Model-to-data comparison of a transport+hydrodynamics model of heavy ion collisions using Bayesian statistics and Gaussian emulators

Jussi Auvinen (Duke U.)

in collaboration with Iu. Karpenko, J. Bernhard and S. A. Bass

INT, Seattle
September 26, 2016

Calibrating model to experimental data

Model parameters (input): $\vec{x} = (x_1, ..., x_n)$

$(\tau_0,\ R_{\text{trans}},\ R_{\text{long}},\ \eta/s,\ \epsilon_C)$

$\Downarrow$

Model output $\vec{y} = (y_1, ..., y_m) \Leftrightarrow$ Experimental values $\vec{y}^{\text{exp}}$

$(N_{\text{ch}},\ \langle p_T \rangle,\ v_2, ...)$

Goal: Find the "true" values of the input parameters, for which $\vec{x}^* \Rightarrow \vec{y}^{\text{exp}}$.
Determine the level of uncertainty associated with the proposed values

Bayes' theorem

Given a set $X = \{\vec{x}_k\}_{k=1}^N$ of points in parameter space and a corresponding set $Y = \{\vec{y}_k\}_{k=1}^N$ of points in observable space,

$$\boxed{P(\vec{x}^*|X, Y, \vec{y}^{\text{exp}}) \propto P(X, Y, \vec{y}^{\text{exp}}|\vec{x}^*)P(\vec{x}^*)}$$

- $P(\vec{x}^*|X, Y, \vec{y}^{\text{exp}})$ is the *posterior* probability distribution of $\vec{x}^*$ for given $(X, Y, \vec{y}^{\text{exp}})$
- $P(\vec{x}^*)$ is the *prior* probability distribution (simplest case: ranges of parameter values)
- $P(X, Y, \vec{y}^{\text{exp}}|\vec{x}^*)$ is the *likelihood* of $(X, Y, \vec{y}^{\text{exp}})$ for given $\vec{x}^*$ (to be determined with statistical analysis)

Likelihood function

$$P(X, Y, \vec{y}^{\,\mathsf{exp}}|\vec{x}^*) = \exp\left(-\tfrac{1}{2}(\vec{y}^* - \vec{y}^{\,\mathsf{exp}})^T \Sigma^{-1} (\vec{y}^* - \vec{y}^{\,\mathsf{exp}})\right),$$

where

- Σ is the covariance matrix
- $\vec{y}^*$ is model output for the input parameter point $\vec{x}^*$

However:
1 hybrid simulation run requires $\approx$ 5 hours, 50 events produced
$\approx$ 100 000 events needed $\Rightarrow$ 2 000 runs
$\Rightarrow \mathcal{O}(10^4)$ CPU hours for one evaluation of $\vec{y}^*$!
$\Rightarrow$ Need a way to predict model output for arbitrary input parameter point
$\Rightarrow$ Model emulation using Gaussian processes

Gaussian process

http://dan.iel.fm/george

Set Y of values, corresponding to set X of points in parameter space, has a multivariate normal distribution if

$$Y \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\mu} = \mu(X) = \{\mu(x_1), ..., \mu(x_N)\}$ is the mean and

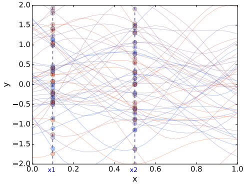
$$\boldsymbol{\Sigma} = \sigma(X, X) = \begin{pmatrix} \sigma(\vec{x}_1, \vec{x}_1) & \cdots & \sigma(\vec{x}_1, \vec{x}_N) \\ \vdots & \ddots & \vdots \\ \sigma(\vec{x}_N, \vec{x}_1) & \cdots & \sigma(\vec{x}_N, \vec{x}_N) \end{pmatrix}$$

is the covariance matrix with covariance function $\sigma(\vec{x}, \vec{x}')$

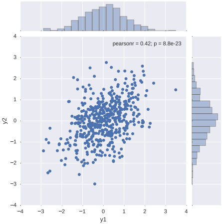
- Model-dependent choice; constant, linear, exponential, periodic, ...
- Restrictions: Needs to be symmetric and positive semidefinite

Gaussian process

Stochastic process: A parameterized collection of random variables $\{r_t\}_{t \in T}$ (T possibly infinite). E.g. random walk over time.



Gaussian process: A stochastic process, in which every finite set $\{r_t\}$ is a multivariate Gaussian random variable.



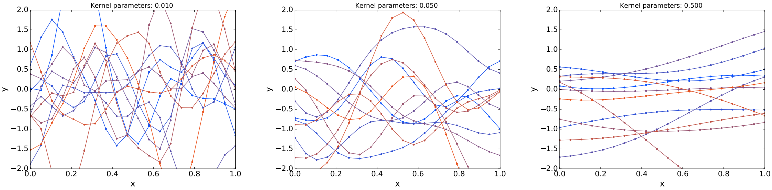
$\mu(X) \equiv 0 \Rightarrow$ GP fully defined by the covariance function $\sigma(\vec{x}, \vec{x}')$. Choice: Squared-exponential

$$\sigma(\vec{x}, \vec{x}') = \theta_0 \exp\left(-\sum_{i=1}^{n} \frac{(x_i - x_i')^2}{2\theta_i^2}\right) + \theta_{\text{noise}} \delta_{\vec{x}\vec{x}'}$$

Gaussian process

Drawing samples from a Gaussian process:

- Define a vector of N points, $\vec{x} = (x_1, ..., x_N)$, on which to evaluate the GP
- Compute covariance matrix $\Sigma_{ij} = \sigma(x_i, x_j)$
- Compute Cholesky decomposition $\Sigma = SS^T$
- The vector $\vec{y} = \mu(\vec{x}) + S\vec{u}$, $u_i \sim N(0, 1)$, defines a GP sample



Gaussian process

The joint distribution of k observations $Y_o = (y(x_{o1}), ..., y(x_{ok}))$ and q predictions $Y_p = (y(x_{p1}), ..., y(x_{pq}))$ is

$$\begin{pmatrix} Y_p \\ Y_o \end{pmatrix} = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{p,p} & \Sigma_{p,o} \\ \Sigma_{o,p} & \Sigma_{o,o} \end{pmatrix} \right),$$

resulting to a conditional probability distribution $p(Y_p|Y_o) \sim \mathcal{N}(\bar{\mu}, \bar{\Sigma})$ with posterior mean (prediction based on known values)

$$\bar{\mu}(X_p) = \Sigma_{p,o}\Sigma_{o,o}^{-1}Y_o$$

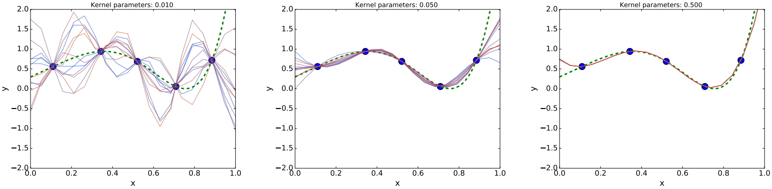
and posterior variance $\bar{\Sigma} = \Sigma_{p,p} - \Sigma_{p,o}\Sigma_{o,o}^{-1}\Sigma_{o,p}$.

For an observation point $x_o \in X_o$:

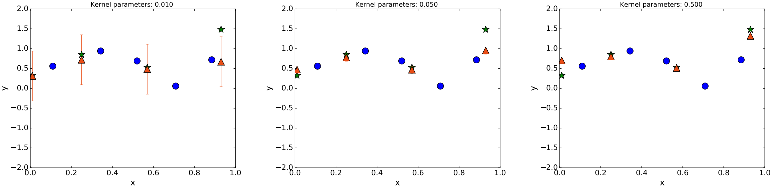
- posterior mean $\bar{\mu}(x_o) = y_o$
- posterior variance $\bar{\Sigma} = 0$

Gaussian process

GP conditioned on known values:



Conditioned GP predictions:

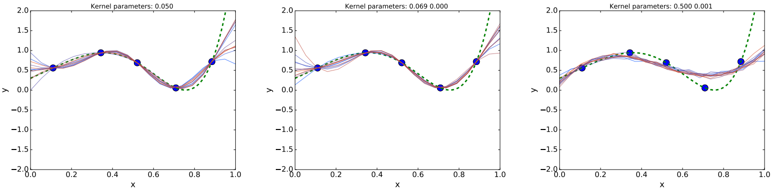


Gaussian process

The *hyperparameters* $\vec{\theta} = (\theta_0, \theta_1, ...\theta_n, \theta_{\text{noise}})$ are not known a priori and must be estimated from the given data ("empirical Bayes")

$\Rightarrow$ emulator training: Maximize the marginal likelihood (aka "evidence")

$$\log P(Y|X,\vec{\theta}) = \underbrace{-\frac{1}{2}Y^T\mathbf{\Sigma}^{-1}(X,\vec{\theta})Y}_{\text{data fit}} \underbrace{-\frac{1}{2}\log|\mathbf{\Sigma}(X,\vec{\theta})|}_{\text{complexity penalty}} \underbrace{-\frac{k}{2}\log(2\pi)}_{\text{normalization}}$$



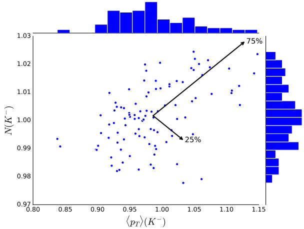
Principal component analysis

m observables $\Rightarrow$ m Gaussian processes

However, m can be up to $\mathcal{O}(100)$ at top RHIC energies and at the LHC!
Number of emulators can be reduced with principal component analysis:

N simulation points, m observables $\Rightarrow$ N x m data matrix Y

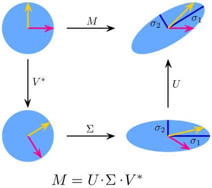
- Goal: Find orthonormal matrix P such that the covariance matrix $S = \frac{1}{N} Z^T Z$ is diagonalized for $Z = YP$
- Solution: Columns of P (principal components) are eigenvectors of $Y^T Y$ (directions of maximal variance)



Singular value decomposition

Singular value decomposition: $Y = U\Sigma V^T$

- Σ is a diagonal matrix containing the singular values
- U and V^T are orthogonal matrices containing the left- and right-singular vectors, respectively



Wikipedia

- Eigenvalue decomposition of $Y^T Y$ becomes

$$Y^T Y = V\Sigma^2 V^T$$

  - Singular values in Σ are square roots of eigenvalues λ_i of $Y^T Y$
  - Right singular vectors in V^T are eigenvectors of $Y^T Y$
  - $V^T Y^T Y V = \Sigma^2 = \frac{1}{N} Z^T Z \Rightarrow Z = \sqrt{N} Y V$

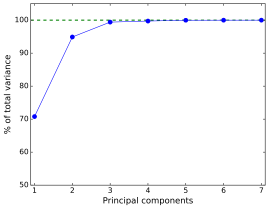
PCA and dimensional reduction

Fraction of variance explained by principal component p_q: $\text{Var}(p_q) = \frac{\lambda_q}{\sum_{i=1}^{m} \lambda_i}$

- $\lambda_1 > \lambda_2 > ... > \lambda_q > ... > \lambda_m$
  $\Rightarrow \text{Var}(p_q) \approx 0$ starting from some $i < q < m$
  $\Rightarrow$ Reduced-dimension transformation

$$Z_q = \sqrt{N} Y V_q$$

- Select the number of principal components which together explain desired fraction of total variance; often only a few PCs are needed to explain 99% of the variance



Box-Cox transformation

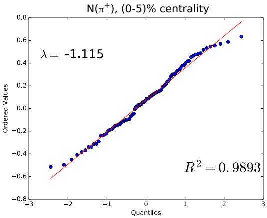
PCA assumes that mean and variance are sufficient statistics to describe the distribution of model output
Many times data is skewed, which reduces the quality of principal component analysis

Try to fix the skew with Box-Cox transformation $y \to y^{(\lambda)}$:

G.E.P. Box and D.R. Cox, Journal of the Royal Statistical Society B, 26, 211 (1964)

$$y^{(\lambda)} = \begin{cases} (y^\lambda - 1)/\lambda & : \lambda \neq 0 \\ \log y & : \lambda = 0 \end{cases}$$

- y dimensionless $\Rightarrow$ Scale with experimental values y^{exp} first
- Assumes $y > 0$; shift if necessary
- Check against normal distribution after transformation (probability plot, QQ plot)



Likelihood function

The likelihood function used in MCMC:

$$\exp\left(-\frac{1}{2}\sum_{i=1}^{q}\lambda_i\frac{(z_i^* - z_i^{\text{exp}})^2}{(\sigma z_i^{\text{exp}})^2 + \Sigma_i^*}\right)$$

- λ_i is the variance explained by ith principal component
- z_i^* is the emulator prediction for ith principal component at the input parameter point $\vec{x}^*$
- $\vec{z}^{\text{exp}}$ is the experimental data transformed to principal component space
- Σ_i^* is the predictive variance (emulator uncertainty)
- $\sigma = 0.1$ is global estimate for all other uncertainties (experimental sys and stat errors etc.)

Markov Chain Monte Carlo

"emcee": D. Foreman-Mackey *et al.*, Publ. Astron. Soc. Pacific **125**, 306 (2013), arXiv:1202.3665

The posterior distribution is sampled with Markov Chain Monte Carlo (MCMC) method

- Random walk in parameter space, where each step is accepted or rejected based on a relative likelihood (calculated in terms of principal components)
- Converges to posterior distribution as number of steps $N \to \infty$
- Acceptance fraction a_f of steps measures the quality of random walk
  - $a_f \sim 0 \Rightarrow$ walker "stuck"
  - $a_f \sim 1 \Rightarrow$ purely random walk
  - aim for 0.2-0.5
- Autocorrelation time = Number of steps between independent samples
  "Burn-in" takes a few autocorrelations,
  gathering enough samples $\sim \mathcal{O}(10)$ autocorrelations

Analysis procedure

Scale with experimental values $\Rightarrow$ Unitless quantities of the order ($\mathcal{O}(1)$)

$\Downarrow$

Verify normal distribution of observables
(apply a transformation if necessary)

$\Downarrow$

Apply weights
Center the data

$\Downarrow$

Principal component analysis $\Rightarrow$ Determine required number of Gaussian processes

$\Downarrow$

Train the emulator(s)

$\Downarrow$

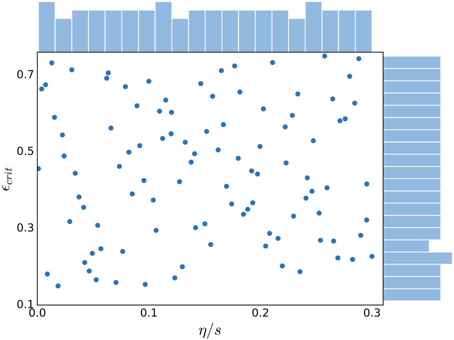
Calibrate on experimental data by running MCMC

Model results

Investigated parameter ranges

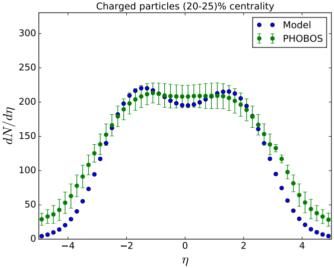
Sample points evenly over whole parameter space using Latin hypercube method

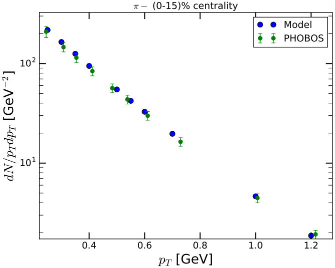
- Shear viscosity over entropy density η/s: 0.001 - 0.4
- Transport-to-hydro transition time τ_0: 0.4 - 3.1 fm
- Transverse Gaussian smearing of particles R_{trans}: 0.2 - 2.2 fm
- Longitudinal Gaussian smearing of particles R_{long}: 0.2 - 2.2 fm
- Hydro-to-transport transition energy density ϵ_C: 0.15 - 0.75 GeV/fm^3



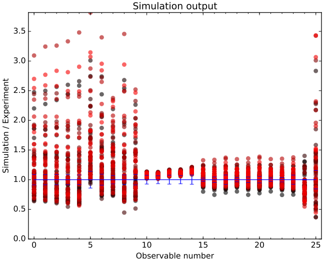
Investigated observables

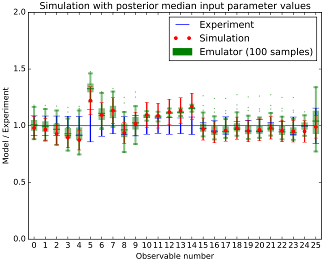
- Charged particles at midrapidity N_{ch}
- Charged particle pseudorapidity distribution $dN_{\text{ch}}/d\eta$
- Number of π, K, p, Ω at midrapidity
- Mean transverse momentum $\langle p_T \rangle$ for π, K, p
- Transverse momentum spectra dN/dp_T for π, K, p
- Charged particle elliptic flow $v_2\{\text{EP}\}$

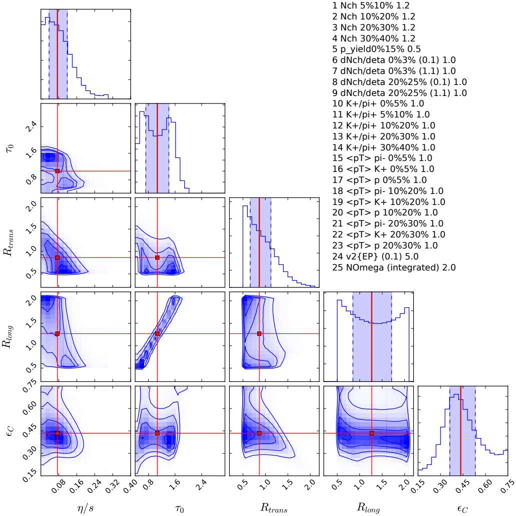




Results at 62.4 GeV

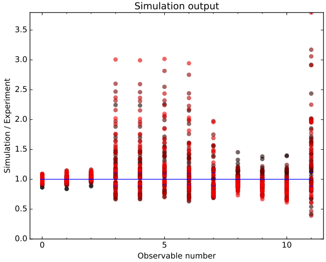


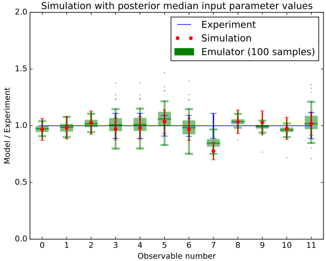


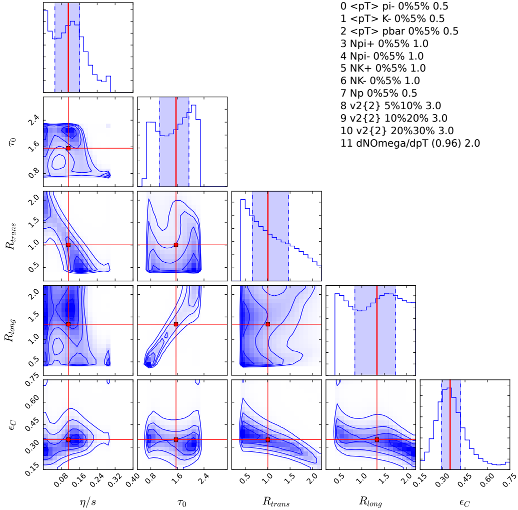


0 Nch 0%5% 1.2
1 Nch 5%10% 1.2
2 Nch 10%20% 1.2
3 Nch 20%30% 1.2
4 Nch 30%40% 1.2
5 p_yield0%15% 0.5
6 dNch/deta 0%3% (0.1) 1.0
7 dNch/deta 0%3% (1.1) 1.0
8 dNch/deta 20%25% (0.1) 1.0
9 dNch/deta 20%25% (1.1) 1.0
10 K+/pi+ 0%5% 1.0
11 K+/pi+ 5%10% 1.0
12 K+/pi+ 10%20% 1.0
13 K+/pi+ 20%30% 1.0
14 K+/pi+ 30%40% 1.0
15 <pT> pi- 0%5% 1.0
16 <pT> K+ 0%5% 1.0
17 <pT> p 0%5% 1.0
18 <pT> pi- 10%20% 1.0
19 <pT> K+ 10%20% 1.0
20 <pT> p 10%20% 1.0
21 <pT> pi- 20%30% 1.0
22 <pT> K+ 20%30% 1.0
23 <pT> p 20%30% 1.0
24 v2{EP} (0.1) 5.0
25 NOmega (integrated) 2.0

Results at 39 GeV

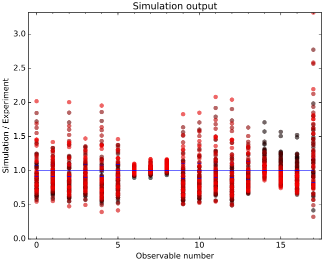


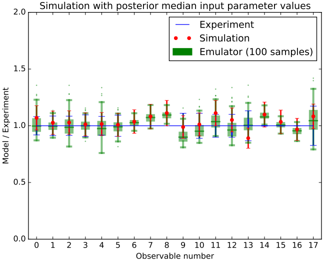


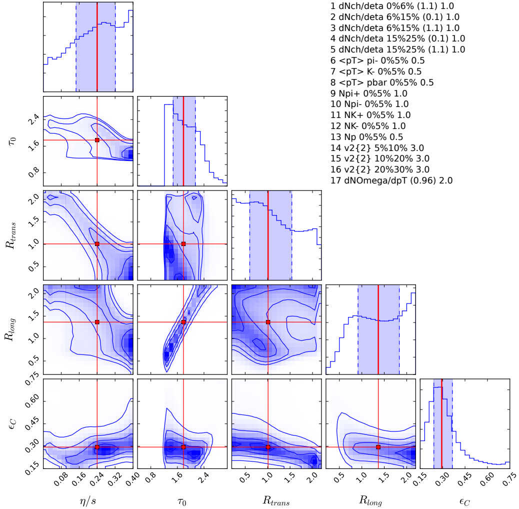


0 <pT> pi- 0%5% 0.5
1 <pT> K- 0%5% 0.5
2 <pT> pbar 0%5% 0.5
3 Npi+ 0%5% 1.0
4 Npi- 0%5% 1.0
5 NK+ 0%5% 1.0
6 NK- 0%5% 1.0
7 Np 0%5% 0.5
8 v2{2} 5%10% 3.0
9 v2{2} 10%20% 3.0
10 v2{2} 20%30% 3.0
11 dNOmega/dpT (0.96) 2.0

Results at 19.6 GeV

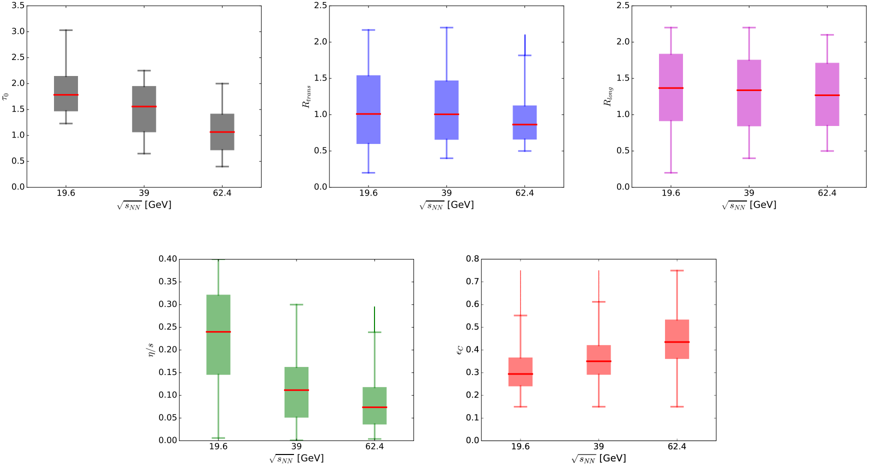






0 dNch/deta 0%6% (0.1) 1.0
1 dNch/deta 0%6% (1.1) 1.0
2 dNch/deta 6%15% (0.1) 1.0
3 dNch/deta 6%15% (1.1) 1.0
4 dNch/deta 15%25% (0.1) 1.0
5 dNch/deta 15%25% (1.1) 1.0
6 <pT> pi- 0%5% 0.5
7 <pT> K- 0%5% 0.5
8 <pT> pbar 0%5% 0.5
9 Npi+ 0%5% 1.0
10 Npi- 0%5% 1.0
11 NK+ 0%5% 1.0
12 NK- 0%5% 1.0
13 Np 0%5% 0.5
14 v2{2} 5%10% 3.0
15 v2{2} 10%20% 3.0
16 v2{2} 20%30% 3.0
17 dNOmega/dpT (0.96) 2.0

Parameter dependence on collision energy



Summary

- Bayesian analysis provides a rigorous method for simultaneous estimation of both the best-fit values and the associated uncertainties for the parameters of heavy ion collision models
- Gaussian processes allow the emulation of complex models, making it possible to investigate multidimensional parameter spaces within reasonable computational effort
- Using posterior median values for the hybrid model gives good agreement with experimental data
- Posterior distributions still rather wide
  - Initial state needs stronger constraints
  - Refine uncertainty estimates in likelihood function; use reported error estimates from experiments for each observable:
    $(\sigma z_i^{\mathsf{exp}})^2 + \Sigma_i^* \to \Sigma_i^{\mathsf{exp}} + \Sigma_i^*$
    Correlated uncertainties between different observables (non-diagonal elements in Σ^{exp})?