Bayesian parameter estimation in heavy ion collisions: the soft sector

Duke
UNIVERSITY

Jean-François Paquet (Duke University),
with Derek Everett, Weiyao Ke and Dan Liyanage

July 21, 2020

JETSCAPE

3rd JETSCAPE School

Lecturing live from Durham, NC

Plan for the summer school

Today (Tuesday):

- ∼1h lecture: general introduction to Bayesian parameter estimation
- ∼2h hands-on session
  (all in Jupyter Python notebooks, which can be run either online or on local machine)
- For today, ask questions on Slack in the **#bayesian-paquet** channel

Tomorrow (Wednesday):

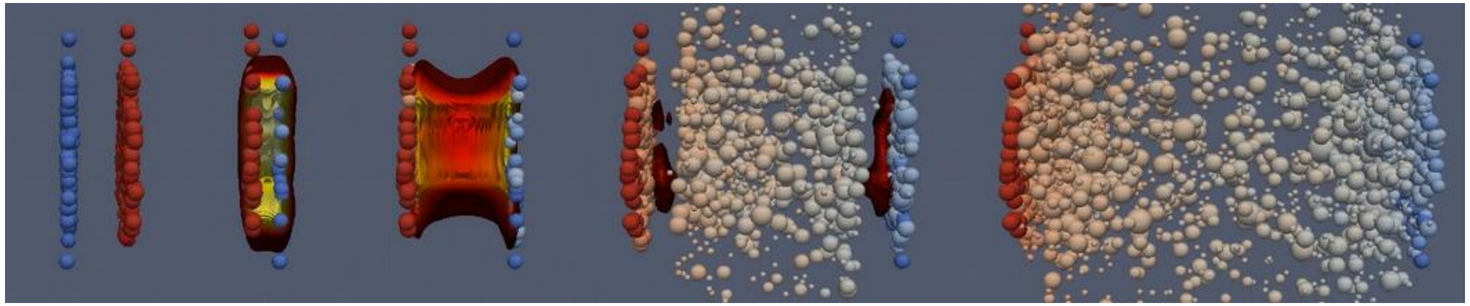
- ∼30min lecture by Weiyao Ke on the Trento ansatz of initial condition
- ∼2h30 hands-on session by Weiyao Ke:
  complete example of Bayesian parameter estimation with Trento as model

Thursday and Friday:

- Applications of Bayesian parameter estimation for the “hard sector” by Yi Chen

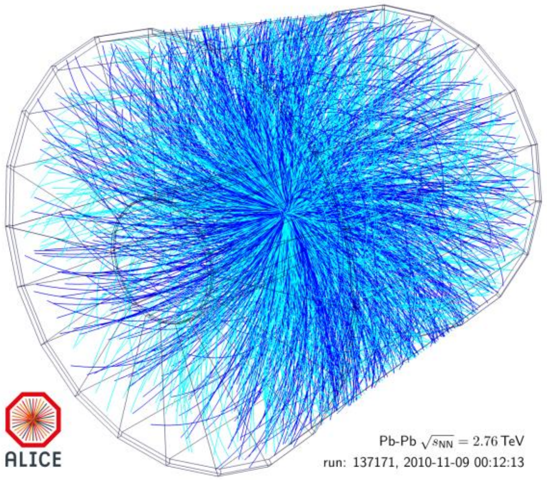
Big picture

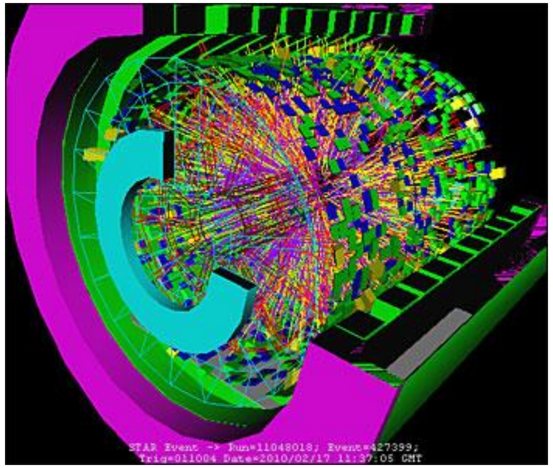
- We have a model of some physical process, say a relativistic heavy ion collision



Initial stage → **Hydrodynamics** → **Cooper-Frye** → **SMASH**

- We have experimental measurements of this same process

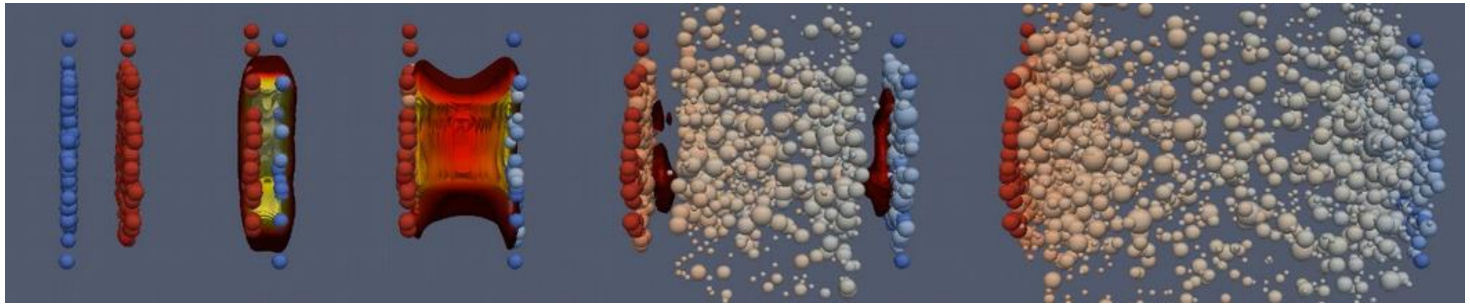




What can we learn about the model from the measurements?

Big picture

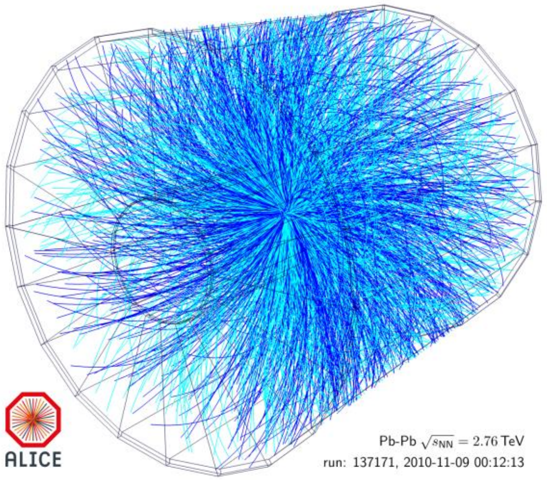
- We have a model of some physical process, say a relativistic heavy ion collision

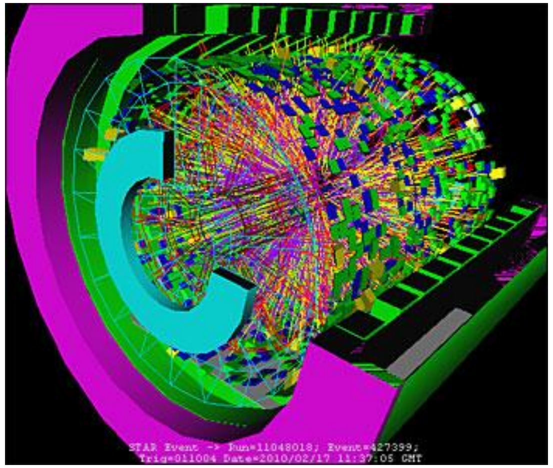


Initial stage → Hydrodynamics → Cooper-Frye → SMASH

Model parameters

- We have experimental measurements of this same process





What can we learn about the model parameters from the measurements?

Big picture: questions

Constraining a model parameters systematically from observations/measurements

Questions:

- Haven't we all been comparing models with data all our lives?
  Why do I suddenly need Bayesian parameter estimation?
- What type of constraints can we get? $parameter \pm \Delta parameter$?
- How do I input theoretical knowledge in such an analysis?
- Applications of Bayesian parameter estimation for the experimental community?
- Do I need to know the prediction of my model for every single values of model parameters?

Why Bayesian parameter estimation?

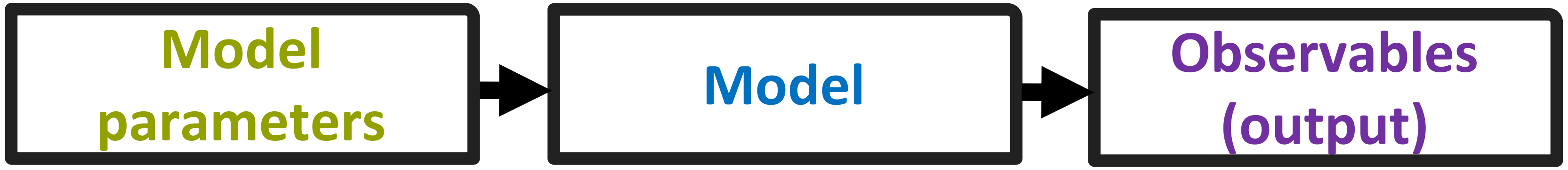
Forward and inverse problems

- Think of physics modelling as having **three main ingredients**:

  1. A **model**: initial conditions + hydrodynamics + SMASH
  2. Model **parameters**: initial condition parameters, shear and bulk viscosities, …
  3. Model **outputs (observables)**: hadron multiplicity, v_n's, R_{AA}, …

"Forward problem"

- Given model parameters, what are the model outputs/observables?
- Well-defined question no matter how non-linear the model is



Forward and inverse problems

- Think of physics modelling as having **three main ingredients**:

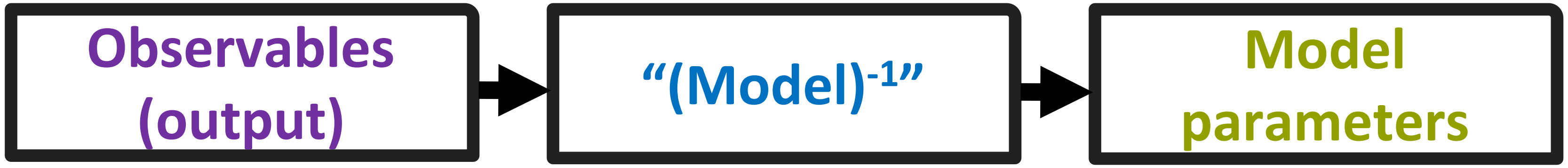
    1. A **model**: initial conditions + hydrodynamics + SMASH
    2. Model **parameters**: initial condition parameter, shear and bulk viscosities, …
    3. Model **outputs (observables)**: hadron multiplicity, v_n's, R_{AA}, …

“Forward problem”

- Given model parameters, what are the model outputs/observables?

“Inverse problem”

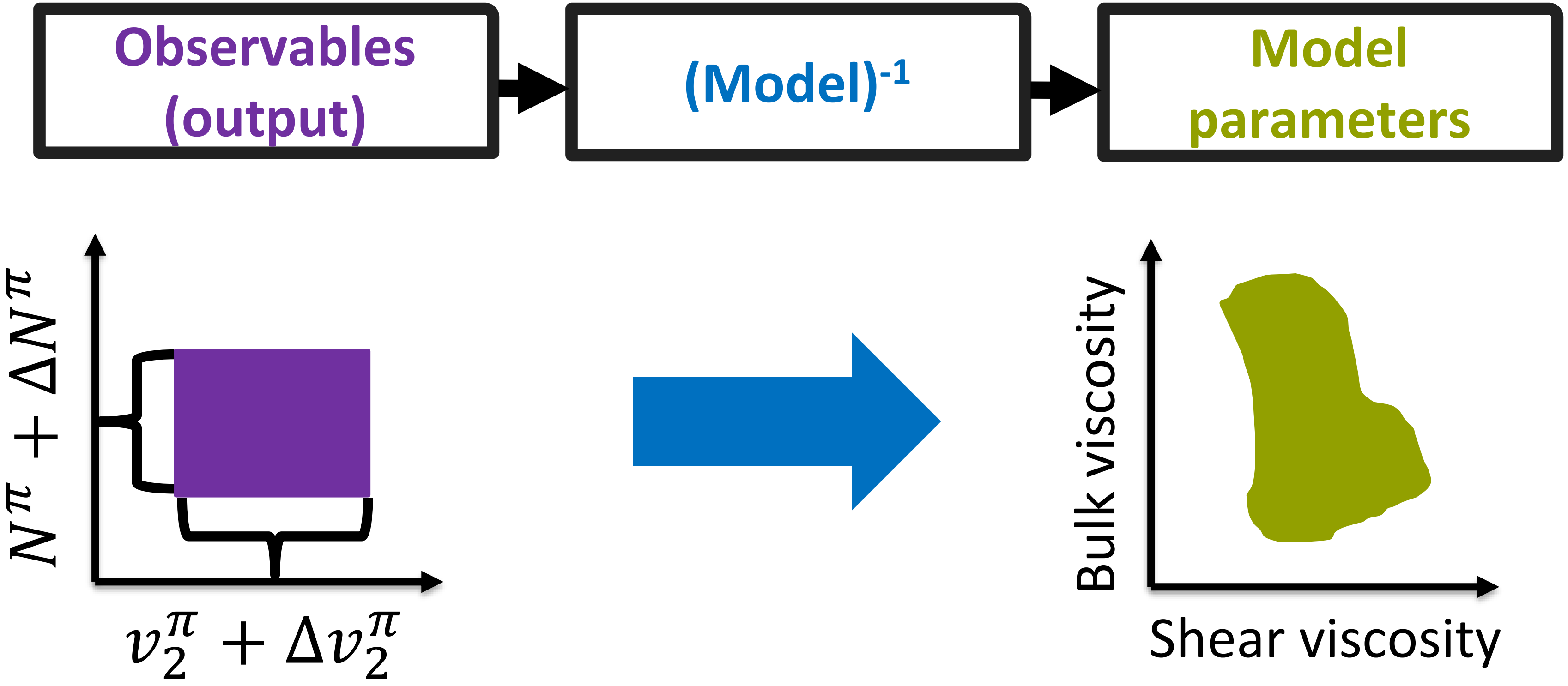
- What are the model parameters consistent with given model outputs/observables?
- Generally ill-defined problem, unless the model is very simple



The inverse problem: mapping observables to parameters

“Inverse problem”

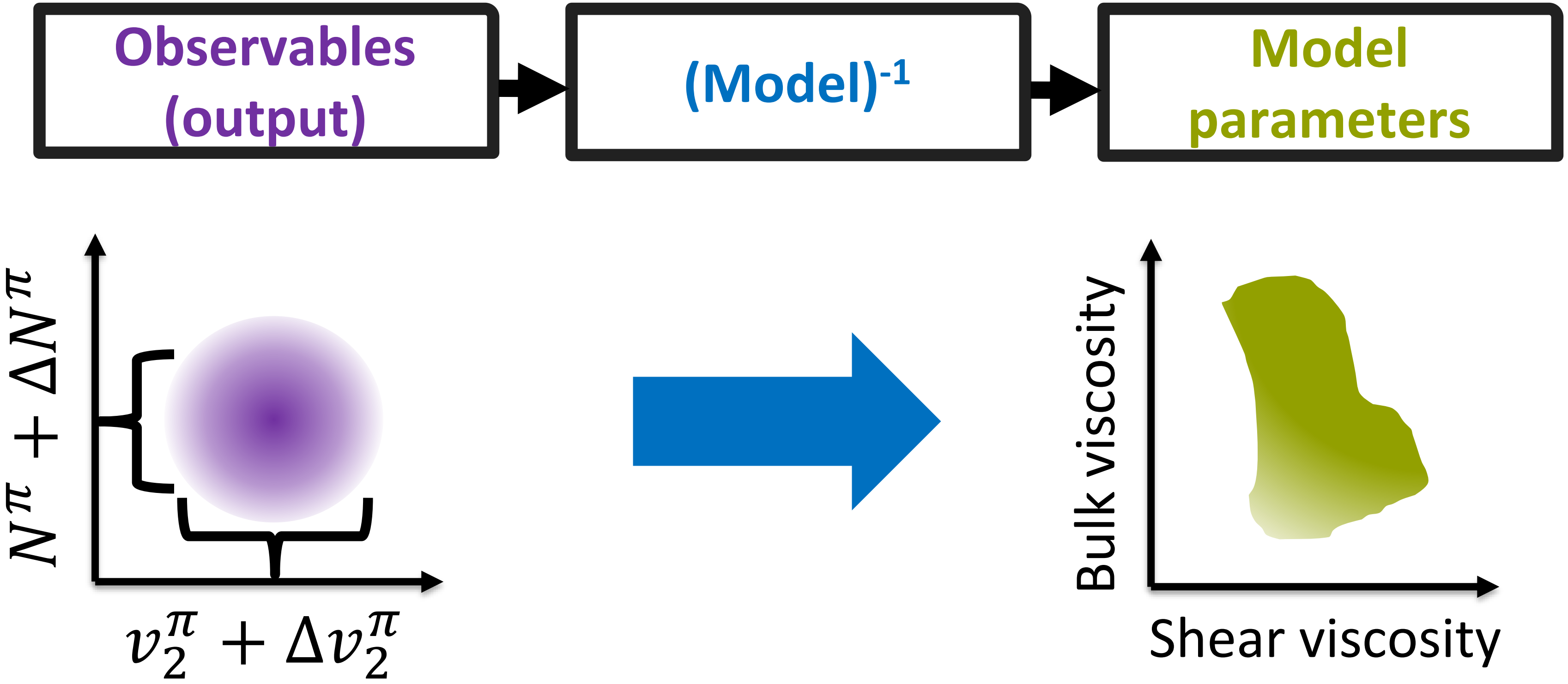
- What are the model parameters consistent with given model outputs/observables?
- <u>Generally ill-defined problem</u>



The inverse problem: mapping probability distributions

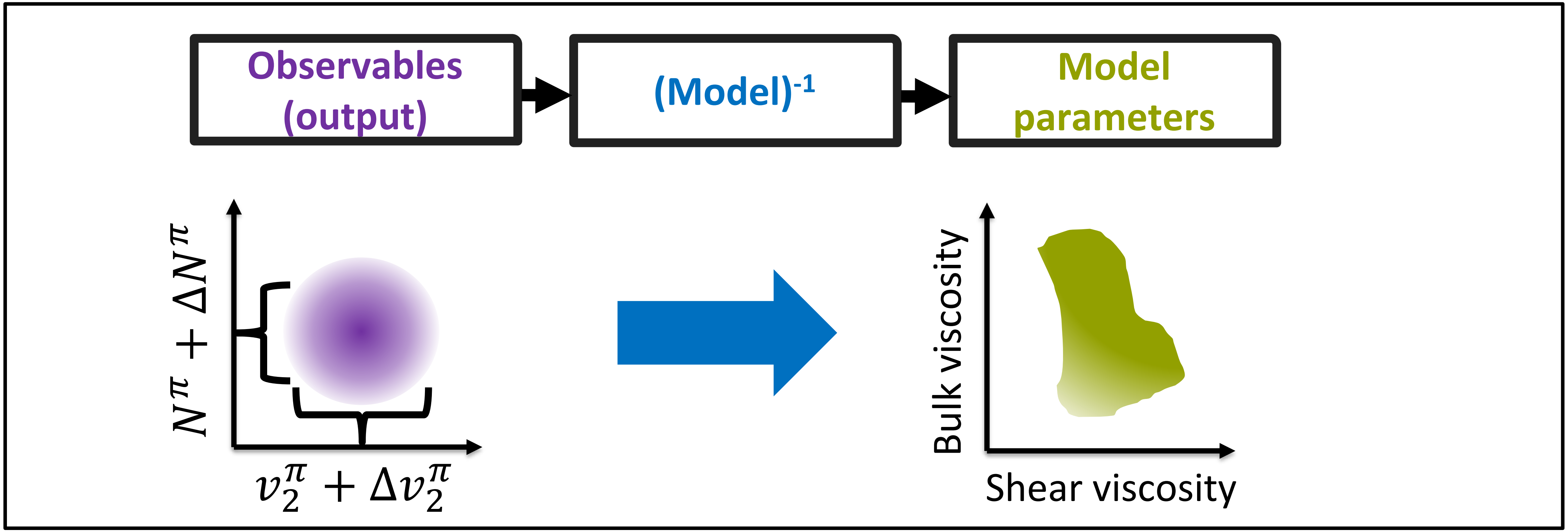
"Inverse problem"

- What are the model parameters consistent with given model outputs/observables?
- Generally ill-defined problem



Question: when to use Bayesian parameter estimation?

- Rule of thumb:
  Models with large numbers of parameters that must be constrained with large number of measurements
- Also very useful if few data points but strong theoretical guidance
- Can be used even for simple problems, and returns the “intuitive” answer



Simple introduction to Bayesian parameter estimation

Measurements as probability distributions

- Data are often averages over ensembles of events, quoted as
  i) a mean value;
  ii) a statistical uncertainty on this mean;
  iii) additional systematic uncertainties
  (not normally distributed)

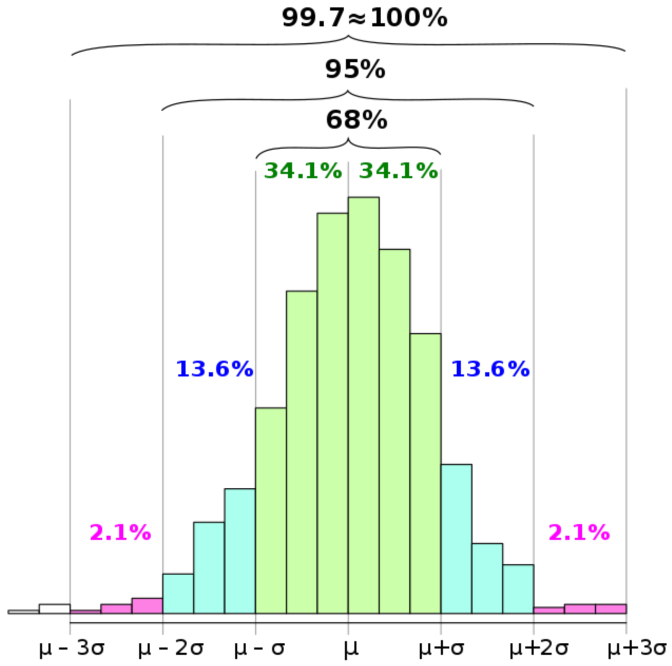


Fig. ref: https://commons.wikimedia.org/wiki/File:Empirical_rule_histogram.svg

Summary of neutron lifetime measurements

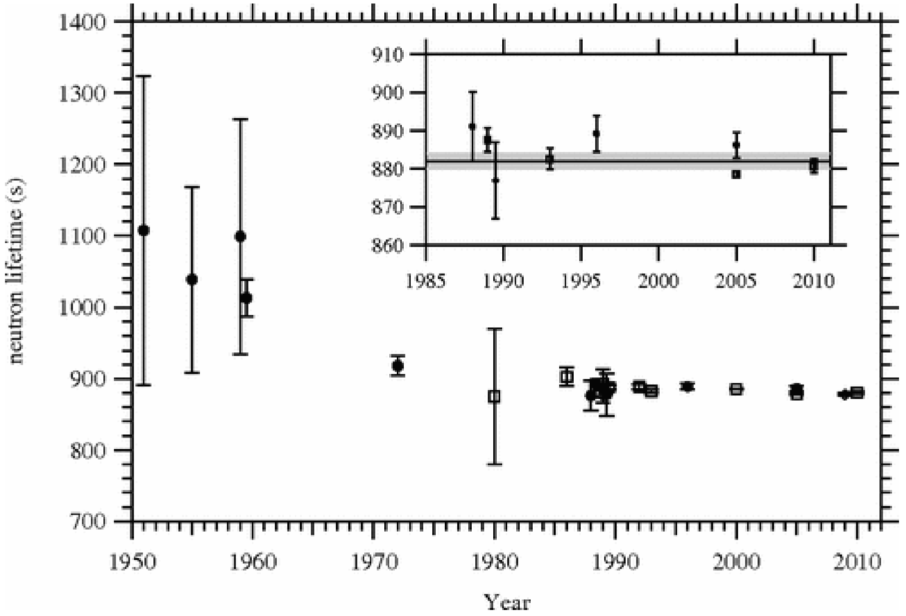


Fig. ref: Wietfeldt and Greene (2011) RMP83:1173

Measurements as probability distributions

- Data are often averages over ensembles of events, quoted as
  i) a mean value;
  ii) a statistical uncertainty on this mean;
  iii) additional systematic uncertainties
  (not normally distributed)

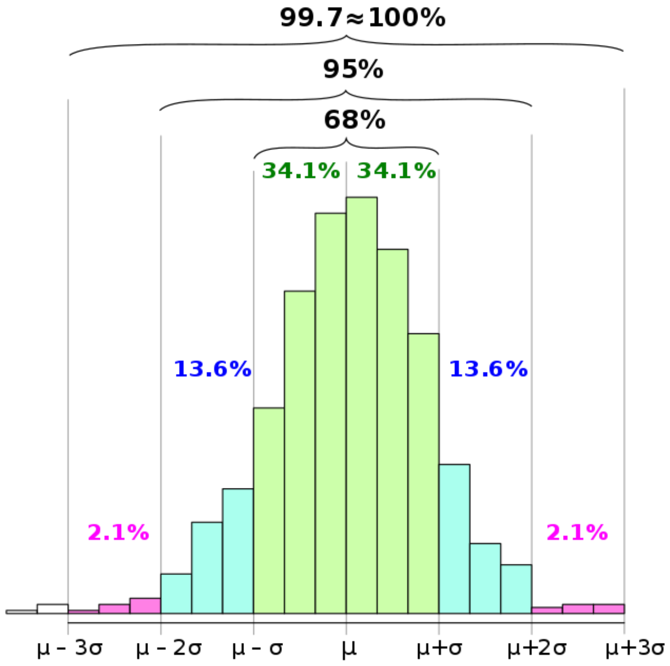


Fig. ref: https://commons.wikimedia.org/wiki/File:Empirical_rule_histogram.svg

Our simplifying assumption:
all uncertainties are normally distributed

$$\rho_d(A) \propto exp[-(A-d)^T\, C_E^{-1}(A-d)/2]$$

where:

"A" is the observable label (e.g. "v_2");

"d" is the mean of the measurements;

"C_E " is the covariance matrix encoding experimental uncertainties

Model predictions as probability distributions

Models have uncertainties too:

- Statistical: from averaging over collisions, or finite number of particles
- Numerical: can be reduced arbitrarily small, but should they?

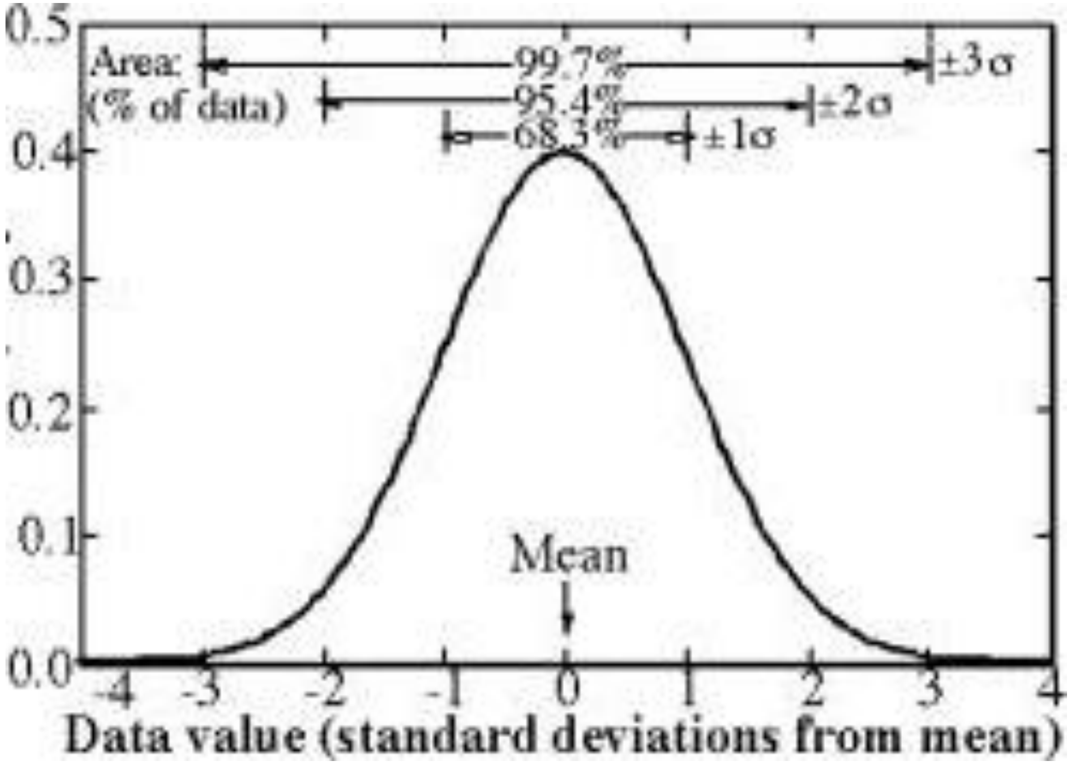


Fig. ref: https://upload.wikimedia.org/wikipedia/commons/4/47/SmFig2_3.jpg

Model predictions as probability distributions

Models have uncertainties too:

- Statistical: from averaging over collisions, or finite number of particles
- Numerical: can usually be reduced arbitrarily small, but should they?

Model predictions as probability distribution,
normal distribution again (for simplicity):

$$prob(A|p) \propto \frac{exp\left[-\left(A - g(p)\right)^T C_T^{-1}(p)\left(A - g(p)\right)/2\right]}{\sqrt{\det[C_T(p)]}}$$

where

“A”: observable label (e.g. “v_2”);

“g(p)”: model prediction for observable “A” given parameters “p”;

“$C_T\ (p)$“: covariance matrix encoding theoretical uncertainties

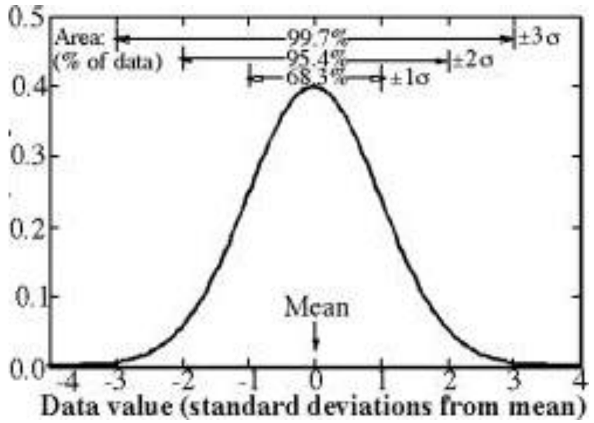


Fig. ref: https://upload.wikimedia.org/wikipedia/commons/4/47/SmFig2_3.jpg

Likelihood: probability of data given model parameters

- Folding
  (i) the probability distribution for the model; with
  (ii) the probability distribution for the data; yields
  (iii) the "likelihood",
  the probability for the data given the model parameters

$$\mathbf{prob}(\boldsymbol{d}|\boldsymbol{p}) \propto \frac{\boldsymbol{exp}\left[-\frac{\left(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{p})\right)^{\boldsymbol{T}}\left(\boldsymbol{C_E}+\boldsymbol{C_T}\right)^{-\mathbf{1}}\left(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{p})\right)}{\mathbf{2}}\right]}{\sqrt{\det[C_E+C_T(p)]}} \propto \boldsymbol{exp}(-\boldsymbol{\chi}^{\mathbf{2}}/\mathbf{2})$$

- Most numerically-challenging part of Bayesian parameter estimation:
  evaluating $\boldsymbol{exp}\left(-\boldsymbol{\chi}^{\mathbf{2}}/\mathbf{2}\right)$
  because:
  (i) the model "$g(p)$" can have a complicated dependence on the parameters
  (ii) the problem can be high-dimensional

Inverse problem: probability of model parameters given data

$\mathbf{prob(p)}$	$\times$	$\mathbf{prob}(\boldsymbol{d}\|\boldsymbol{p})$	$=$	$\boldsymbol{prob}(\boldsymbol{p},\boldsymbol{d})$	$=$	$\boldsymbol{prob}(\boldsymbol{d})$	$\times$	$\boldsymbol{prob}(\boldsymbol{p}\|\boldsymbol{d})$
Prior	$\times$	**Likelihood**	$=$	**Joint**	$=$	**Evidence**	$\times$	**Posterior**

Inputs → **Outputs**

Adapted from:
Bayesian Methods in Cosmology, edited by Michael P. Hobson, et al., Cambridge University Press, 2009.
Chapter "Foundations and algorithms", by John Skilling

- The posterior is what we are after: probability of model parameters given data

$$\boldsymbol{Posterior} \propto \boldsymbol{Prior} \times \boldsymbol{Likelihood}$$

$$\mathbf{prob}(\boldsymbol{p}|\boldsymbol{D}) \propto \boldsymbol{prob}(\boldsymbol{p}) \times \boldsymbol{prob}(\boldsymbol{D}|\boldsymbol{p})$$

- In our case: $\mathbf{prob}(\boldsymbol{p}|\boldsymbol{D}) \propto \boldsymbol{prob}(\boldsymbol{p}) \times \boldsymbol{exp}(-\boldsymbol{\chi}^{2}/\mathbf{2})$

- Note about the “evidence”:
we work under the assumption that the model describes the data reasonably well

The prior

$$\mathrm{prob}(p) \times \mathrm{prob}(d|p) = prob(p,d) = prob(d) \times prob(p|d)$$

Prior × Likelihood = Joint = Evidence × Posterior

Inputs → Outputs

- Our posterior is: $\propto prob(p) \times exp\left[-\frac{(d-g(p))^T(C_E+C_T)^{-1}(d-g(p))}{2}\right] / \sqrt{\det[C_E + C_T(p)]}$

Prior

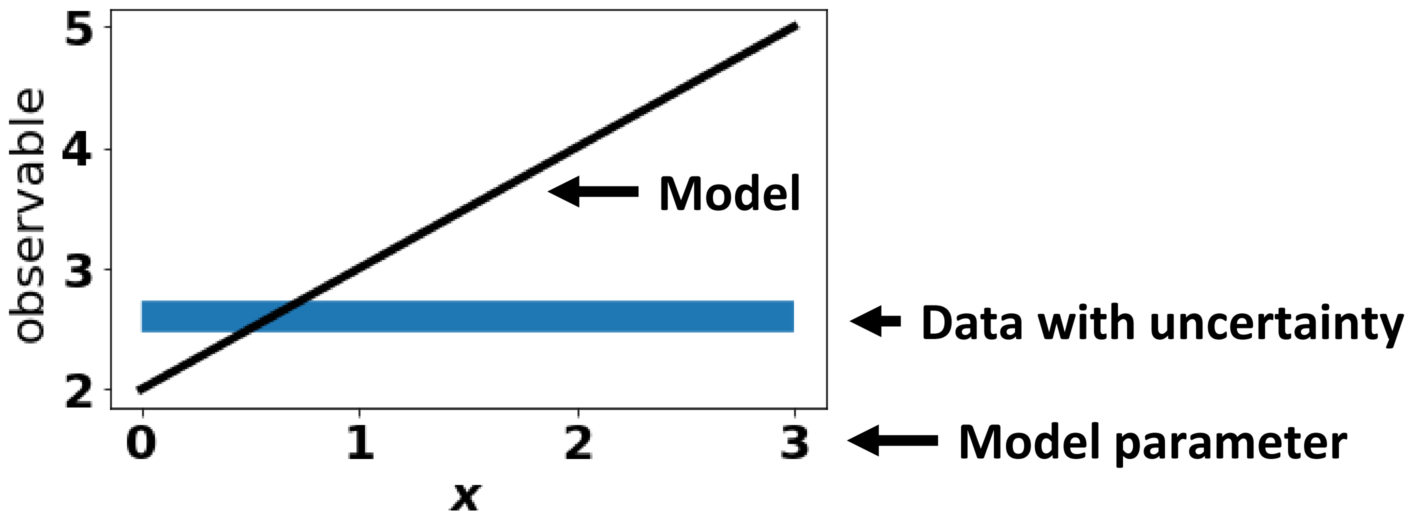
Priors encode previous knowledge about the model parameters

- Positivity: viscosities, nucleon width, … must be positive
- Knowledge from other experiments: “nucleon widths > 2 fm” are unlikely…
- Knowledge from theory
- Self-consistency of model

(Very) Simple examples of Bayesian parameter estimation

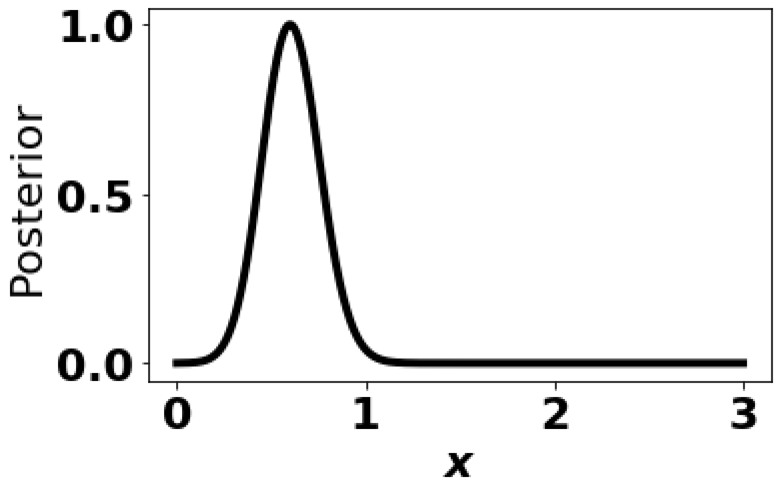
Simple example: one parameter, one observable, linear model

- Say your model is $observable(x) = 2 + x$
- Say the measured observable is $d = 2.6 \pm 0.13$ (5% relative uncertainty)



Priors:
I assume a uniform prior on "x"
$\boldsymbol{prob(p) = Step\ (0 < x < 3)}$

- Result:

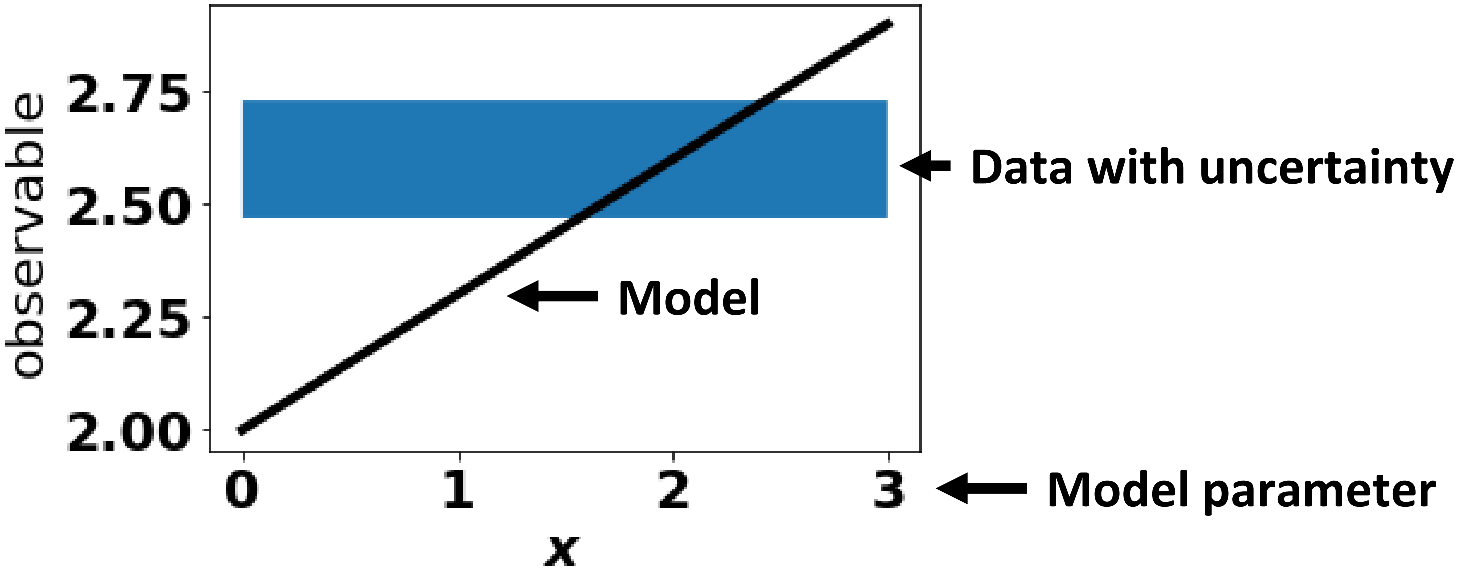


$$\mathbf{posterior}(\boldsymbol{p}|\boldsymbol{D}) \propto \boldsymbol{exp}\left[-\frac{(\boldsymbol{d}-2-\boldsymbol{x})^2}{2(0.05\boldsymbol{d})^2}\right]$$

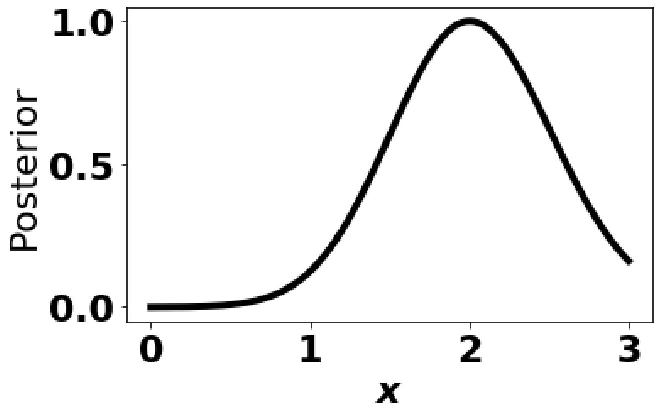
- **Mean of x:** $(\boldsymbol{d}-2)$
- **Uncertainty on x:** $0.05 \times \boldsymbol{d}$ **(same as data)**

Simple example: one parameter, one observable, linear model

- Say your model is $observable(x) = 2 + 0.3\,x$
- Say the measured observable is $d = 2.6 \pm 0.13$ (5% relative uncertainty)



- Result:



$$\boldsymbol{exp}\left[-\frac{(d-2-0.3\,x)^2}{2(0.05d)^2}\right] = \boldsymbol{exp}\left[-\frac{\left((d-2)/0.3-\ x\right)^2}{2(0.05d/0.3)^2}\right]$$

- **Mean of x:** $(\mathbf{d}-2)/0.3$
- **Uncertainty on x:** $\frac{(\text{exp.uncert})}{(\textit{slope of model})} = \frac{0.05\,d}{0.3}$

Bayesian parameter estimation in higher dimensions

General case

$$\text{Posterior}(\boldsymbol{p}|\boldsymbol{d}) \propto \boldsymbol{prior}(\boldsymbol{p}) \times \boldsymbol{exp}\left[-\frac{\left(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{p})\right)^{\boldsymbol{T}}(\boldsymbol{C_E}+\boldsymbol{C_T})^{-\boldsymbol{1}}\left(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{p})\right)}{\boldsymbol{2}}\right] / \sqrt{\det[C_E + C_T(p)]}$$

Inverse problem can become complicated very quickly:

- Many parameters: posterior becomes a high-dimensional probability distribution
- Non-linearity in model: correlated parameters, …
- Many observables
- Theoretical and experimental uncertainties can be complex:
  - statistical uncertainties,
  - systematic uncertainties correlated across p_T bins, centrality, observables, …
- Prior theoretical knowledge on parameters

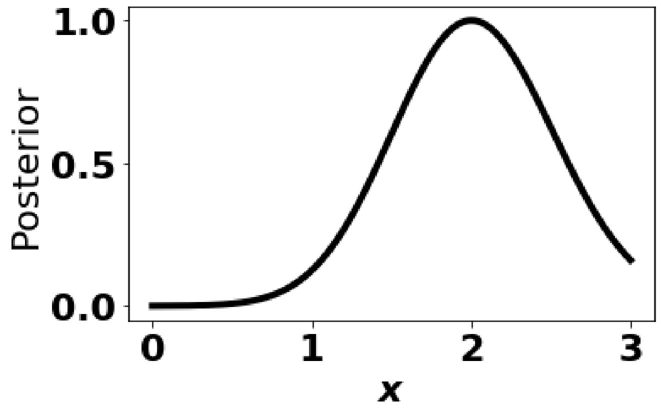
Can all be accounted methodically for in Bayesian parameter estimations

Multiple parameters: visualizing the posterior

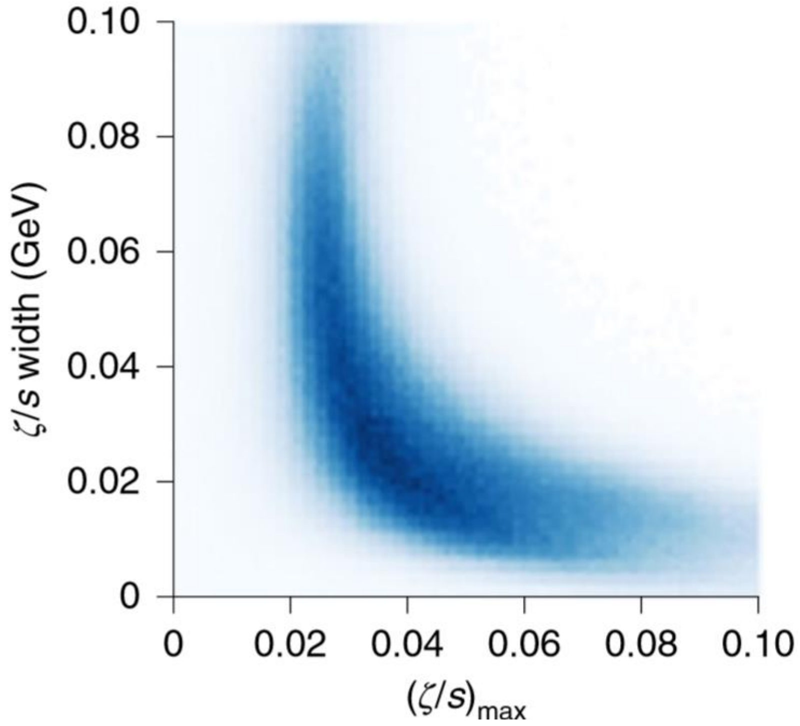
- The posterior is a probability distribution of the same dimension as the number of parameters:

$$\boldsymbol{Posterior}(\boldsymbol{p}|\boldsymbol{d}) \propto \boldsymbol{exp}\left[-\frac{\left(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{p})\right)^{\boldsymbol{T}}(\boldsymbol{C_E}+\boldsymbol{C_T})^{-\boldsymbol{1}}\left(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{p})\right)}{\boldsymbol{2}}\right] / \sqrt{\det[C_E + C_T(p)]}$$

- One-dimensional case is easy to visualize



- Two-dimensional case is easy enough as well:



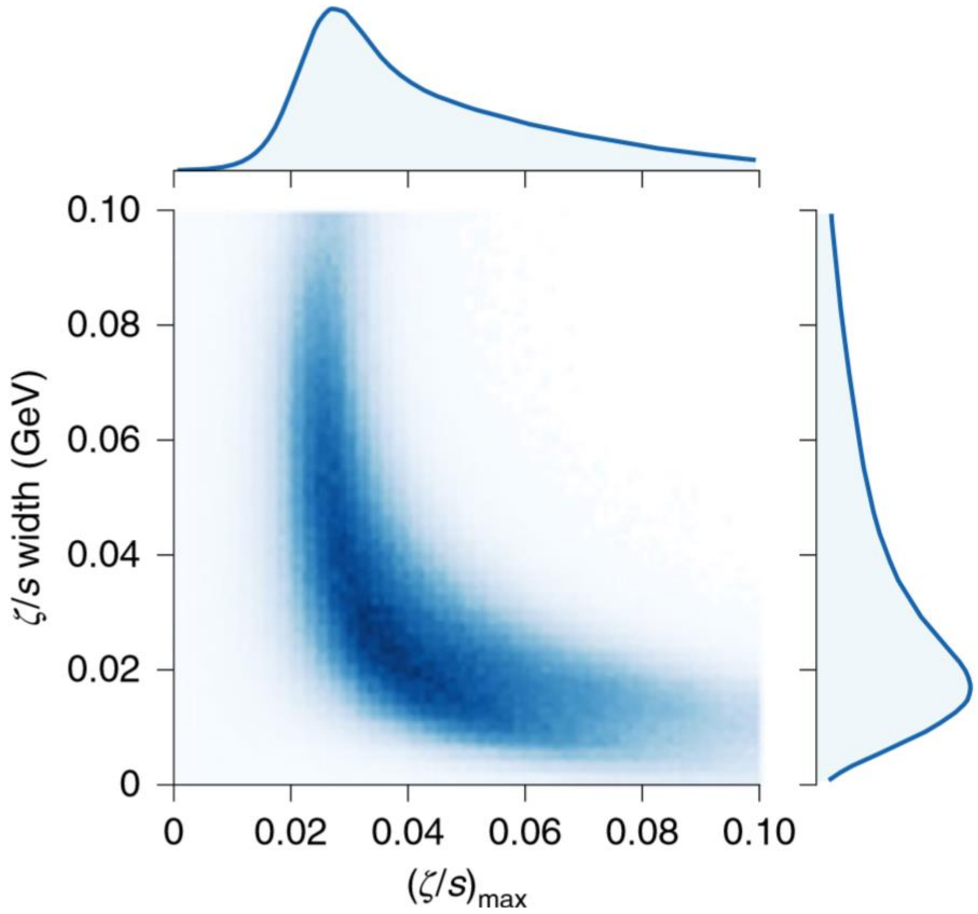
Ref.: Bernhard, Moreland & Bass (2019) Nature Phys.15,11:1113

Multiple parameters: marginalizing the posterior

- The posterior is a probability distribution of the same dimension as the number of parameters:

$$\boldsymbol{Posterior(p|d)} \propto \frac{\boldsymbol{exp}\left[-\frac{(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{p}))^{T}(\boldsymbol{C_E}+\boldsymbol{C_T})^{-1}(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{p}))}{2}\right]}{\sqrt{\det[C_E+C_T(p)]}}$$

1D projection of $(\zeta/s)_{max}$



- Even with two parameters, we may want to "project" in 1D: marginalize the distribution

$$\boldsymbol{Marginal\ posterior(p_1)} = \int \boldsymbol{dp_2\ Posterior(p_1, p_2)}$$

- Can be generalized in arbitrary dimensions
- Especially useful for "nuisance parameters"

(Example on the right is actually marginalized in 2D from a 14-dimensional posterior)

Ref.: Bernhard, Moreland & Bass (2019) Nature Phys.15,11:1113

Multiple parameters: credible intervals

- The posterior is a probability distribution of the same dimension as the number of parameters:

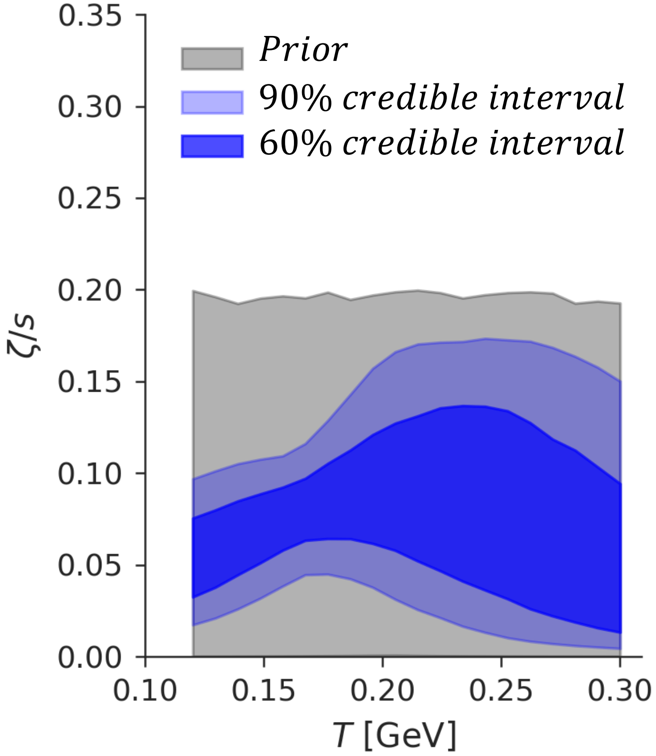
$$\boldsymbol{Posterior(p|d)} \propto \frac{\boldsymbol{exp}\left[-\frac{(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{p}))^{\boldsymbol{T}}(\boldsymbol{C_E}+\boldsymbol{C_T})^{-1}(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{p}))}{2}\right]}{\sqrt{\det[C_E+C_T(p)]}}$$

- Example:
  This figure visualizes 4 parameters

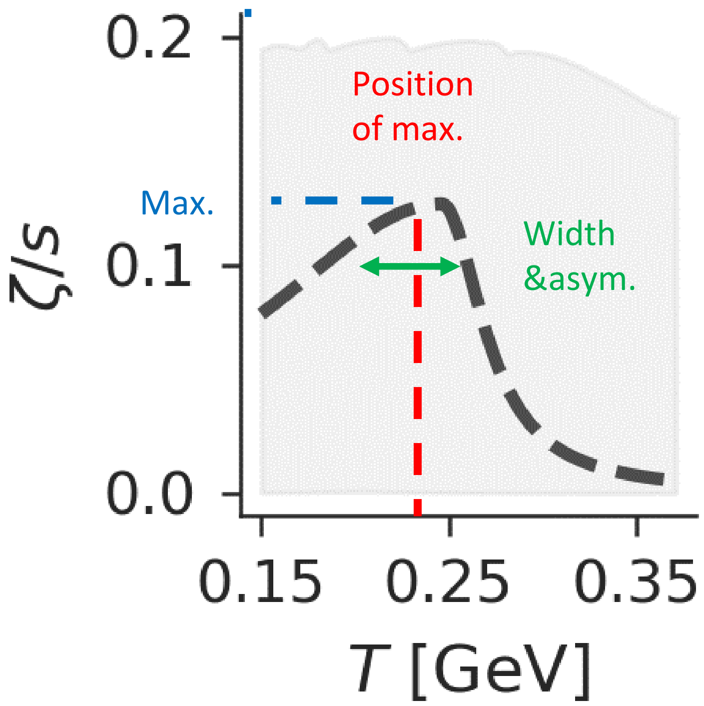
  The temperature-dependence of the specific bulk viscosity ζ/s, parametrized with 4 parameters

  The credible intervals show the higher probability regions of the posterior

(Example on the right is marginalized over other parameters as well)



Ref.: JETSCAPE Collaboration



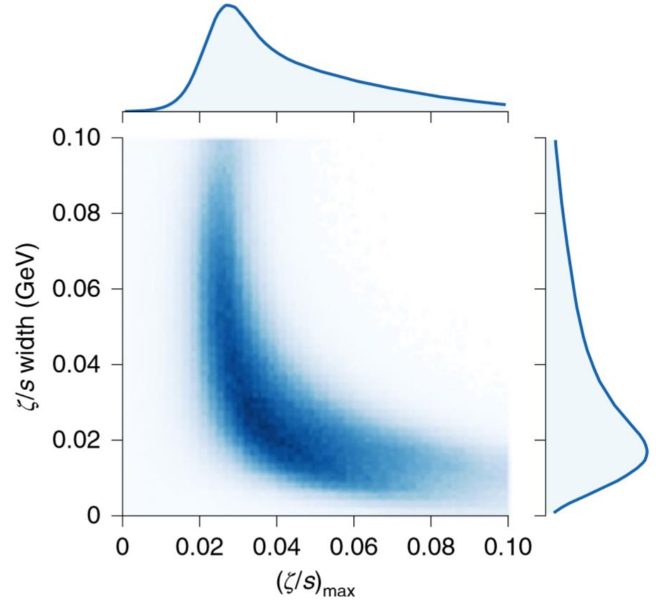
Sampling the posterior

$$\boldsymbol{Posterior(p|d)} \propto \frac{\boldsymbol{exp}\left[-\frac{(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{p}))^{\boldsymbol{T}}(\boldsymbol{C_E}+\boldsymbol{C_T})^{-\boldsymbol{1}}(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{p}))}{\boldsymbol{2}}\right]}{\sqrt{\det[C_E+C_T(p)]}}$$

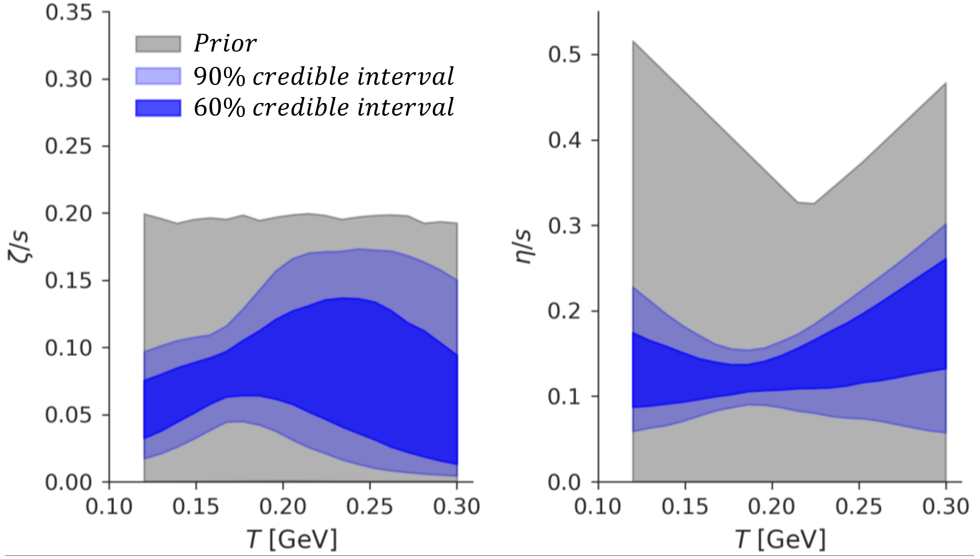
- How does one sample (or marginalize over) a high-dimensional probability?

  Monte Carlo sampling

- Generally performed with different types of Markov chain Monte Carlo (MCMC) methods



Ref.: Bernhard, Moreland & Bass (2019) Nature Phys.15,11:1113

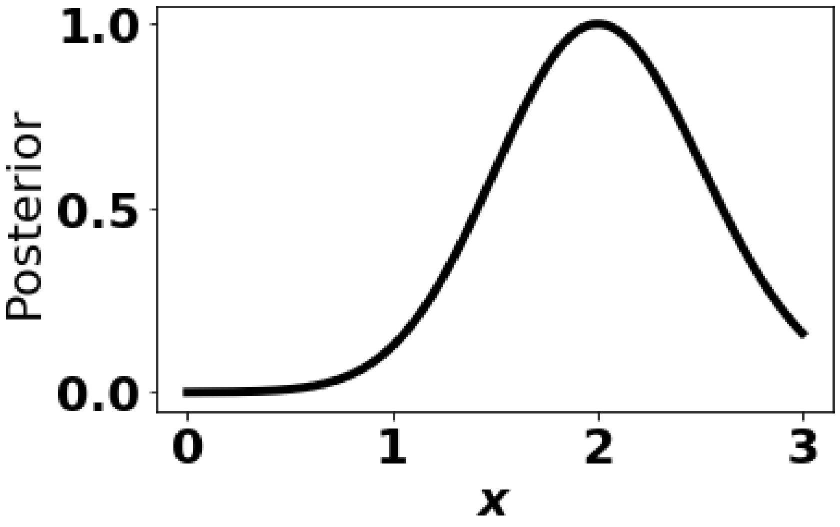


Ref.: JETSCAPE Collaboration

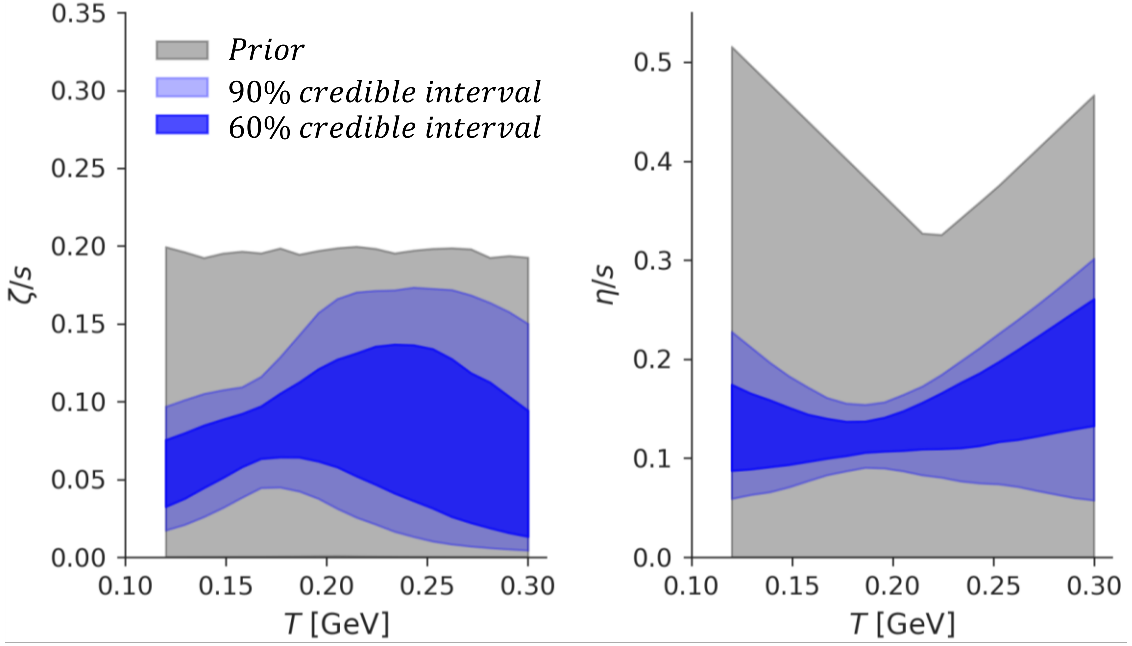
Questions?

Question: type of constraints

- <u>Probabilistic</u>



Ref.: JETSCAPE Collaboration



- If the resulting posterior is Gaussian, it can be summarized by a mean and a standard deviation
- If non-Gaussian, the median can be more robust than the mean

- When parameters are correlated, visualizing the posterior can require some creativity

Question: theoretical input?

- Theoretical insights can be included in the model (for example, lattice calculations of the equation of state)

  or

  it can be included in the parameter priors

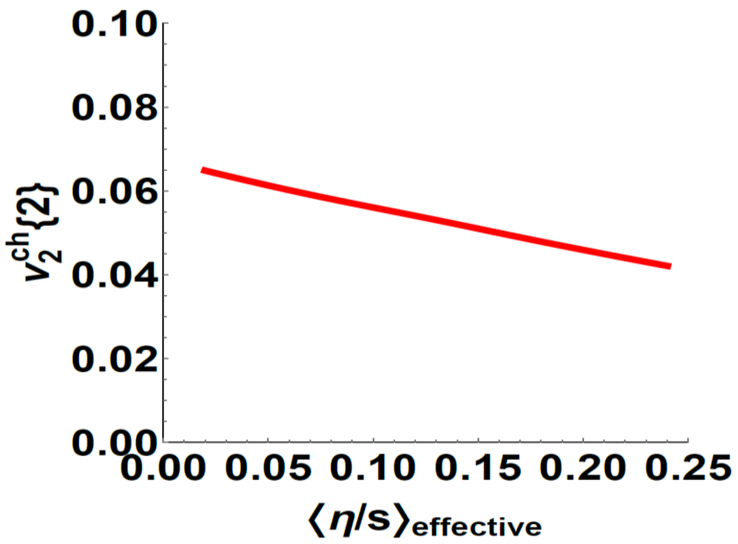
Practical aspect: emulators

Question: what if my model is "slow"?

$$\boldsymbol{Posterior} \propto \boldsymbol{exp}\left[-\frac{(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{p}))^{\boldsymbol{T}}(\boldsymbol{C_E}+\boldsymbol{C_T})^{-1}(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{p}))}{\boldsymbol{2}}\right] / \sqrt{\det[C_E + C_T(p)]}$$

- "g(p)" is our model.
  For a multi-observable model, think of "g(p)" as a vector containing all the observables

- In practice, the model should have a (relatively) smooth dependence on the parameters



- **Solution: <u>emulators</u>**

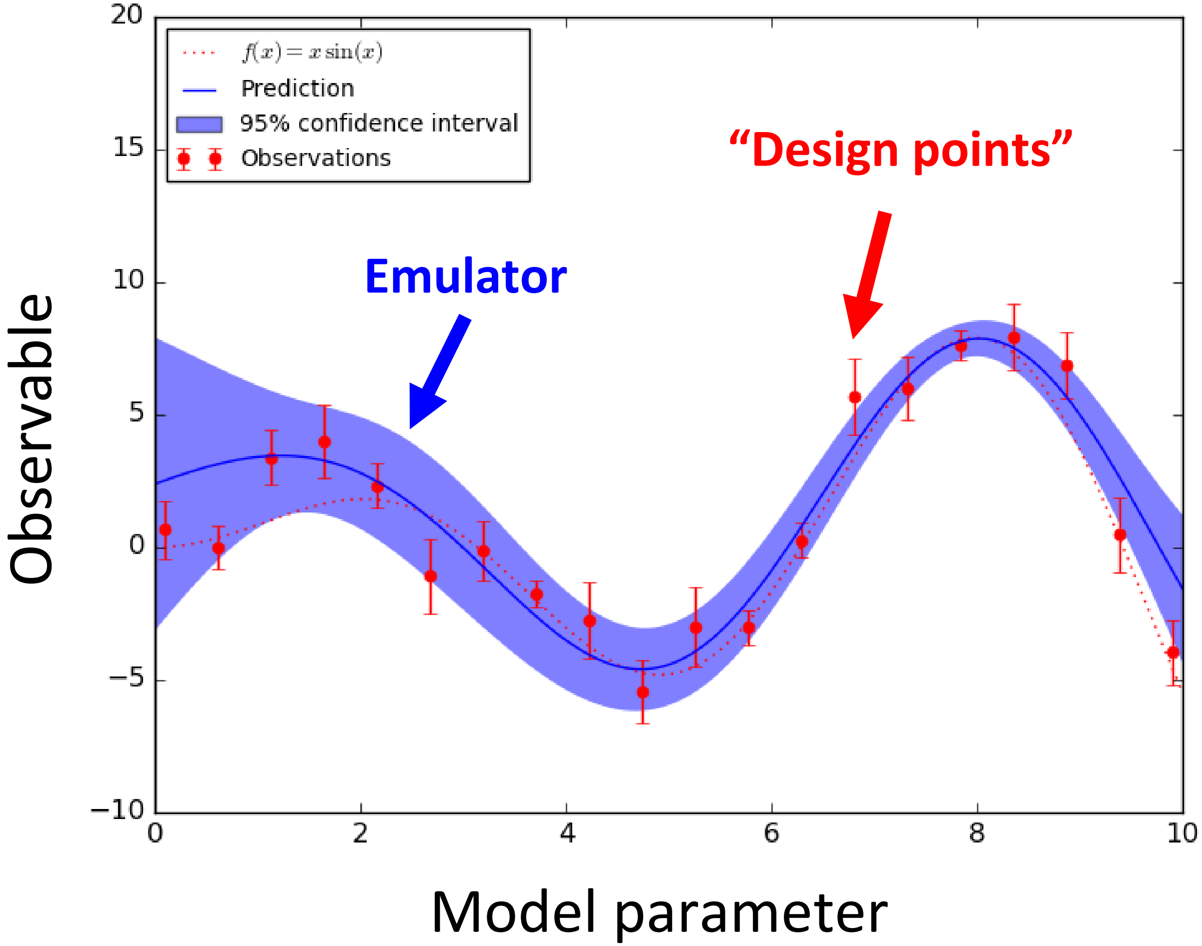
- Recall that we already view the prediction of the model "g(p)" as a probability distribution (which includes theoretical uncertainties)

Emulators

- Emulators are <u>probability distributions</u> that mimics the <u>model outputs'</u> dependence on the parameters

- Constrained by the model (and model uncertainty) at sample points: "design points"

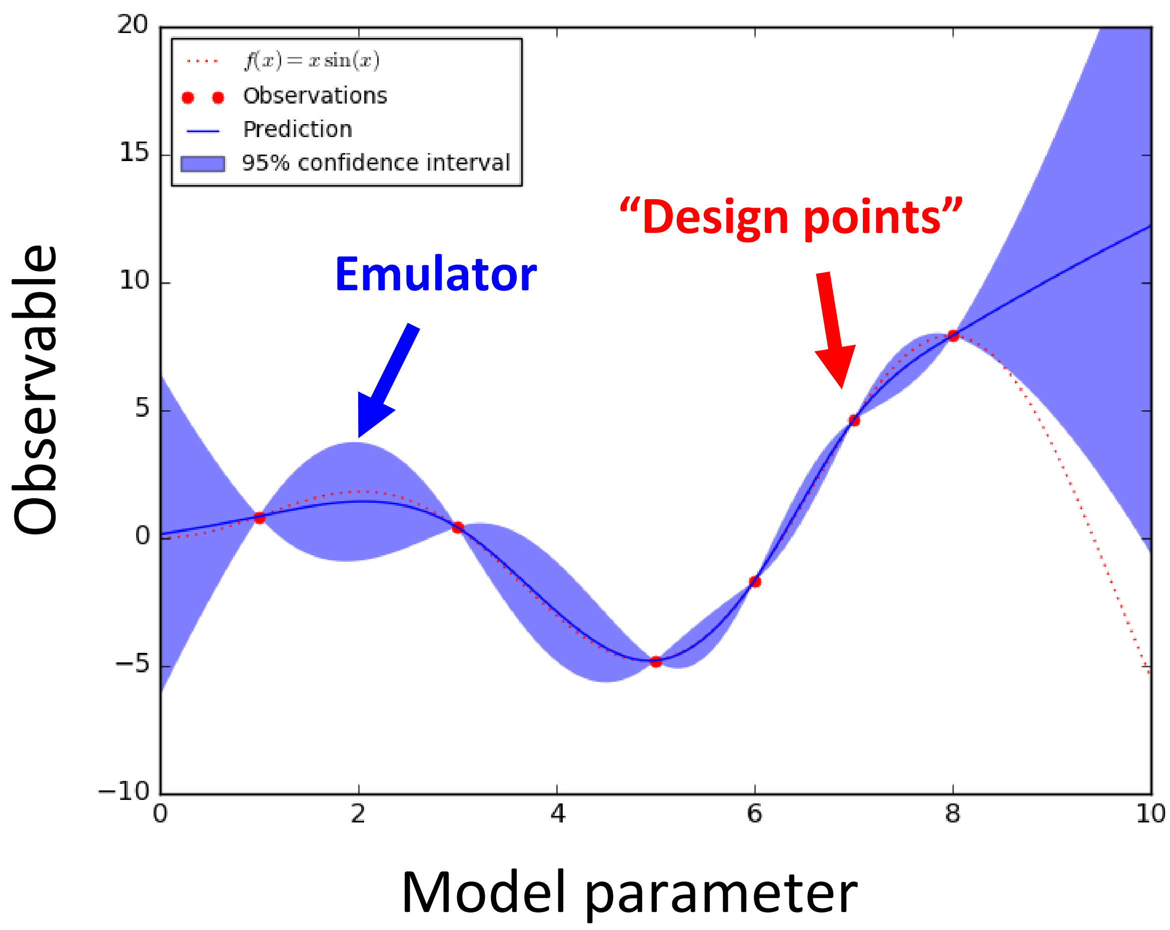
Fig. ref.: https://scikit-learn.org/0.17/auto_examples/gaussian_process/plot_gp_regression.html



Emulators

- Emulators are probability distribution that mimics the model outputs' dependence on the parameters
- Constrained by the model (and model uncertainty) at sample points: "design points"
- Good emulators, like Gaussian process emulators, can estimate the interpolation uncertainty: extremely important!

Fig. ref.: https://scikit-learn.org/0.17/auto_examples/gaussian_process/plot_gp_regression.html



Emulators: important additional information

- Parameter space:
  Range of parameters over which the emulator can mimic the output of the model
  (the parameter space is sampled with the Latin hypercube algorithm)

- The emulator is not for the model itself,
  but for the model outputs (the observables: $v_2, dN/dy, R_{AA}$, …)

- For practical reasons, it’s not even the model observables that are emulated:
  It is linear combinations of model parameters
  (identified through principal component analysis)

- An independent Gaussian process emulator describes each major linear combinations of observables (“principal components”)

Gaussian process emulators: more information

Excellent material from previous JETSCAPE Schools by Jake Coleman and Weiyao Ke

- Jake Coleman lecturing at the 2018 JETSCAPE School in Berkeley
  https://indico.bnl.gov/event/3958/
  https://sites.google.com/a/lbl.gov/jetscape2018/home/school-material/school-preparation

- Weiyao Ke lecturing at the 2019 JETSCAPE School in Texas A&M
  https://indico.bnl.gov/event/5031/page/115-school-material (“School preparation - Statistical Analysis”)
  https://github.com/JETSCAPE/STAT/blob/master/WS_Theory_Exercises.pdf
  https://github.com/keweiyao/BayesExample/blob/master/example.ipynb

Closure tests as examples of Bayesian parameter estimation?

Idea behind closure tests

- Bayesian parameter estimations are typically performed against data

Experimental uncertainties

$$\boldsymbol{Posterior}(\boldsymbol{p}|\boldsymbol{d}) \propto \boldsymbol{exp}\left[-\frac{\left(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{p})\right)^{T}(\boldsymbol{C_E}+\boldsymbol{C_T})^{-1}\left(\boldsymbol{d}-\boldsymbol{g}(\boldsymbol{p})\right)}{\boldsymbol{2}}\right] / \sqrt{\det[C_E + C_T(p)]}$$

Experimental measurements

Idea behind closure tests

- Bayesian parameter estimations are typically performed against data

Use a chosen uncertainty as "experimental uncertainty"

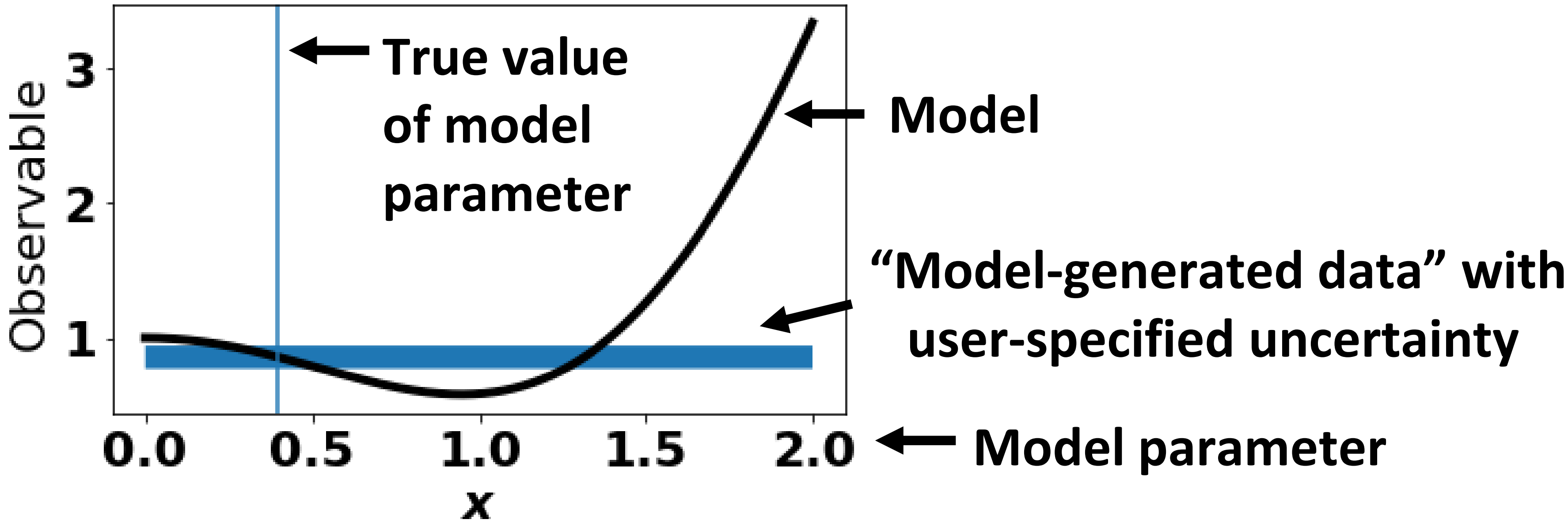
$$\boldsymbol{Posterior}(\boldsymbol{p}|\boldsymbol{d}) \propto \boldsymbol{exp}\left[-\frac{\left(\boldsymbol{g}(\tilde{p})-\boldsymbol{g}(\boldsymbol{p})\right)^{T}\left(\boldsymbol{C}_{E}+\boldsymbol{C}_{T}\right)^{-1}\left(\boldsymbol{g}(\tilde{p})-\boldsymbol{g}(\boldsymbol{p})\right)}{\boldsymbol{2}}\right]/\sqrt{\det[C_E+C_T]}$$

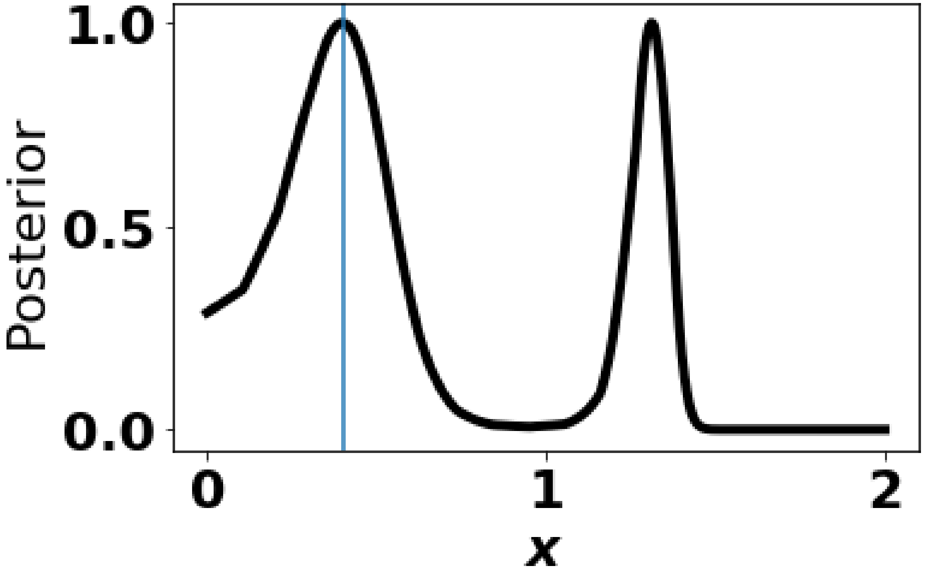
Use a model calculation as "data"

But there are benefits to performing Bayesian parameter estimation on calculations

- If "fake data" "**d=**$\boldsymbol{g}(\tilde{p})$" is generated from the model with a set of parameters "$\{\tilde{p}\}$", we know what the results of the Bayesian analysis should be
- Can be used to:
  - validate various aspects of the analysis: the emulator, the sampling of the posterior, …
  - better understand how different parameters depend on model observables
  - better understand the impact of different uncertainties

Closure tests





- May not recover the parameter exactly, because:
  - Theoretical and “data” uncertainties
  - Emulator uncertainties
  - Degeneracies in model

Question:
how will measurement XYZ help constrain parameter ABC?

Different approaches

- Parameter sensitivity of observables:
  vary parameters, see if observable is sensitive to them

- Closure tests:
  vary observable, see effect on parameter estimation

Summary

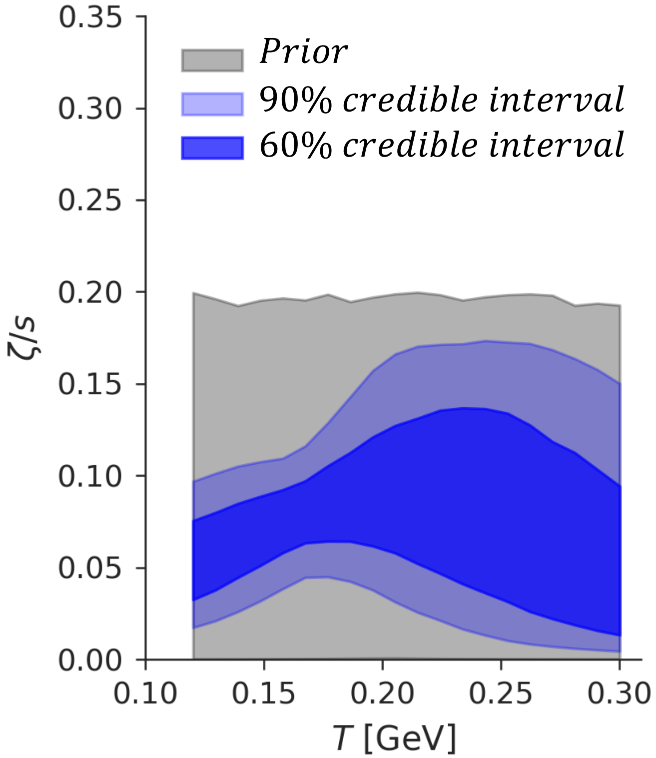
Summary: Bayesian parameter estimations

- "Forward problem": Model outputs from parameters; straightforward
- "Inverse problem": Parameters from the model outputs; challenging

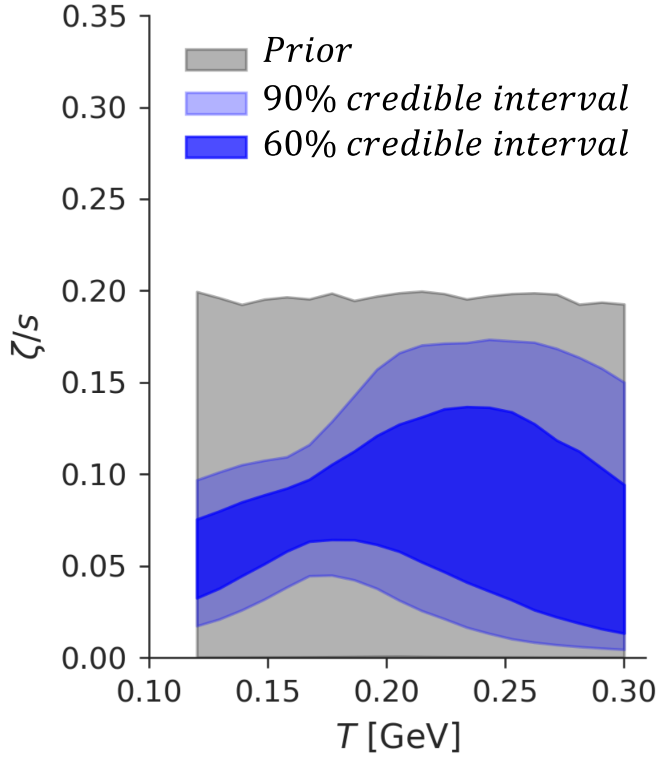
- Bayesian parameter estimation:
  methodical comparison of experimental measurements with model

- Can handle:
  - non-linearity in model
  - large number of observables
  - complex experimental uncertainties
  - theoretical uncertainties
  - theoretical input through priors

- Give probabilistic constraints



Outlook: Bayesian parameter estimations



- Methodical comparison of experimental measurements with model

- Know when to use Bayesian parameter estimation

  - The model must describe the data reasonably well (if not, consider the “evidence” and “model comparisons”)

  - Must understand the range of parameters to be probed (minimum “prior” information)

  - Verify beforehand that the observables depend on the parameters

  - Verify beforehand that the model is well-behaved across the parameter space

- Wide range of applicability

Acknowledgements

JETSCAPE Collaboration and the SIMS working group.

Special thanks to Derek Everett (OSU), Weiyao Ke (UC Berkeley & LBNL) and Dan Liyanage (OSU) for crucial input and feedback on this talk

This work was supported by:

Computational resources from XSEDE & TACC

XSEDE
Extreme Science and Engineering Discovery Environment

J.-F.P. was supported by the U.S. Department of Energy (DOE) under award number DE-FG02-05ER41367 and by the National Science Foundation (NSF) under award number ACI-1550300

Questions?