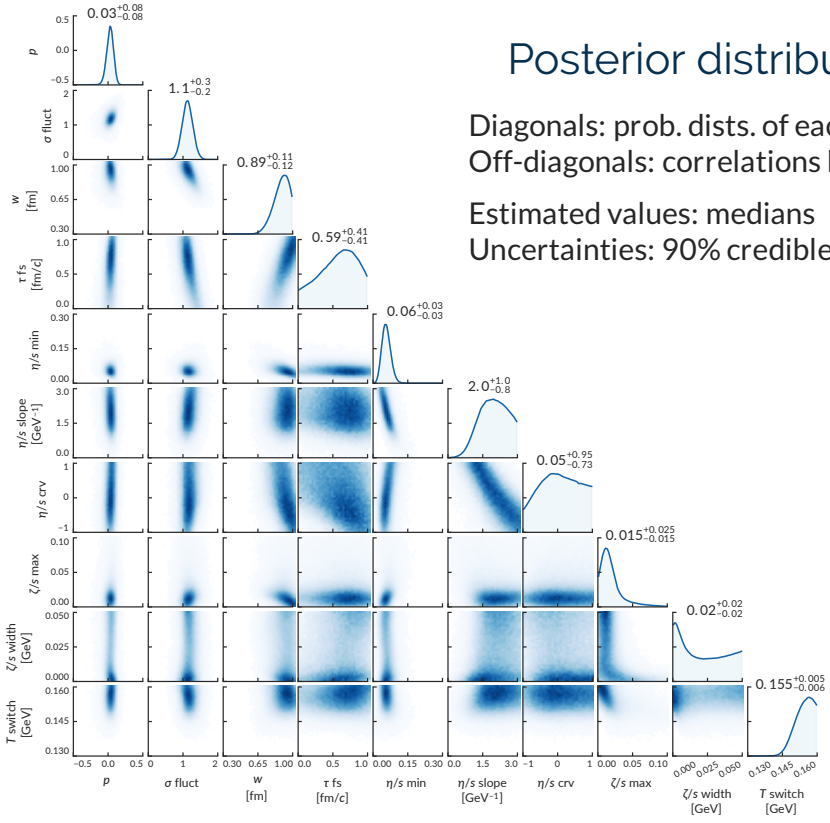
Introduction to Bayesian analysis

Jussi Auvinen

Institute of Physics Belgrade

Karpacz ~~Winter~~ School
June 2022

Success of Bayesian approach

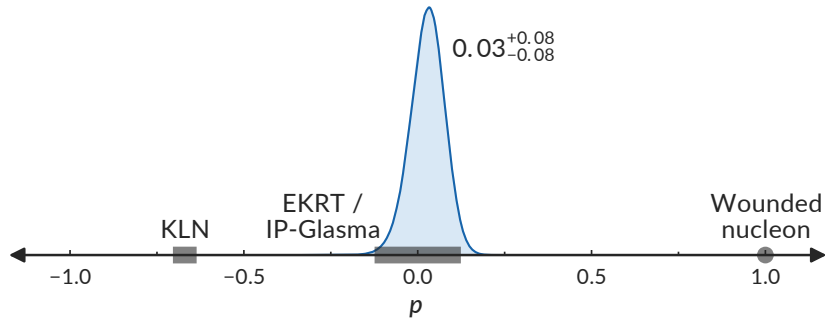


Posterior distribution

Diagonals: prob. dists. of each param.
Off-diagonals: correlations b/w pairs

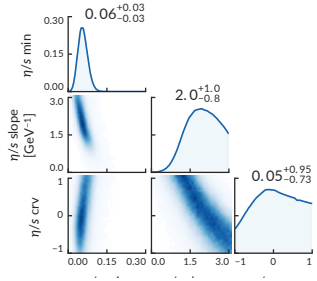
Estimated values: medians
Uncertainties: 90% credible intervals

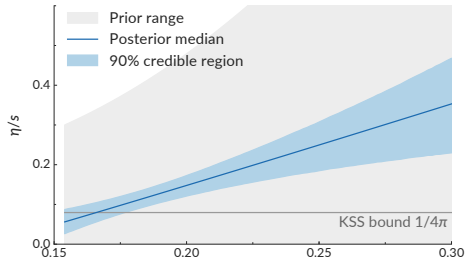
Initial state



Shear viscosity

$$(\eta/s)(T) = (\eta/s)_{\text{min}} + (\eta/s)_{\text{slope}}(T - T_c) \times \left(\frac{T}{T_c}\right)^{(\eta/s)_{\text{crv}}}$$





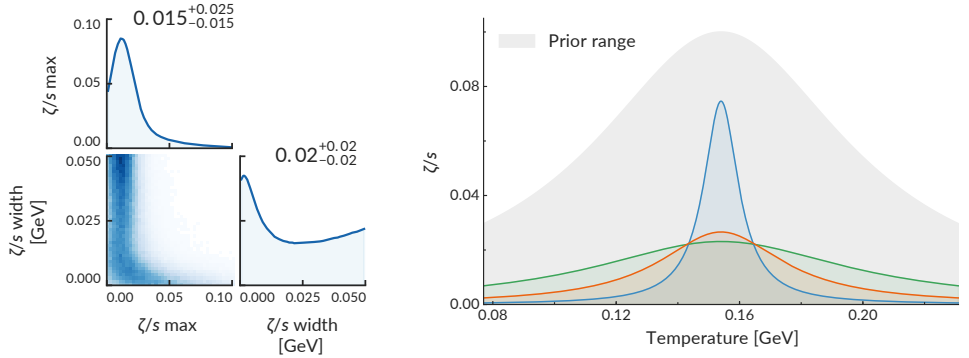
Jonah Bernhard, Quark Matter 2017

See also Nature Phys. 15, 1113 (2019)

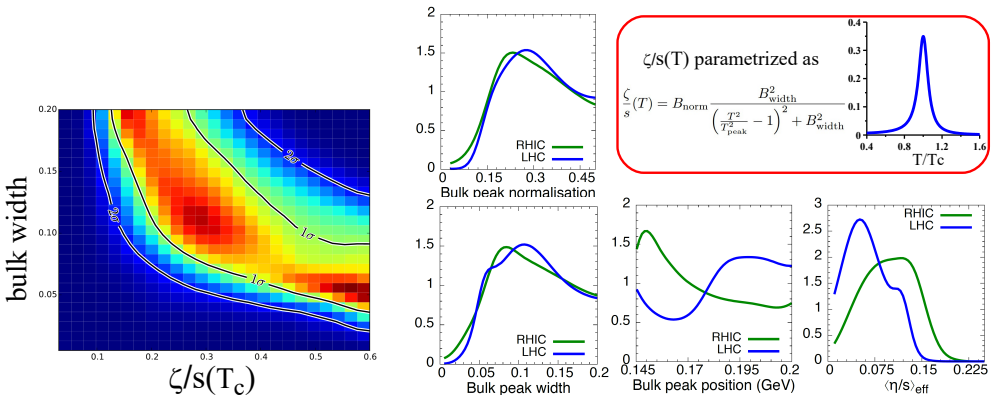
Success of Bayesian approach

Bulk viscosity

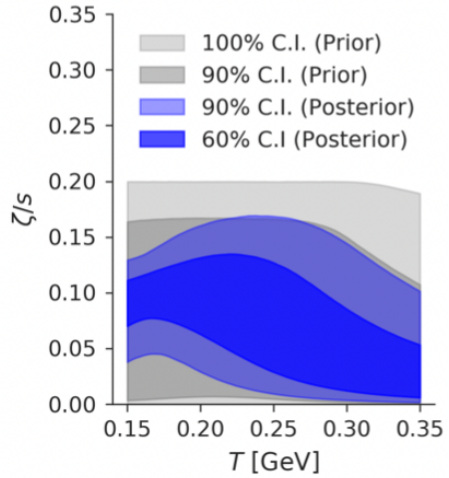
$$(\zeta/s)(T) = \frac{(\zeta/s)_{\max}}{1 + \left(\frac{T - T_c}{(\zeta/s)_{\text{width}}}\right)^2}$$



Jonah Bernhard, QM2017

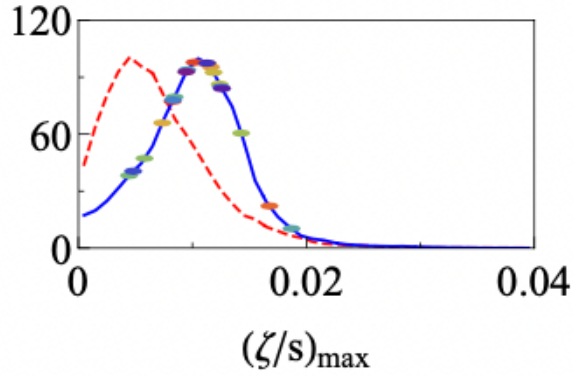


Gabriel Denicol, QM2017



JETSCAPE, PRC 103, 054904 (2021)

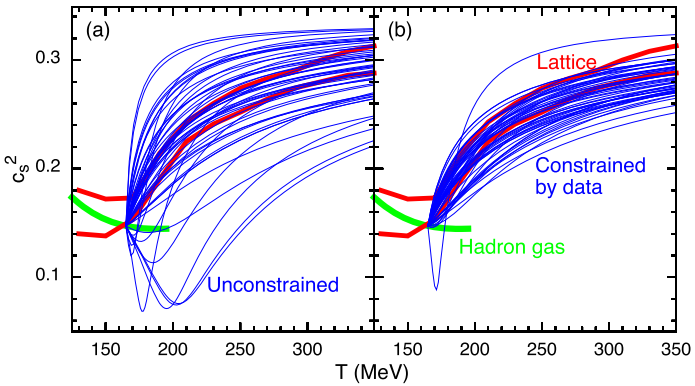
$0.0102^{+0.0064}_{-0.0082}$ and $0.0070^{+0.0061}_{-0.0070}$ (90% CI)

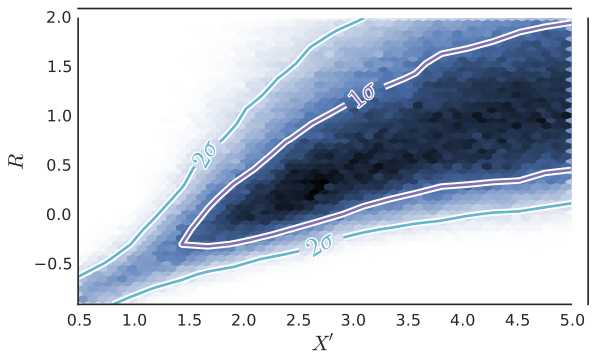


Nijs, van der Schee, arxiv:2110.13153

Success of Bayesian approach

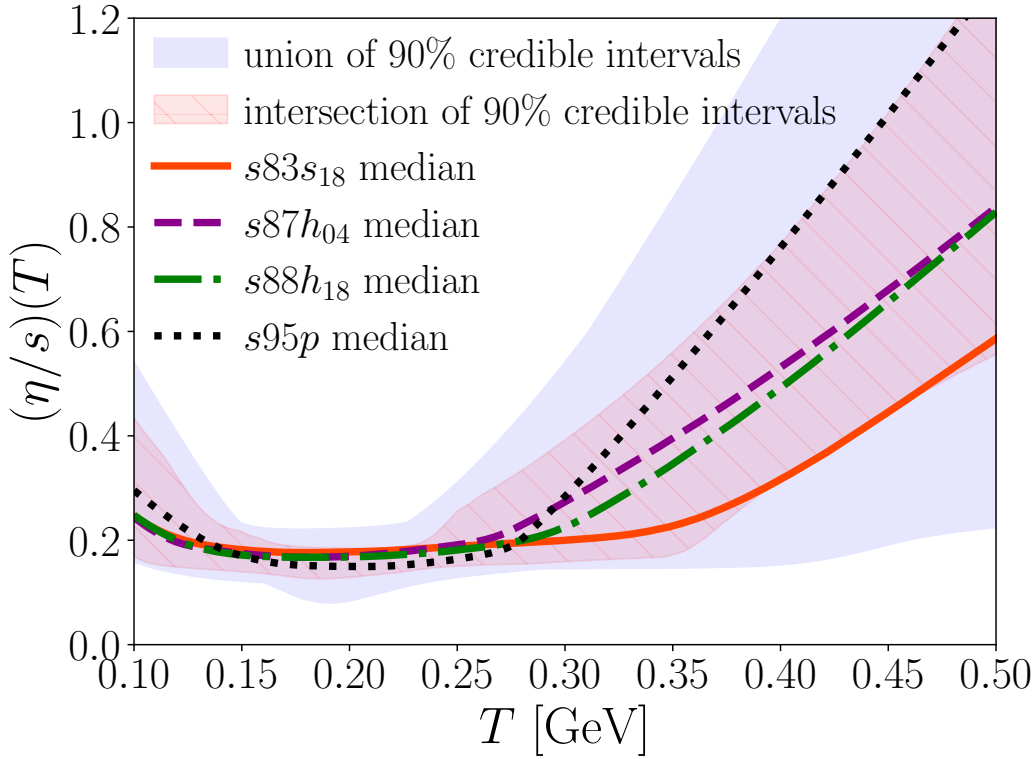
Equation of state





Pratt et al., PRL114, 202301 (2015)

η/s sensitivity to EoS



Auvinen et al., PRC 102, 044911 (2020)

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

$$\Downarrow$$

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

We'd like to find the most probable parameter values $\vec{x}$ given $\vec{y}^{\,\mathrm{exp}}$, which is a conditional probability:

$$P(\vec{x}|\vec{y}^{\,\mathrm{exp}}) = \frac{P(\vec{x} \wedge \vec{y}^{\,\mathrm{exp}})}{P(\vec{y}^{\,\mathrm{exp}})}$$

Likewise the probability of the observation $\vec{y}^{\,\mathrm{exp}}$ on the condition of $\vec{x}$ is

$$P(\vec{y}^{\,\mathrm{exp}}|\vec{x}) = \frac{P(\vec{x} \wedge \vec{y}^{\,\mathrm{exp}})}{P(\vec{x})}$$

Combining the two conditions gives

$$P(\vec{x}|\vec{y}^{\,\mathrm{exp}}) = \frac{P(\vec{y}^{\,\mathrm{exp}}|\vec{x})P(\vec{x})}{P(\vec{y}^{\,\mathrm{exp}})}$$

Bayes' theorem

Posterior probability $\propto$ Likelihood $\cdot$ Prior knowledge

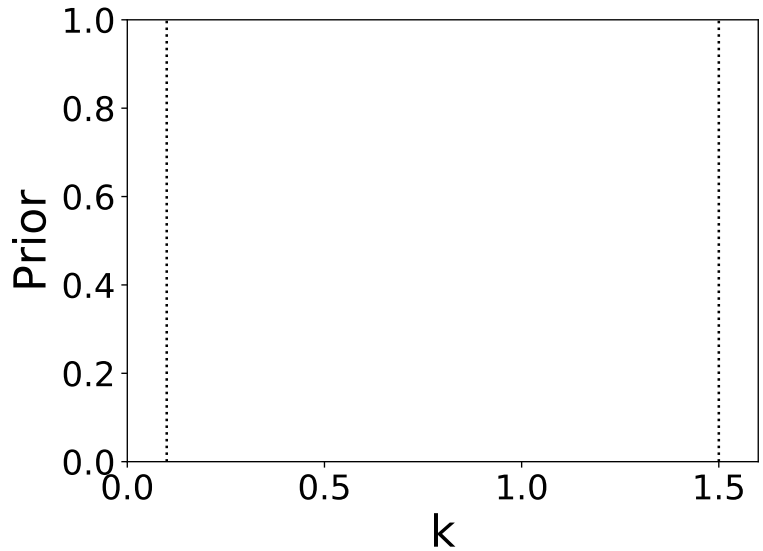
$$P(\vec{x}|\vec{y}^{\,\mathrm{exp}}) = \frac{1}{\mathcal{E}}\mathcal{L}(\vec{y}^{\,\mathrm{exp}}|\vec{x})p(\vec{x})$$

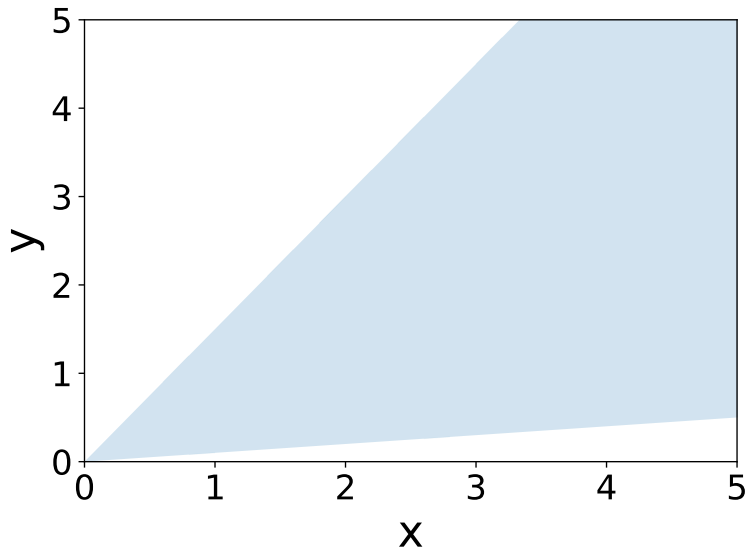
- $P(\vec{x}|\vec{y}^{\,\mathrm{exp}})$: Posterior probability of $\vec{x}$ on the condition of observation $\vec{y}^{\,\mathrm{exp}}$
- $\mathcal{L}(\vec{y}^{\,\mathrm{exp}}|\vec{x})$: Likelihood of observation $\vec{y}^{\,\mathrm{exp}}$ for a given $\vec{x}$
- $p(\vec{x})$: Prior probability of $\vec{x}$ before data comparison
- $\mathcal{E} = \int_{\vec{x}} \mathcal{L}(\vec{y}^{\,\mathrm{exp}}|\vec{x})p(\vec{x})$: Bayes evidence (normalization)

Likelihood function

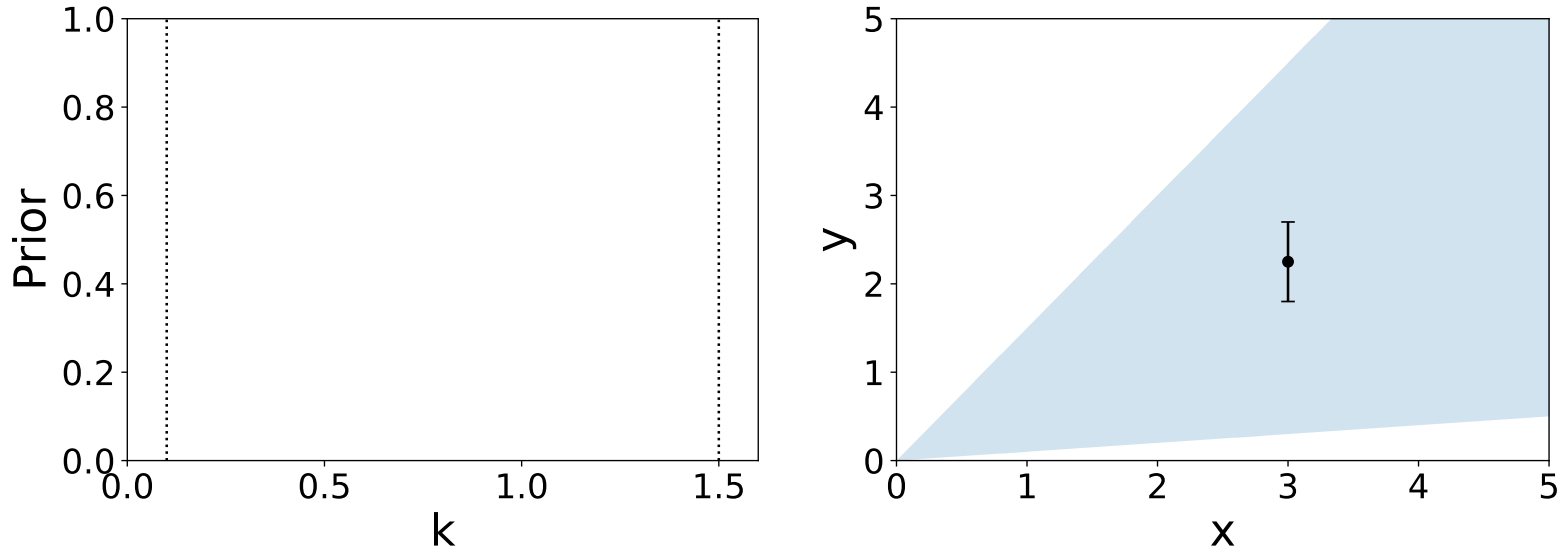
Simple example: $y_k(x) = kx$, determine linear coefficient k

Prior: k is between 0.1 and 1.5, uniform probability distribution

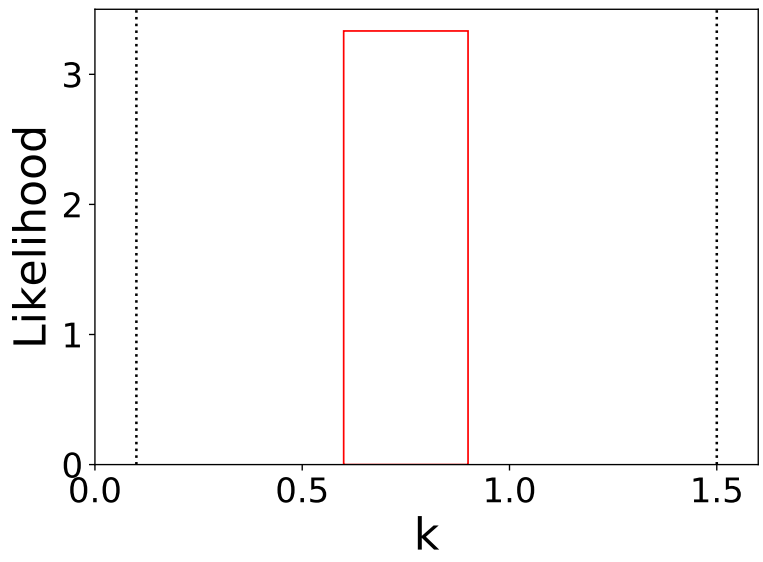


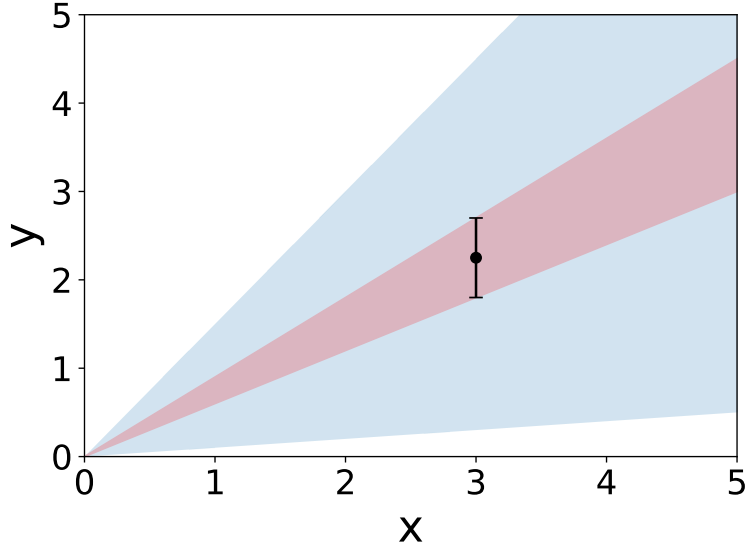


Observation A: $y_A = 2.25 \pm 0.45$ at $x_A = 3.0$

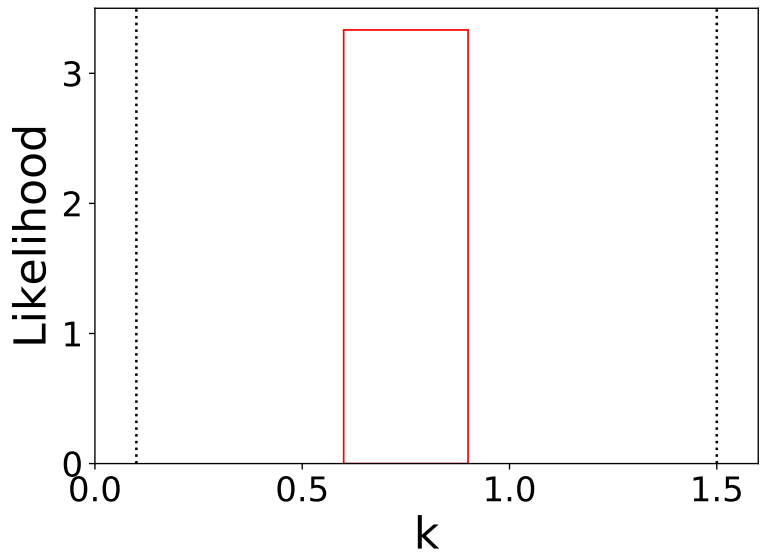


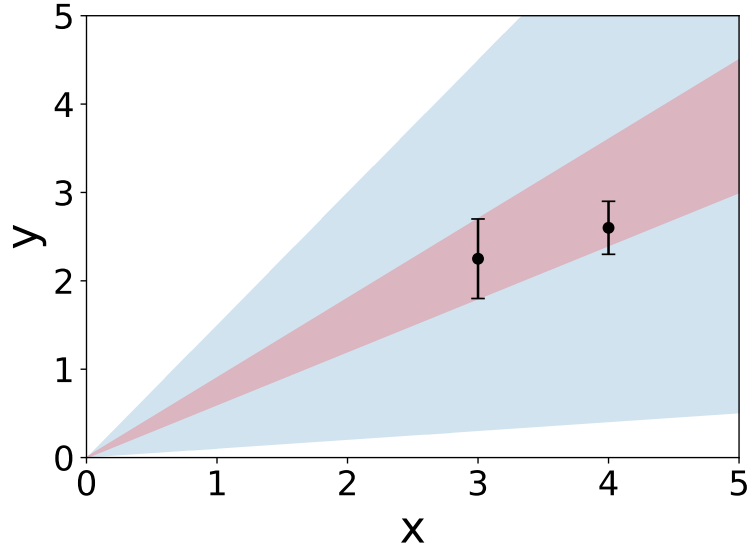
Simplistic likelihood: $\mathcal{L}(k) =$ constant if
$y_A - \sigma_A < y_k(x_A) < y_A + \sigma_A$, 0 otherwise



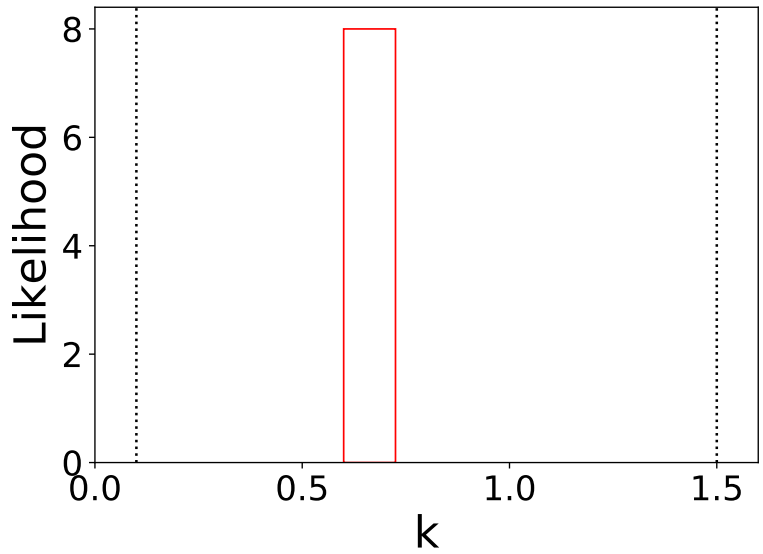


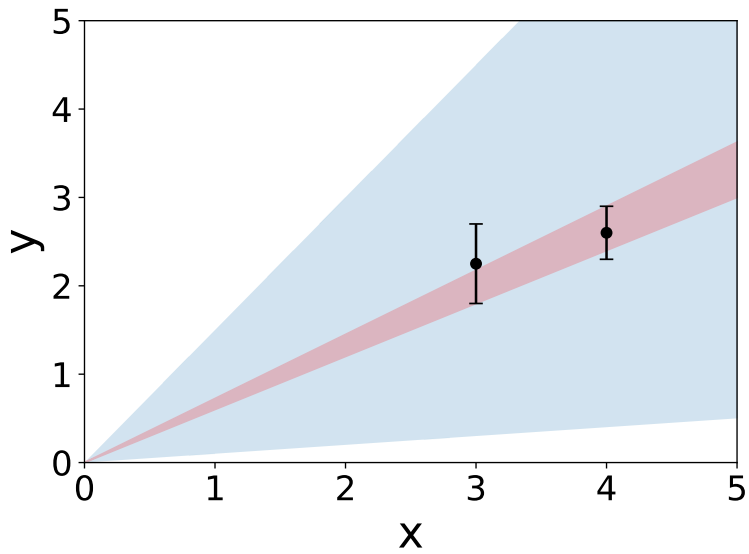
Add a more accurate observation B: $y_B = 2.6 \pm 0.3$ at $x_B = 4.0$





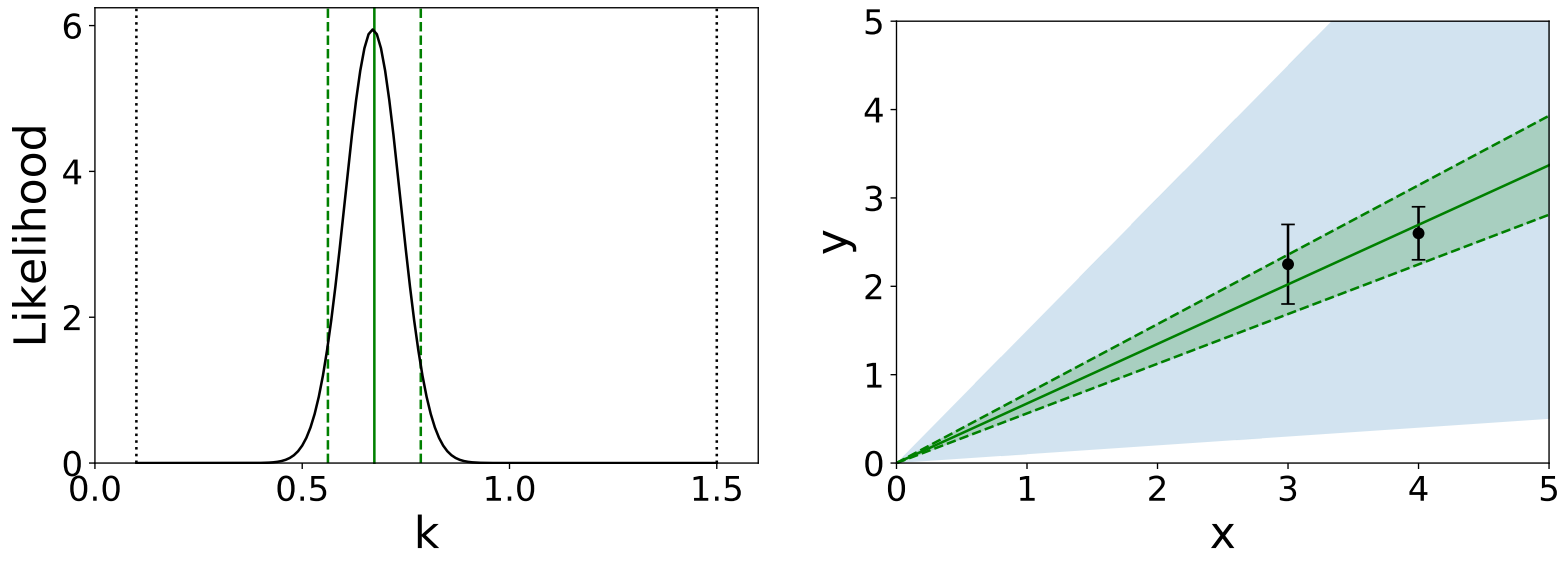
Problem: Product of step functions accounts only for the intersection of the two measurements





More typical choice for the likelihood:

$$\mathcal{L}(k) = \frac{1}{N} \exp\left(-\frac{1}{2}\left[\frac{(y_k(x_A)-y_A)^2}{\sigma_A^2} + \frac{(y_k(x_B)-y_B)^2}{\sigma_B^2}\right]\right)$$



Dashed lines: 5th and 95th percentiles, demarcating the **90% credible interval**. Solid line: 50th percentile, indicating the **median**

Mahalanobis distance

We would like to generalize the likelihood function for arbitrary dimensions.

This can be conveniently done using Mahalanobis distance

$$d_M(\vec{a}, \vec{b}) = \sqrt{(\vec{a} - \vec{b})^T \Sigma^{-1} (\vec{a} - \vec{b})}.$$

where Σ is the covariance matrix containing the uncertainties (in the previous example $\Sigma = \text{diag}(\sigma_A^2, \sigma_B^2)$)

We can then define the likelihood as

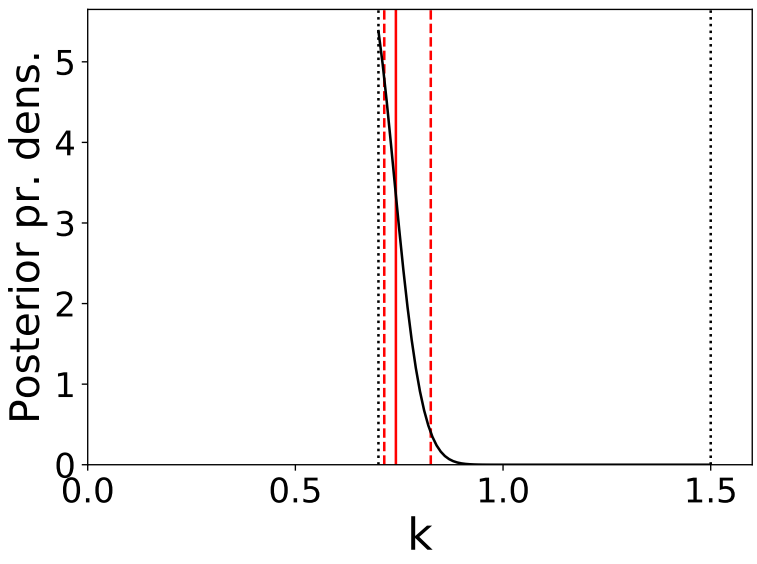
$$\mathcal{L} = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{d_M^2}{2}\right)$$

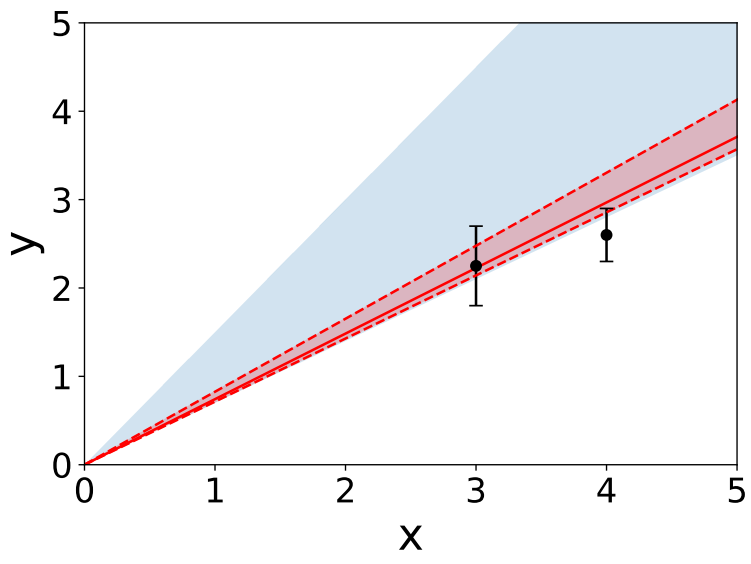
where the determinant $|2\pi\Sigma|$ functions as a penalty factor for estimates which have large uncertainties

Caveat of uniform prior probability

Uniform prior $\Rightarrow$ Shape of posterior $\sim$ Likelihood (**Data-driven posterior**)

However, while uniform priors are non-informative within the defined interval, they become very informative at the edges!





Conjugate prior

In cases where we can write our likelihood as a normal distribution with a known, constant variance σ^2

$$\mathcal{L}(y) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - y)^2\right] \propto \exp\left[-\frac{n}{2\sigma^2}(\langle y\rangle - y)^2\right],$$

its conjugate prior is a normal distribution

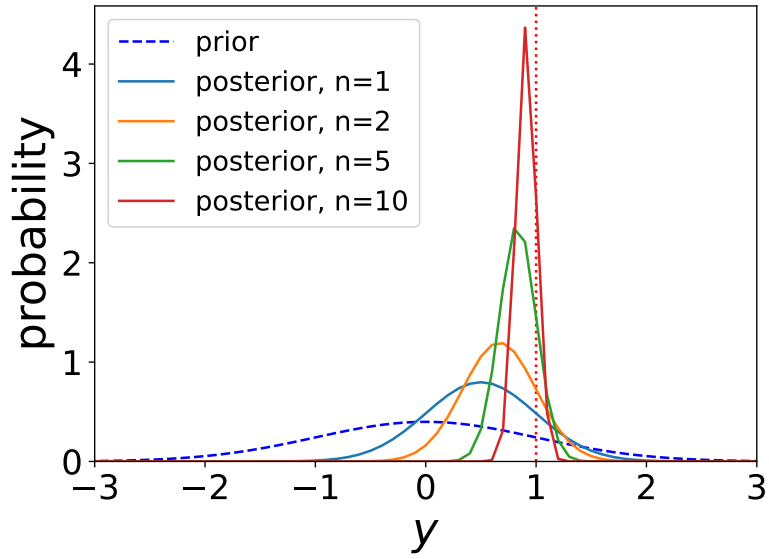
$$p(y) \propto \exp\left[-\frac{1}{2\sigma_p^2}(y - \mu_p)^2\right].$$

This means the posterior distribution belongs to same distribution family as the prior (i.e. normal):

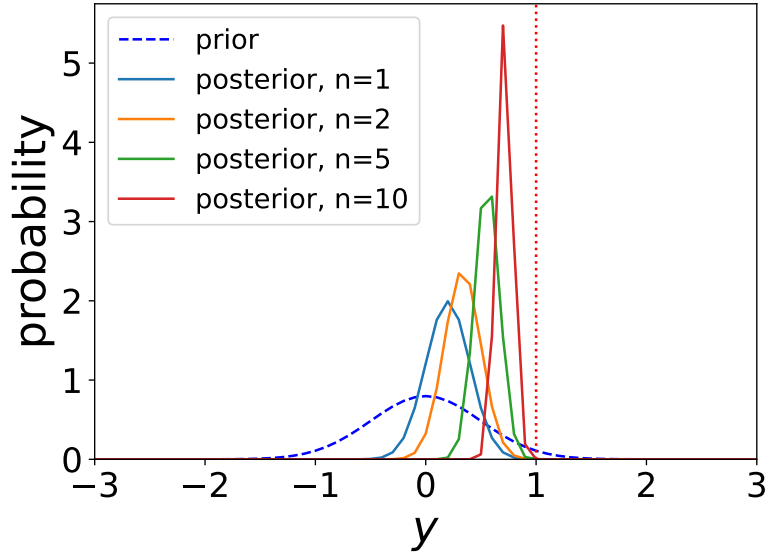
$$P(y) \propto \exp\left[-\frac{1}{2\sigma_n^2}(y - \mu_n)^2\right]$$

with $\sigma_n^2 = \frac{1}{n\sigma^{-2}+\sigma_p^{-2}}$ and $\mu_n = \sigma_n^2\left(\frac{\langle y\rangle}{\sigma^2/n} + \frac{\mu_p}{\sigma_p^2}\right)$.

Prior: $\mu_p = 0.0$, $\sigma_p = 1.0$,
measurement (likelihood):
$\langle y \rangle = 1.0$, $\sigma = 1.0$



Prior: $\mu_p = 0.0$, $\sigma_p = 0.5$,
measurement (likelihood):
$\langle y \rangle = 1.0$, $\sigma = 1.0$



Bayes factor

The evidence $\mathcal{E} = \int_{\vec{x}} \mathcal{L}(\vec{y}^{\,\mathrm{exp}}|\vec{x})p(\vec{x})$ can usually be ignored when determining best-fit parameters for a model, as we're most interested in ratios of likelihood

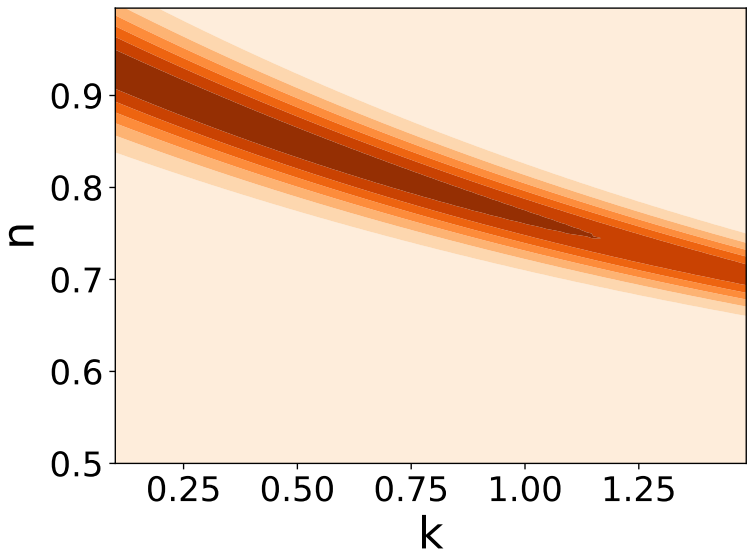
It can, however, be used to compare the credibility of two competing models M_A and M_B:

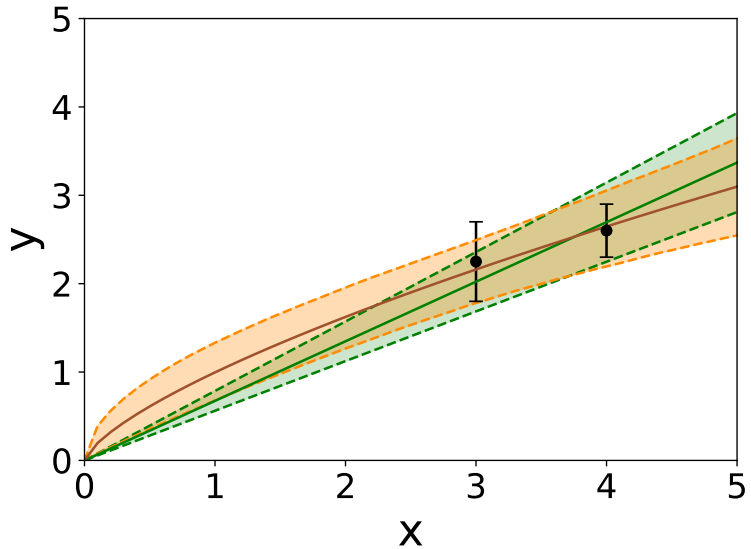
$$\text{Bayes factor } B_{12} = \frac{\mathcal{E}_A}{\mathcal{E}_B}\frac{p(M_A)}{p(M_B)}$$

$B_{AB} \gg 1 \Rightarrow$ model A preferred over model B. If we have no prior preference, $p(M_A) = p(M_B)$ and the Bayes factor is simply the ratio of evidences:

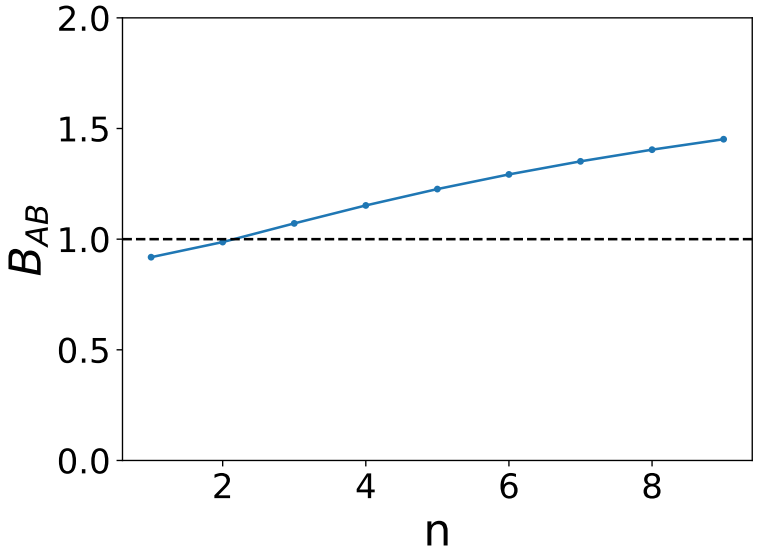
$$B_{AB} = \frac{\mathcal{E}_A}{\mathcal{E}_B} = \frac{\int_{\vec{x_A}} \mathcal{L}_A(\vec{y}^{\,\mathrm{exp}}|\vec{x_A})p_A(\vec{x_A})}{\int_{\vec{x_B}} \mathcal{L}_2(\vec{y}^{\,\mathrm{exp}}|\vec{x_B})p_B(\vec{x_B})}$$

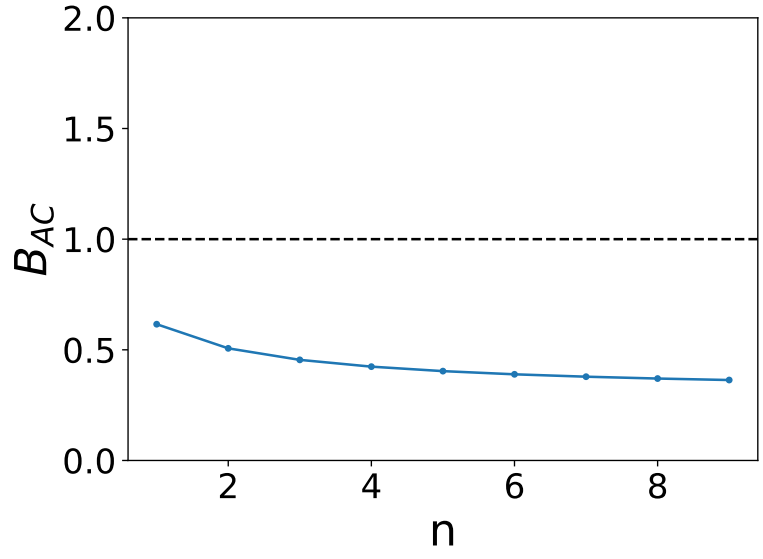
Example: 1-parameter model $M_1 : y = kx$
vs. 2-parameter model $M_2 : y = kx^n$
Bayes factor $B_{12} \approx 0.8 \Rightarrow$ No significant preference





Example: Measurement (likelihood): $\langle y \rangle = 1.0$, $\sigma = 1.0$
Model A prior: $\mu_A = 0.0$, $\sigma_A = 1.0$
Model B prior: $\mu_B = 0.0$, $\sigma_B = 0.5$
Model C prior: $\mu_C = 1.0$, $\sigma_C = 0.5$



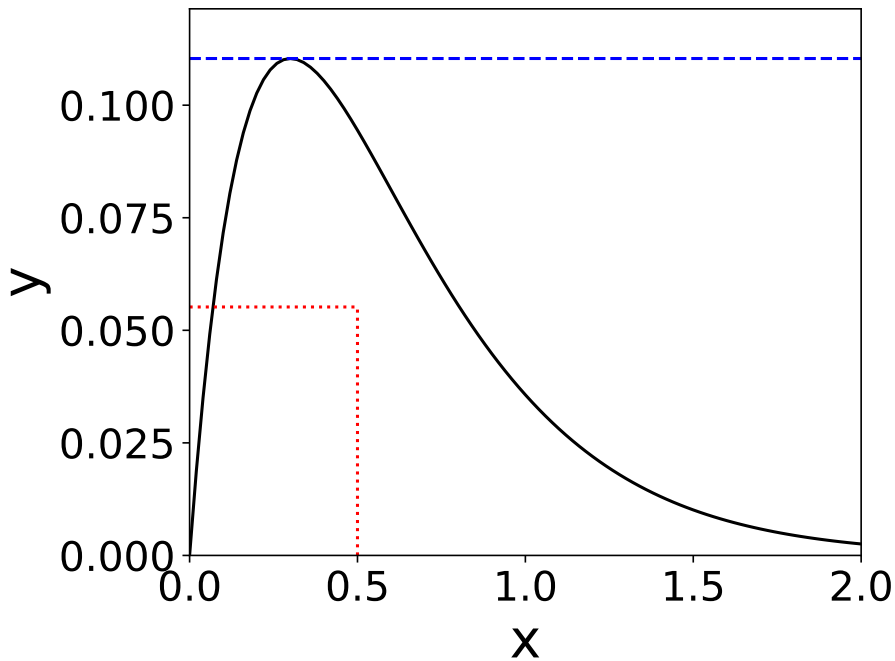


Often in practical applications with several model parameters, the exact expression for posterior probability distribution becomes difficult to obtain

$\Rightarrow$ Use **Markov chain Monte Carlo** method to sample the posterior

Monte Carlo sampling of x from a distribution $f(x)$:

- Pick a random value x_r from the uniform range $[x_{\text{min}}, x_{\text{max}}]$
- Pick a random value f_r from the uniform range $[f_{\text{min}}, f_{\text{max}}]$
- If $f(x_r) \geq f_r$, accept x_r



... What if we don't know the exact expression for f?

Markov Chain Monte Carlo

Stochastic process (discrete time): Indexed collection of random variables $\{R(t_i)\}_{i=1}^{N} \equiv \{R_i\}_{i=1}^{N}$

Markov chain: A stochastic process where the future states depend only on the present state (Markov property or "memorylessness")

$$p[R_{n+1} = r | R_1 = r_1, ..., R_n = r_n] = p[R_{n+1} | R_n = r_n]$$

Examples:

- Discrete time: 1-d random walk starting at 0 with possible steps -1 and 1
- Continuous time: Brownian motion

The goal: Generate samples $x_i \to x_{i+1}$ such that the distribution of samples $d(x) = m(x)/n$ ($m(x)$ being the number of samples in a region around x) converges to the posterior distribution $P(x)$ as $n \to \infty$

To ensure convergence, we require **detailed balance** (reversibility)

$$p(x_i \to x_{i+1})P(x_i) = p(x_{i+1} \to x_i)P(x_{i+1}) \Rightarrow \frac{p(x_i \to x_{i+1})}{p(x_{i+1} \to x_i)} = \frac{P(x_{i+1})}{P(x_i)}$$

We decompose the transition probability $p(x_i \to x_{i+1})$ to a **proposal distribution** Q and **acceptance probability** A:

$$p(x_i \to x_{i+1}) = Q(x_i \to x_{i+1})A(x_i \to x_{i+1})$$

For A to satisfy detailed balance, we must have

$$\frac{A(x_i \to x_{i+1})}{A(x_{i+1} \to x_i)} = \frac{P(x_{i+1})}{P(x_i)} \frac{Q(x_{i+1} \to x_i)}{Q(x_i \to x_{i+1})}$$

Metropolis-Hastings algorithm:

Metropolis acceptance criterion

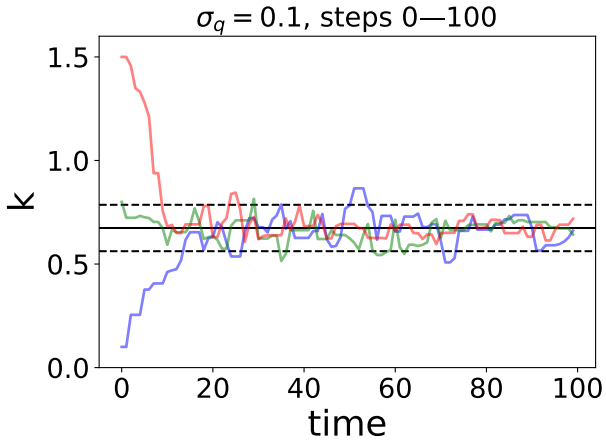
$$A(x_i \to x_{i+1}) \equiv \min\left\{1, \frac{P(x_{i+1})}{P(x_i)} \frac{Q(x_{i+1} \to x_i)}{Q(x_i \to x_{i+1})}\right\}$$

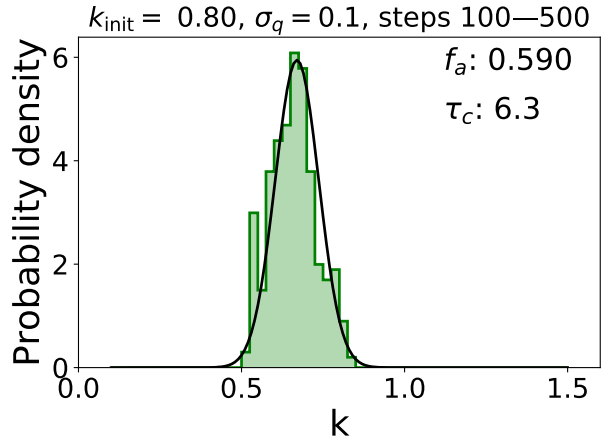
1. Draw a proposal for $x_i \to x'_{i+1}$ from Q
2. Compute acceptance probability $A(x_i \to x'_{i+1})$
3. Pick a random number r from uniform range $[0, 1]$
4. If $A(x_i \to x'_{i+1}) > r$, accept the proposed move and set $x_{i+1} = x'_{i+1}$. Otherwise set $x_{i+1} = x_i$
5. Set $i = i + 1$ and repeat the process

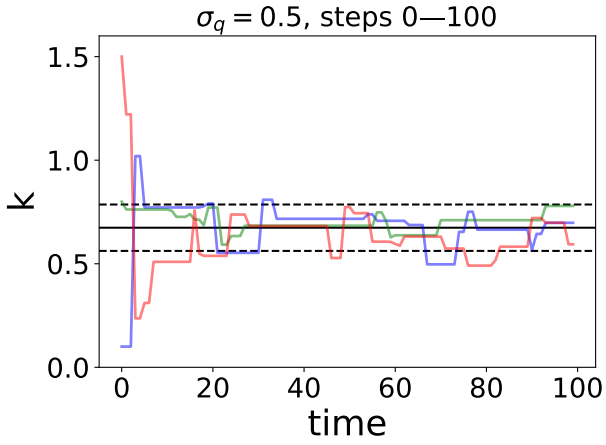
Example: 1-parameter model with step proposals from the normal distribution

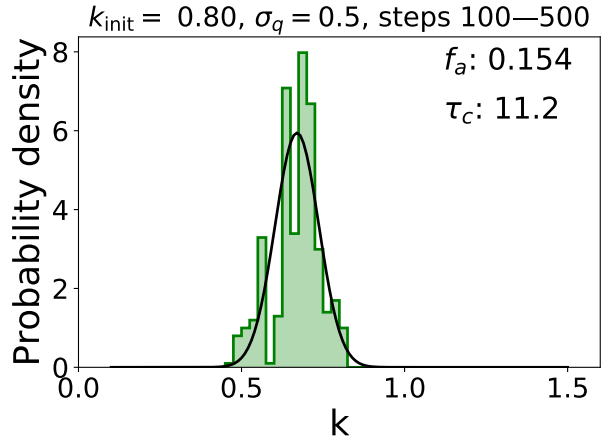
$$Q(k', k_0) = \frac{1}{\sqrt{2\pi\sigma_q}} e^{-(k'-k_0)^2/(2\sigma_q^2)},$$

$$A(k_0 \to k') = \min\left\{1, \frac{\mathcal{L}(k')}{\mathcal{L}(k)} \frac{Q(k_0,k')}{Q(k',k_0)}\right\}$$









MCMC has a "burn-in" period that depends on the initialisation of the chain and the proposal distribution Q — plot the trace (chain values vs time) to estimate the number of steps to discard from the start of the chain

Any given sample k_i will be correlated with the previous sample k_{i-1} and the following sample k_{i+1}, and the strength of the correlation depends on the acceptance probability A (rejecting the proposed new position means k_i and k_{i+1} will be maximally correlated). These correlations reduce the sampling quality of MCMC

Ways to quantify chain quality:

Acceptance fraction $f_a = \frac{N(\text{accepted})}{N(\text{proposed})}$.

- $f_a \ll 1$: Strong correlations, poor sampling quality
- $f_a \approx 1$: Distribution of samples is Q instead of P
- Rule of thumb: $0.2 < f_a < 0.5$

Autocorrelation time

- Autocovariance function with respect to lag T for a chain of M samples: $C(T) \approx \frac{1}{M-T} \sum_{n=1}^{M-T} (x_{T+n} - \langle x \rangle)(x_n - \langle x \rangle)$
- Autocorrelation time $\tau_c = 1 + 2 \sum_{T=1}^{M_c} \frac{C(T)}{C(0)}$
  where $M_c = \min \{m > K\tau_c(m)\}_{m<M}$ for a constant $K \approx 5$

Parallel tempering

If the posterior distribution has multiple peaks, using parallel tempered MCMC (PTMCMC) might provide better results (at the cost of computation time)

As the name suggests, PTMCMC runs several chains in parallel, each chain sampling a different modified posterior defined by inverse temperature $\beta = 1/T$ ("thermal noise"):

$$P_\beta(\vec{x}|\vec{y}^{\,\mathrm{exp}}) = [\mathcal{L}(\vec{y}^{\,\mathrm{exp}}|\vec{x})]^\beta p(\vec{x})$$

Two chains swap walker positions at pre-determined intervals with probability

$$A_{i,j} = \min\left\{\left(\frac{\mathcal{L}(x_i)}{\mathcal{L}(x_j)}\right)^{\beta_j-\beta_i}, 1\right\}$$

$\beta \to 0\,(T \to \infty) \Rightarrow P_\beta \to p$
$\beta \to 1\,(T \to 1) \Rightarrow P_\beta \to P$

We can obtain the full integral over the parameter space (i.e. the evidence $\mathcal{E}$) using parallel tempered MCMC and thermodynamic integration

Let's redefine the evidence as a function of inverse temperature β:

$$\mathcal{E}(\beta) = \int_{\vec{x}} [\mathcal{L}(\vec{y}^{\,\mathrm{exp}}|\vec{x})]^{\beta} p(\vec{x}) \equiv \int_{\vec{x}} \mathcal{L}^{\beta}(\vec{x}) p(\vec{x})$$

This satisfies the differential equation

$$\frac{\mathrm{d}(\log \mathcal{E})}{\mathrm{d}\beta} = \frac{1}{\mathcal{E}(\beta)} \int_{\vec{x}} \log[\mathcal{L}(\vec{x}) \mathcal{L}^{\beta}(\vec{x})] p(\vec{x}) \equiv \langle \log \mathcal{L} \rangle_{\beta}$$

Integrating over β gives the evidence estimate:

$$\log \mathcal{E}(1) = \int_0^1 d\beta \, \langle \log \mathcal{L} \rangle_{\beta}$$

Gaussian process

Markov chain Monte Carlo requires a large number of model evaluations for the likelihood function.

However, it can take hundreds of CPU hours to simulate a heavy ion collision event.

$\Rightarrow$ Need a method for producing quick approximations for the given model parameter values
$\Rightarrow$ Model emulation using Gaussian processes

Reminder: stochastic process in discrete time: Indexed collection of random variables $\{R(t_i)\}_{i=1}^{N}$.

Gaussian process: A stochastic process, in which every finite set $\{Y_i\}_{i=1}^{N}$ is a multivariate Gaussian (or "normal") random variable $\mathcal{N}(\mu, \Sigma)$, where

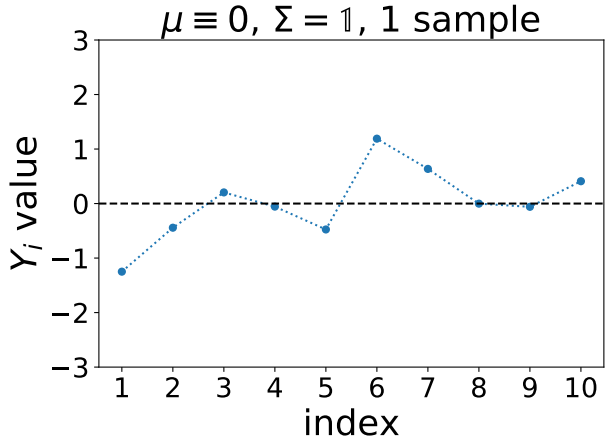
$$\mu = \{\mu_1, ..., \mu_N\}$$

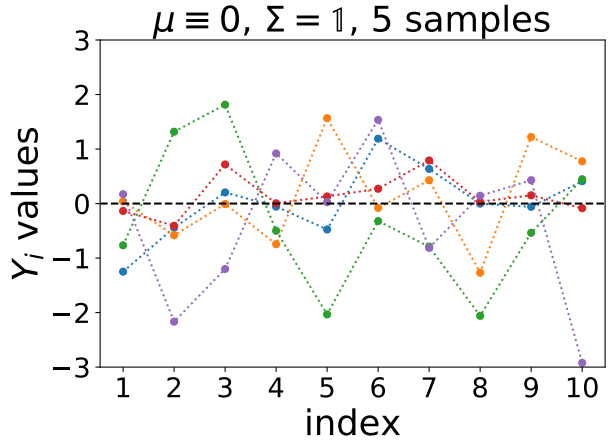
is the mean and

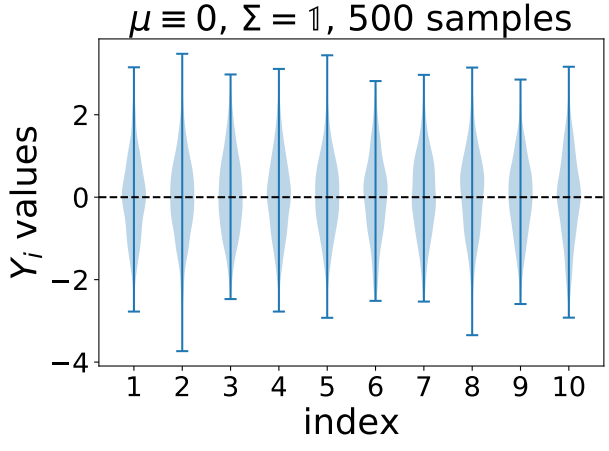
$$\Sigma = \begin{pmatrix} v(1,1) & \cdots & v(1,N) \\ \vdots & \ddots & \vdots \\ v(N,1) & \cdots & v(N,N) \end{pmatrix}$$

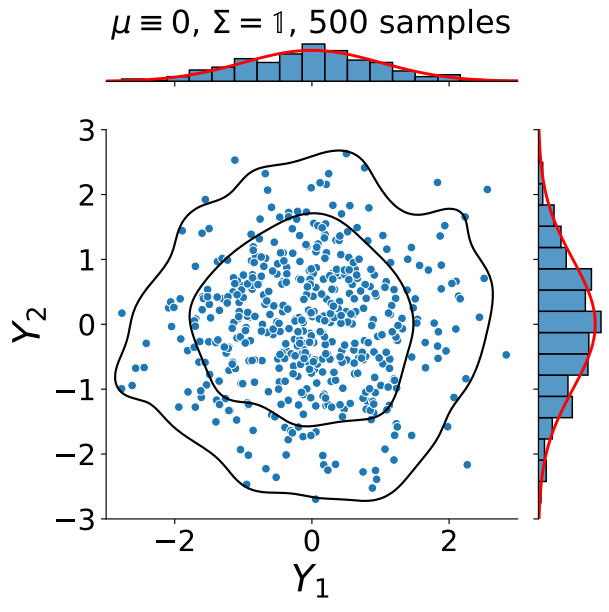
is the covariance matrix. The covariance function (or "kernel") $v(i,j)$ is symmetric ($v(i,j) = v(j,i)$) and positive semidefinite ($v(i,j) \geq 0$).

Gaussian process examples

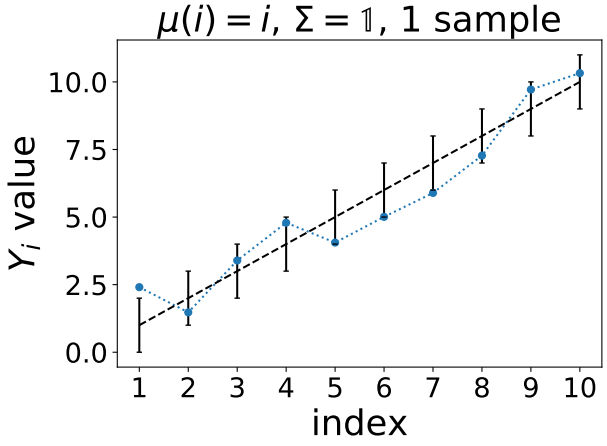


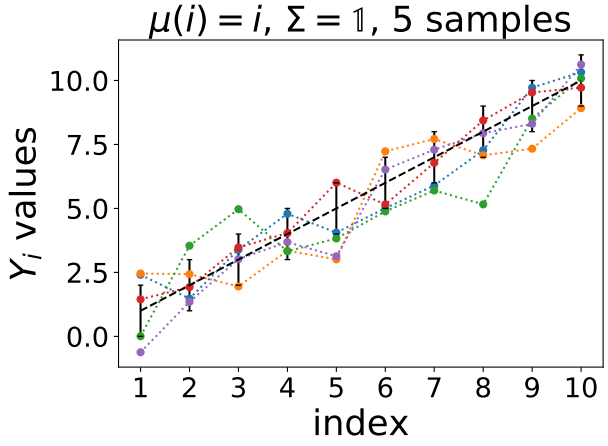


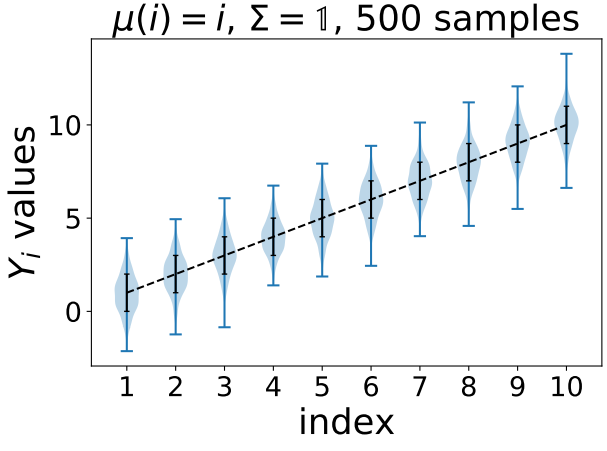


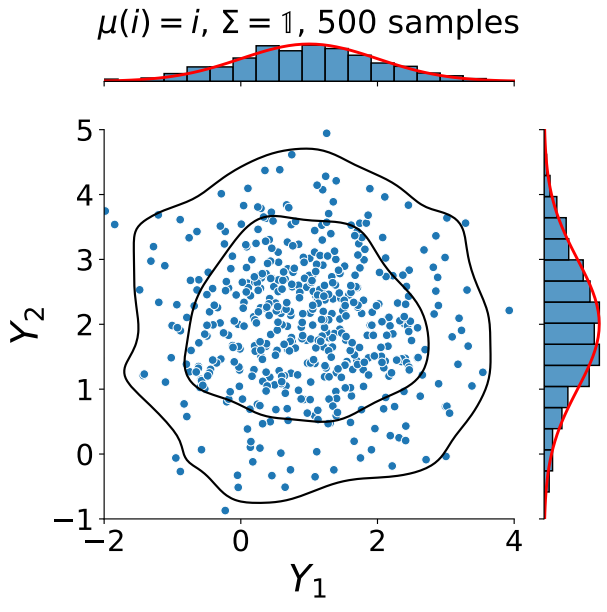


Gaussian process examples

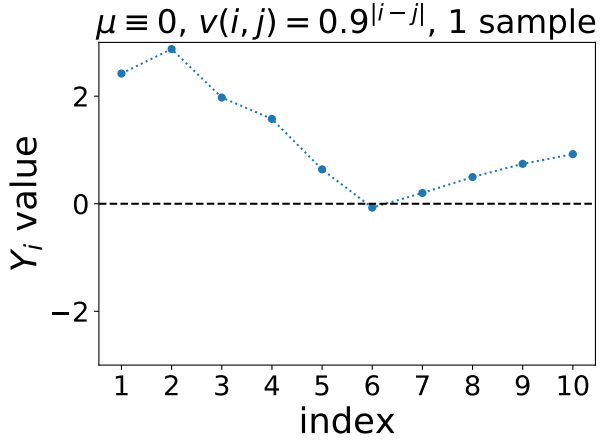


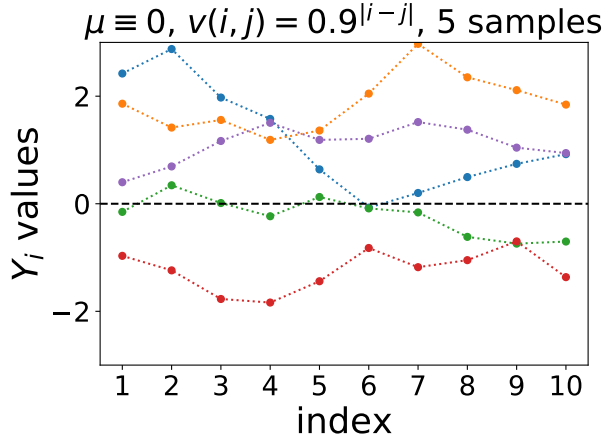


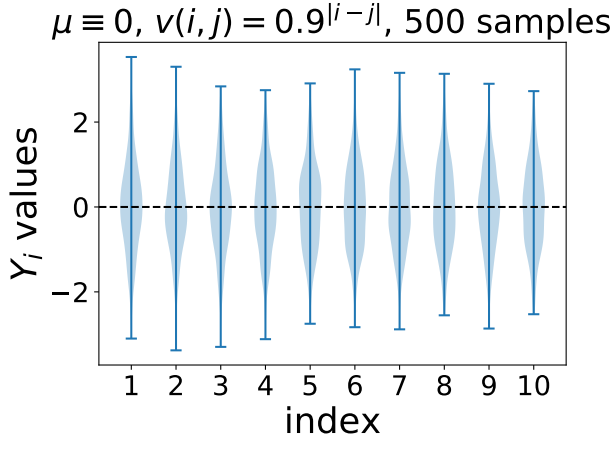


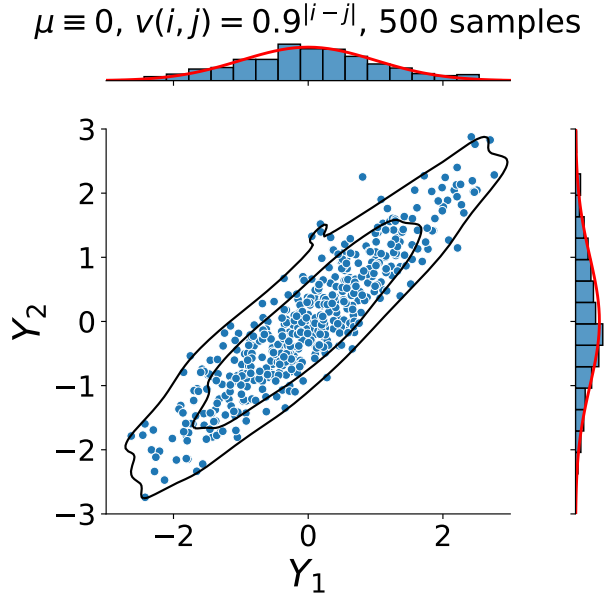


Gaussian process examples









Multivariate normal random variables

Multivariate normal random variables have the following important property:

Let $A \sim \mathcal{N}(\mu, \Sigma)$ be a n-dimensional normal random vector. If $B = c + TA$ is an affine transformation with m-dimensional vector c and a constant $m \times n$-dimensional matrix T, then $B \sim \mathcal{N}(c + T\mu, T\Sigma T^T)$.

Let's denote a joint distribution for normal random vectors A and B as

$$J = \begin{pmatrix} A \\ B \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \begin{pmatrix} \Sigma_A & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_B \end{pmatrix} \right)$$

Let us then define a transformation matrix T (assuming Σ_B^{-1} exists!)

$$T = \begin{pmatrix} I & -\Sigma_{AB}\Sigma_B^{-1} \\ 0 & I \end{pmatrix}$$

Now the affine transformation $J' = T(J - \mu_J)$ is also a joint normal distribution:

$$\begin{aligned} J' = \begin{pmatrix} A' \\ B' \end{pmatrix} &= \begin{pmatrix} A - \mu_A - \Sigma_{AB}\Sigma_B^{-1}(B - \mu_B) \\ B - \mu_B \end{pmatrix} \\ &\sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_A - \Sigma_{AB}\Sigma_B^{-1}\Sigma_{BA} & 0 \\ 0 & \Sigma_B \end{pmatrix} \right) \end{aligned}$$

$$\begin{pmatrix} A' \\ B' \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_A - \Sigma_{AB}\Sigma_B^{-1}\Sigma_{BA} & 0 \\ 0 & \Sigma_B \end{pmatrix} \right)$$

A' and B' are jointly normal and uncorrelated $\Rightarrow$ A' is independent of B'

$\Rightarrow$ A' is independent of B

$\Rightarrow$ Conditional probability distribution of A' with fixed $B = b_0$:

$$A'|_{B=b_0} \sim \mathcal{N}(0, \Sigma_A - \Sigma_{AB}\Sigma_B^{-1}\Sigma_{BA})$$

$A = A' + \mu_A + \Sigma_{AB}\Sigma_B^{-1}(b_0 - \mu_B)$, where $\mu_A + \Sigma_{AB}\Sigma_B^{-1}(b_0 - \mu_B)$ is a constant

$$\Rightarrow \boxed{A|_{B=b_0} \sim \mathcal{N}(\mu_A + \Sigma_{AB}\Sigma_B^{-1}(b_0 - \mu_B), \Sigma_A - \Sigma_{AB}\Sigma_B^{-1}\Sigma_{BA})}$$

To summarise: if A and B have the joint probability

$$\begin{pmatrix} A \\ B \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \begin{pmatrix} \Sigma_A & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_B \end{pmatrix} \right)$$

then A has a conditional mean

$$A|_{B=b_0} = \mu_A + \Sigma_{AB}\Sigma_B^{-1}(b_0 - \mu_B)$$

and conditional covariance

$$v_A|_{B=b_0} = \Sigma_A - \Sigma_{AB}\Sigma_B^{-1}\Sigma_{BA}$$

for a fixed $B = b_0$.

If we have a function f with a set of known values $Y = \{f(x_i)\}_{i=1}^N$, we can use this result for predicting function values at other points x' with the associated uncertainty!

Gaussian process regression

The problem: We need to know the covariance matrix for the given data Y.

Typical choice for the covariance function: squared-exponential with variable amplitude θ_A, correlation length θ_L and a noise term θ_{noise}:

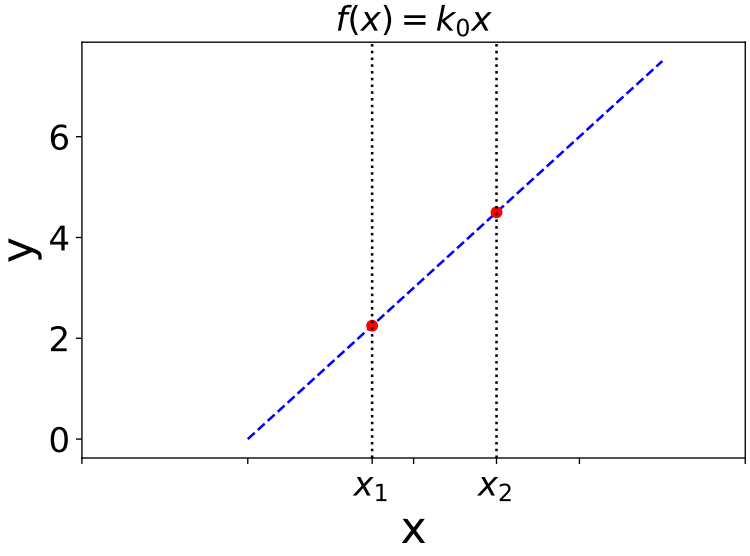
$$v(x, x') = \theta_A \exp\left(-\frac{(x - x')^2}{2\theta_L^2}\right) + \theta_{\text{noise}}\delta_{xx'}$$

The *hyperparameters* $\vec{\theta} = (\theta_A, \theta_L, \theta_{\text{noise}})$ are not known a priori and have to be estimated from the model data Y by maximising the log-likelihood

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

This is known as emulator training.

1-D example



$$Y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} f(x_1) \\ f(x_2) \end{pmatrix}$$

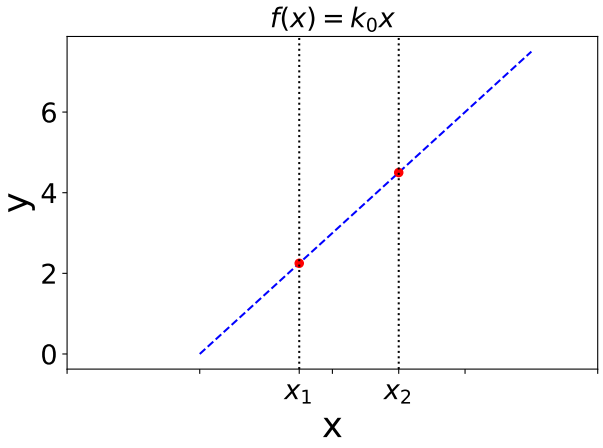
$$\Sigma_Y = \begin{pmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{pmatrix}$$

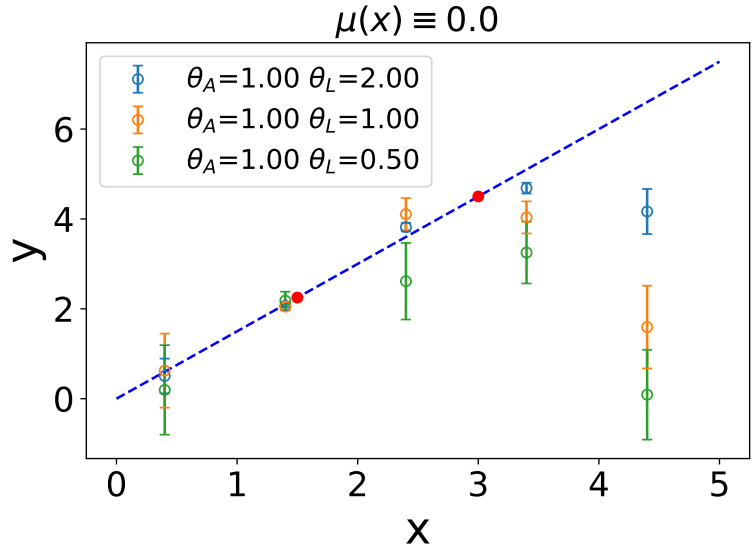
$$v_{ij} = \theta_A \exp\left(-\frac{(x_i - x_j)^2}{2\theta_L^2}\right)$$

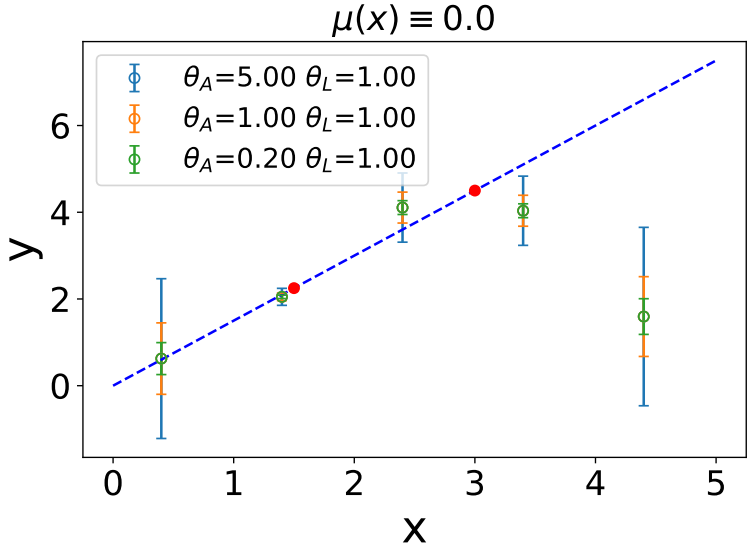
$$|\Sigma_Y| = v_{11}v_{22} - v_{12}^2 = \theta_A^2\left[1 - \exp\left(-\frac{(x_1 - x_2)^2}{2\theta_L^2}\right)\right] > 0$$

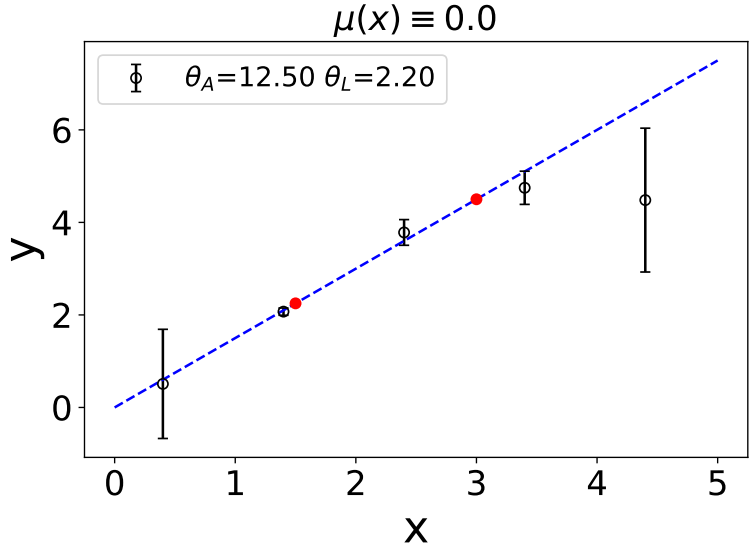
$$\Sigma_Y^{-1} = \frac{1}{|\Sigma_Y|}\begin{pmatrix} v_{22} & -v_{12} \\ -v_{12} & v_{11} \end{pmatrix}$$

1-D example


$f(x) = k_0 x$
y
x
x_1
x_2

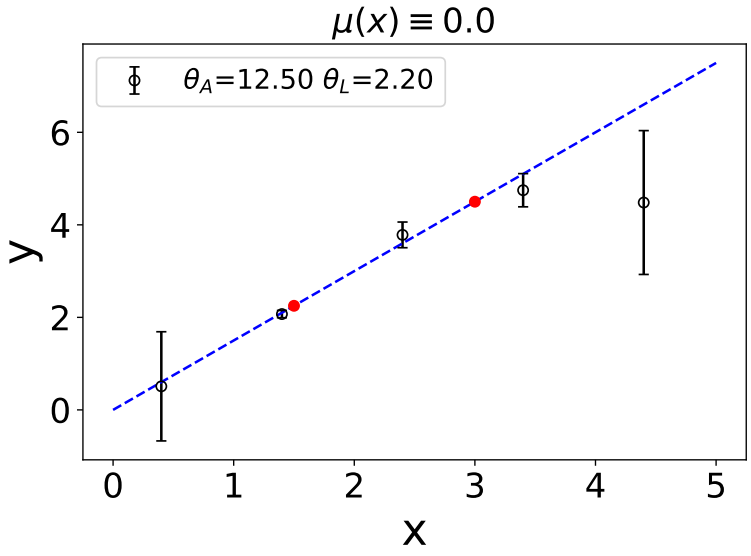

$\mu(x) \equiv 0.0$
θ_A=1.00 θ_L=2.00
θ_A=1.00 θ_L=1.00
θ_A=1.00 θ_L=0.50
y
x

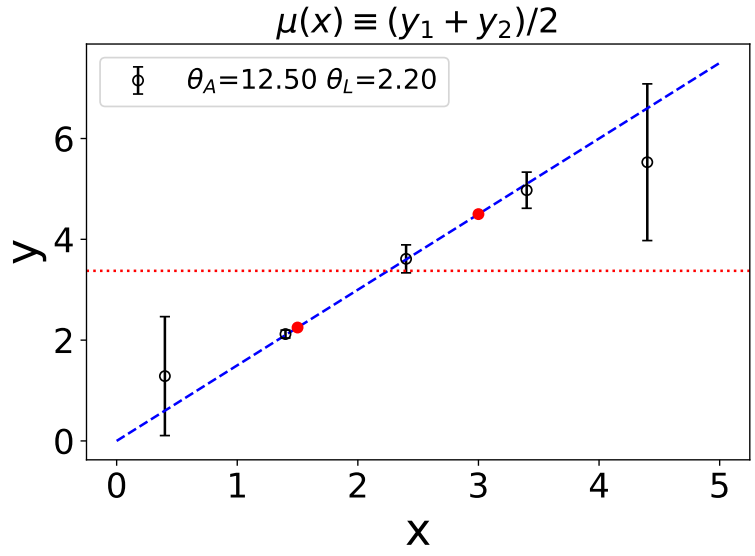

$\mu(x) \equiv 0.0$
θ_A=5.00 θ_L=1.00
θ_A=1.00 θ_L=1.00
θ_A=0.20 θ_L=1.00
y
x

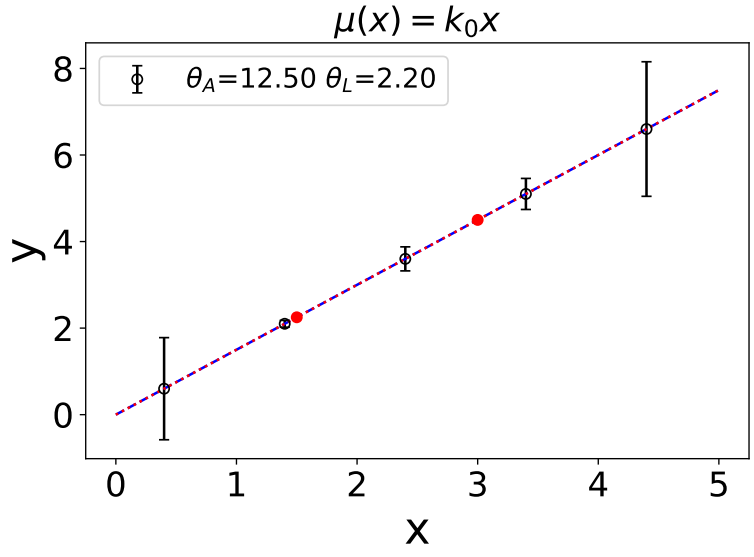

$\mu(x) \equiv 0.0$
θ_A=12.50 θ_L=2.20
y
x

1-D example

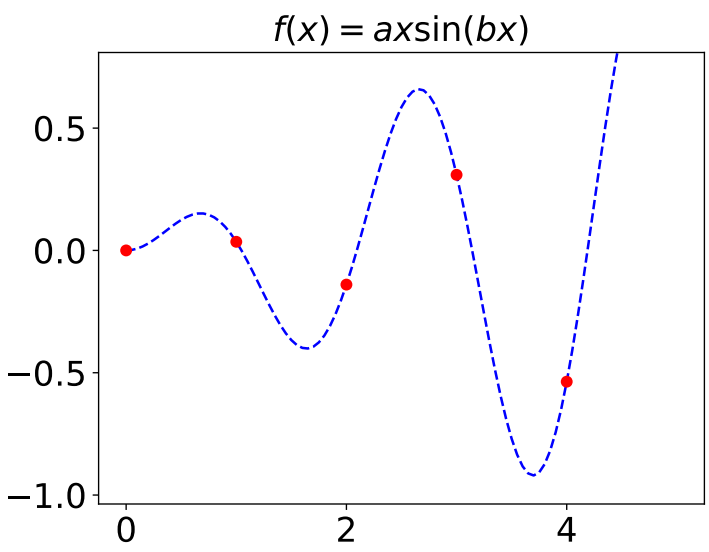
Effect of the mean function

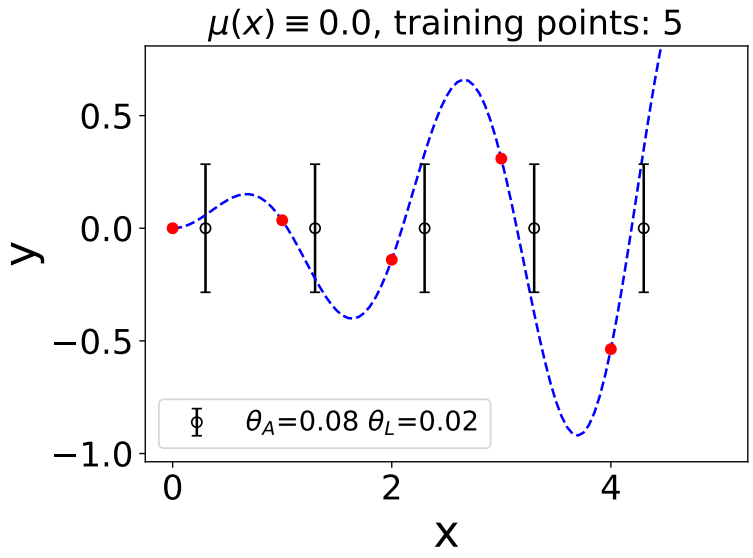


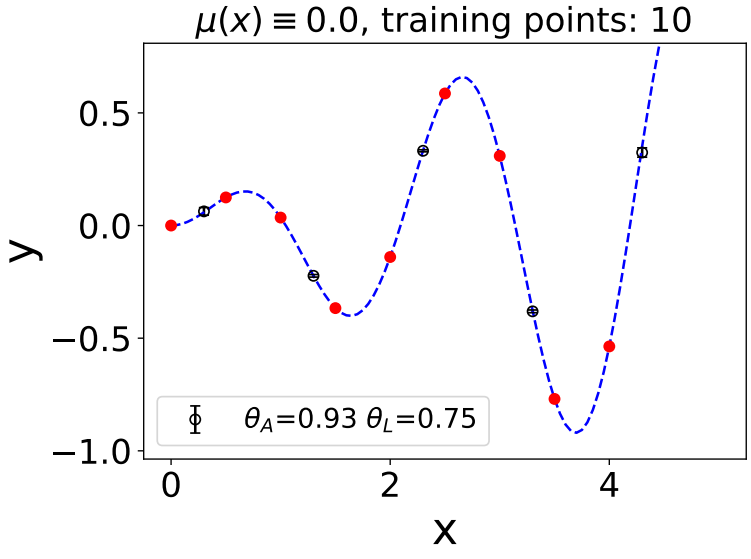


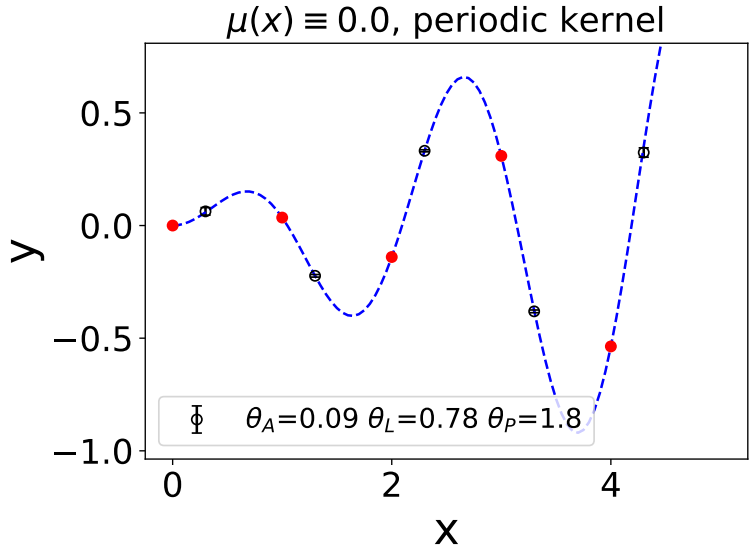


1-D example


$f(x) = ax\sin(bx)$
0.5
0.0
−0.5
−1.0
0
2
4


$\mu(x) \equiv 0.0$, training points: 5
0.5
0.0
−0.5
−1.0
y
θ_A=0.08 θ_L=0.02
0
2
4
x


$\mu(x) \equiv 0.0$, training points: 10
0.5
0.0
−0.5
−1.0
y
θ_A=0.93 θ_L=0.75
0
2
4
x


$\mu(x) \equiv 0.0$, periodic kernel
0.5
0.0
−0.5
−1.0
y
θ_A=0.09 θ_L=0.78 θ_P=1.8
0
2
4
x

Let the set of known function values be $Y = \{f(x_i)\}_{i=1}^{N}$ and the prediction point be $\hat{y} = f(\hat{x})$.

- Since the point $\hat{x}$ is correlated with all data points $\{x_i\}_{i=1}^{N}$, the ordering of points x_i (i.e. the indexing) in the set Y does not matter
- x_i and $\hat{x}$ can be vectors with arbitrary many dimensions. The squared-exponential covariance function can be generalized to M dimensions by introducing a new correlation length for each dimension:

$$v(\hat{x}, x_i) = \theta_A \exp\left(-\sum_{k=1}^{M} \frac{(\hat{x}_k - (x_i)_k)^2}{2(\theta_L)_k^2}\right) + \theta_{\text{noise}}\delta_{\hat{x}x_i}$$

$\Rightarrow$ Gaussian process regression is immediately applicable for a parameter space with arbitrary many dimensions!

Training data selection

As the GP prediction variance increases quickly in the absence of nearby conditioning points, the points in the dataset $\{(x_i, f(x_i))\}$ should be evenly distributed over the investigated parameter space.

Most straightforward solution: A grid over the parameter space. However, the number of points required increases quickly with higher dimensions

More economical solution: Latin hypercube sampling

- For N samples, partition the parameter range into N equal-probability intervals
- Sample parameter values within the intervals, optionally maximising the minimum distance between samples (by iteration)
- Rule of thumb: At least 10 samples per dimension

Emulator validation

We want to use Gaussian processes as surrogate models or "emulators" which can provide quick estimates for model output with quantified uncertainty. In such cases we rarely can compare the GP predictions to analytical results.

How can we assure ourselves that the emulator is working correctly?

Two sets of data are needed:

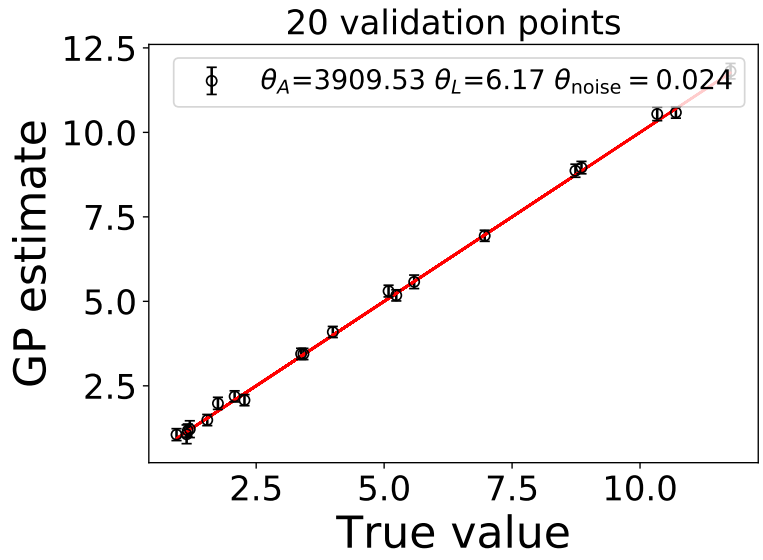
- Training set used for conditioning and tuning the GP emulator
- Validation set, consisting of points which are not in the training set

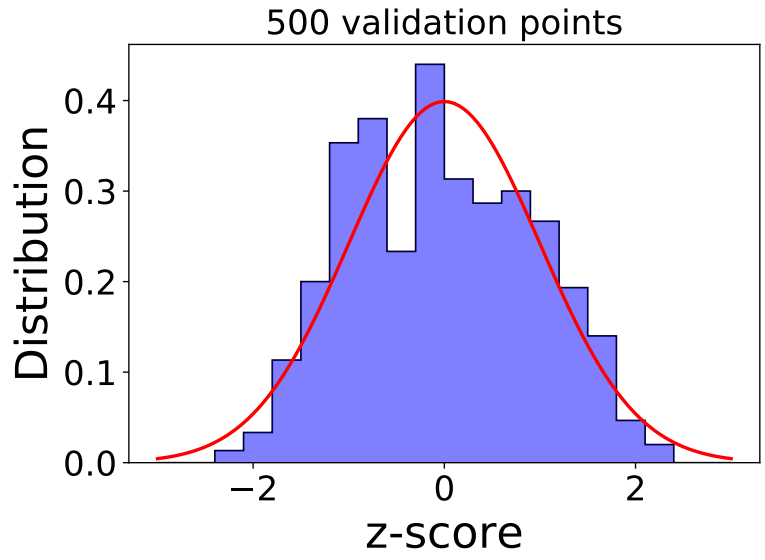
For a small number of samples ($\mathcal{O}(10)$), we can try to assess the emulator quality simply by plotting the validation set predictions against the true values from the model

If we have a larger number of samples ($\mathcal{O}(100)$), we can plot the z-scores:

$$z_i = \frac{y_i - \hat{y}_i}{\sqrt{v_i}}.$$

Since the emulator prediction consists of conditional mean $\hat{y}$ and variance v, we expect the distribution of z-scores to approach the normal distribution $\mathcal{N}(0, 1)$ as the sample size increases





While Gaussian processes are able to handle multidimensional input parameter spaces $\vec{x}$, the prediction $\hat{y}(\vec{x})$ is always a scalar

m observables $\Rightarrow$ m Gaussian processes needed for model emulation. However, in heavy ion collisions with multiple different particle types and centrality classes, m can be up to $\mathcal{O}(100)$

As the calculation of emulator predictions involves potentially high-dimensional matrix multiplication, the computational costs can become notable

Principal component analysis

Let's represent the model output with a n x m matrix M, where n is the number of simulation points and m the number of observables

In preparation for the principal component analysis, we need to normalize the data columns (either with their sample standard deviations or with the corresponding experimental values) to obtain dimensionless quantities of similar scale, and center by subtracting the mean of each observable from the elements of each column

Note: PCA assumes that variance is a good measure of the spread of data. If this is not the case (if the data is skewed, for example), further transformation methods are needed, such as Box-Cox transformation

$$y' = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{for} \lambda \neq 0 \\ \log(y) & \text{for} \lambda = 0 \end{cases}$$

We want to find an eigenvalue decomposition of the m x m covariance matrix $M^T M$:

$$M^T M = V \Lambda V^T,$$

where Λ is the diagonal matrix containing the eigenvalues $\lambda_1, ..., \lambda_m$ and V is an orthogonal matrix containing the eigenvectors. This decomposition is found by factorising M via singular value decomposition:

$$M = USV^T$$

- n x n orthogonal matrix U contains the left-singular vectors of M, which are eigenvectors of MM^T
- n x m rectangular diagonal matrix S contains the singular values of M (square roots of $M^T M$ eigenvalues)
- m x m orthogonal matrix V contains the right-singular vectors of M (eigenvectors of $M^T M$) — these are the principal components (PCs)

Principal component space is rotated and stretched in the direction of the first principal component(s). First principal component represents the direction of largest variance in the output space, second PC the direction of second largest variance, etc.

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

The transformation of a vector $\vec{y}$ from observable space to a vector $\vec{z}$ in the reduced-dimension principal component space is

$$\vec{z} = \vec{y}\, V_k,$$

while for matrices (such as the covariance matrix) the transformation is

$$\Sigma_z = V_k^T\, \Sigma_y\, V_k.$$

To compare an emulator prediction $\vec{z}^{\,\mathrm{GP}}$ against physical observables, we use the inverse transformation

$$\vec{y}^{\,\mathrm{GP}} = \vec{z}^{\,\mathrm{GP}}\, V_k^T.$$

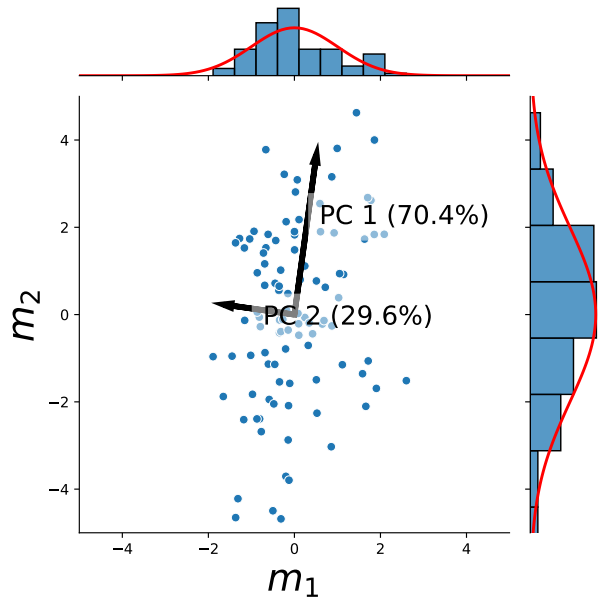
Likelihood function with GP and PCA

$$\frac{1}{|2\pi(\Sigma_{\text{exp}} + \Sigma_{GP})|} \exp\left(-\frac{1}{2}(\vec{z}^{*}_{GP} - \vec{z}_{\text{exp}})(\Sigma_{\text{exp}} + \Sigma_{GP})^{-1}(\vec{z}^{*}_{GP} - \vec{z}_{\text{exp}})^{T}\right)$$

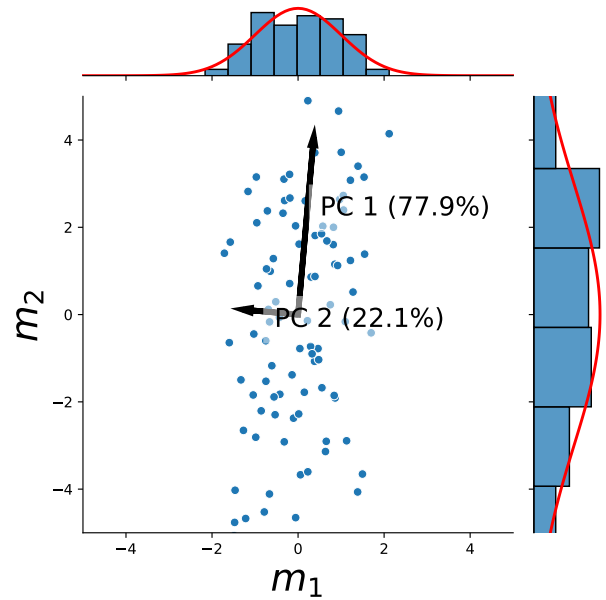
- $\vec{z}^{*}_{GP}$ is the emulator prediction at the input parameter point $\vec{x}^{*}$
- $\vec{z}_{\text{exp}}$ is the experimental data transformed to principal component space
- Σ_{GP} is the predictive variance (emulator uncertainty)
- Σ_{exp} is the experimental error squared, transformed to PC space

Example: Model output M consisting of two column vectors m_1 and m_2 sampled from a bivariate normal distribution (number of samples $n = 100$)

$$\begin{pmatrix} m_1 & m_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 & 0 \end{pmatrix}, \begin{pmatrix} v_1 & 0.5 \\ 0.5 & v_2 \end{pmatrix} \right)$$



$v_1 = 1.0, v_2 = 2.0$, 1:2 ratio



$v_1 = 1.0, v_2 = 3.0$, 1:3 ratio

Example: Model output M consisting of 10 column vectors m_1 ... m_{10} sampled from a multivariate normal distribution (number of samples $n = 200$)

$$\begin{pmatrix} m_1 & \cdots & m_{10} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 & \cdots & 0 \end{pmatrix}, \begin{pmatrix} v_{1,1} & \cdots & v_{1,10} \\ \vdots & \ddots & \vdots \\ v_{10,1} & \cdots & v_{10,10} \end{pmatrix} \right)$$

where $v_{i,i} = 10 \cdot 0.5^{(i-1)}$ and $v_{i,j} = v_{i,i} v_{j,j} \cdot 0.5^{|i-j|}$ $(i \neq j)$

Variance fractions for first 6 PCs: 0.583, 0.359, 0.0327, 0.00832, 0.00729, 0.00496 ($\geq 99.5\%$ of total variance)

Average transformation error with dimension reduced to 6 ($\tilde{y}_i = (y_i V_6) V_6^T$)

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\sqrt{(\tilde{y}_i - y_i)^2}}{|y_i|} \approx 9\%$$

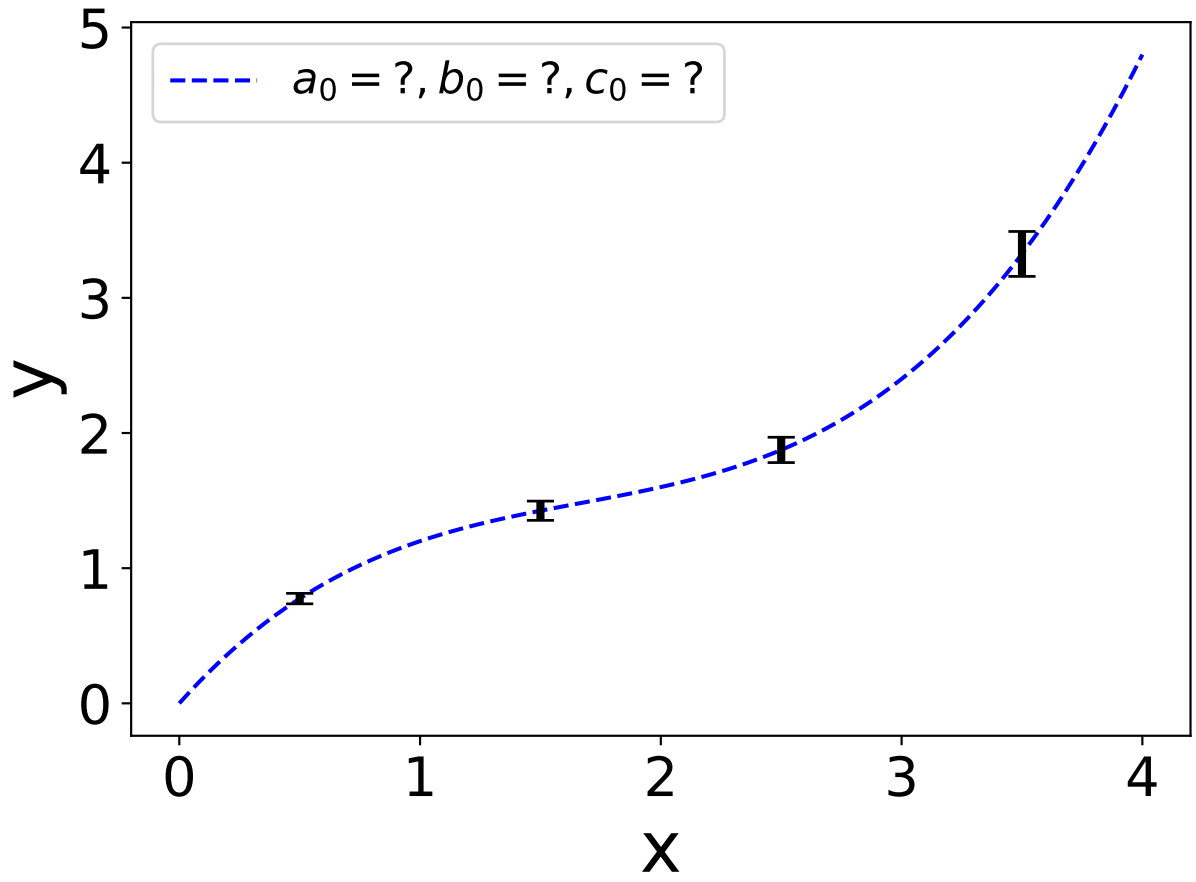
Closure testing

The final check before performing calibration to experimental data is closure testing, i.e. using pre-calculated model data (such as the validation data) as target value. This checks

- if the posterior probability distribution peaks at the original known parameter values (how credible is the analysis result)
- how wide is the distribution (how constraining is the dataset used in the analysis)

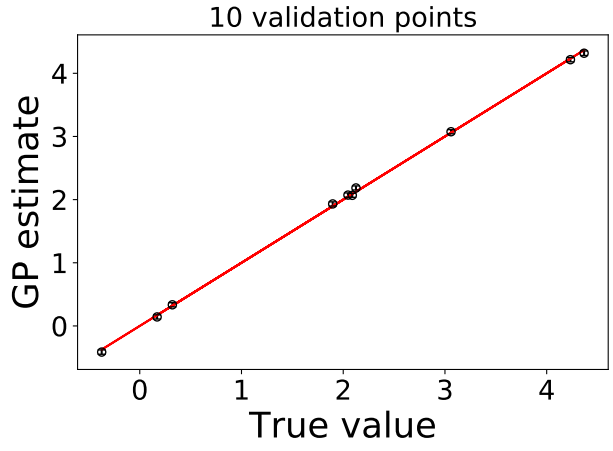
Example

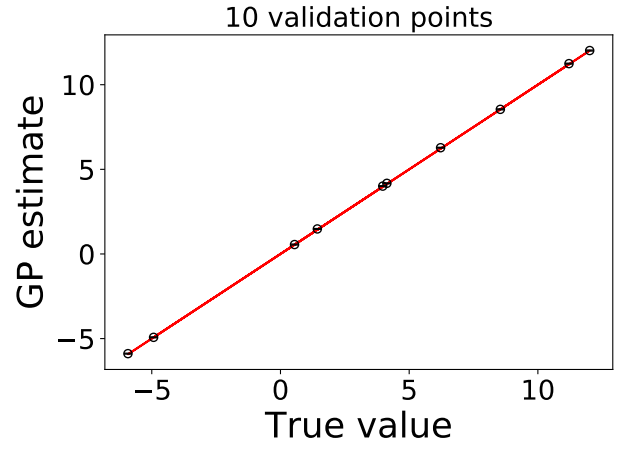
Three-parameter model $y(x; a, b, c) = ax - bx^2 + cx^3$ calibration to measured values y_0 at $x_0 = (0.5, 1.5, 2.5, 3.5)$ with 5% errors



Let's assume uniform prior distributions with ranges $a \in [0.0, 10.0]$, $b \in [0.0, 5.0]$, $c \in [0.0, 1.0]$. We sample 30 training points and 10 validation points from this cuboid.
4 measurements $\Rightarrow$ 4 GP emulators. GP emulator kernels shall be squared-exponentials with 3 correlation lengths

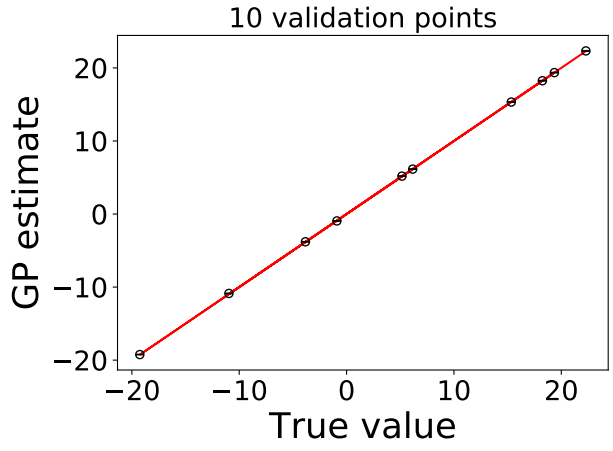
$x = 0.5$

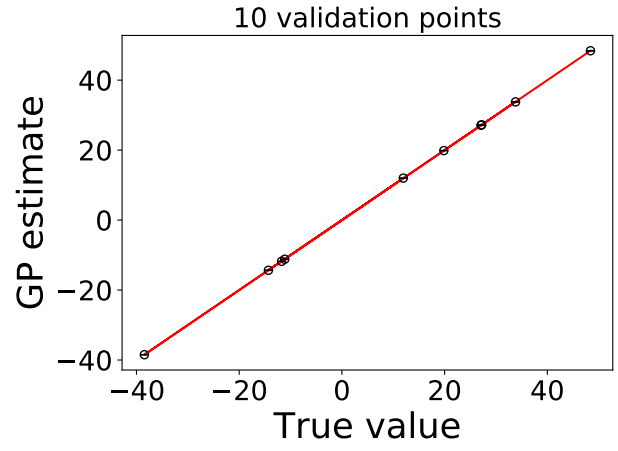




$x = 1.5$

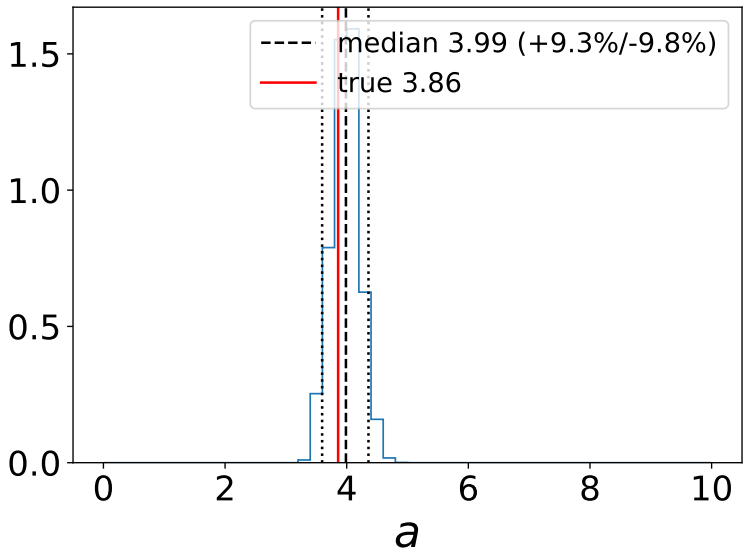
$x = 2.5$

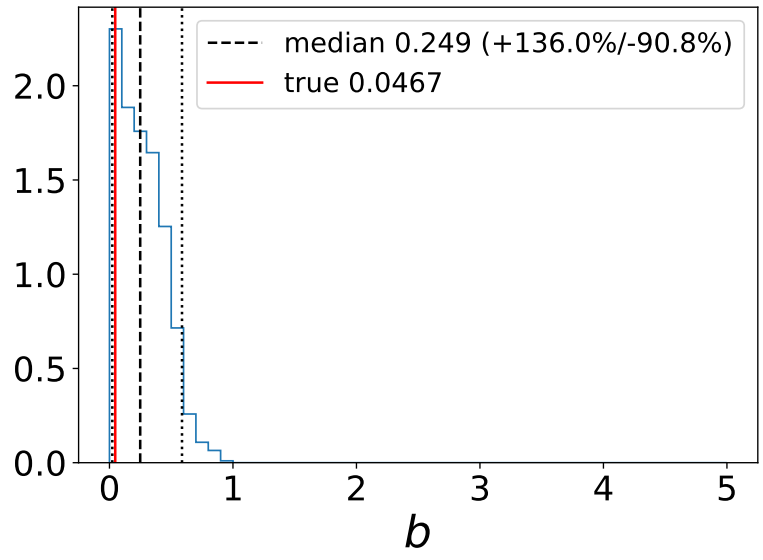


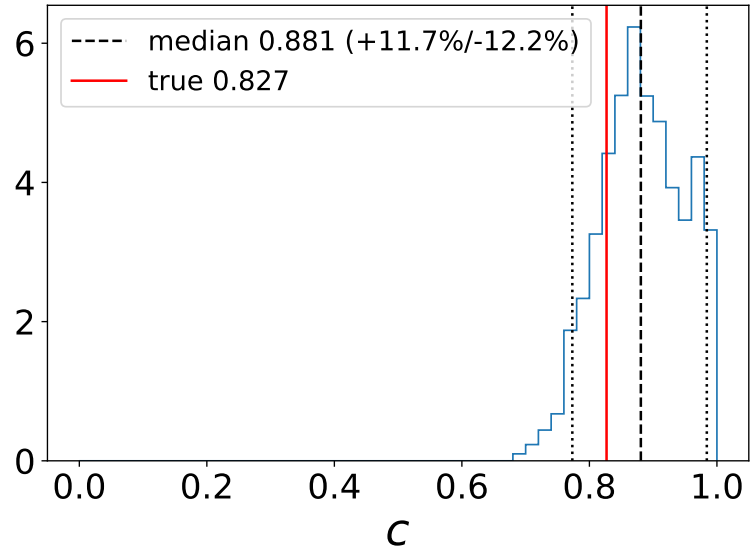


$x = 3.5$

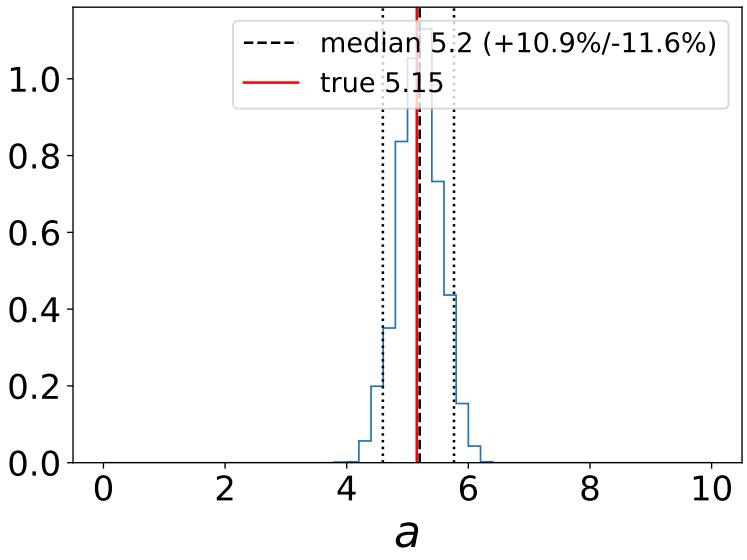
Closure test on GP + MCMC calibration using validation point 1

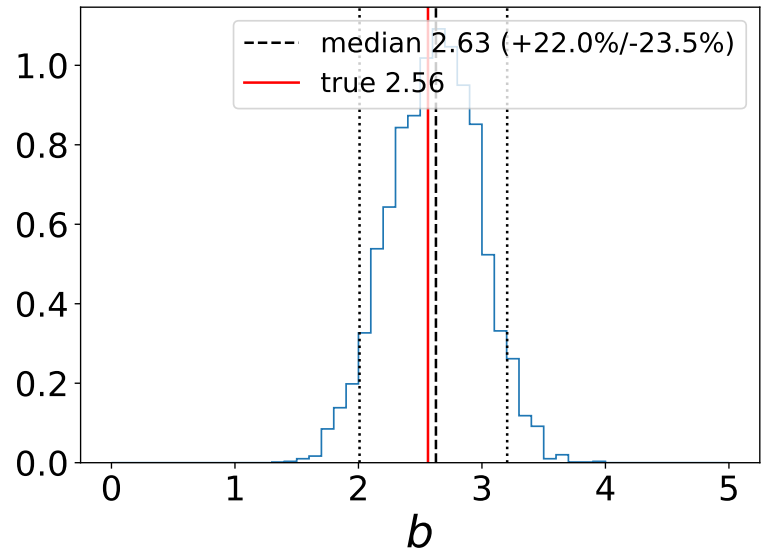


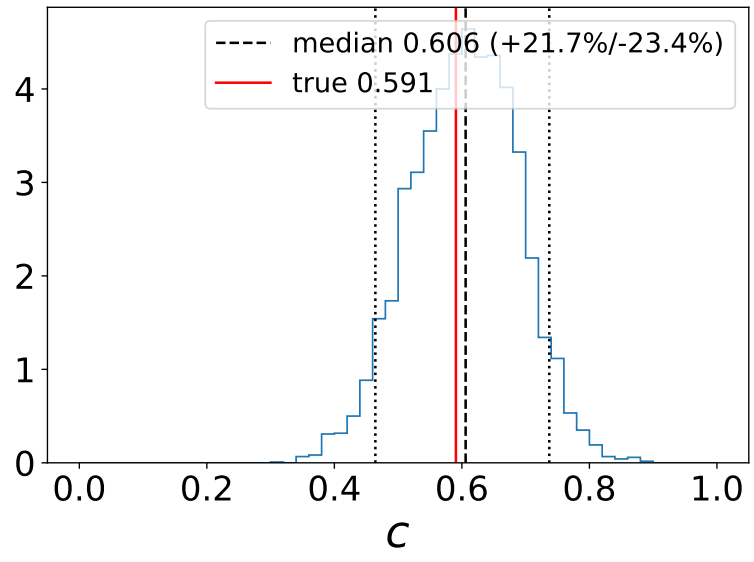




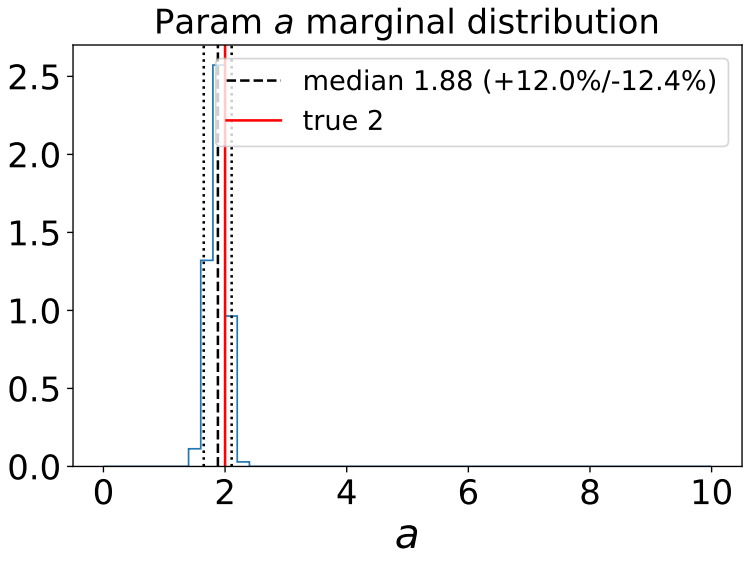
Closure test on GP + MCMC calibration using validation point 5

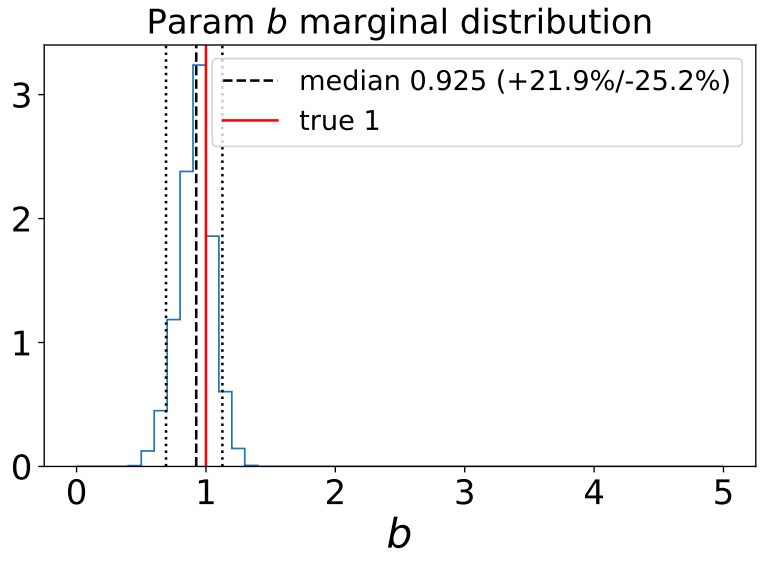


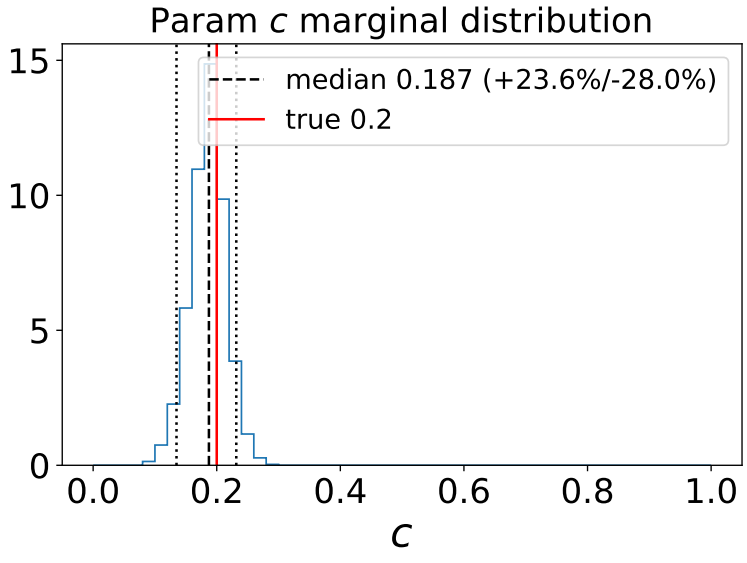


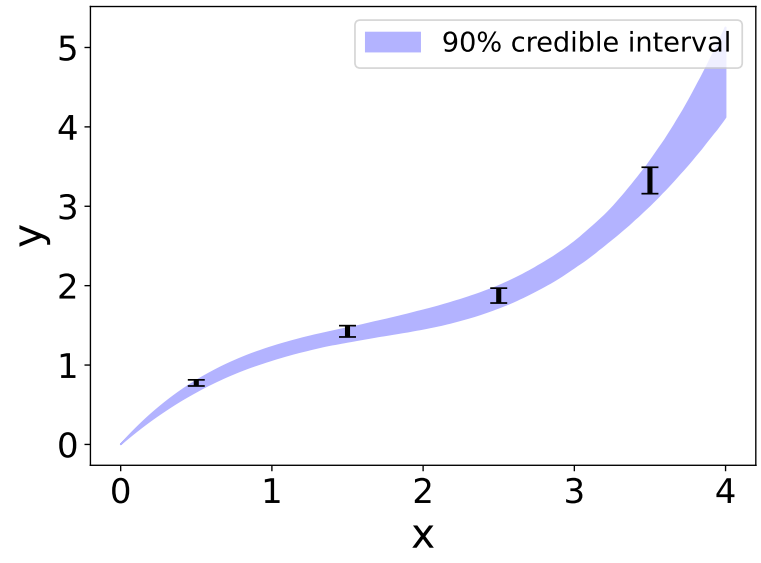


Calibration on measurements









Analysis procedure

Choose prior distribution and target observables

$\Downarrow$

Produce training data for emulator conditioning

$\Downarrow$

Determine required number of Gaussian processes with PCA

$\Downarrow$

Condition the emulators on training data and validate with separate data

$\Downarrow$

Closure testing

$\Downarrow$

Calibrate on experimental data

Example analysis: η/s sensitivity to EoS

Temperature dependence of η/s of strongly interacting matter: effects of the equation of state and the parametric form of $(\eta/s)(T)$

Jussi Auvinen,[1,*] Kari J. Eskola,[2,3] Pasi Huovinen,[1,4] Harri Niemi,[2,3] Risto Paatelainen,[5] and Péter Petreczky[6]

[1] *Institute of Physics Belgrade, 11080 Belgrade, Serbia*
[2] *University of Jyvaskyla, Department of Physics, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland*
[3] *Helsinki Institute of Physics, P.O. Box 64, FI-00014 University of Helsinki, Finland*
[4] *Institute of Theoretical Physics, University of Wroclaw, 50-204 Wrocław, Poland*
[5] *Theoretical Physics Department, CERN, CH-1211 Geneve 23, Switzerland*
[6] *Physics Department, Brookhaven National Laboratory, Upton, NY 11973, USA*

PRC 102, 044911 (2020), arxiv:2006.12499

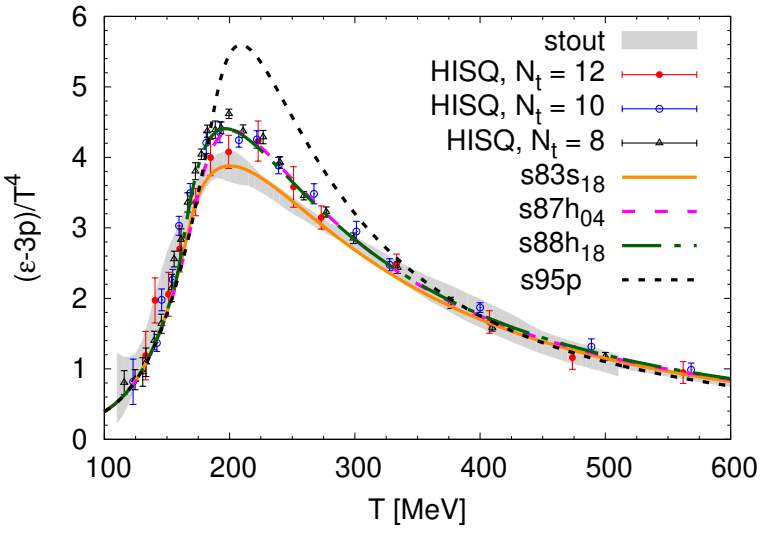
Baseline EoS: P. Huovinen and P. Petreczky, Nucl. Phys. A **837**, 26 (2010)

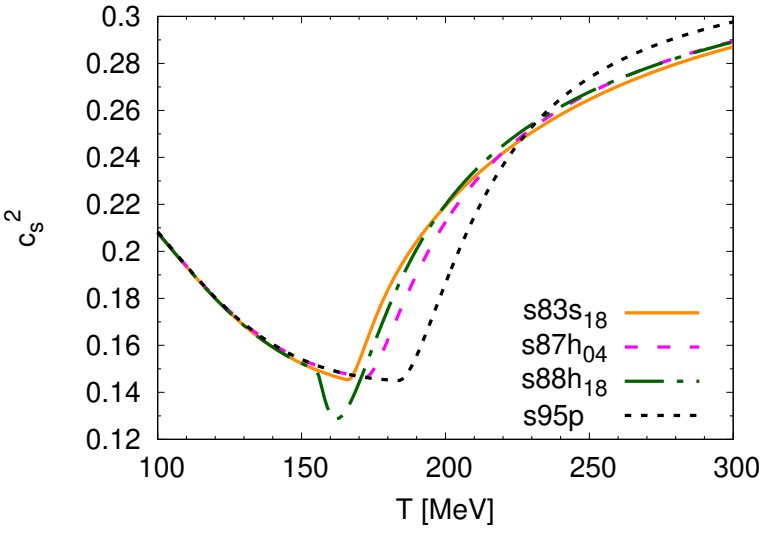
- **s95p**: $M < 2$ GeV hadron gas particles from 2004 PDG summary tables, lattice data used for fitting: Phys. Rev. D **80**, 014504 (2009); Phys. Rev. D **77**, 014511 (2008)

New parametrizations: J. Auvinen, K. J. Eskola, P. Huovinen, H. Niemi, R. Paatelainen, P. Petreczky,

PRC **102**, 044911 (2020) [arXiv:2006.12499 [nucl-th]]

- $s83s_{18}$: HG: 2018 PDG summary, lattice: stout action PLB **730**, 99 (2014)
- $s87h_{04}$: HG: 2004 PDG, lattice: HISQ PRD **90**, 094503 (2014), PRD **97**, 014510 (2018)
- $s88h_{18}$: HG: 2018 PDG, lattice: HISQ PRD **90**, 094503 (2014), PRD **97**, 014510 (2018)





EKRT+hydrodynamics model

- Initial energy density from the EKRT minijet saturation model

  Paatelainen et al., Phys. Rev. C **87**, no. 4, 044904 (2013); Phys. Lett. B **731**, 126 (2014)

  $$e(\vec{r}_T, \tau_s(\vec{r}_T)) = \frac{K_{\rm sat}}{\pi}[p_{sat}(\vec{r}_T, K_{\rm sat})]^4;\ \tau_s(\vec{r}_T) = 1/p_{\rm sat}(\vec{r}_T, K_{\rm sat})$$

  For each centrality class, produce a number of energy density profiles, convert to entropy density via EoS, and average over events

- 2+1D viscous hydrodynamics with piecewise linear temperature dependence on shear viscosity coefficent η/s Niemi et al., PRC **93**, 024907 (2016)

  $$(\eta/s)(T) = S_{\rm HG}(T_{\rm H} - T) + (\eta/s)_{\rm min},\ T < T_{\rm H}$$
  $$(\eta/s)(T) = (\eta/s)_{\rm min},\ T_{\rm H} \leq T \leq T_{\rm H} + W_{\rm min}$$
  $$(\eta/s)(T) = S_{\rm QGP}(T - T_{\rm H} - W_{\rm min}) + (\eta/s)_{\rm min},\ T > T_{\rm H} + W_{\rm min}$$

- Kinetic decoupling temperature $T_{\rm dec}$ and chemical freeze-out temperature $T_{\rm chem}$ are also free parameters

Bayesian analysis

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

$(K_{\rm sat},\ (\eta/s)_{\rm min},\ T_{\rm H},\ W_{\rm min}\ S_{\rm HG},\ S_{\rm QGP},\ T_{\rm dec},\ T_{\rm chem}\ )$

$\Downarrow$

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

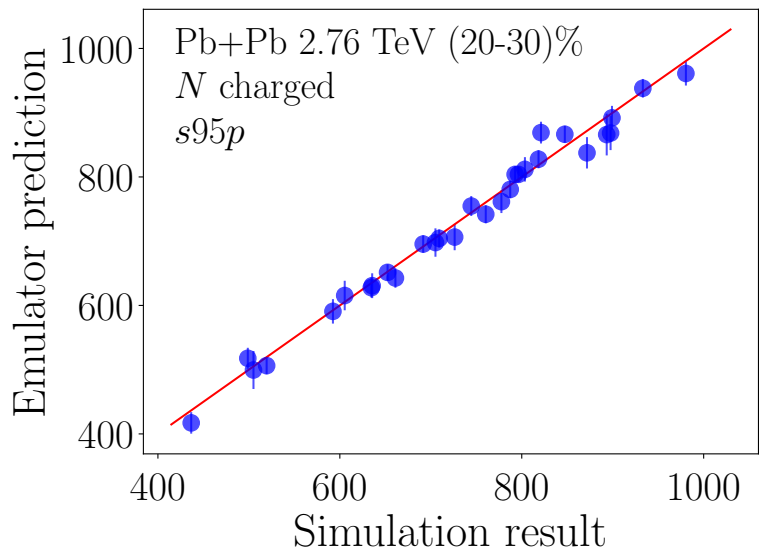
- charged particle multiplicity at midrapidity, $dN_{\rm ch}/d\eta$ and 4-particle cumulant p_T-averaged elliptic flow, $v_2\{4\}$, in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV and Pb+Pb collisions at $\sqrt{s_{NN}} = 2.76$ and $\sqrt{s_{NN}} = 5.02$ TeV
- Identified particle multiplicity dN_i/dy, and average transverse momenta $\langle p_T \rangle_i$, of pions (π^+), kaons (K^+) and protons (p) in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV and in Pb+Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV

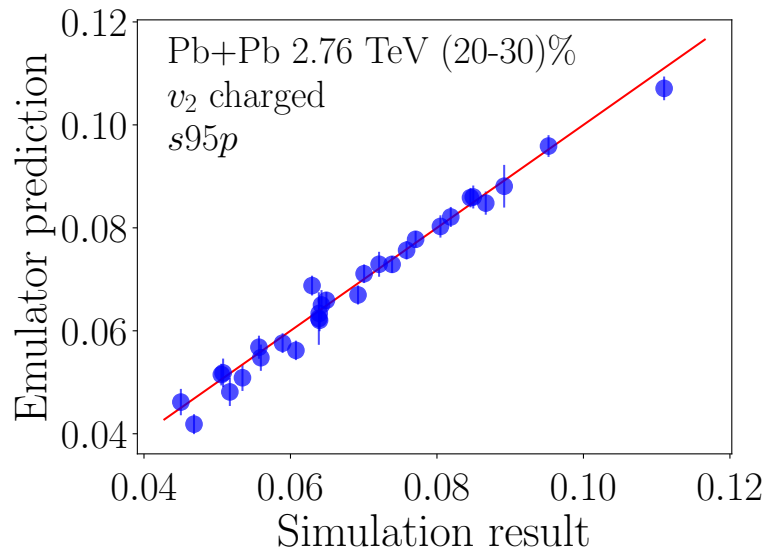
Centralities: (10–20)%, (20–30)%, (30–40)%, (40–50)% and (50–60)%.

Bayesian analysis

90 observables $\Rightarrow$ Reduce to 6 principal components for GP emulation

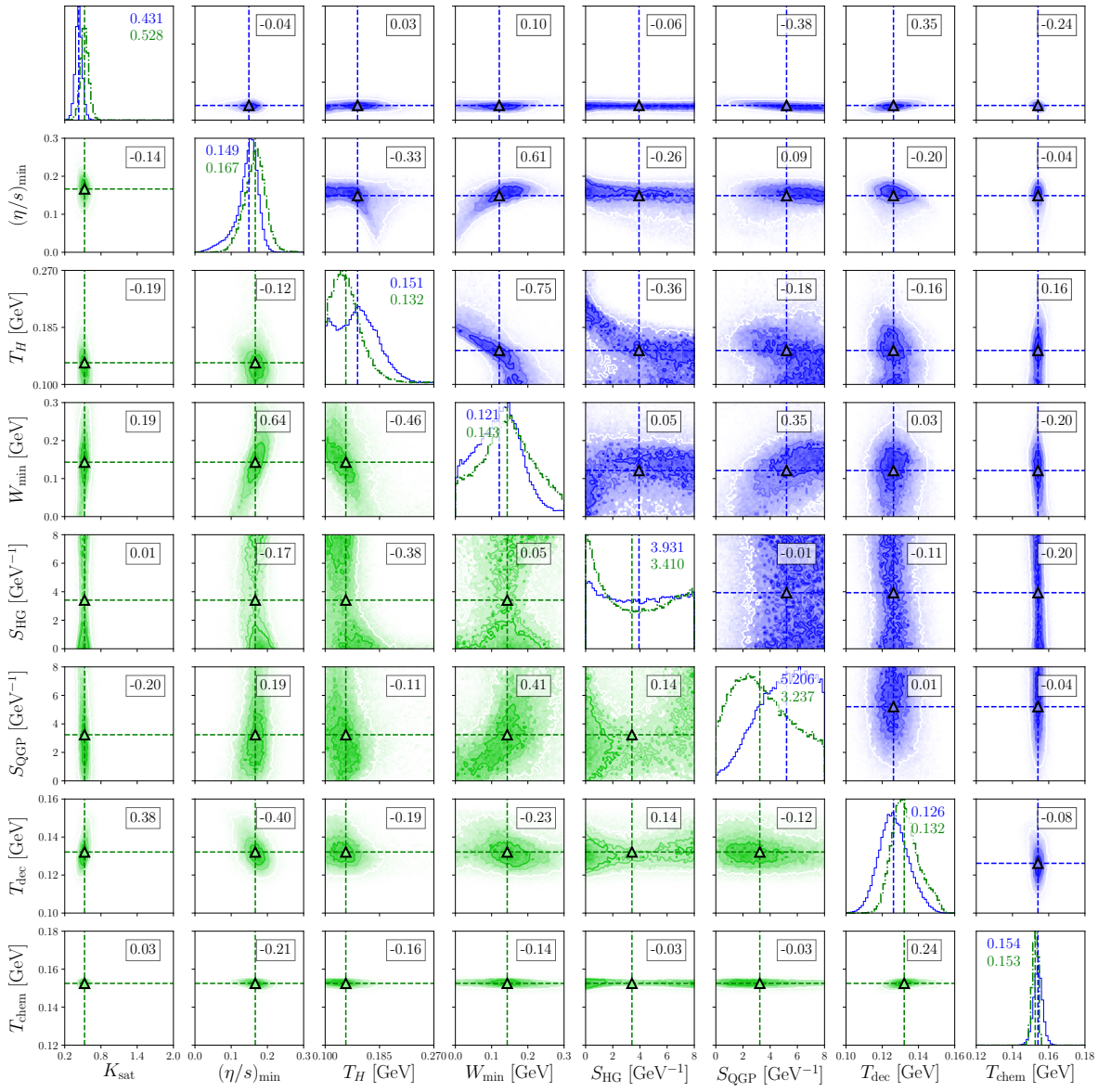
GP training with 170 points, validation with 30 points





Posterior probability

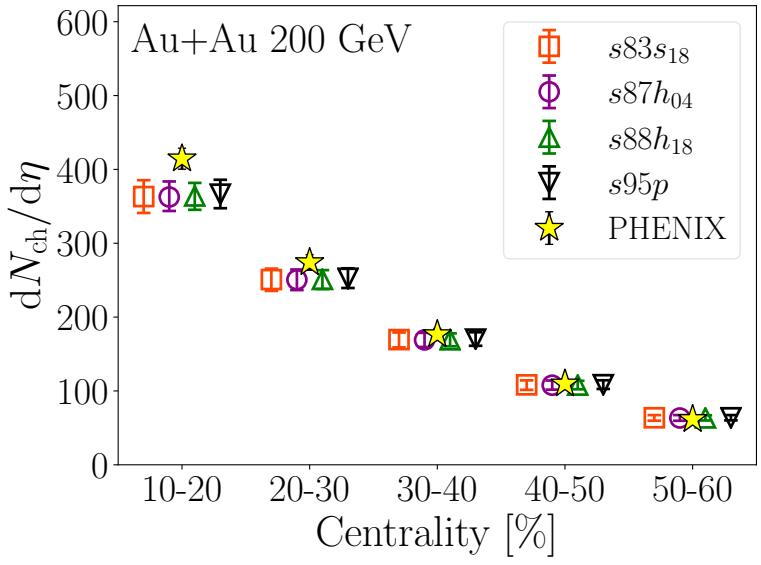
s88h$_{18}$



s95p

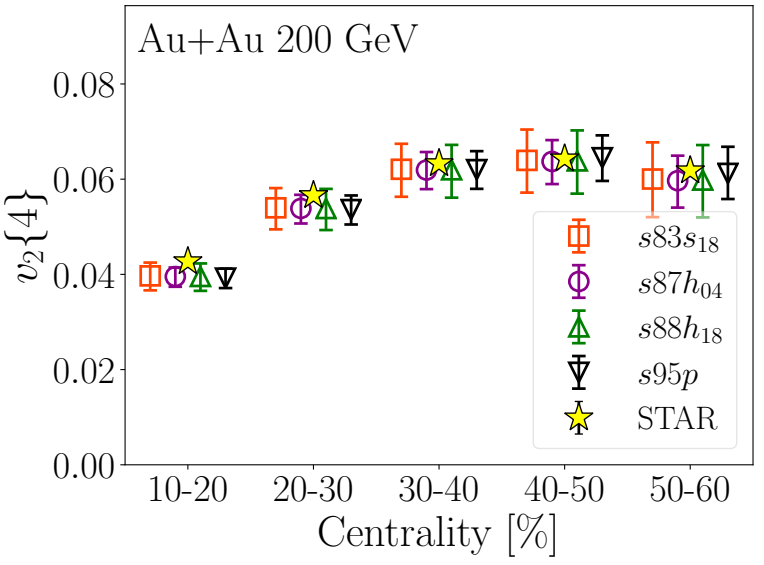
Observables at $\sqrt{s_{NN}} = 200$ GeV

Charged particle yield



Experimental data: STAR, PRC 79, 034909 (2009)

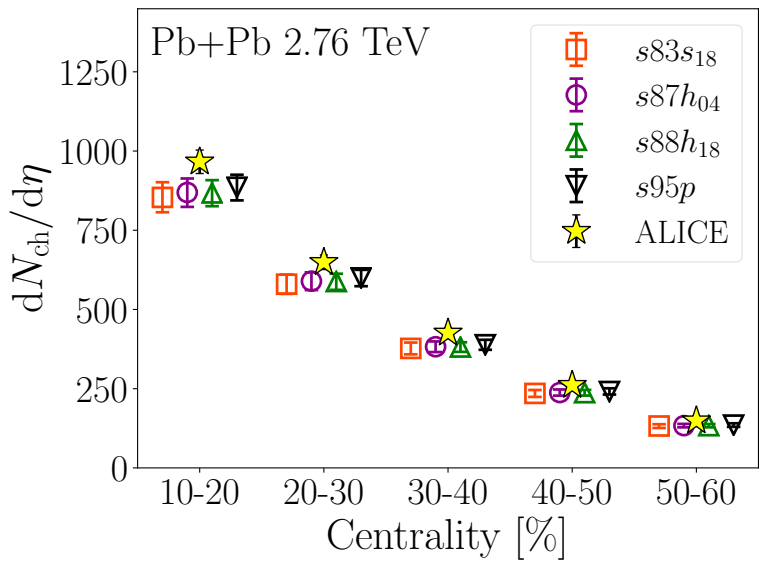
Elliptic flow $v_2\{4\}$



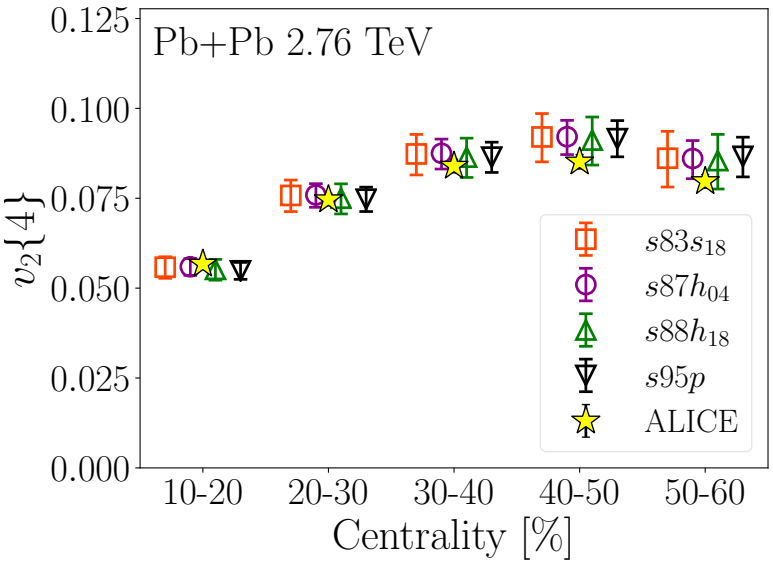
Experimental data: STAR, PRC 72, 014904 (2005)

Observables at $\sqrt{s_{NN}} = 2.76$ TeV

Charged particle yield



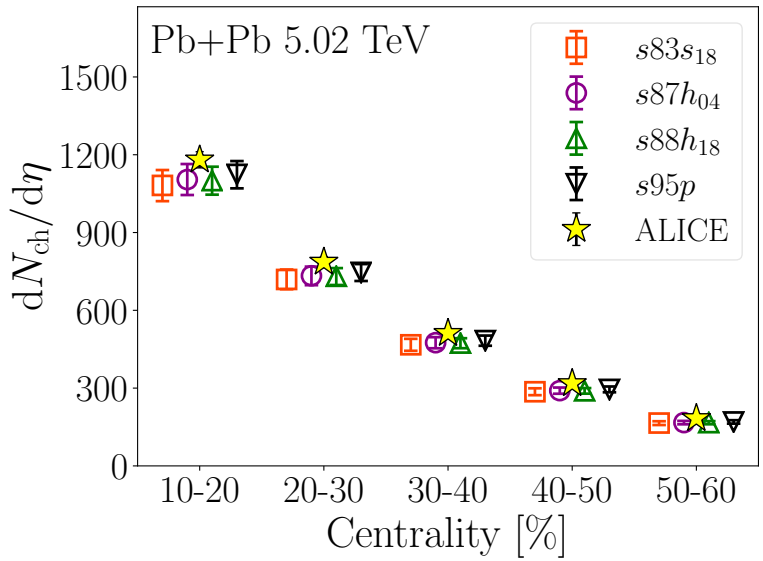
Elliptic flow $v_2\{4\}$



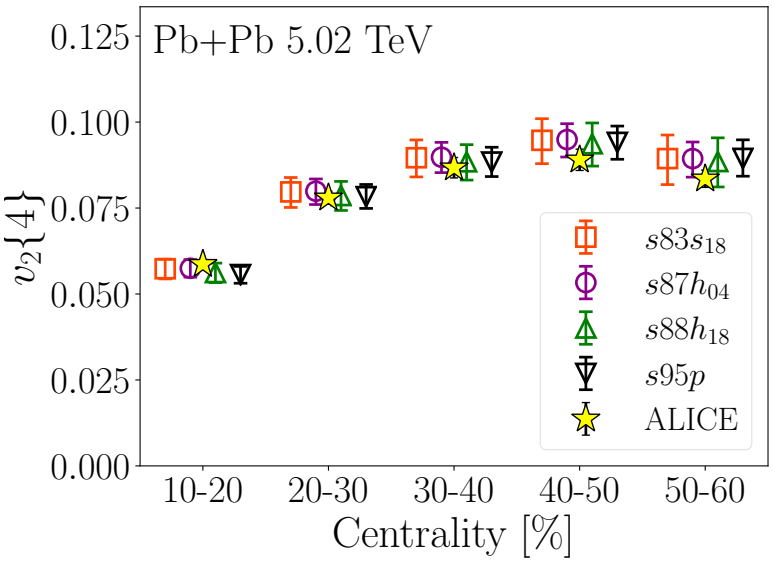
Experimental data: ALICE PRL 106, 032301 (2011) Experimental data: ALICE PRL 116, 132302 (2016)

Observables at $\sqrt{s_{NN}} = 5.02$ TeV

Charged particle yield



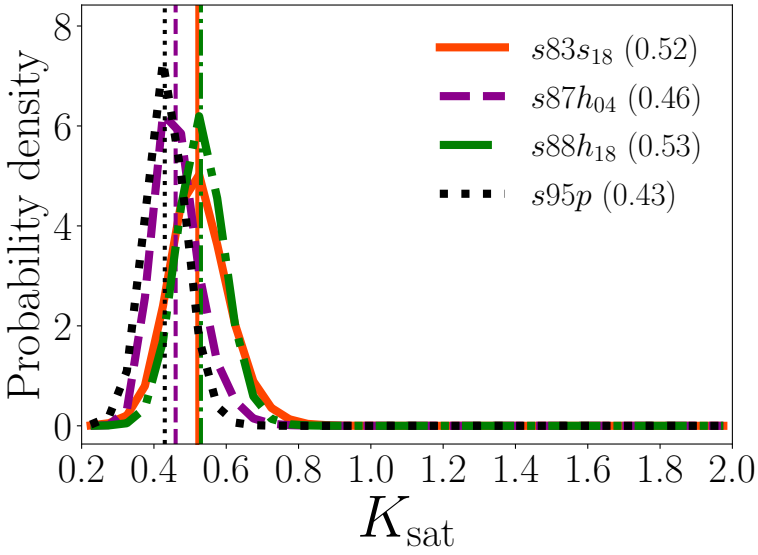
Elliptic flow $v_2\{4\}$



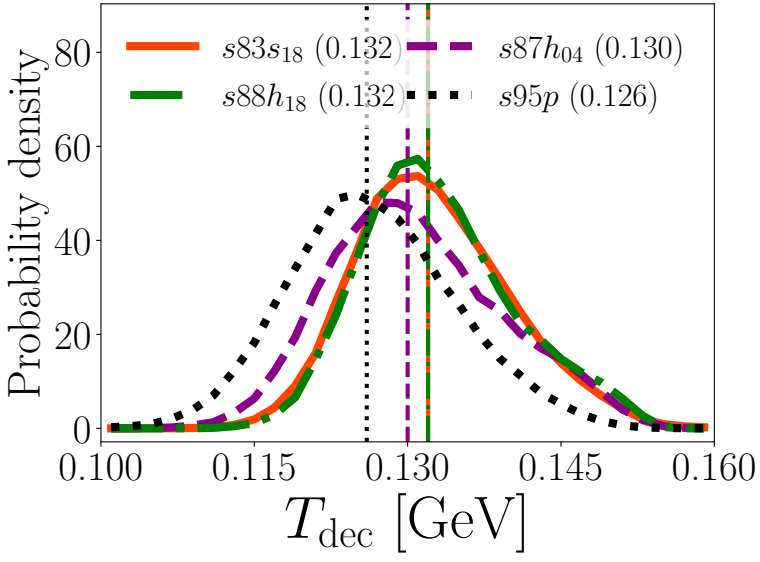
Experimental data: ALICE PRL 116, 222302 (2016)

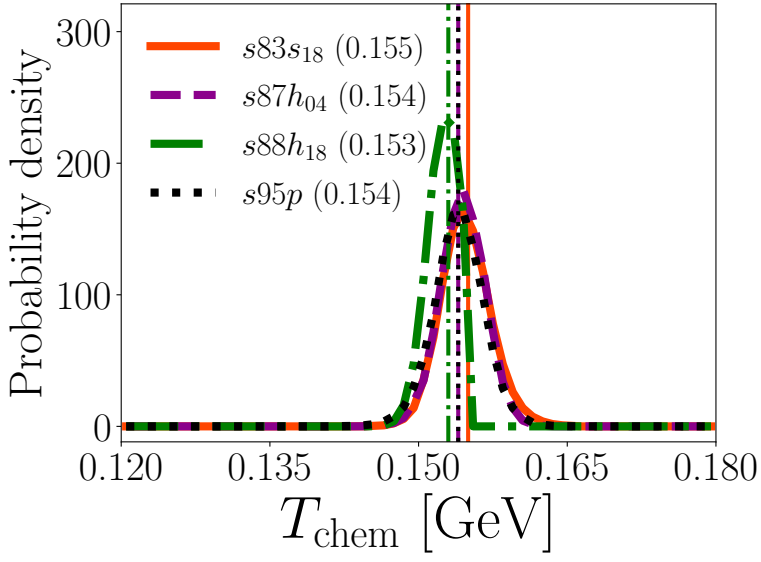
Experimental data: ALICE PRL 116, 132302 (2016)

EoS posterior distribution comparisons

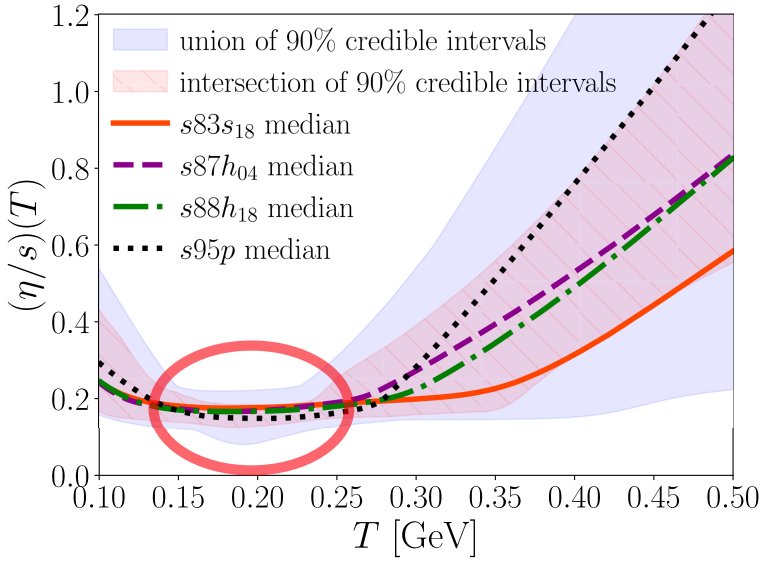


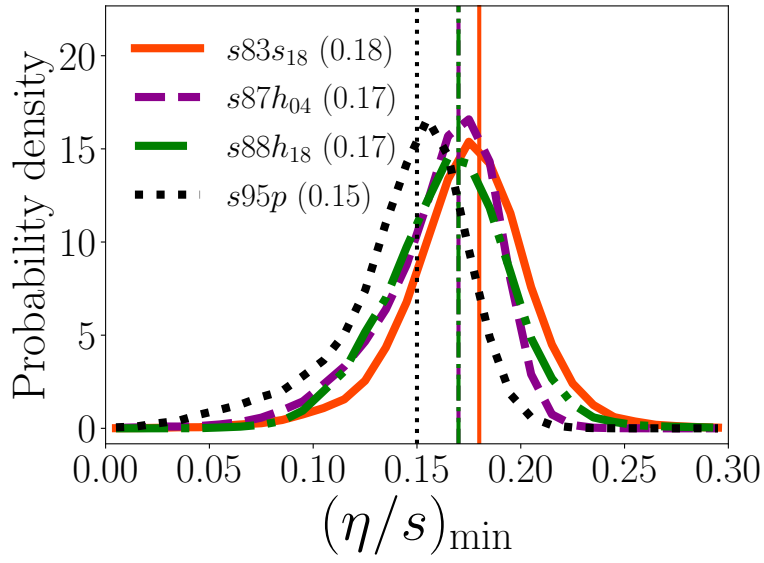
- "Nuisance" parameters (= not directly related to $(\eta/s)(T)$) quite well constrained
- Not much difference between EoSs





Posterior $(\eta/s)(T)$





- Tightest constraints on η/s in the temperature range $T \approx 150$–220 MeV, where η/s is approximately constant
- All EoSs: $0.08 < \eta/s < 0.23$
- $s83s_{18}$ and $s88h_{18}$: $0.12 < \eta/s < 0.23$
- Differences between equations of state within uncertainties

Example analysis: JETSCAPE transport coefficients

Multi-system Bayesian constraints on the transport coefficients of QCD matter

D. Everett,[1] W. Ke,[2,3] J.-F. Paquet,[4] G. Vujanovic,[5] S. A. Bass,[4] L. Du,[1] C. Gale,[6] M. Heffernan,[6] U. Heinz,[1] D. Liyanage,[1] M. Luzum,[7] A. Majumder,[5] M. McNelis,[1] C. Shen,[5,8] Y. Xu,[4] A. Angerami,[9] S. Cao,[5] Y. Chen,[10,11] J. Coleman,[12] L. Cunqueiro,[13,14] T. Dai,[4] R. Ehlers,[13,14] H. Elfner,[15,16,17] W. Fan,[4] R. J. Fries,[18,19] F. Garza,[18,19] Y. He,[20] B. V. Jacak,[2,3] P. M. Jacobs,[2,3] S. Jeon,[6] B. Kim,[18,19] M. Kordell II,[18,19] A. Kumar,[5] S. Mak,[12] J. Mulligan,[2,3] C. Nattrass,[13] D. Oliinychenko,[3] C. Park,[6] J. H. Putschke,[5] G. Roland,[10,11] B. Schenke,[21] L. Schwiebert,[22] A. Silva,[13] C. Sirimanna,[5] R. A. Soltz,[5,9] Y. Tachibana,[5] X.-N. Wang,[20,2,3] and R. L. Wolpert[12]
(The JETSCAPE Collaboration)

PRC 103, 054904 (2021), arxiv:2011.01430

JETSCAPE multistage model (18 parameters):

- Initial transverse energy density profile $\bar{\epsilon}(x, y)$ from Trento
- Free streaming until $\tau_{fs}(\tau_R, \alpha) = \tau_R \left(\frac{\langle \bar{\epsilon} \rangle}{4.0 \mathrm{GeV/fm}^3} \right)^{\alpha}$
- 2+1d viscous hydrodynamics with both shear and bulk viscous effects
- Hadron transport

Norm. Pb-Pb 2.76 TeV	N[2.76 TeV]	[10, 20]	temperature of (η/s) kink	T_η	[0.13, 0.3] GeV
Norm. Au-Au 200 GeV	N[0.2 TeV]	[3, 10]	(η/s) at kink	$(\eta/s)_{\mathrm{kink}}$	[0.01, 0.2]
generalized mean	p	[–0.7, 0.7]	low temp. slope of (η/s)	a_{low}	[–2, 1] GeV^{-1}
nucleon width	w	[0.5, 1.5] fm	high temp. slope of (η/s)	a_{high}	[–1, 2] GeV^{-1}
min. dist. btw. nucleons	d^3_{min}	[0, 1.7^3] fm^3	shear relaxation time factor	b_π	[2, 8]
multiplicity fluctuation	σ_k	[0.3, 2.0]	maximum of (ζ/s)	$(\zeta/s)_{\mathrm{max}}$	[0.01, 0.25]
free-streaming time scale	τ_R	[0.3, 2.0] fm/c	temperature of (ζ/s) peak	T_ζ	[0.12, 0.3] GeV
free-streaming energy dep.	α	[–0.3, 0.3]	width of (ζ/s) peak	w_ζ	[0.025, 0.15] GeV
particlization temperature	T_{sw}	[0.135, 0.165] GeV	asymmetry of (ζ/s) peak	λ_ζ	[–0.8, 0.8]

TABLE I. A list of all priors used (see Sec. III for the definitions of the model parameters). All prior distributions are assumed to be uniform and nonzero within the range quoted, and zero outside. The Table does not exhibit the step functions that ensure non-negativity of the shear viscosity at all temperatures (see Eq. (25)).

Side note about priors

> It is essential that the construction of the prior distribution should not be informed by the same data that will be used in performing parameter estimation. In particular, the posterior of earlier analyses that used the same data sets should not in any way be used as a prior for a new analysis: it would be an attempt to use the same measurements twice, as well as being likely inconsistent given differences in the models.

However: Using previously obtained posterior distribution is appropriate when updating the existing analysis using new data points

Viscous corrections at particlization

Grad ($\tilde{f} = 1 \mp f$):

$$\delta f_i^{\mathrm{Grad}} = f_{\mathrm{eq},i}\tilde{f}_{\mathrm{eq},i}\left[\Pi(A_T m_i^2 + A_E(u\cdot P)^2) + A_\pi \pi^{\mu\nu}\Delta^{\alpha\beta}_{\mu\nu}P_\alpha P_\beta\right]$$

Chapman-Enskog (CE):

$$\delta f_i^{/mathrmCE} = f_{\mathrm{eq},i}\tilde{f}_{\mathrm{eq},i}\left[\frac{\Pi}{\beta_\Pi}\left(\frac{(u\cdot P)\mathcal{F}}{T^2} - \frac{P\cdot\Delta P}{3(u\cdot P)T}\right) + \frac{\pi_{\mu\nu}\Delta^{\mu\nu}_{\alpha\beta}P^\alpha P^\beta}{2\beta_\pi(u\cdot P)T}\right]$$

Pratt-Torrieri-Bernhard (PTB):

$$f^{\mathrm{PTB}} = \frac{\mathcal{Z}_\Pi}{\det B}\left[\exp\left(\frac{\sqrt{|P'|^2 + m^2}}{T}\right) \pm 1\right]^{-1}$$

with $P_i = B_{ij}P'_j$, $B_{ij} = (1+\lambda_\Pi)\delta_{ij}\frac{\pi_{ij}}{2\beta_\pi}$ where λ_Π is adjusted to match the total pressure of the system

Experimental data

Pb+Pb at $\sqrt{s_{NN}} = 2760$ GeV:

- $dN_{\rm ch}/d\eta$, $dE_T/d\eta$, $\delta p_T/p_T$ in (0-70)% centrality
- dN/dy, $\langle p_T \rangle$ for pions, kaons, protons in (0-70)% centrality
- $v_2\{2\}$ in (0-70)% centrality and $v_n\{2\}$ ($n = 3, 4$) in (0-50)% centrality

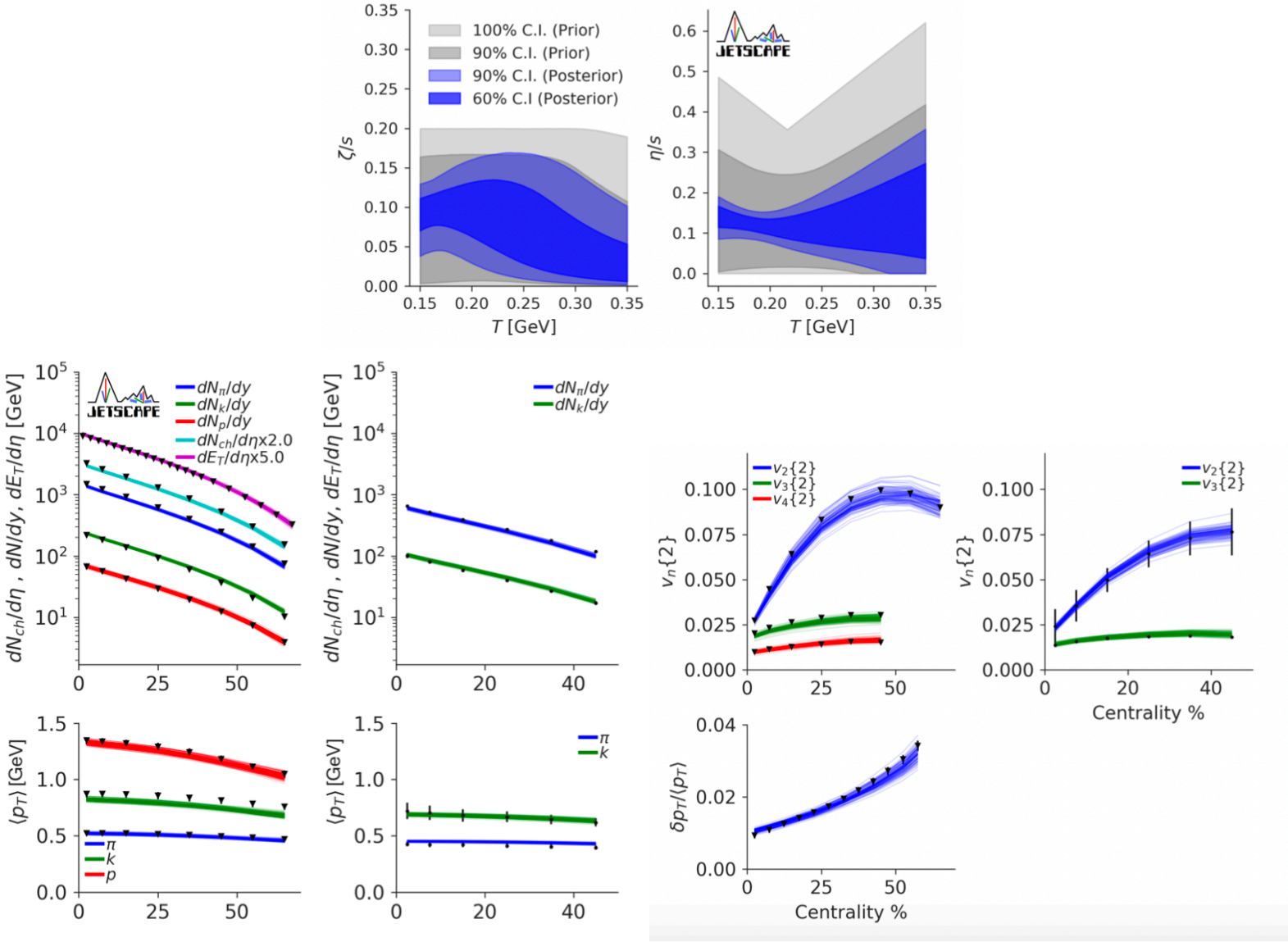
Total of 123 data points reduced to 10 principal components (98% of total variance)

Au+Au at $\sqrt{s_{NN}} = 200$ GeV:

- dN/dy, $\langle p_T \rangle$ for pions, kaons in (0-50)% centrality
- $v_n\{2\}$ ($n = 2, 3$) in (0-50)% centrality

Total of 29 data points reduced to 6 principal components (>98% of total variance)

Emulator training points: 500


100% C.I. (Prior)
90% C.I. (Prior)
90% C.I. (Posterior)
60% C.I (Posterior)
ζ/s
η/s
T [GeV]
JETSCAPE
dN_π/dy
dN_k/dy
dN_p/dy
dN_ch/dηx2.0
dE_T/dηx5.0
dN_ch/dη , dN/dy , dE_T/dη [GeV]
⟨p_T⟩ [GeV]
π
k
p
v_2{2}
v_3{2}
v_4{2}
v_n{2}
δp_T/⟨p_T⟩
Centrality %

Bayes factor

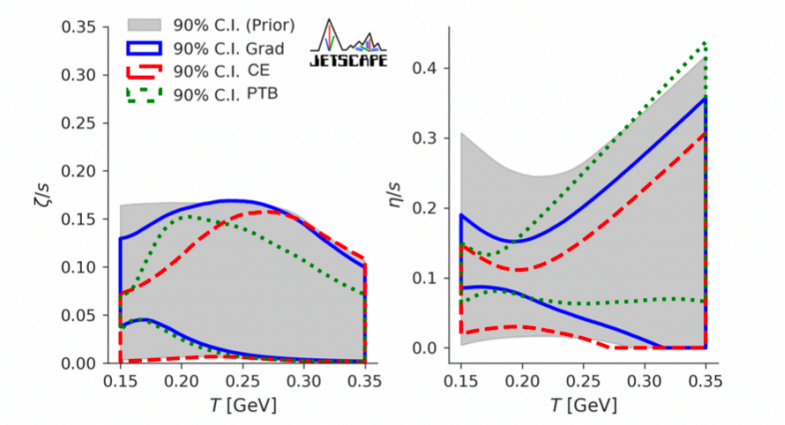


FIG. 9. The 90 % credibility intervals for the prior (gray shaded area) and for the posteriors (colored outlines) of the specific bulk (left) and shear (right) viscosities, for three viscous correction models: Grad (blue), Chapman-Enskog (CE, red) and Pratt-Torrieri-Bernhard (PTB, green). The Pratt-Torrieri-McNelis (PTM) posterior is not shown, but is nearly identical with the Chapman-Enskog result.

Model A	Model B	$\ln B_{A/B}$
Grad	CE	8.2 ± 2.3
Grad	PTB	1.4 ± 2.5
PTB	CE	6.8 ± 2.4

TABLE IV. A table of the logarithm of the Bayes factor $\ln B_{A/B}$ for each pair of viscous correction models and its integration uncertainty for the Grad, Chapman-Enskog (CE) and Pratt-Torrieri-Bernhard (PTB) viscous correction models.

Other comparisons:

- Temperature dependent η/s (A) vs. no T-dependence (B): $\ln B_{A/B} = -0.2 \pm 2.4$
- Nonzero η/s (A) vs. $(\eta/s) \equiv 0$ (B): $\ln B_{A/B} = 11.7 \pm 2.6$

Example analysis: Trajectum

A transverse momentum differential global analysis of Heavy Ion Collisions

Govert Nijs,[1,2,*] Wilke van der Schee,[3,†] Umut Gürsoy,[2,‡] and Raimond Snellings[4,5,§]

[1] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[2] *Institute for Theoretical Physics and Center for Extreme Matter and Emergent Phenomena, Utrecht University, 3584 CC Utrecht, The Netherlands*
[3] *Theoretical Physics Department, CERN, CH-1211 Genève 23, Switzerland*
[4] *Institute for Gravitational and Subatomic Physics (GRASP), Utrecht University, 3584 CC Utrecht, The Netherlands*
[5] *Nikhef, 1098 XG Amsterdam, The Netherlands*

PRL 126, 202301 (2021), arxiv:2010.15130

Bayesian analysis of heavy ion collisions with the heavy ion computational framework *Trajectum*

Govert Nijs,[1,2] Wilke van der Schee,[3] Umut Gürsoy,[2] and Raimond Snellings[4,5]

[1] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[2] *Institute for Theoretical Physics and Center for Extreme Matter and Emergent Phenomena, Utrecht University, 3584 CC Utrecht, The Netherlands*
[3] *Theoretical Physics Department, CERN, CH-1211 Genève 23, Switzerland*
[4] *Nikhef, 1098 XG Amsterdam, The Netherlands*
[5] *Institute for Gravitational and Subatomic Physics (GRASP), Utrecht University, 3584 CC Utrecht, The Netherlands*

PRC 103, 054909 (2021), arxiv:2010.15134

Trajectum framework (21 parameters):

- Initial transverse energy density profile $\bar{\epsilon}(x, y)$ from Trento
- Free streaming with effective velocity v_{fs} until τ_{fs}
- 2+1d viscous hydrodynamics with temperature dependent shear and bulk coefficients $(\eta/s)(T)$ and $(\zeta/s)(T)$ with additional variable transport coefficient ratios $\tau_\Pi \delta^2 sT/\zeta$, $\tau_\pi sT/\eta$, $\tau_{\pi\pi}/\tau_\pi$
- Hadron transport

Experimental data

Pb+Pb:

- $dN_{\mathrm{ch}}/d\eta$ at 2.76 and 5.02 TeV
- $dE_T/d\eta$, $\delta p_T/p_T$ at 2.76 TeV
- dN/dy, $\langle p_T \rangle$, dN/dp_T for pions, kaons, protons at 2.76 TeV
- $v_n\{k\}$ for 2.76 and 5.02 TeV
- $v_n(p_T)$ for pions, kaons, protons at 2.76 and 5.02 TeV

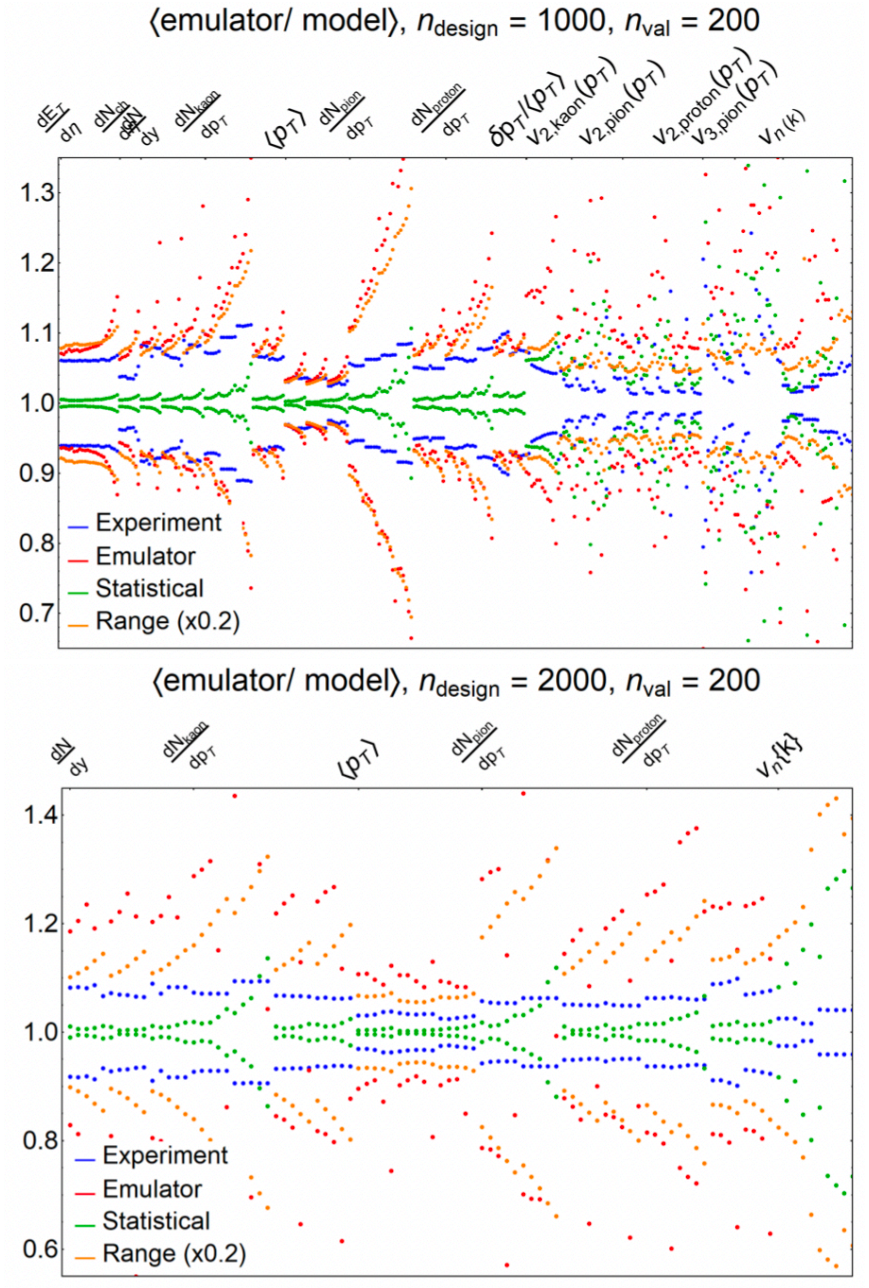
Total of 418 data points reduced to 25 principal components

p+Pb at 5.02 TeV:

- $\langle p_T \rangle$, dN/dpT for pions, kaons, protons
- $\bar{v}_n\{k\} = \mathrm{sgn}(v_n\{k\}^k)|v_n\{k\}|$

Total of 96 data points reduced to 25 principal components

Emulator validation



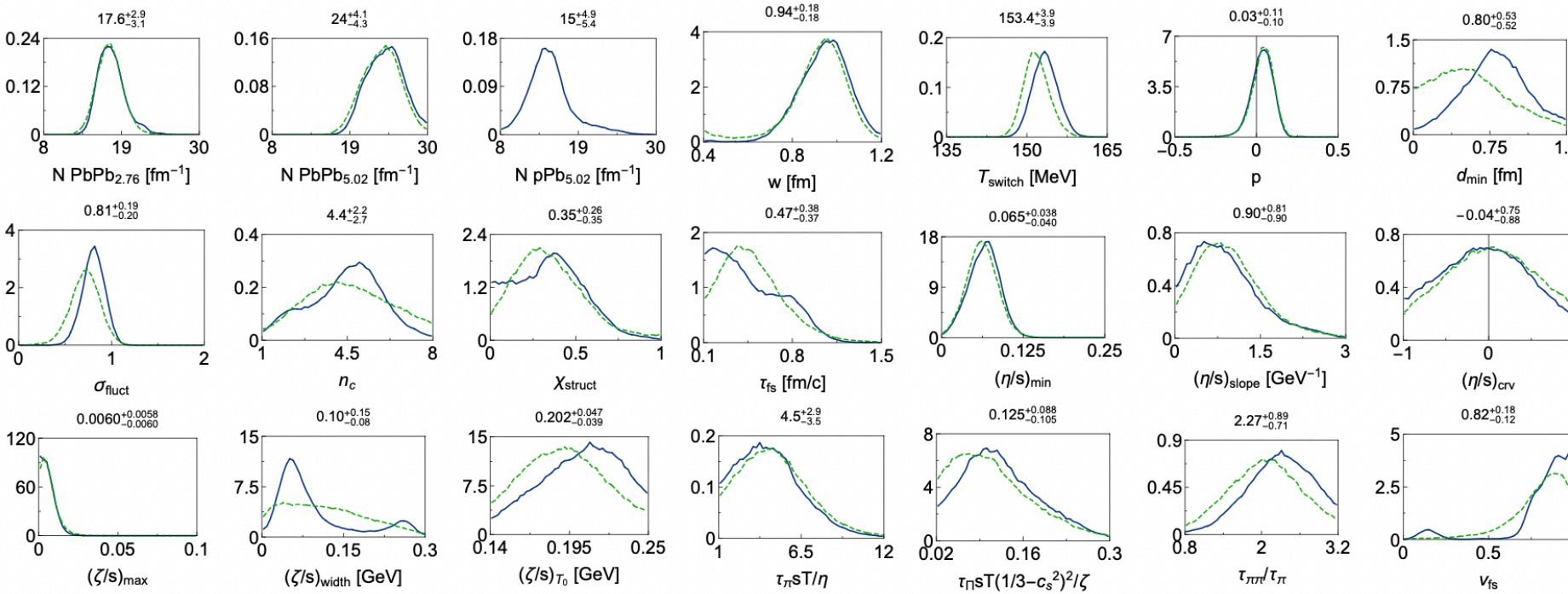
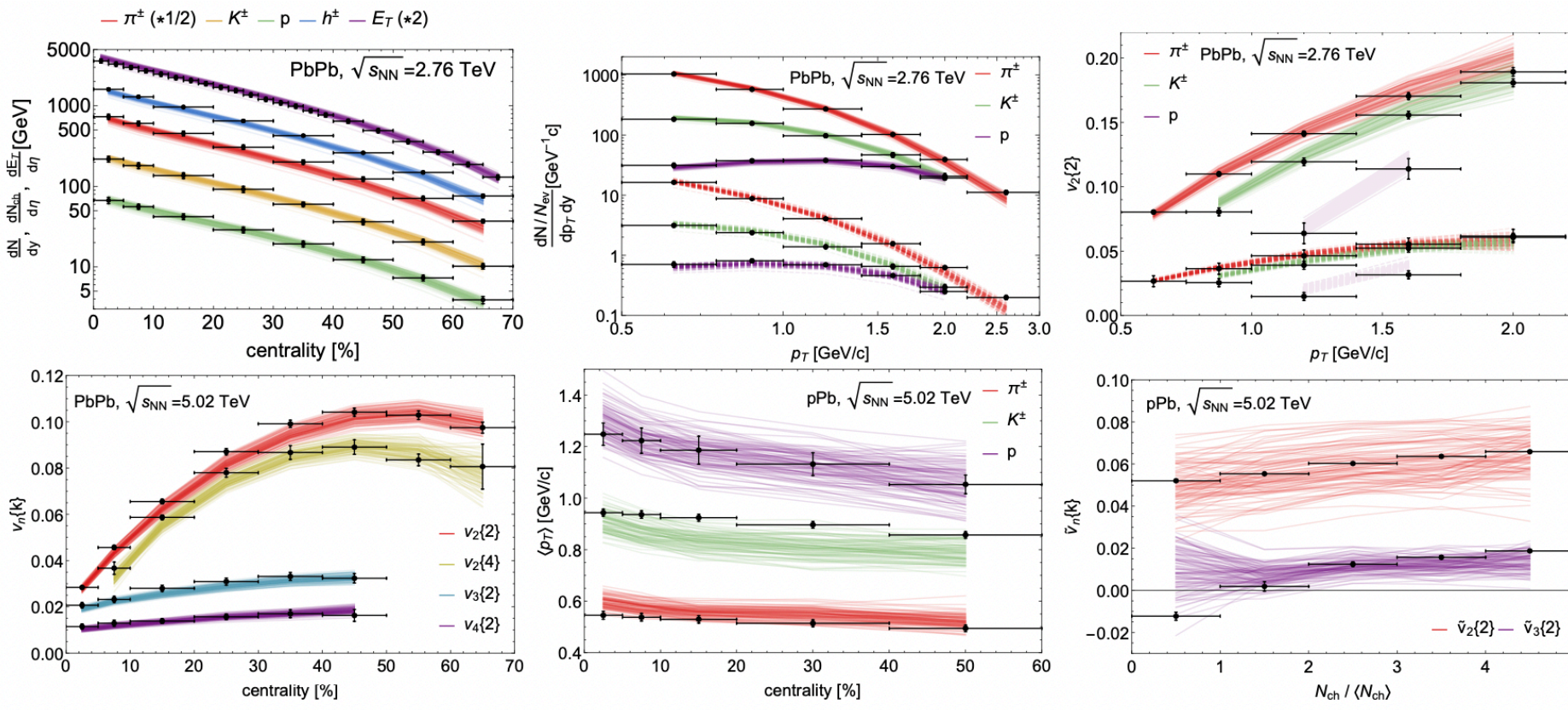


FIG. 1. Posterior distributions for all model parameters fitted to PbPb and pPb (solid) or PbPb only (dashed, not applicab to pPb norm) data. Values indicate the expectation values with the 90% highest posterior density credible interval.


π± (*1/2)
K±
p
h±
E_T (*2)
PbPb, √sNN =2.76 TeV
dN/dy, dNch/dη, dET/dη [GeV]
centrality [%]
PbPb, √sNN =2.76 TeV
π±
K±
p
dN/Nev[GeV−1c]/dpT dy
pT [GeV/c]
PbPb, √sNN =2.76 TeV
π±
K±
p
v2{2}
pT [GeV/c]
PbPb, √sNN =5.02 TeV
vn{k}
v2{2}
v2{4}
v3{2}
v4{2}
centrality [%]
pPb, √sNN =5.02 TeV
π±
K±
p
⟨pT⟩ [GeV/c]
centrality [%]
pPb, √sNN =5.02 TeV
ṽn{k}
ṽ2{2}
ṽ3{2}
Nch / ⟨Nch⟩