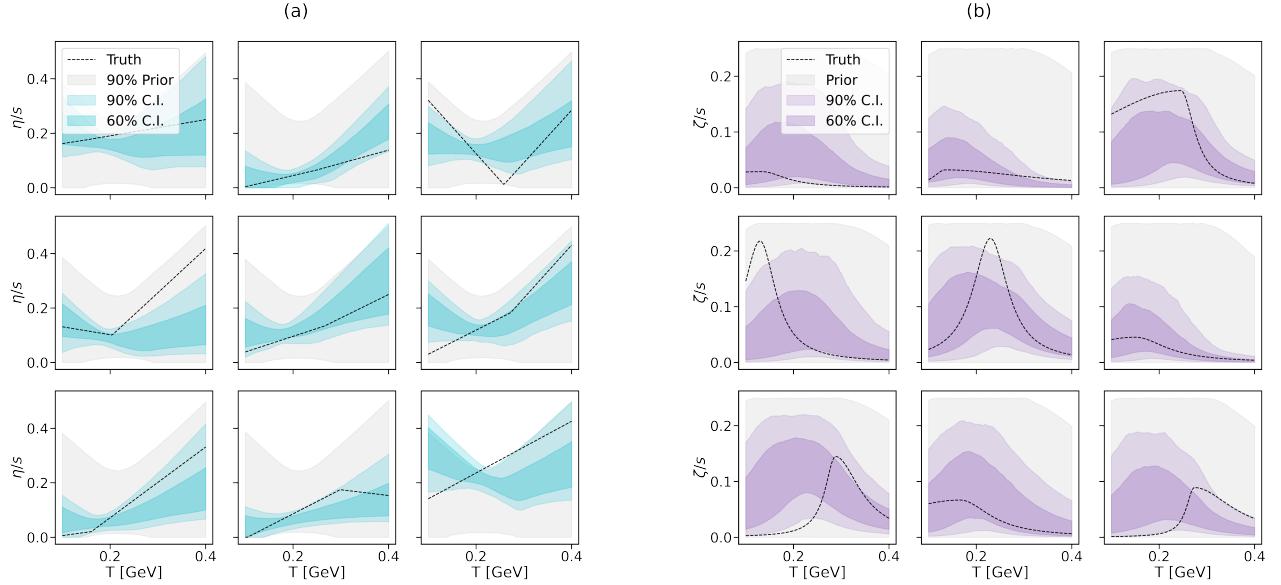


FIG. 5. Closure tests for temperature dependent specific bulk and specific shear viscosity.

parameter	prior range
N	[10, 30]
p	[-0.7, 0.7]
w [fm]	[0.5, 1.5]
d_{min} [fm]	[0.0, 1.7]
σ_k	[0.3, 2.0]
T_{sw}	[0.135, 0.165]
R	[0.3, 1]
$T_{\eta/s, \text{kink}}$ [GeV]	[0.13, 0.3]
$(\eta/s)_{\text{kink}}$	[0.01, 0.2]
a_{high} [GeV^{-1}]	[-1, 2]
a_{low} [GeV^{-1}]	[-2, 1]
$(\zeta/s)_{\text{max}}$	[0.01, 0.25]
T_ζ [GeV]	[0.12, 0.3]
w_ζ [GeV]	[0.025, 0.15]
λ_ζ	[-0.8, 0.8]

TABLE I. We use uniform priors for all the model parameters. The prior ranges for each of the parameters are shown in the table.

ratio between longitudinal and transverse pressure is unique to the VAH model. We assume the minimum p_L/p_\perp ratio is 0.3 and can it can increase until the pressures are equal where $R = p_L/p_\perp = 1$. The pressure ratio, R is assumed to be uniformly distributed between these bounds prior to be conditioned on the experimental data.

We use the multivariate normal likelihood (Eq. 19) introduced in Sec. III. The experimental uncertainty and the simulation uncertainty are added together to get the total uncertainty that appear in the likelihood.

The posterior can be found combining the likelihood with the prior according to the Bayes rule (Eq 18). We analyze the 15 dimensional posterior probability distribution by sampling it using Markov Chain Monte Carlo (MCMC) technique [60, 61]. More specifically, in this work we choose to use

the Parallel Tempering MCMC technique mainly due to it's robustness in sampling multi model posterior distributions [62–64]. We use 500 temperatures evenly distributed between 0 and 1000 as the temperature ladder. We use 100 randomly initialized chains for each temperature in the temperature ladder. The first 1000 steps that are taken are considered burn in steps and gets discarded. The next 5000 steps are considered to be samples from the posterior and we use a thinning factor of 10 (save only every 10th sample) to get the final posterior samples.

A. Posterior for model parameters

The joint marginal distribution plot of the posterior for the parameters in the model that are not related to the viscosity are shown in Fig. 6. The join marginal distributions are generated by projecting the posterior to two dimensions. The projection is done by simply integrating the posterior over all other model parameters except two. On the diagonal, we have the marginal distributions corresponding to each of the model parameters. The model parameters that maximize the posterior are called Maximum a Posteriori (MAP) and they are shown in blue dotted lines on the diagonal.

The blue color corresponds to the uniform prior distribution density and the red color shows the density of the posterior distribution. In a situation where experimental measurements do not have good constraining power to infer the model parameters we would expect the prior (blue) and posterior (red) distributions to have a larger overlap with each other. These type of situations where essentially the information in the prior is returned to the posterior without a significant change is called "returning of prior". From the marginal distributions we observe that two of our model parameters, the initial pressure ratio (R) and the minimum distance between nucleons (d_{min})

are not well constrained in our model by the experimental data.

We observe that the normalization parameter of T_{RENTo}, N , has an interesting correlation with the initial pressure ratio parameter, R , which gives rise to the double peak structure in the marginal posterior distribution of the N . The normalization parameter scales the initial energy density deposition profile from T_{RENTo} which is then used to initialize the hydrodynamic variables of the VAH module. A large value of the normalization corresponds to a larger initial energy density in the hydrodynamic stage and this directly increase the particle yields and transverse energy in the final state observables. The R parameter is related to the VAH stage and it sets the $R = p_L/p_\perp$ initial pressure ratios. Increase in the value of R corresponds to increase of initial longitudinal pressure and thus to a decrease of the initial transverse pressure. The decrease of initial transverse pressure directly results in decrease of particle yields, transverse energy and mean momenta in the final state mid rapidity observables. The competing effect of R and N when fitting the experimental observables is clearly visible in the joint marginal distribution of the left bottom corner in Fig. 6. We can see that there exist two main regions where the posterior probability is high for both N and R . In the smaller region where both $R (\geq 0.75)$ and $N (\geq 25)$ have large values and more broader region where N is small (≤ 25) but R is uniformly distributed.

The nucleon width parameter w is constrained to a value around 1 fm. We note that there is an interesting relationship between w , N and R which is clearly visible in the joint marginal distribution of nucleon width w and normalization N . The joint marginal distribution between N and w shows that increasing the value of R results in increasing values for both N and w . This is due to the fact that increasing R reduces the particle yields, transverse energy and mean momenta as we discussed earlier and increasing the value of N and w has the opposite effect on the same observables. Increasing nucleon width w results in generating larger hot spots in the initial energy density profile which contributes to increase in final state particle yields, transverse energy and mean momenta. It is clear that to fit the experimental data with a higher initial pressure ratio R and to compensate the lower particle yields, transverse energy, mean moment produced, both normalization N and nucleon width w have to increase.

For the VAH model the harmonic mean parameter p from T_{RENTo} (Eq.3) seems to prefer negative values closer to zero. We find this interesting since all previous Bayesian parameter inferences that were done using a free streaming pre-hydrodynamic stage clearly showed a preference for the p value to be zero [7–17]. In our work we have found that the constraints obtained for p might be sensitive to the assumptions in the pre-hydrodynamic model that is being used. Also we observe in the joint marginal distribution of p with N , the value of p tends to decrease as the normalization N increases. This is possibly due to the fact that the larger values of N are correlated with the larger values of nucleon width w which try to compensate for increasing R values as discussed in the last paragraph. The large nucleon width w value tend to reduce the harmonic flows by introducing larger hot spots and making the initial energy deposition profile smooth. To compen-

sate for this effect the harmonic mean parameter in T_{RENTo}, p tends to be negative as the negative values of the p parameter increase all flow harmonics [20].

The same effect as above drives the σ_k which is the standard deviation of the fluctuation of intensity of hot spots in the initial energy profile (Eq. 1) to larger values for large R values as seen in the joint posterior distribution plot between σ_k and R . The larger σ_k values increase the fluctuations in the initial energy density profile which cause the harmonic flows to increase. Large R values which corresponds to large nucleon width w tends to decrease the harmonic flows and σ_k can compensate that by increasing it's value and increasing the harmonic flows.

The minimum distance between nucleons parameter d_{\min} is unconstrained in our Bayesian parameter estimation. This has also been observed in past Bayesian parameter estimation studies done with the T_{RENTo} model [17]. The switching temperature T_{sw} where the QGP gets converted into hadrons is constrained to a lower temperature than in previous studies [15, 17]. From the marginal posterior for the T_{sw} we observe that the prior lower bound we have for the switching temperature needs to be further lowered in future Bayesian parameter studies with the VAH model.

1. Posterior for temperature dependent specific shear and bulk viscosity

In this work we performed for the first time the Bayesian parameter estimation of the QGP viscosity using more realistic viscous anisotropic hydrodynamic simulation model for relativistic heavy ion collisions. The posterior for temperature dependent specific shear and bulk viscosity of QGP inferred using experimental data for Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV are shown in Fig. 7. The constraints for the specific shear viscosity have not improved and are similar to recent phenomenological Bayesian parameter studies although constraint for specific bulk viscosity has largely improved [15, 17].

SIMS have a toy model. mismatch between EOS. VAH is better. No artificially high bulk viscous pressure. Read Dereks thesis about initial positive bulk pressure which relaxes to negative values. What about initial bulk pressure for VAH??Major difference is SIMS failed to constrain bulk viscosity at large temperature but we are able to do that in our work....(Maybe also include KL divergence plot to measure info gain?)

VII. MODEL SENSITIVITY

VIII. PREDICTING TRANSVERSE MOMENTUM FLUCTUATION OBSERVABLE

IX. CONCLUSION

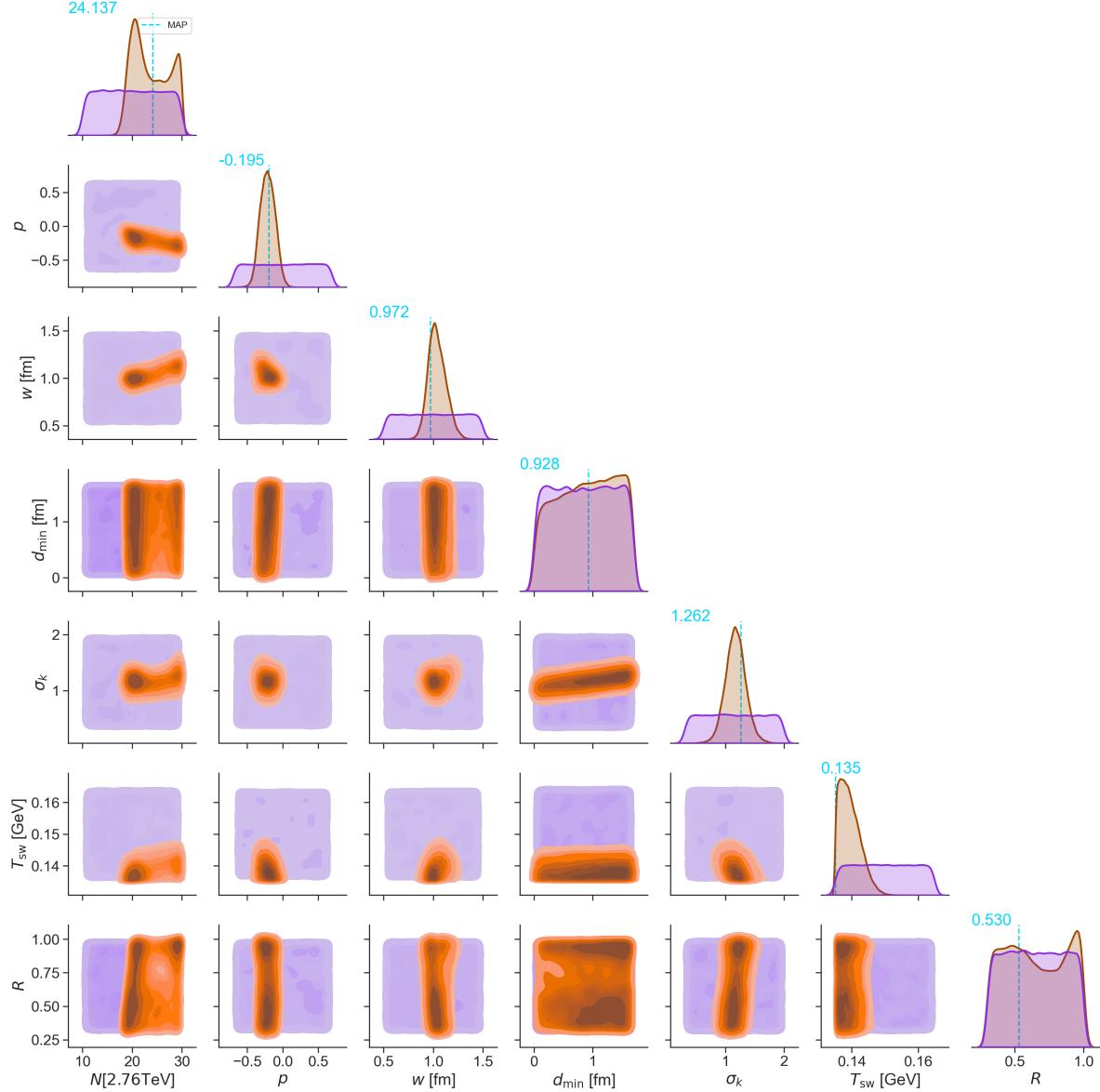


FIG. 6. Joint marginal distribution of the posterior for all model parameters except the viscosities. We posterior is found for the viscous anisotropic hydrodynamics model with PTMA viscous corrections at the particlization and using the experimental data for Pb-Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76 \text{ TeV}$. Posterior distribution is in red and uniform prior distribution is shown in blue.

Appendix A: Gaussian process fitting procedure

In practice, the parameters of the covariance function $k(\mathbf{x}, \mathbf{x}')$ are estimated by maximizing the log likelihood shown in Eq. (A3) and then plugged into the predictive Eq. (21) for emulation (see [54] for further details on plug-in predictors).

We now integrate the data obtained from the full model simulations. Suppose from the stochastic simulation we have noisy outputs $\mathbf{y} = (y_1, \dots, y_n)$ at parameters $\mathbf{x}_1, \dots, \mathbf{x}_n$.

$$y_i = f(\mathbf{x}_i) + \epsilon_i, \quad \epsilon_i \stackrel{i.i.d.}{\sim} N(0, \gamma^2), \quad (\text{A1})$$

where ϵ_i represents statistical uncertainty, *i.i.d.* stands for “in-

dependent and identically distributed”, and $N(0, \gamma^2)$ denotes a Gaussian normal distribution with zero mean and variance γ^2 . Using this relation the probability of observing the data given the model parameter value to be \mathbf{x} is,

$$\Pr(\mathbf{y}|\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{K} + \gamma^2 \mathbf{I}_n|}} \exp\left\{-\frac{1}{2} \mathbf{y}^\top (\mathbf{K} + \gamma^2 \mathbf{I}_n)^{-1} \mathbf{y}\right\} \quad (\text{A2})$$

Thus the log likelihood (\mathbf{L}) is found to be,

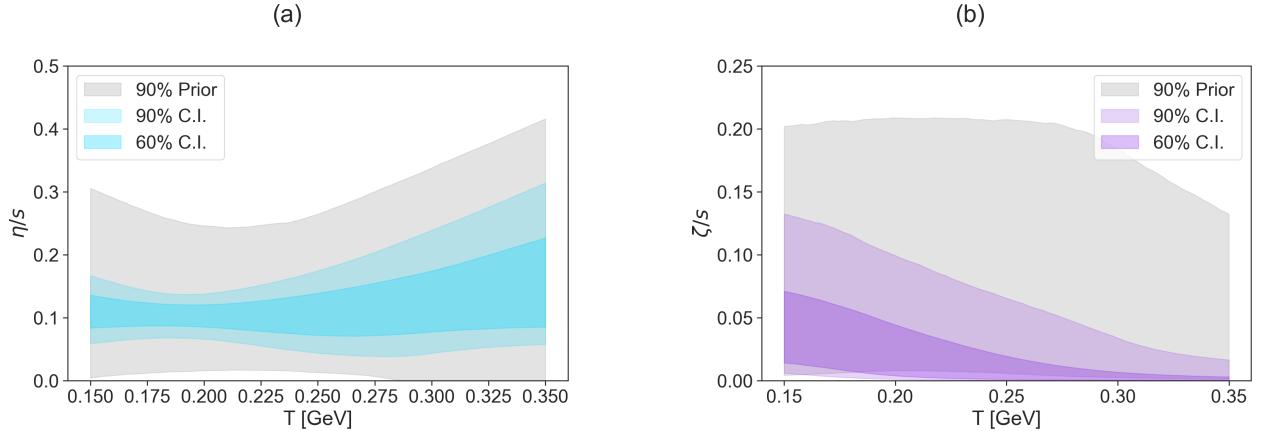


FIG. 7. Posterior for the Temperature dependent specific shear (left) and specefic bulk (right) viscosity found for the viscous anisotropic hydrodynamics model with PTMA viscous corrections at the particlization and using the experimental data for Pb-Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV. The grey bands show the 90% prior interval. The colored regions corresponds to 90% (light) and 60% (dark) posterior credible intervals.

$$\log \mathbf{L} = -\frac{1}{2}\mathbf{y}^\top(\mathbf{K} + \gamma^2\mathbf{I}_n)^{-1}\mathbf{y} - \frac{1}{2}\log(|\mathbf{K} + \gamma^2\mathbf{I}_n|) - \frac{n}{2}\log(2\pi) \quad (\text{A3})$$

-
- [1] A. D. Linde, *Reports on Progress in Physics* **42**, 389 (1979).
- [2] J. C. Collins and M. J. Perry, *Phys. Rev. Lett.* **34**, 1353 (1975).
- [3] K. Yagi, T. Hatsuda, and Y. Miike, Cambridge Monogr. Part. Phys. Nucl. Phys. Cosmol. **23**, 1 (2005).
- [4] M. Gyulassy and L. McLerran, *Nucl. Phys. A* **750**, 30 (2005), arXiv:nucl-th/0405013.
- [5] L. Adamczyk et al. (STAR Collaboration), *Phys. Rev. C* **96**, 044904 (2017).
- [6] W. Busza, K. Rajagopal, and W. van der Schee, *Annual Review of Nuclear and Particle Science* **68**, 339 (2018), <https://doi.org/10.1146/annurev-nucl-101917-020852>.
- [7] H. Petersen, C. Coleman-Smith, S. A. Bass, and R. Wolpert, *J. Phys. G* **38**, 045102 (2011), arXiv:1012.4629 [nucl-th].
- [8] J. Novak, K. Novak, S. Pratt, J. Vredevoogd, C. E. Coleman-Smith, and R. L. Wolpert, *Phys. Rev. C* **89**, 034917 (2014), arXiv:1303.5769 [nucl-th].
- [9] E. Sangaline and S. Pratt, *Phys. Rev. C* **93**, 024908 (2016), arXiv:1508.07017 [nucl-th].
- [10] J. E. Bernhard, P. W. Marcy, C. E. Coleman-Smith, S. Huzurbazar, R. L. Wolpert, and S. A. Bass, *Phys. Rev. C* **91**, 054910 (2015), arXiv:1502.00339 [nucl-th].
- [11] J. E. Bernhard, J. S. Moreland, S. A. Bass, J. Liu, and U. Heinz, *Phys. Rev. C* **94**, 024907 (2016), arXiv:1605.03954 [nucl-th].
- [12] J. S. Moreland, J. E. Bernhard, and S. A. Bass, *Phys. Rev. C* **101**, 024911 (2020), arXiv:1808.02106 [nucl-th].
- [13] J. E. Bernhard, J. S. Moreland, and S. A. Bass, *Nature Phys.* **15**, 1113 (2019).
- [14] D. Everett et al. (JETSCAPE Collaboration), *Phys. Rev. Lett.* **126**, 242301 (2021).
- [15] G. Nijs, W. van der Schee, U. Gürsoy, and R. Snellings, *Phys. Rev. Lett.* **126**, 202301 (2021), arXiv:2010.15130 [nucl-th].
- [16] G. Nijs, W. van der Schee, U. Gürsoy, and R. Snellings, *Phys. Rev. C* **103**, 054909 (2021), arXiv:2010.15134 [nucl-th].
- [17] D. Everett et al. (JETSCAPE Collaboration), *Phys. Rev. C* **103**, 054904 (2021).
- [18] T. Nunes da Silva, D. Chinellato, M. Hippert, W. Serenone, J. Takahashi, G. S. Denicol, M. Luzum, and J. Noronha, *Phys. Rev. C* **103**, 054906 (2021), arXiv:2006.02324 [nucl-th].
- [19] M. Mc Nelis, D. Bazow, and U. Heinz, *Computer Physics Communications* **267**, 108077 (2021).
- [20] J. S. Moreland, J. E. Bernhard, and S. A. Bass, *Phys. Rev. C* **92**, 011901(R) (2015).
- [21] <https://github.com/Duke-QCD/trento.git>.
- [22] D. Everett et al. (JETSCAPE), *Phys. Rev. C* **103**, 054904 (2021), arXiv:2011.01430 [hep-ph].
- [23] H. Song and U. Heinz, *Phys. Lett. B* **658**, 279 (2008), arXiv:0709.0742 [nucl-th].
- [24] R. Anishetty, P. Koehler, and L. McLerran, *Phys. Rev. D* **22**, 2793 (1980).
- [25] J. D. Bjorken, *Phys. Rev. D* **27**, 140 (1983).
- [26] E. Molnár, H. Niemi, and D. H. Rischke, *Phys. Rev. D* **93**, 114025 (2016).
- [27] A. Bazavov, P. Petreczky, and J. H. Weber, *Phys. Rev. D* **97**, 014510 (2018).
- [28] M. Mc Nelis, D. Bazow, and U. Heinz, *Phys. Rev. C* **97**, 054912 (2018).
- [29] P. Huovinen and H. Petersen, *Eur. Phys. J. A* **48**, 171 (2012), arXiv:1206.3371 [nucl-th].
- [30] M. Mc Nelis, D. Everett, and U. Heinz, *Comput. Phys. Commun.* **258**, 107604 (2021), arXiv:1912.08271 [nucl-th].
- [31] <https://github.com/derekeverett/iS3D>.
- [32] F. Cooper and G. Frye, *Phys. Rev. D* **10**, 186 (1974).

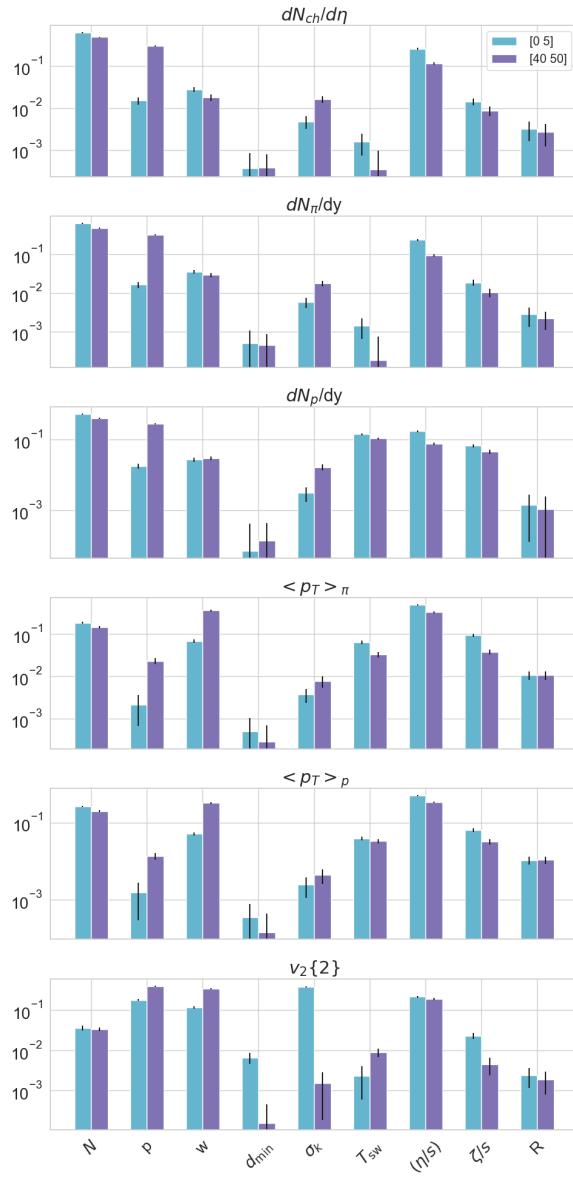


FIG. 8. First order Sobol sensitivity for VAH model

- [33] M. Mc Nelis and U. Heinz, *Phys. Rev. C* **103**, 064903 (2021).
- [34] G. S. Denicol, T. Kodama, T. Koide, and P. Mota, *Phys. Rev. C* **80**, 064901 (2009).
- [35] S. Pratt and G. Torrieri, *Phys. Rev. C* **82**, 044901 (2010), arXiv:1003.0413 [nucl-th].
- [36] C. Nonaka and S. A. Bass, *Phys. Rev. C* **75**, 014902 (2007), arXiv:nucl-th/0607018 [nucl-th].
- [37] T. Hirano, U. Heinz, D. Kharzeev, R. Lacey, and Y. Nara, *Phys. Rev. C* **77**, 044909 (2008), arXiv:0710.5795 [nucl-th].
- [38] H. Petersen, J. Steinheimer, G. Burau, M. Bleicher, and

- H. Stocker, *Phys. Rev. C* **78**, 044901 (2008), arXiv:0806.1695 [nucl-th].
- [39] H. Song, S. A. Bass, and U. Heinz, *Phys. Rev. C* **83**, 024912 (2011), arXiv:1012.0555 [nucl-th].
- [40] U. Heinz, C. Shen, and H. Song, *AIP Conf. Proc.* **1441**, 766 (2012), arXiv:1108.5323 [nucl-th].
- [41] H. Song, S. A. Bass, and U. Heinz, *Phys. Rev. C* **89**, 034919 (2014).
- [42] X. Zhu, F. Meng, H. Song, and Y.-X. Liu, *Phys. Rev. C* **91**, 034904 (2015), arXiv:1501.03286 [nucl-th].
- [43] S. Ryu, J.-F. Paquet, C. Shen, G. Denicol, B. Schenke, S. Jeon, and C. Gale, *Phys. Rev. C* **97**, 034910 (2018), arXiv:1704.04216 [nucl-th].
- [44] <https://github.com/smash-transport/smash>.
- [45] K. Aamodt et al. (ALICE), *Phys. Rev. Lett.* **106**, 032301 (2011), arXiv:1012.1657 [nucl-ex].
- [46] J. Adam et al. (ALICE), *Phys. Rev. C* **94**, 034903 (2016), arXiv:1603.04775 [nucl-ex].
- [47] B. Abelev et al. (ALICE), *Phys. Rev. C* **88**, 044910 (2013), arXiv:1303.0737 [hep-ex].
- [48] K. Aamodt et al. (ALICE), *Phys. Rev. Lett.* **107**, 032301 (2011), arXiv:1105.3865 [nucl-ex].
- [49] D. Sivia and J. Skilling, *Data analysis: a Bayesian tutorial* (OUP Oxford, 2006).
- [50] D. Liyanage, Y. Ji, D. Everett, M. Heffernan, U. Heinz, S. Mak, and J.-F. Paquet, *Phys. Rev. C* **105**, 034910 (2022).
- [51] V. R. Joseph, D. Wang, L. Gu, S. Lyu, and R. Tuo, *Technometrics* **61**, 297 (2019).
- [52] R. L. Iman, J. M. Davenport, and D. K. Zeigler, (1980).
- [53] C. E. Rasmussen, “Gaussian processes in machine learning,” in *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra*, edited by O. Bousquet, U. von Luxburg, and G. Rätsch (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004) pp. 63–71.
- [54] T. J. Santner, B. J. Williams, W. I. Notz, and B. J. Williams, *The Design and Analysis of Computer Experiments* (Springer, 2003).
- [55] R. B. Gramacy, *Surrogates: Gaussian Process Modeling, Design and Optimization* (Chapman Hall/CRC, Boca Raton, Florida, 2020).
- [56] B. Zhang, R. B. Gramacy, L. Johnson, K. A. Rose, and E. Smith, “Batch-sequential design and heteroskedastic surrogate modeling for delta smelt conservation,” [Online]. Available from: <https://arxiv.org/abs/2010.06515> (2020).
- [57] A. Sauer, R. B. Gramacy, and D. Higdon, “Active learning for deep gaussian process surrogates,” (2020).
- [58] V. R. Joseph, T. Dasgupta, R. Tuo, and C. F. J. Wu, *Technometrics* **57**, 64 (2015).
- [59] B. B. Abelev et al. (ALICE), *Eur. Phys. J. C* **74**, 3077 (2014), arXiv:1407.5530 [nucl-ex].
- [60] G. Peters, *Statistics in Medicine* **27**, 3213 (2008), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.3240>.
- [61] R. Trotta, *Contemporary Physics* **49**, 71–104 (2008).
- [62] W. D. Vousden, W. M. Farr, and I. Mandel, *Monthly Notices of the Royal Astronomical Society* **455**, 1919 (2015).
- [63] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, *Publications of the Astronomical Society of the Pacific* **125**, 306 (2013).
- [64] <https://github.com/willvousden/ptemcee>.

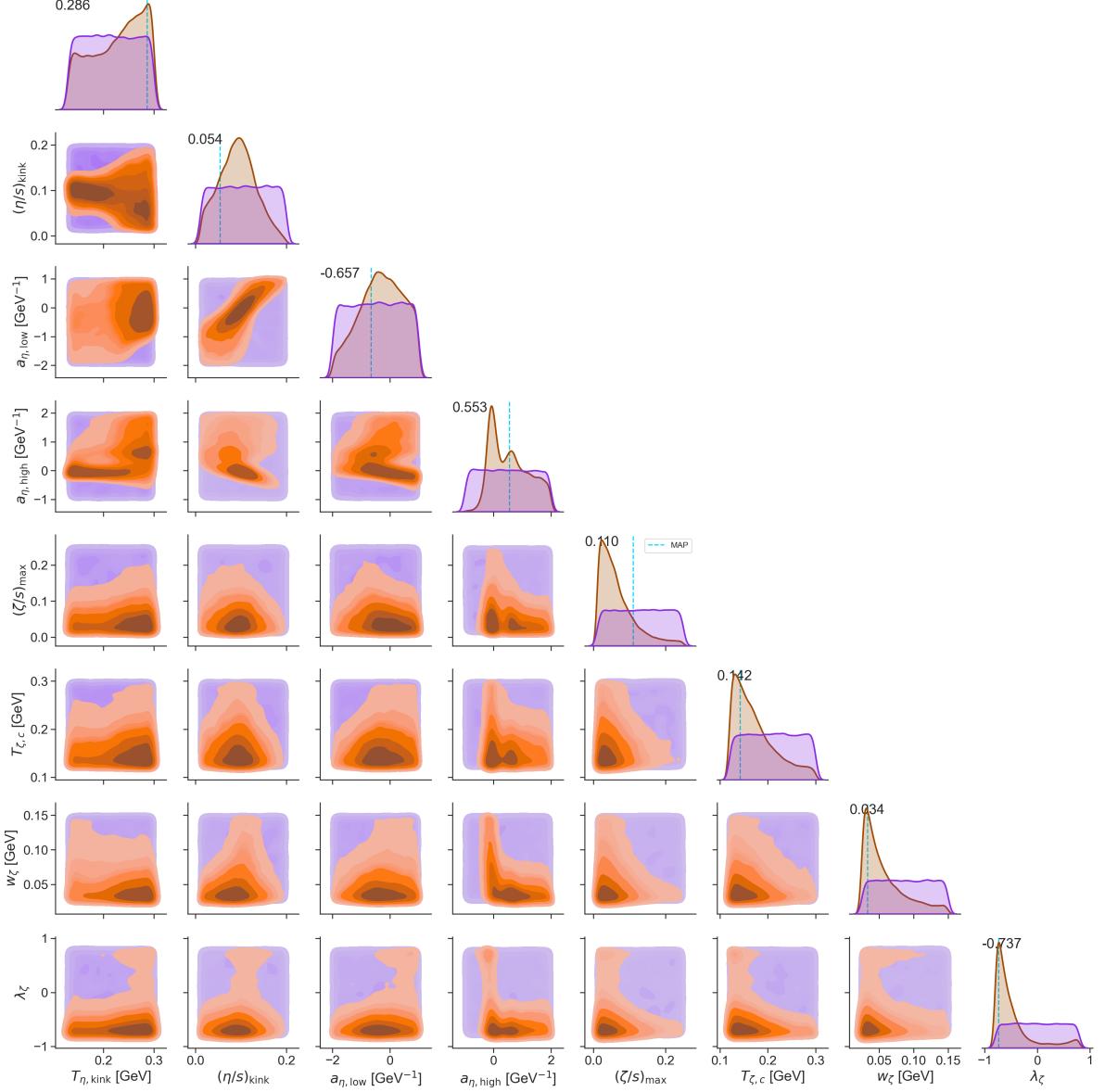


FIG. 9. Posterior for viscosity model parameters.