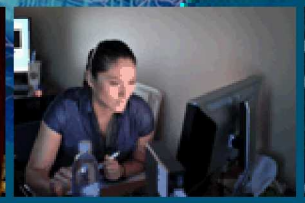
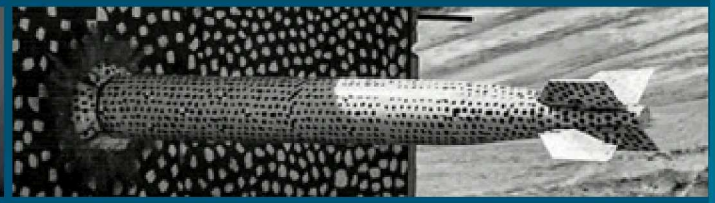
SAND2020-0124C

A Primer on Bayesian Inference and applications to data analysis





PRESENTED BY

Patrick F. Knapp

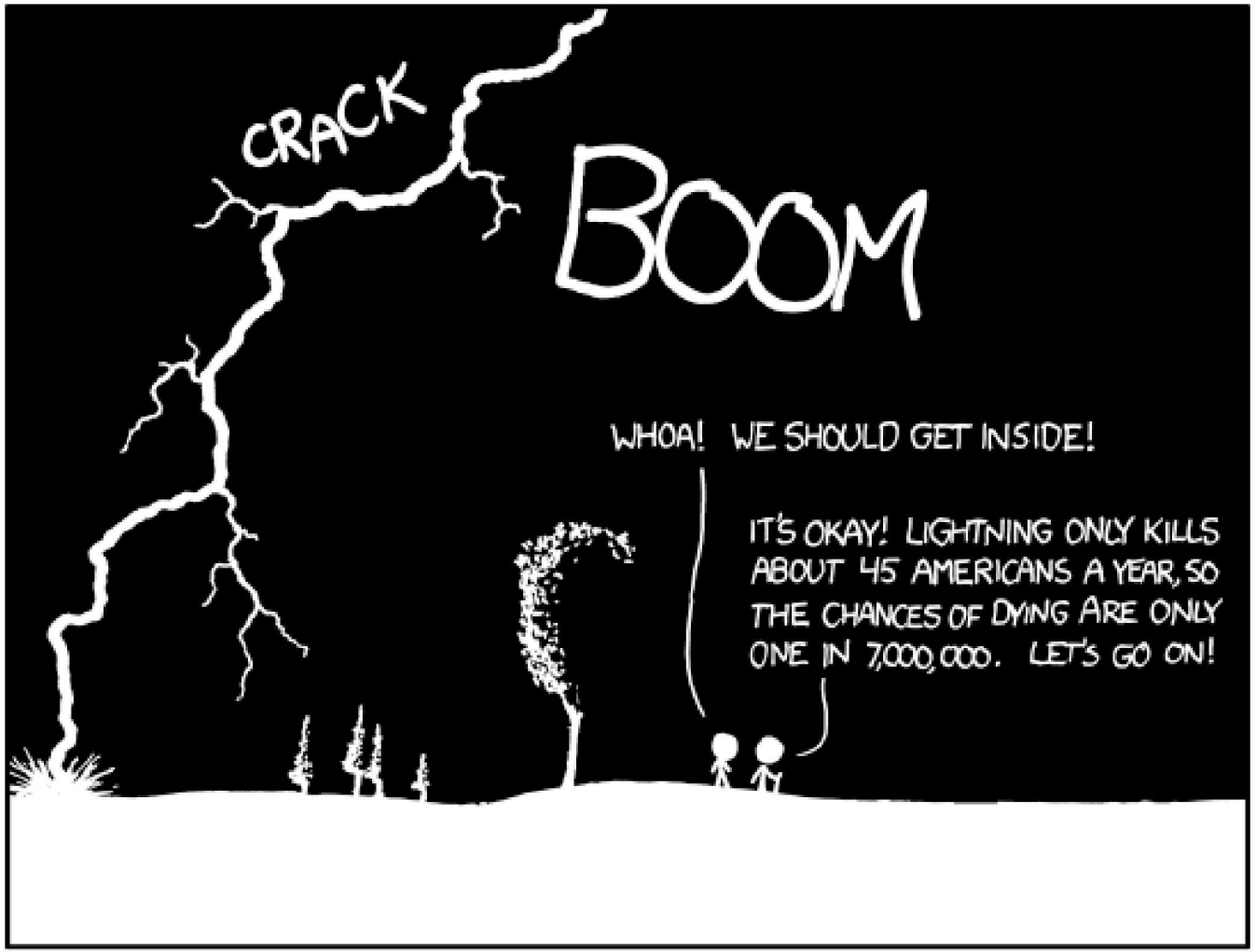
Extreme Physics, Extreme Data Workshop, Leiden, Netherlands

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Outline

- Bayes' Theorem and its implications
- Estimating physical parameters from data
  - Most likely values and confidence intervals
  - Correlations
  - How valuable is my data?
  - Is my model good enough?
- Implementation
  - Packages
  - Using ML/DL to speed up

Remember: All probabilities are conditional! This fact could save your life



Credit: https://xkcd.com/795/

P(lighting | American) = 1.5e-7

P(lighting | outside in a thunderstorm) >>
P(lighting | American) !!

The product rule for conditional probabilities

$$\mathcal{P}(X, Y|I) = \mathcal{P}(X|Y, I)\mathcal{P}(Y|I)$$

Bayes' Theorem

$$\mathcal{P}(\overline{\mathbf{m}}|\overline{\mathbf{x}}, A)\mathcal{P}(\overline{\mathbf{x}}|A) = \mathcal{P}(\overline{\mathbf{x}}|\overline{\mathbf{m}}, A)\mathcal{P}(\overline{\mathbf{m}}|A)$$

The stuff we already know

Bayes' Theorem

$$\mathcal{P}(\overline{\mathbf{m}}|\overline{\mathbf{x}}, A)\mathcal{P}(\overline{\mathbf{x}}|A) = \mathcal{P}(\overline{\mathbf{x}}|\overline{\mathbf{m}}, A)\mathcal{P}(\overline{\mathbf{m}}|A)$$

The stuff we can measure

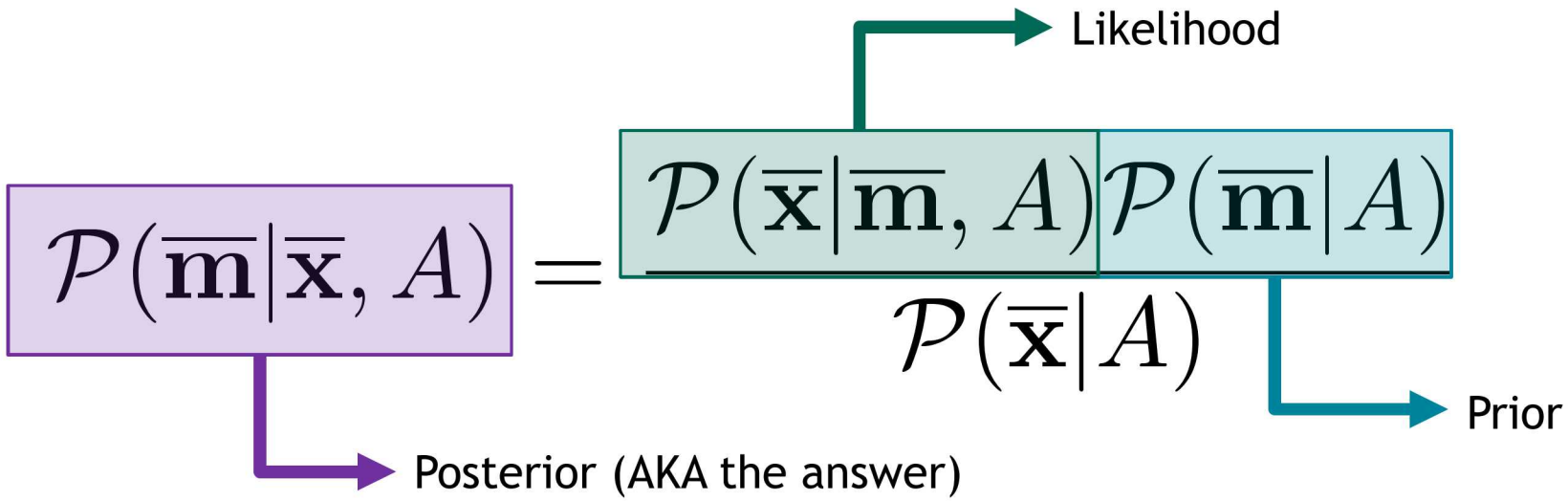
Bayes' Theorem

$$\mathcal{P}(\overline{\mathbf{m}}|\overline{\mathbf{x}}, A)\mathcal{P}(\overline{\mathbf{x}}|A) = \mathcal{P}(\overline{\mathbf{x}}|\overline{\mathbf{m}}, A)\mathcal{P}(\overline{\mathbf{m}}|A)$$

The thing we want to know

Bayes' Theorem

$$\mathcal{P}(\overline{\mathbf{m}}|\overline{\mathbf{x}}, A)\mathcal{P}(\overline{\mathbf{x}}|A) = \mathcal{P}(\overline{\mathbf{x}}|\overline{\mathbf{m}}, A)\mathcal{P}(\overline{\mathbf{m}}|A)$$



The likelihood function is what probabilistically relates our model to our observations

The likelihood function used varies depending on the application

- E.g. Single particle counting with a γ-spectrometer begs for a Poisson distribution
- Signal processing with white noise begs for a normal distribution

Multivariate Normal Distribution

$$\mathcal{P}(\overline{\mathbf{x}}|\overline{\mathbf{m}},\mathcal{A}) \propto \prod_{i=1}^{N} \exp\left(-\frac{(\mathcal{F}_i(\overline{\mathbf{m}})-x_i)^2}{2\sigma_i^2}\right)$$

Poisson Distribution

$$\mathcal{P}(\overline{\mathbf{N}}|\overline{\mathbf{m}},\mathcal{A}) = \prod_{k=1}^{M} \frac{D_k^{N_k} e^{-D_k}}{N_k!} \qquad D_k = f(x_k, \overline{\mathbf{m}})$$

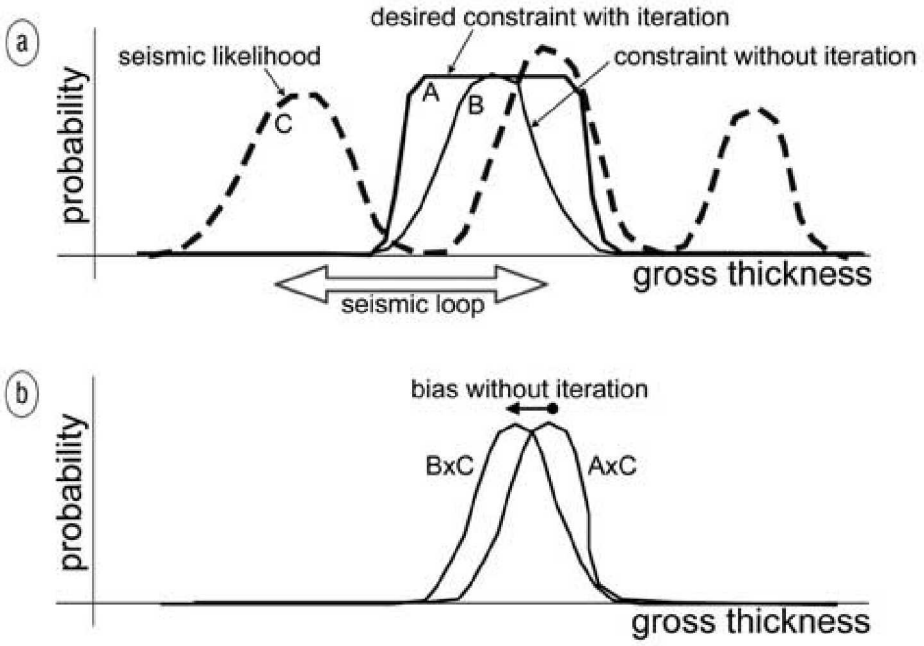
What about the priors?? How do we quantify our prior knowledge (or ignorance)?

There is an overwhelming amount of information out there about choosing priors

- MaxEnt methods for maximizing the ignorance of priors
- Uniform priors are often used as they (not always!) maximize ignorance

Remember: Its OK to use Gaussian priors

- True, they can bias your answer
- But, they are mathematically simple (particularly for the *linear* posterior approximation)
- You can truncate them to enforce hard constraints (e.g. density can't be negative)
- There are methods to de-bias the solution in the event that the prior is more informative than the data



Glinsky et al., **Bayesian inversion whispers,** THE LEADING EDGE MAY 2008

Outline

- Bayes' Theorem and its implications
- Estimating physical parameters from data
  - Most likely values and confidence intervals
  - Correlations
  - How valuable is my data?
  - Is my model good enough?
- Implementation
  - Packages
  - Using ML/DL to speed up

A *linear* approximation gives us deep insight into the problem at hand

In this approximation, finding the most likely value is a minimization problem

The full distribution can be approximated as a taylor series expansion

$$L \equiv \log_e(\mathcal{P}(\overline{\mathbf{m}}|\overline{\mathbf{x}}, A))$$

$$\nabla L\big|_{\overline{\mathbf{m}}_\mathbf{o}} = 0$$

Linear because we retain only the first non-zero term in the expansion

$$L \approx L(\overline{\mathbf{m}}_\mathbf{o}) + \frac{\mathbf{1}}{\mathbf{2}} \sum_{\mathbf{i=1}}^{\mathbf{M}} \sum_{\mathbf{j=1}}^{\mathbf{M}} \frac{\partial^\mathbf{2}\mathbf{L}}{\partial \mathbf{m_i} \partial \mathbf{m_j}}\bigg|_{\overline{\mathbf{m}}_\mathbf{o}} (\mathbf{m_i} - \mathbf{m_{o,i}})(\mathbf{m_j} - \mathbf{m_{o,j}}) + \ldots$$

Rewritten in matrix notation, shows that the posterior is approximated as a gaussian near the maximum

$$\mathcal{P}(\overline{\mathbf{m}}|\overline{\mathbf{x}}, \mathbf{A}) \propto \exp\left(\frac{\mathbf{1}}{\mathbf{2}}(\overline{\mathbf{m}} - \overline{\mathbf{m}}_\mathbf{o})^\mathbf{T} \nabla\nabla\mathbf{L}(\overline{\mathbf{m}}_\mathbf{o})(\overline{\mathbf{m}} - \overline{\mathbf{m}}_\mathbf{o})\right)$$

The covariance matrix is defined by the hessian of L at the optimum

Useful solutions can be found by deterministic optimization routines to find the (e.g. Levenberg-Marquardt)

Calculate the hessian matrix at optimum point

This gives the full covariance matrix defining a multivariate gaussian distribution

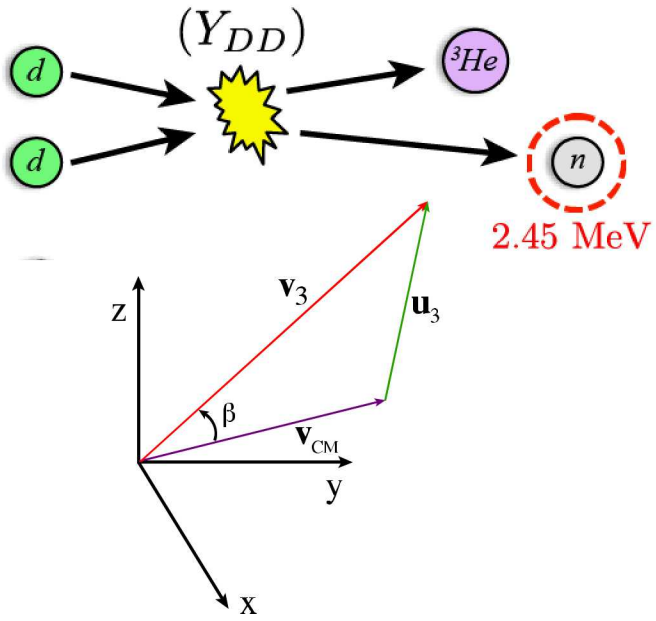
This distribution can be used to

- Calculate the C.I.'s on the model parameters
- Sample the posterior diagnostic "fits" with confidence intervals
- OR to define the initial step size for an MCMC sampler to map the nonlinear posterior

$$\sigma^2_{i,j} = -\left[\nabla\nabla L(\mathbf{\overline{m}_o})\right]^{-1}_{\mathbf{i,j}}$$

Its all in the gradients!

A Simple, but relevant case: Neutron Time of Flight in ICF



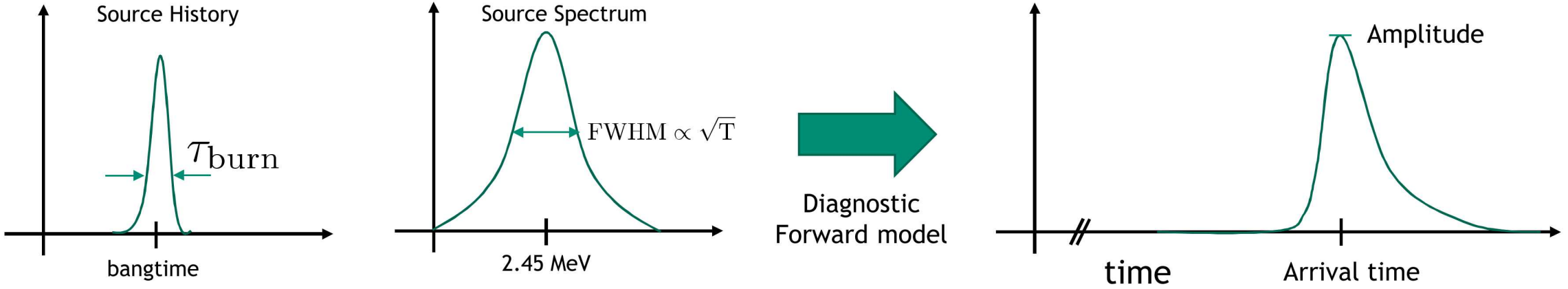
Neutrons carry useful information about the temperature of the plasma producing them

Neutrons are born in the CM frame with 2.45 MeV

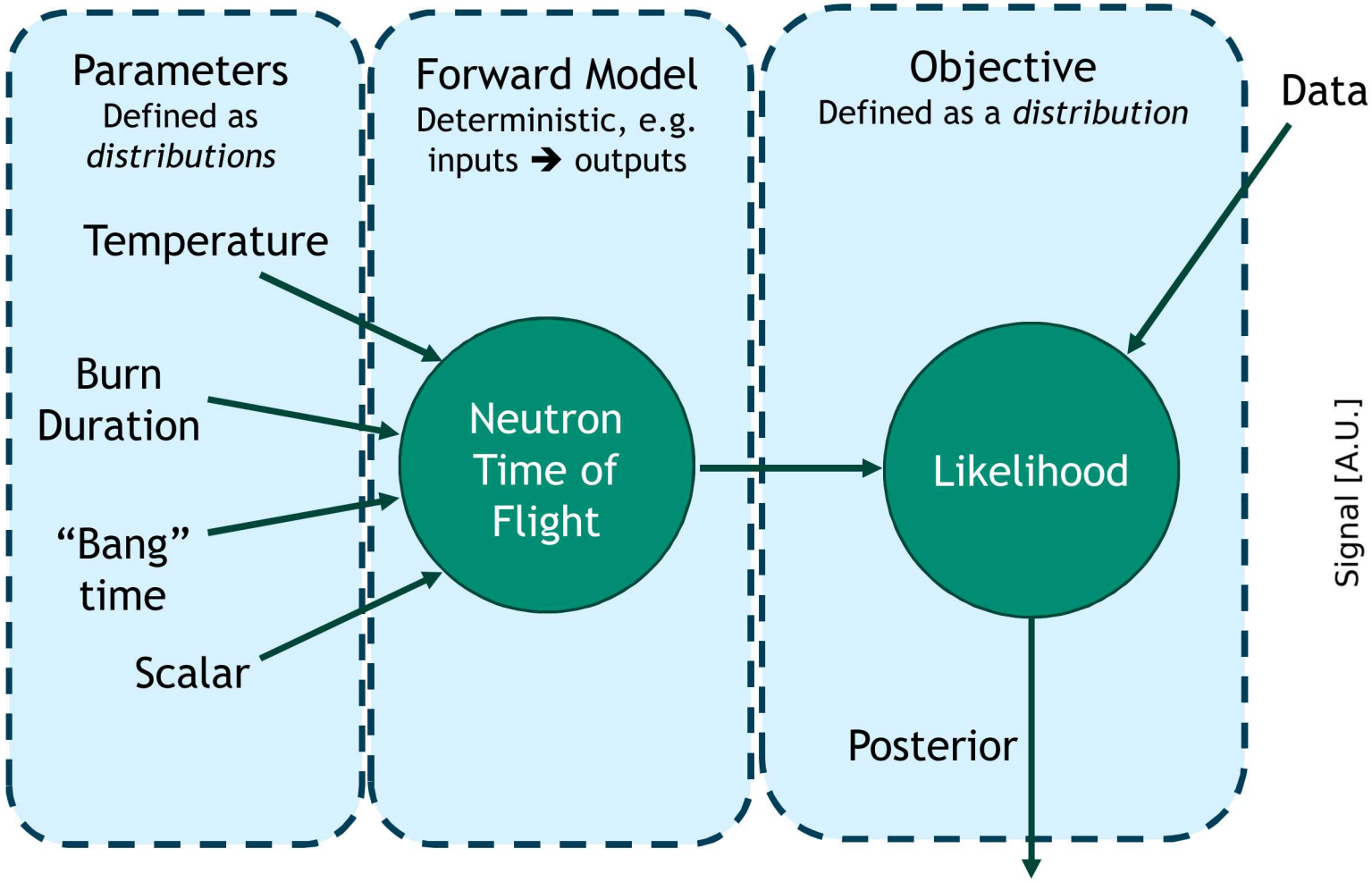
In the lab frame, they have a spread in velocities due to the Maxwellian distribution of reactant velocities

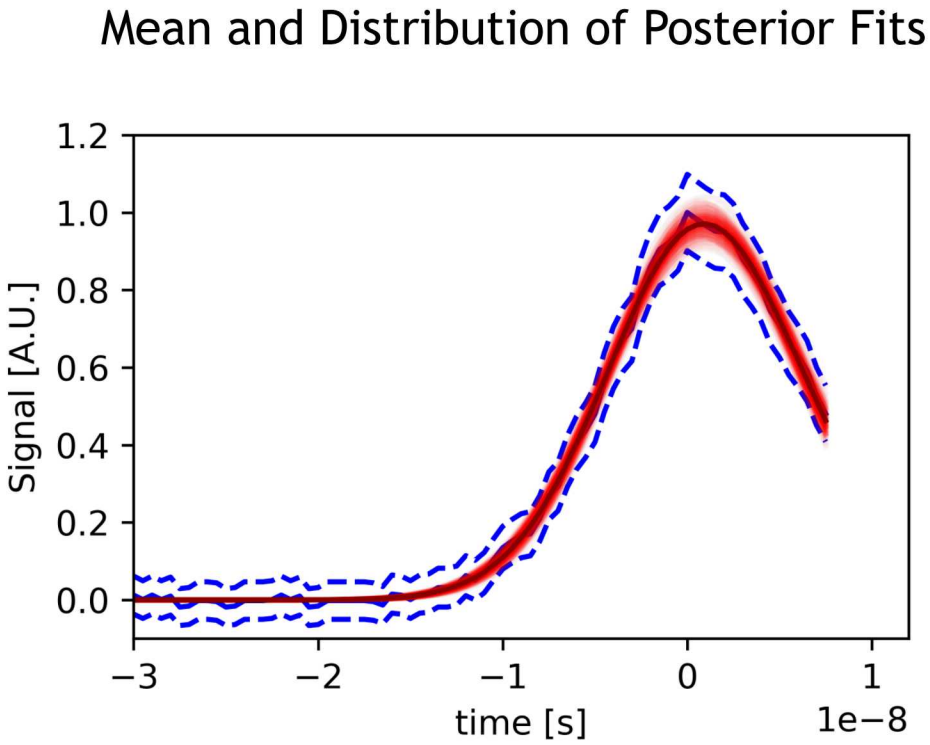
When the burn duration is short this spread in velocities is detected as a spread in arrival time at a detector (nTOF)

We would like to estimate the temperature of the plasma that produced the neutrons with confidence intervals in the presence of additional nuisance parameters



Using Bayes' theorem to estimate parameters of our model

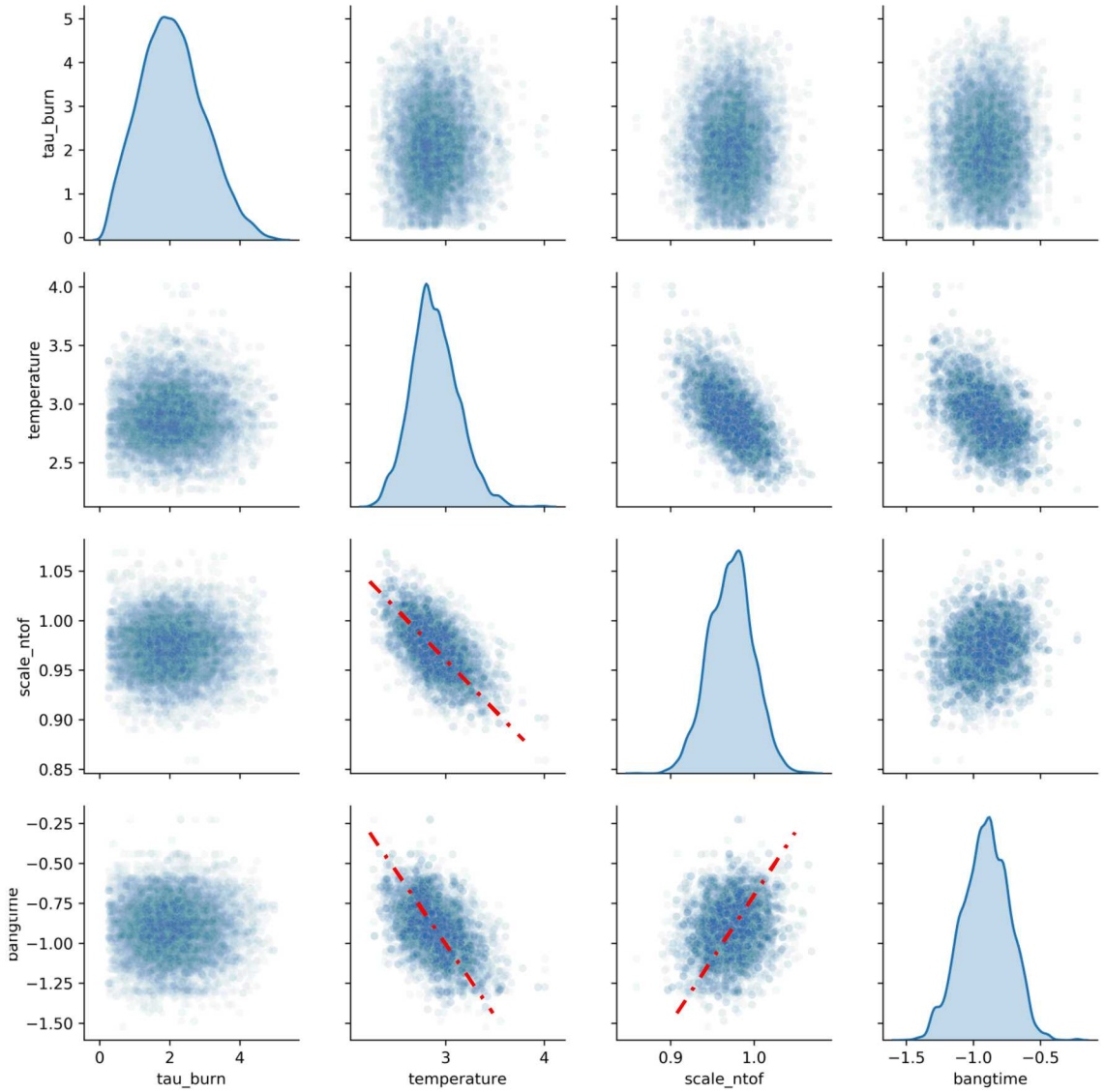




$$\mathcal{P}(\overline{\mathbf{m}}|\overline{\mathbf{x}},\mathbf{A}) \propto \prod_{\mathbf{i}=1}^{\mathbf{N}} \exp\left(-\frac{(\mathcal{F}(\overline{\mathbf{m}})-\mathbf{x_i})^2}{2\sigma_\mathbf{i}^2}\right)\mathcal{N}(\mu_\mathbf{T},\sigma_\mathbf{T})\mathcal{N}(\mu_\tau,\sigma_\tau)\mathcal{N}(\mu_{\mathbf{t_o}},\sigma_{\mathbf{t_o}})\mathcal{N}(\mu_\mathbf{C},\sigma_\mathbf{C})$$

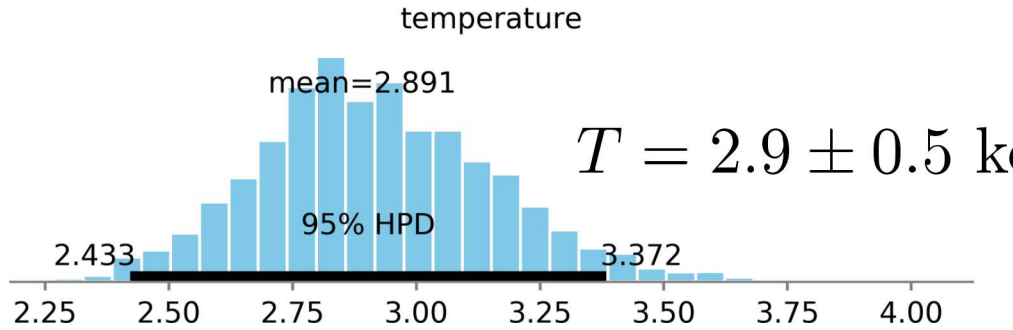
Confidence intervals and Marginal Distributions

The full Posterior for each of the parameters



Marginal posterior distribution for the temperature

$$\mathcal{P}(T|\overline{\mathbf{x}},\mathbf{A}) = \int_{\tau} \mathbf{d}\tau \int_{\mathbf{C}} \mathbf{dC} \int_{\mathbf{t_o}} \mathbf{dt_o} \mathcal{P}(\overline{\mathbf{m}}|\overline{\mathbf{x}},\mathbf{A})$$

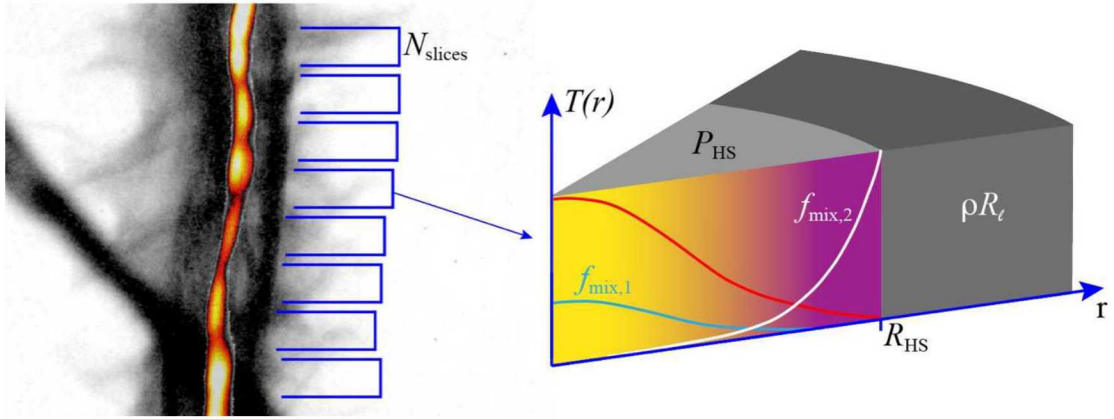


$$T = 2.9 \pm 0.5 \text{ keV}$$

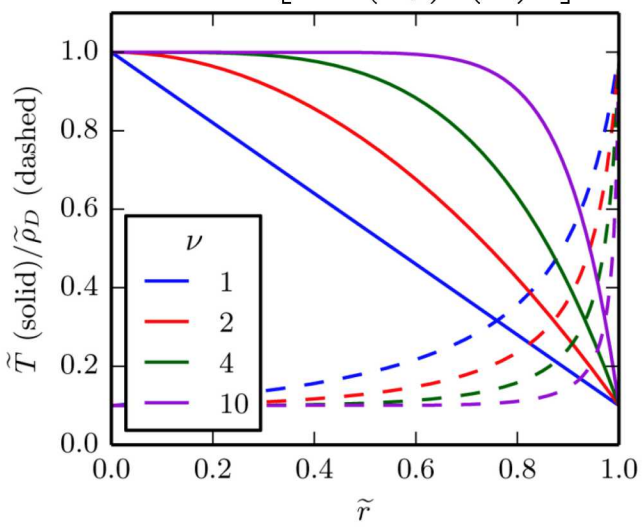
Simply integrate the posterior over the nuisance parameters

Correlations suggest that better knowledge of our instrument would improve our inference of the temperature

Lets examine a more complex model



$$T(r) = T_c\left[1 - \left(\frac{T_w}{T_c}\right)\left(\frac{r}{R}\right)^{\nu}\right]$$



Model Parameters

$$\{T_{\rm i}\} = \{T_{\rm e}\}$$
$$\{\rho R_\ell\}$$
$$\{P_{HS}\}$$
$$\{f_{\rm mix}\}$$
$$\{Z_{\rm mix}\}$$
$$\{R_{\rm HS}\}$$

Assumptions:

- Each slice has its own independent parameters characterizing a static, isobaric hot spot surrounded by a liner
- Ideal gas EOS: $P_{\rm HS} = (1 + \langle Z\rangle) n_{\rm i} k_B T$
- All elements have same burn duration
- Electron and ion temperatures are equal
- X-ray emission is dominated by continuum (BF & FF)

X-ray Emission:

$$\epsilon_\nu = A_{f-f} e^{-\rho R_\ell \kappa_\nu} \tau_{\rm b} P_{\rm HS}^2 \frac{g_{\rm FF}\langle Z\rangle}{(1+\langle Z\rangle)^2} \sum_i f_i \tilde{j}_i \frac{e^{-h\nu/T}}{T^{5/2}}$$

$$\tilde{j}_i \equiv \frac{j_i}{j_{\rm D}} = Z_i^2 + \frac{A_{f-b}}{A_{f-f}} \frac{Z_i^4}{T} e^{RyZ_i^2/T}$$

Neutron Emission:

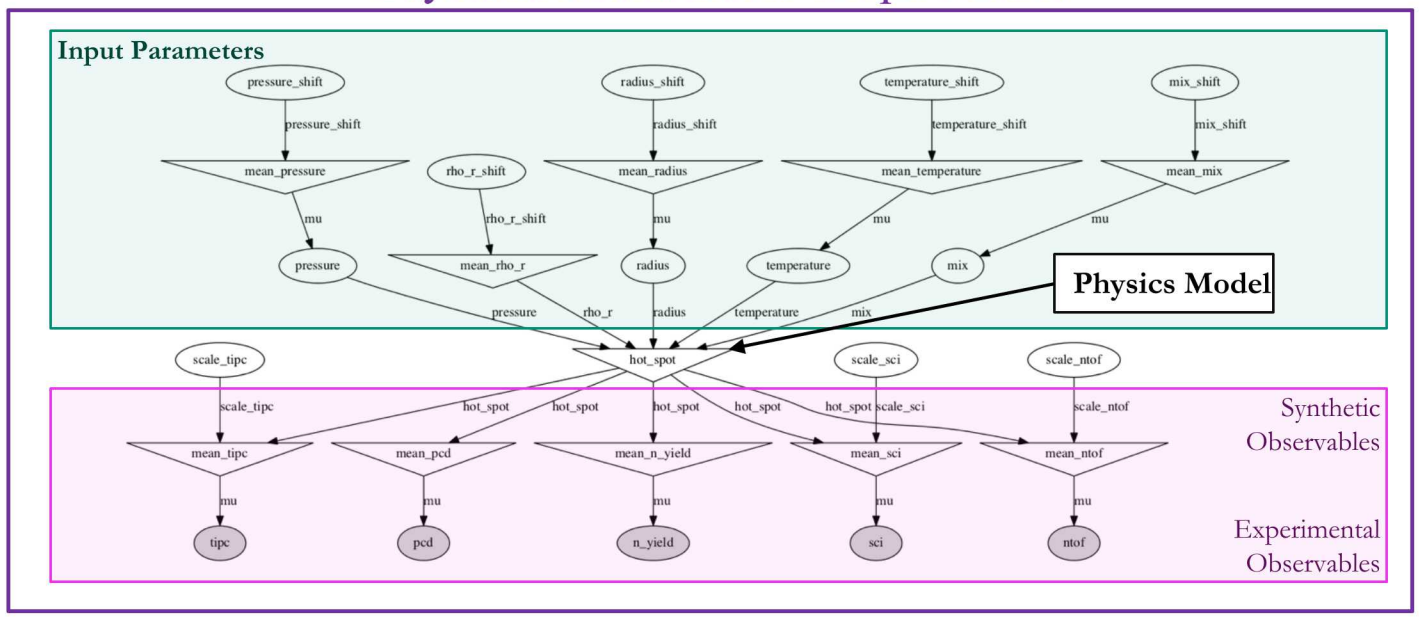
$$\epsilon_E = \frac{P_{\rm HS}^2 \tau_{\rm b}}{1+\delta_{1,2}} \frac{f_1 f_2 \langle\sigma v\rangle}{(1+\langle Z\rangle)^2 T_{\rm i}^2} I_o(E)$$

$$^*I_o(E) = e^{\frac{-2\bar{E}}{\sigma^2}(\sqrt{E}-\sqrt{\bar{E}})^2}$$

*Ballabio et al., NUCLEAR FUSION, Vol. 38, No. 11 (1998)

Analysis is performed using Bayesian Parameter estimation to determine the most likely hotspot parameters

Bayesian Hierarchical Graph Model



- Bayesian parameter estimation is a well-established technique used in a variety of fields*
- Analysis can be used to infer most likely parameters, correlations between model parameters and/or data
- Can compute value of information to determine which data constrain which parameters and how well

- Prior distribution is sampled
- Levenberg-Marquardt optimization (with optional multiple starts) used to determine the MAP solution
  - By assuming a Gaussian form this solution uniquely determines the posterior
- MCMC sampling used to refine the solution and determine if posteriors show any non-linear behavior
- Posterior distribution is sampled to form the posterior diagnostic and model parameter statistics (e.g. mean, confidence interval, etc.)

*U. Von Toussaint, Rev. Mod. Phys. Vol. 83 (2011)

Bayesian Parameter estimation is an iterative process that updates our assumptions based on observables

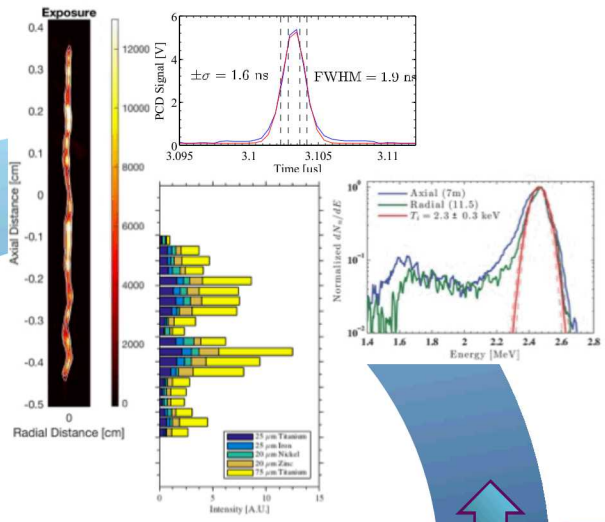
Bayes' Theorem

$$\mathcal{P}(\overline{\mathbf{m}}|\overline{\mathbf{d}},A)=\frac{\mathcal{P}(\overline{\mathbf{d}}|\overline{\mathbf{m}},A)\mathcal{P}(\overline{\mathbf{m}}|A)}{\mathcal{P}(\overline{\mathbf{d}}|A)}$$

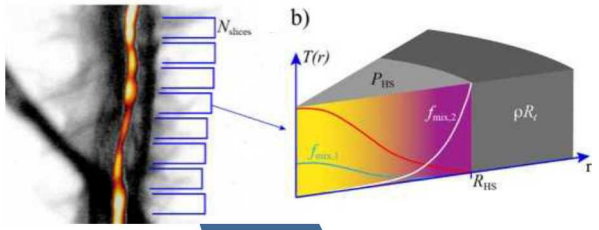
Likelihood

$$\mathcal{P}(\overline{\mathbf{x}}|\overline{\mathbf{m}},\mathcal{A})\propto\prod_{i=1}^{N}\exp\left(-\frac{(\mathcal{F}_i(\overline{\mathbf{m}})-x_i)^2}{2\sigma_i^2}\right)$$

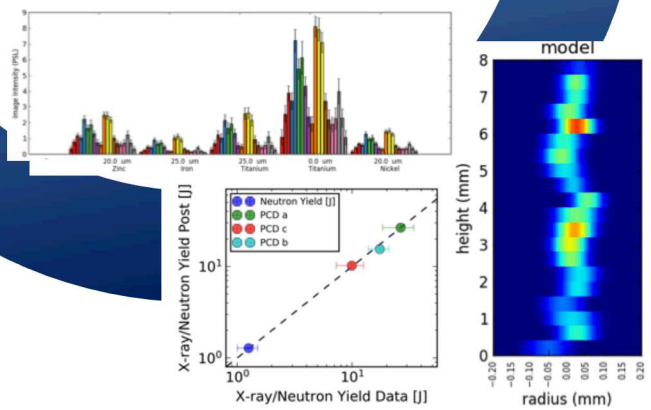
Experimental Data



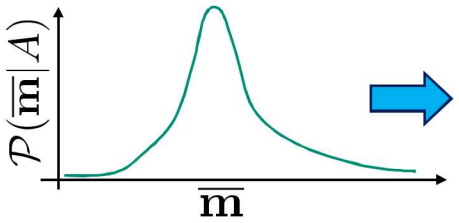
Proposed Stagnation Conditions



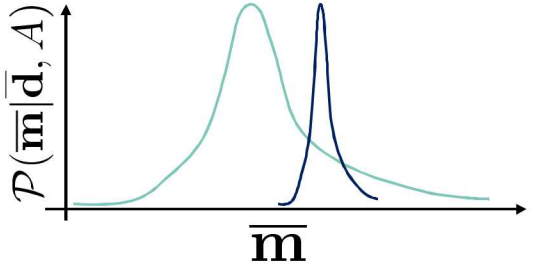
Synthetic Data



Prior Distribution



Posterior Distribution



Model Parameters

$$\overline{\mathbf{m}}=\begin{cases}P_{\mathrm{HS}}\\T\\f_{\mathrm{mix}}\\R_{\mathrm{HS}}\\\rho R_\ell\end{cases}$$

Outputs/Benefits:

- most likely parameter values
- confidence intervals
- correlations
- Value of information

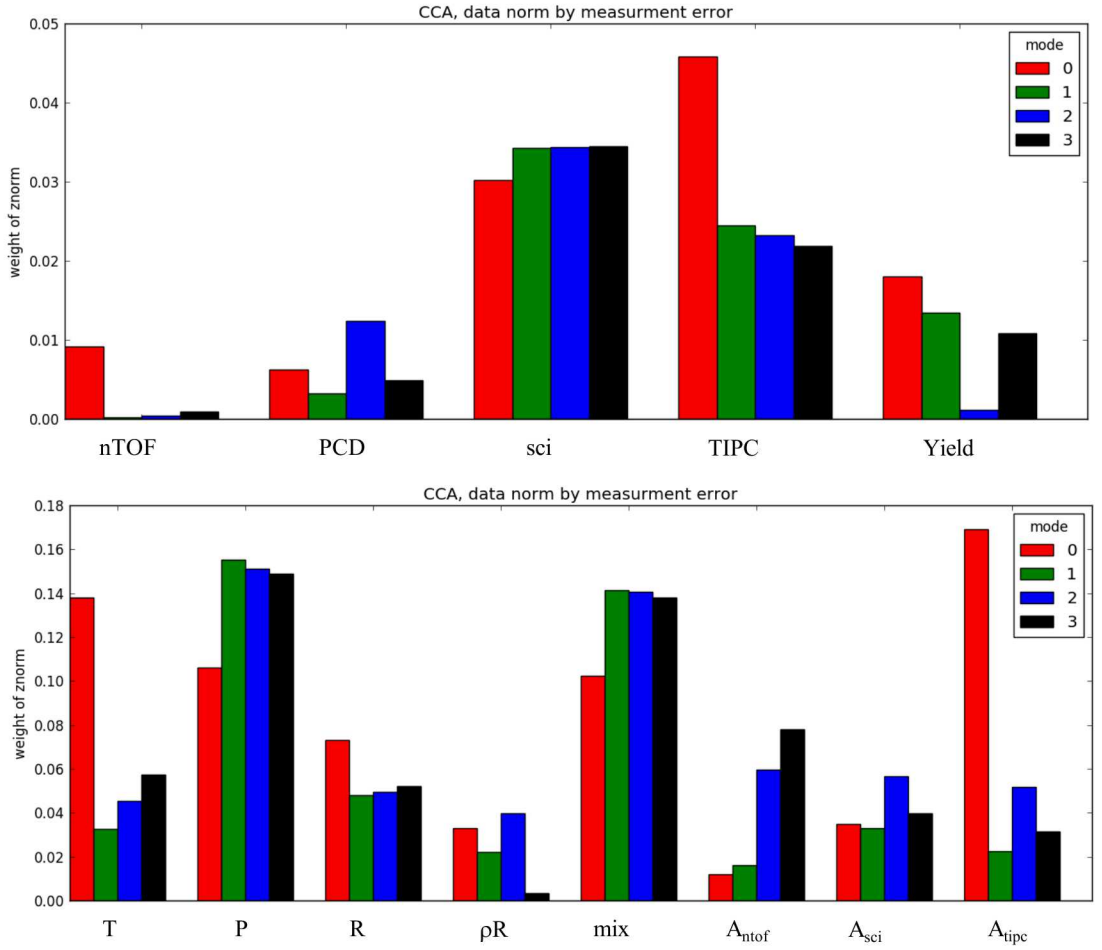
We can use the posterior to determine which data are constraining which model parameters

Cross-variance between model and data

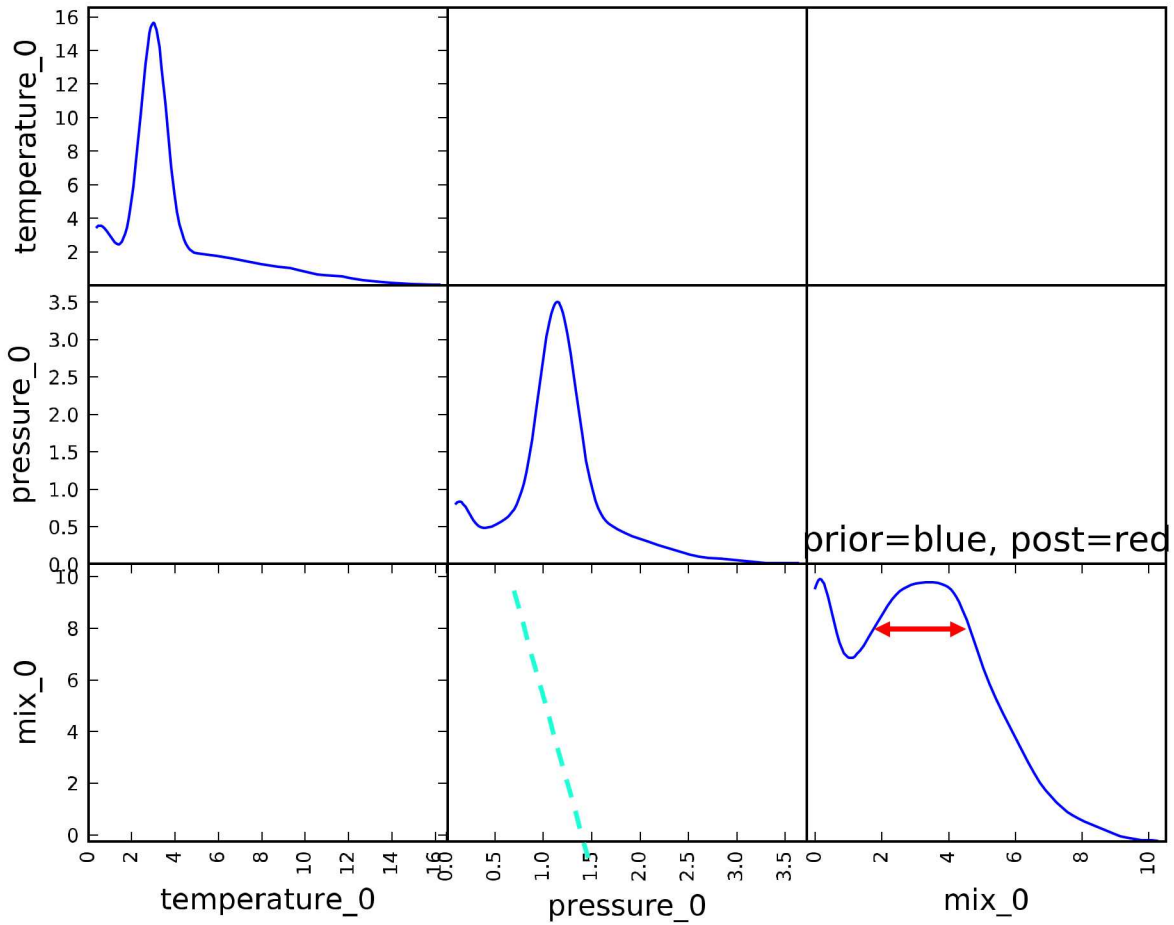
$$C(m,d) \overset{\text{SVD}}{=} U_m \Sigma^{-1}_{md} V^T_d$$

Decomposing the matrix gives the weight of each data and model parameter in the strongest modes

This is a way of quantifying the value of information (VOI)



The posterior pdf's reveal correlations between the model parameters



- Mix is relatively poorly determined
- Significant correlation with the Pressure

X-ray Emission:

$$\epsilon_\nu = A_{f-f} e^{-\rho R_\ell \kappa_\nu \tau_b} P_{\rm HS}^2 \frac{g_{\rm FF}\langle Z\rangle}{(1+\langle Z\rangle)^2} \sum_i f_i j_i \frac{e^{-h\nu/T}}{T^{5/2}}$$

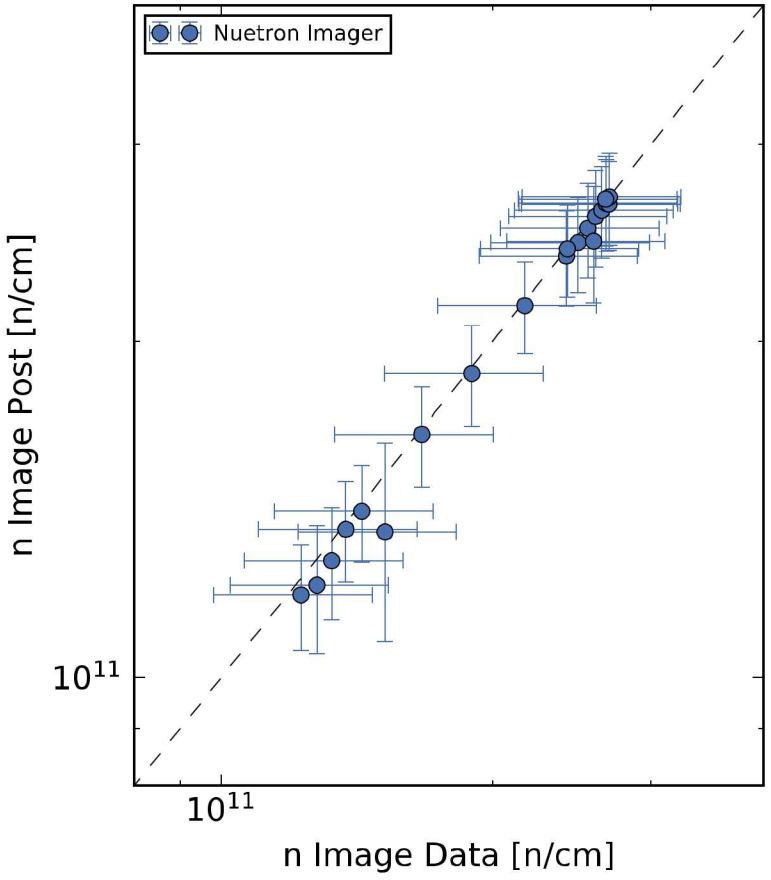
Neutron Emission:

$$\epsilon_E = \frac{P_{\rm HS}^2 \tau_b}{1+\delta_{1,2}} \frac{f_1 f_2 \langle \sigma v\rangle}{(1+\langle Z\rangle)^2 T_i^2} I_o(E)$$

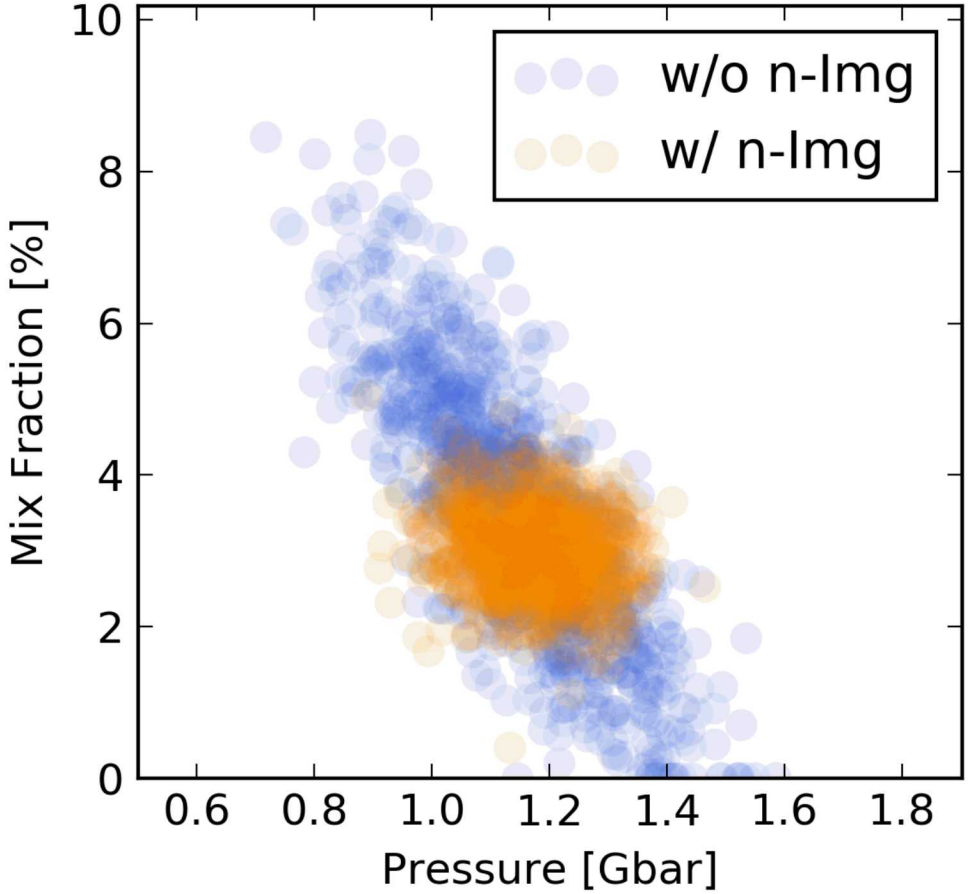
- Neutrons and x-rays have the same dependence on pressure, but not on mix
- We have local and global x-ray measurements, but only global neutron measurements…

Adding new information can affect the correlations in the posterior pdf's, improving our ability to determine certain quantities

Reconstructed 1D Neutron Image Data



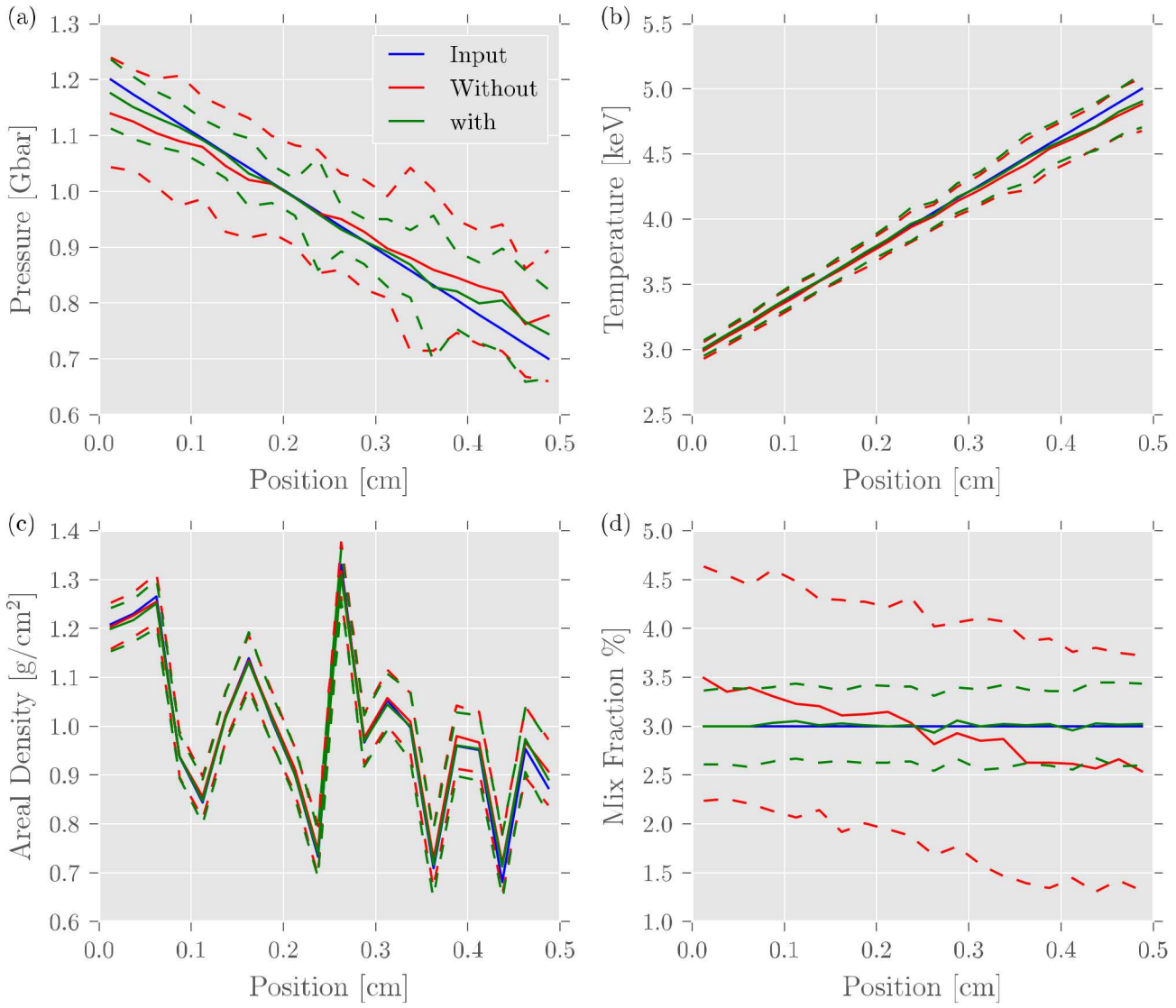
Posterior pdf



- Motivated by diagnostic developments and the previous observation, we implemented a simple 1D neutron image model
- This approach can be formalized and extended to the design and optimization of diagnostics as well as experiments**

**U. Von Toussaint, Rev. Mod. Phys. Vol. 83 (2011)

Adding new information can affect the correlations in the posterior pdf's, improving our ability to determine certain quantities



- As expected we find a dramatic improvement in accuracy and confidence of the mix prediction
- There is an additional improvement in the accuracy of the pressure inference

Finally, we can seek an answer to the question: Is my model good enough?

The traditional way: Model selection

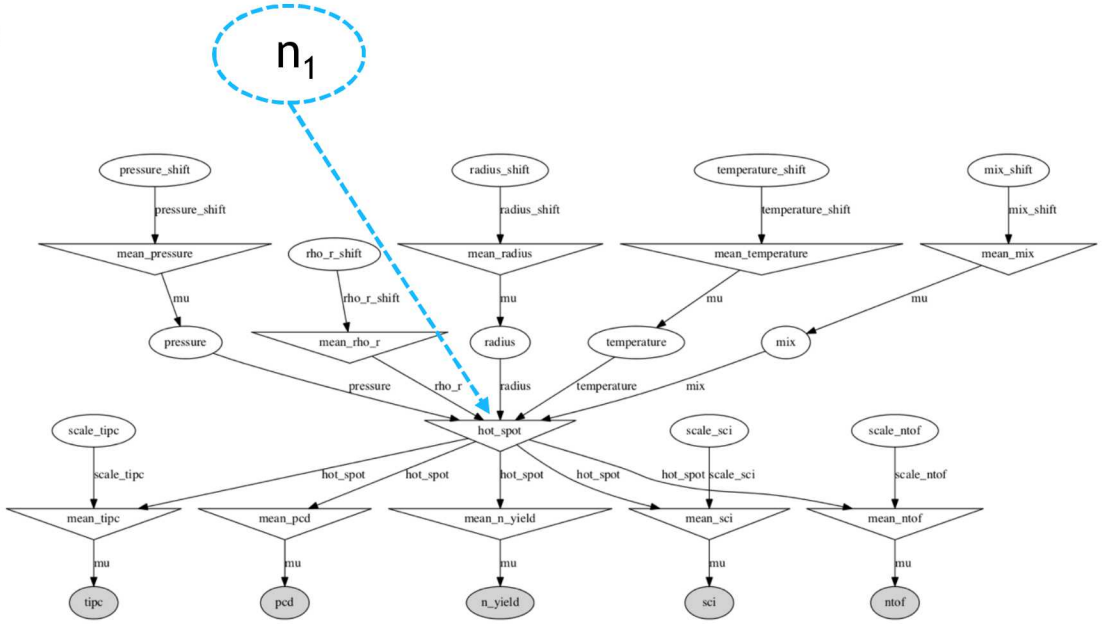
$$Z = \mathcal{P}(\overline{\mathbf{x}}|\mathcal{A}) = \int \mathrm{dm}_1 \ldots \mathrm{dm}_N \mathcal{P}(\overline{\mathbf{x}}|\overline{\mathbf{m}}, \mathcal{A}) \mathcal{P}(\overline{\mathbf{m}}|\mathcal{A})$$

$$O_{ij} = \frac{\mathcal{P}(\mathcal{A}_i)}{\mathcal{P}(\mathcal{A}_j)} \times \frac{Z_i}{Z_j}$$

By calculating the *evidence* term, Z, we can see which model is better supported by the data

A new way: Causal Statistics

What is $P(n_1)$? What does n_1 mean?



This new branch of statistics seeks to evaluate the probability of existence of a new node in the graph network and what this node means

Outline

- Bayes' Theorem and its implications
- Estimating physical parameters from data
  - Most likely values and confidence intervals
  - Correlations
  - How valuable is my data?
  - Is my model good enough?
- Implementation
  - Packages
  - Using ML/DL to speed up

There are many tools out there that can be used to construct these models

Python has many packages available

pymc2, pymc3, (pymc4 under construction right now)

- Significant differences in API between pymc2 and 3
- Basically, they all allow you to specify distributions on your parameters, pass them to a model to perform operations on the parameters, and define statistics to compare model outputs to observations
- Defining these connections constructs a hierarchical graph model of your problem, the likelihood and posterior are defined by this graph, Not by you explicitly
- We use pymc2, but it supports python 2, not 3 so we need to migrate our tools
- We also had to build in additional functionality into pymc (LM optimization, post-processing tools, etc.)

Tensorflow probability has some functionality available

- pymc4 is being built using tf as its backend to enable better integration w/ modern machine learning tools

Other packages include emcee, stan, R has extensive tools available

DAKOTA has some functionality available and works well with HPC and large multiphysics codes

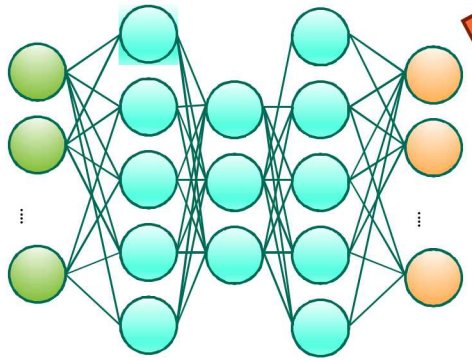
How do we leverage modern ML tools within the Bayesian framework

The most straightforward way to do this is to replace the physics and/or diagnostic models with machine learned surrogates

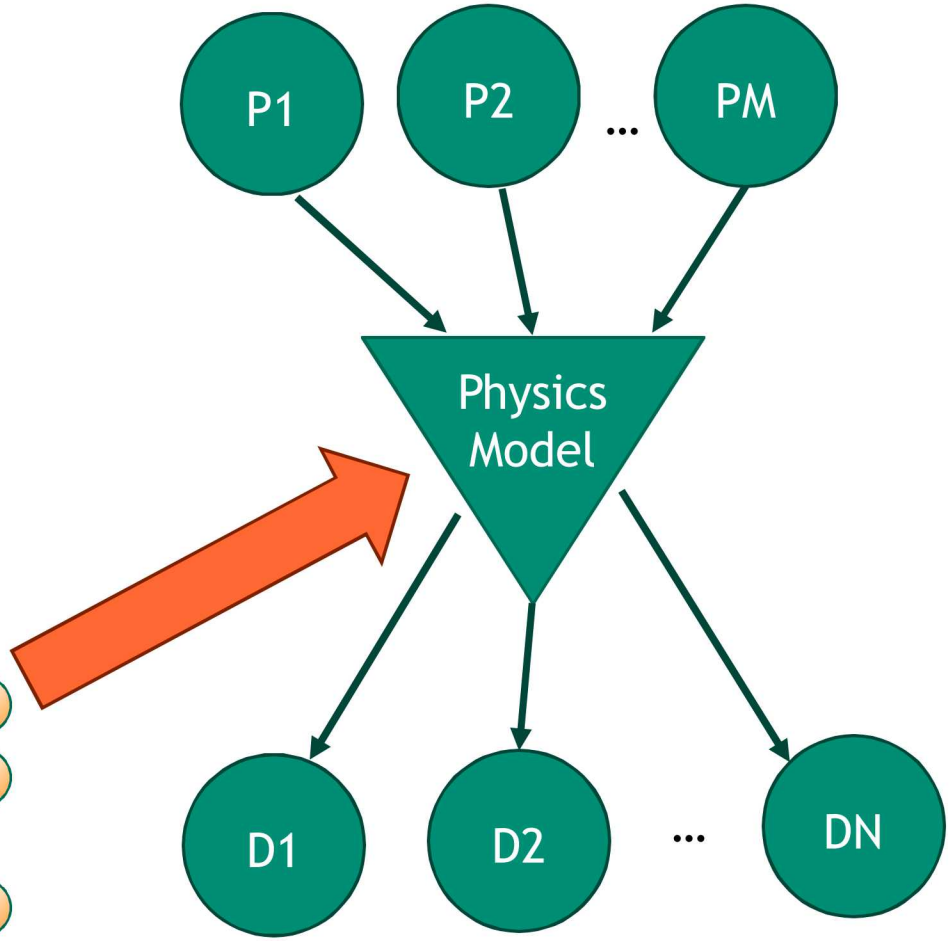
- Conceptually straightforward
- Black-box like behavior
- How do we propagate uncertainties through our new blackbox?

D. Clark

This is most similar to the approach Jim Gaffney uses



DNN surrogate of Multiphysics radhydro code



There are a lot of choices about how to implement surrogates

Abstraction from physics

Computation speed

Full end-to-end surrogate

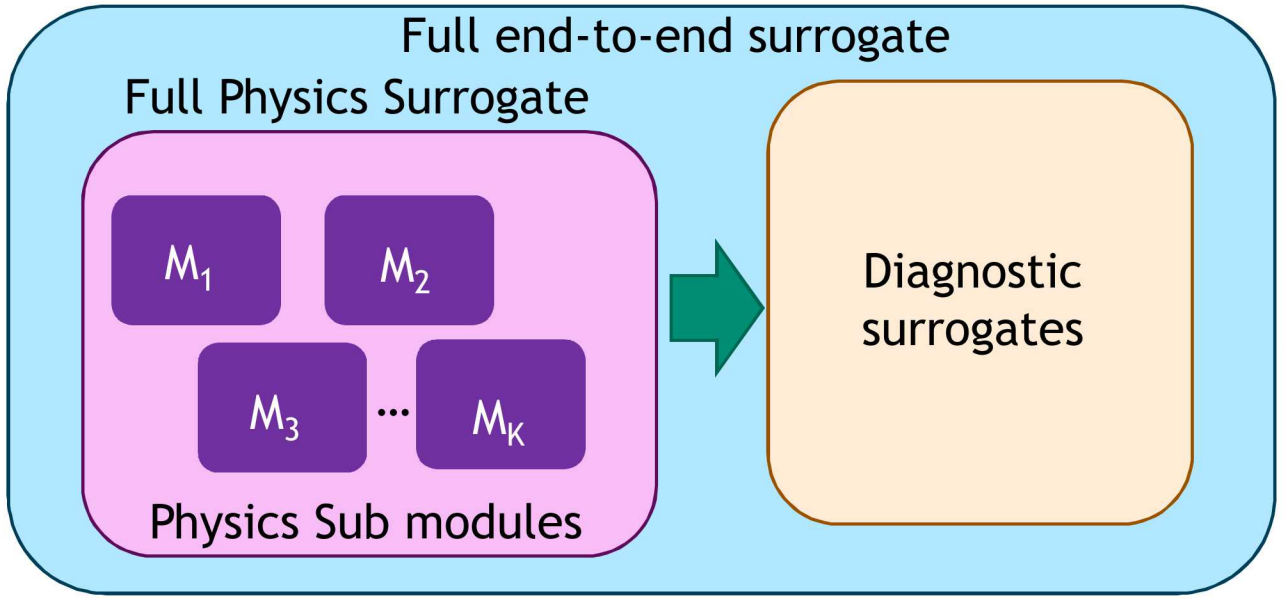
- Experimental inputs to diagnostic outputs, no intermediate steps
- High training cost, risk of extrapolation, diagnostics are directly linked to physics

Physics + Diagnostic Surrogates

- Separate surrogates for physics and diagnostics
- Requires a means to link the two, could be a learned latent space or full blown simulation output
- More flexibility with diagnostics

Surrogate sub-modules

- Surrogates operate in better "confined spaces"
- Have to integrate submodules in a stable and sensible wahy



How do we account for uncertainties in every step of the chain?
We want our surrogates to be fully probabilistic objects to enable fully "Bayesian" analysis