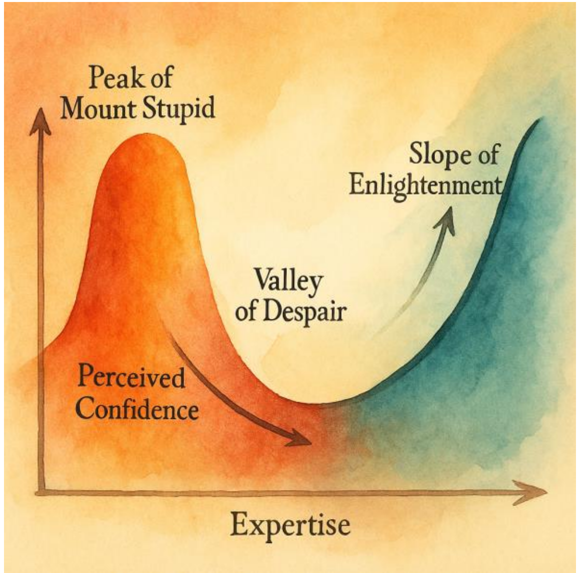
UQ Isn't Sorcery—It's a Toolkit: Understanding Methods and Pitfalls

Beth Lindquist (LANL), Justin Brown (SNL), Jason Bernstein (LLNL)

06/25/25

LA-UR-25-25421

It's easy to know just enough about UQ to be dangerous



ChatGPT artistic rendition of Dunning-Kruger

Many assumptions can be baked into UQ that practitioners need to understand

- Your model is a good approximation of reality
- Your uncertainty model matches the real uncertainty
- You know what kind of uncertainty you're quantifying
- Your data are representative and i.i.d.
- Priors (or sampling schemes) encode real beliefs
- You're interpreting the results correctly
- You can trust the computation

GARBAGE IN = GARBAGE OUT

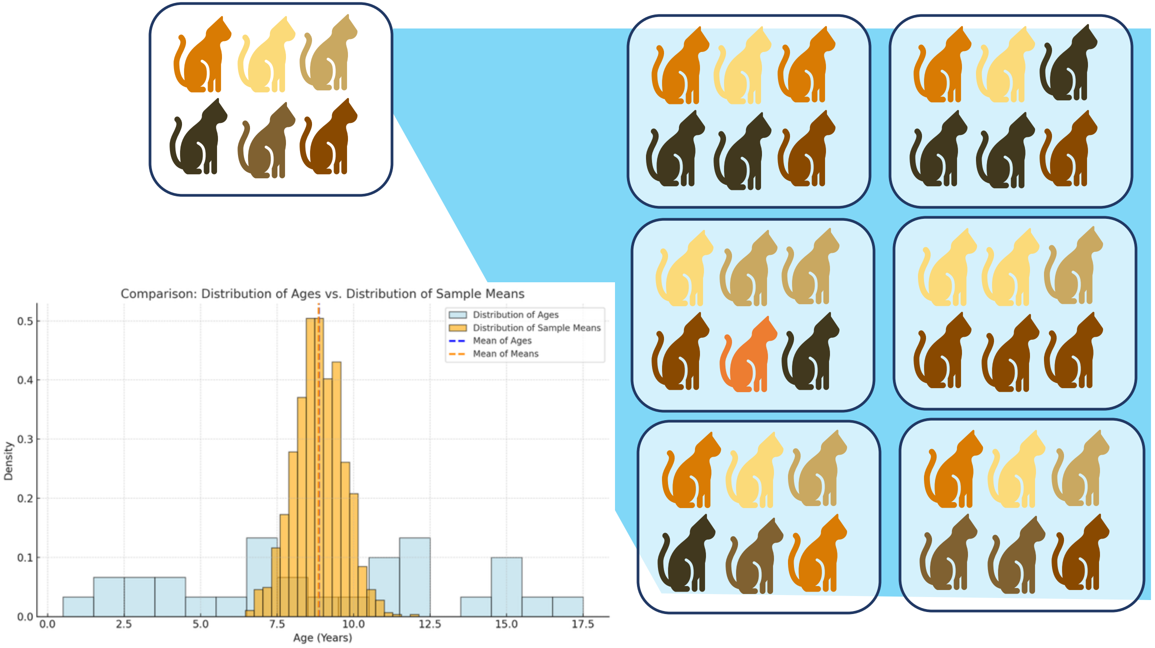
For the remainder of the slides, we will…

- Recapitulate the distinction between bootstrap and Bayesian approaches
- Explore case studies showing how different UQ and modeling choices can significantly affect the inferred parameter distributions
- Switch from talking about wizardry to cats for reasons that will not be explained (with a diversion into house plants)

Bootstrap resampling estimates the uncertainty of a statistic by repeatedly sampling with replacement from the original data

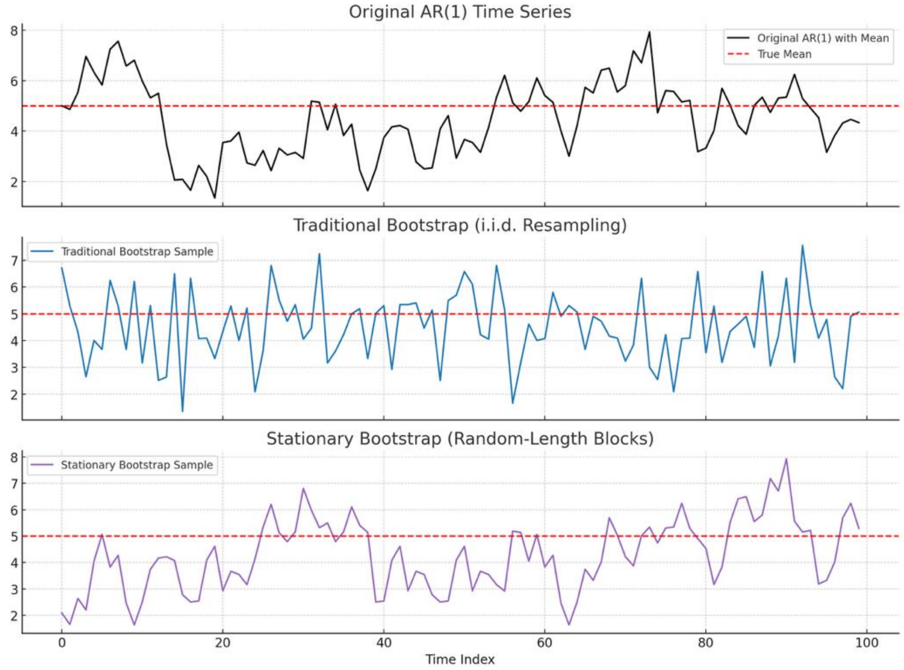
Original sample

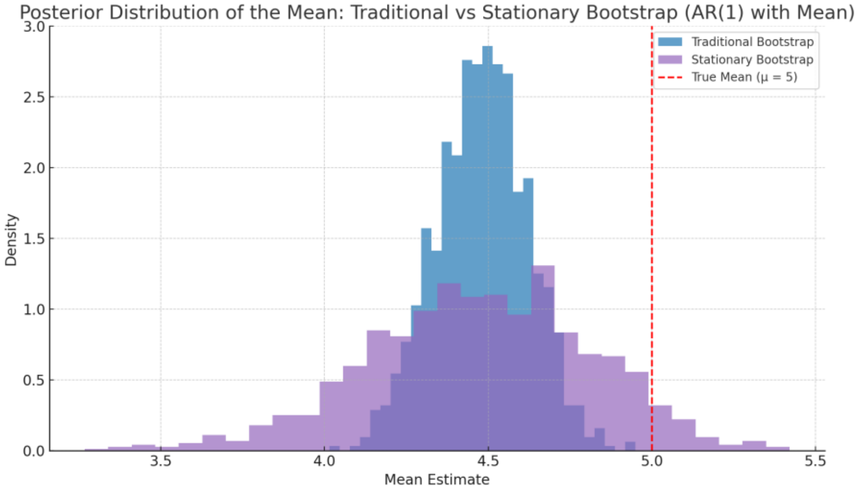
Re-sample with replacement



e.g., compute the mean age of the cats on Aoshima Island in Japan via a non-parametric “traditional” bootstrap

Traditional Bootstrap should not be applied to a single instance of time-dependent data (e.g., measurements of a fluctuating quantity)

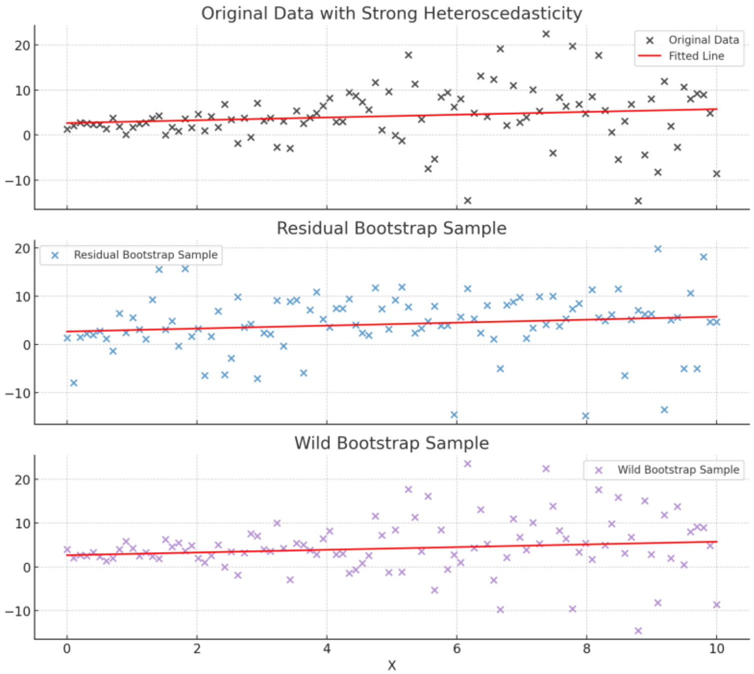


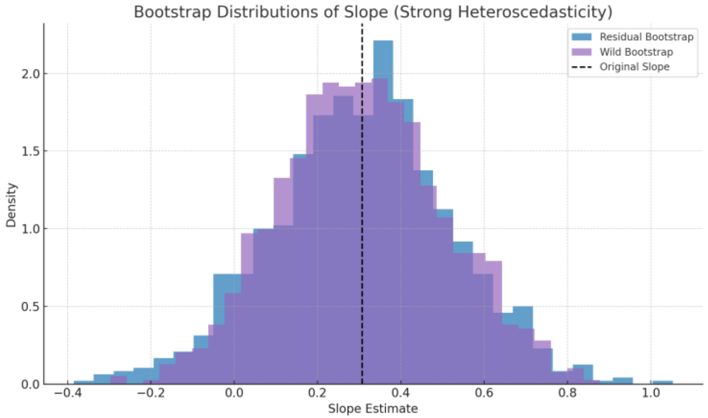


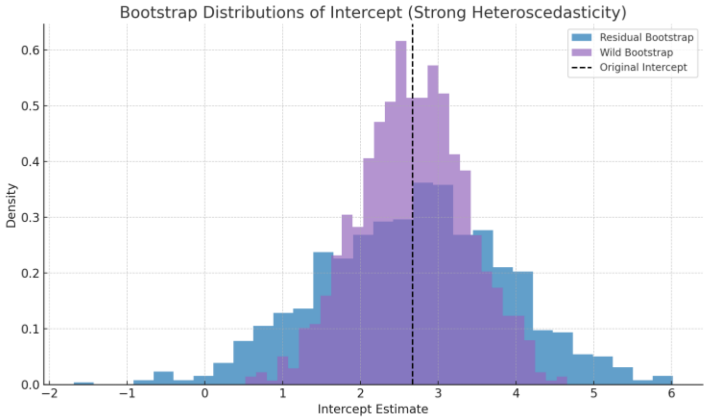
Traditional bootstrap is **over-confident** because it assumes that the data are independent. They are not, and so the effective sample size is smaller than the number of data points.

Los Alamos NATIONAL LABORATORY

Heteroscedasticity is a challenge to many flavors of bootstrap that assume i.i.d. data







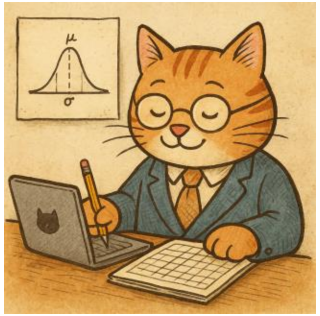
Residual bootstrap is **under-confident** because it assumes that the data are identically distributed. They are not: here, we have less uncertainty around the intercept.

Los Alamos NATIONAL LABORATORY

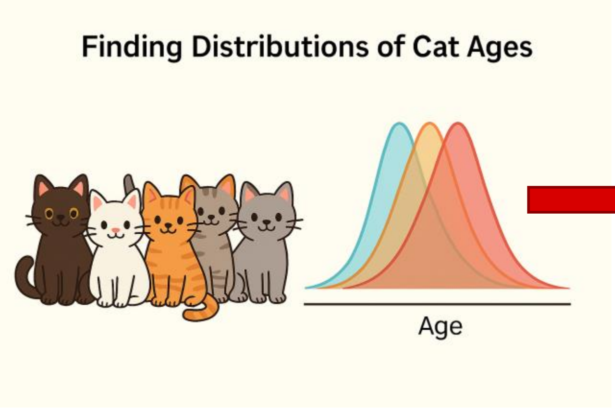
Many flavors of bootstrap: choose one that fits your problem

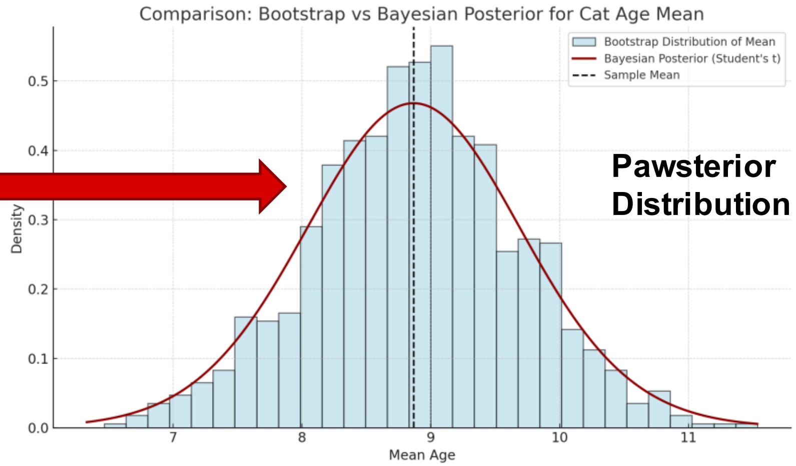
Category	Bootstrap Methods	When to Use	Key Assumptions / Notes
🧪 **Basic / General-Purpose**	- Traditional (Naive) - Stratified - Double Bootstrap - Smoothed	- i.i.d. data - Class imbalance - Small samples or noisy stats - Improved accuracy	- Data points are independent - Stratified preserves class balance - Double = more precise CIs - Smoothed helps for quantiles
🕰️ **Time Series / Dependent Data**	- Block Bootstrap - Moving Block - Circular Block - Stationary Bootstrap	- Stationary time series - Short-range dependencies - Periodic data - Need more natural blocks	- Use blocks to preserve autocorrelation - Stationary uses random-length blocks - Circular avoids edge artifacts
📊 **Regression-Focused**	- Residual - Wild - Pairs	- Linear or generalized regression - Heteroscedastic errors - Non-i.i.d. predictors	- Residual: assumes homoscedasticity - Wild: handles changing variance - Pairs: resamples (X, y) together
🧠 **Bayesian / Model-Based**	- Bayesian Bootstrap - Parametric Bootstrap	- Want posterior-like inference - Known model structure	- Bayesian: resample weights (Dirichlet) - Parametric: simulate from fitted model
🧱 **Structured / Hierarchical / Scalable**	- Cluster Bootstrap - Bag of Little Bootstraps (BLB)	- Clustered/grouped data - Massive datasets	- Cluster: resample entire groups - BLB: approximate bootstrap for large data

Bayesian uncertainty quantification combines prior beliefs with data to produce full probabilistic estimates of model parameters and predictions



Catistician





Pawsterior Distribution

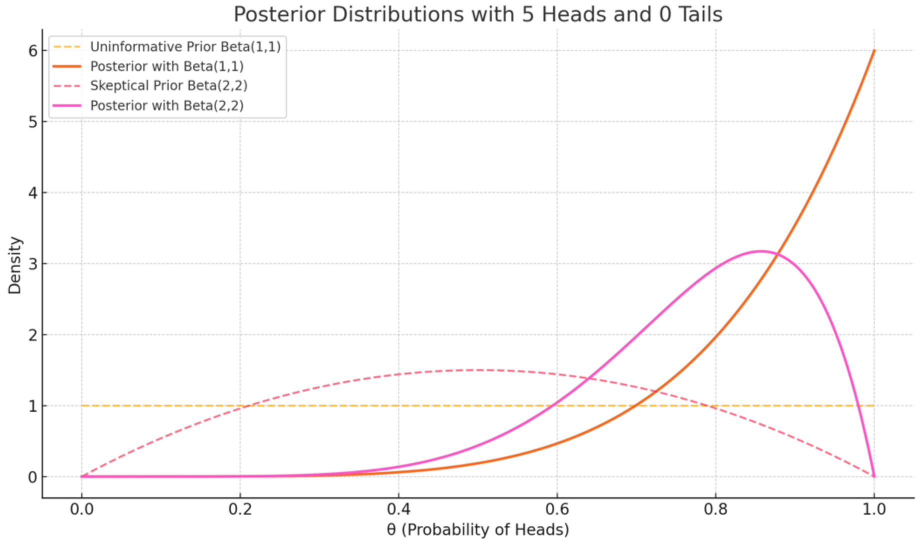
If we choose our prior smartly, we can solve for the distribution of cat mean age analytically:
Normal Inverse Gamma x Normal = Normal Inverse Gamma
Most real-world problems are not solvable analytically. In those cases, we use numerical sampling techniques such as MCMC.

Careful consideration of prior information is required for Bayesian approaches

- It's very popular to use uninformative priors. **Is this appropriate to your problem?**

- Consider a coin that you flip 5 times, and you get heads 5 times. What is the probability of flipping heads with this coin?

- In physics, we are often limited in terms of the amount and quality of experimental data, making appropriate use of prior information extra crucial.



No more cat puns… starting right meow

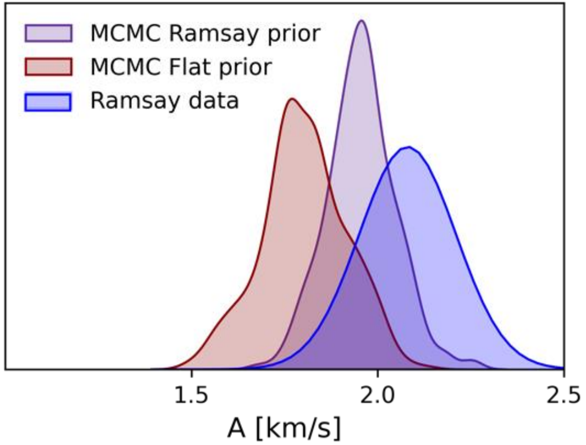
George Pawshington, the first U.S. purr-sident

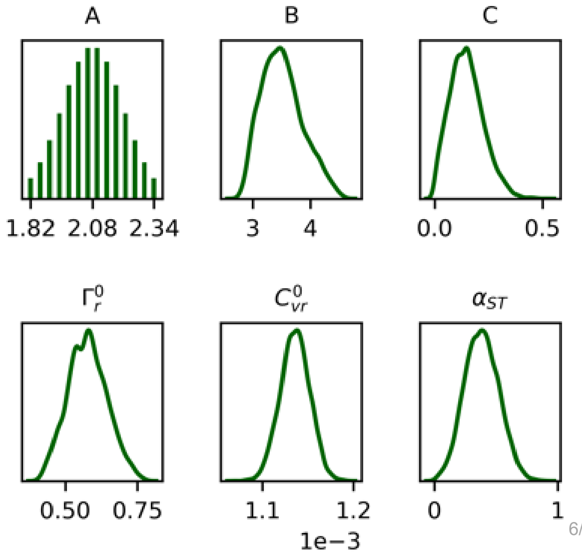
There are a variety of ways to integrate multiple data streams

- In physics, integrating multiple streams of data can allow for better inferences if conflicting data treated appropriately.

- Multiple data sources can be included via:
  - Multiple sources of information in the likelihood
  - Appropriate construction of the prior
  - Modular Bayesian approaches
  - Hierarchical Bayesian methods

- Example to the right is an example of a modular Bayesian strategy



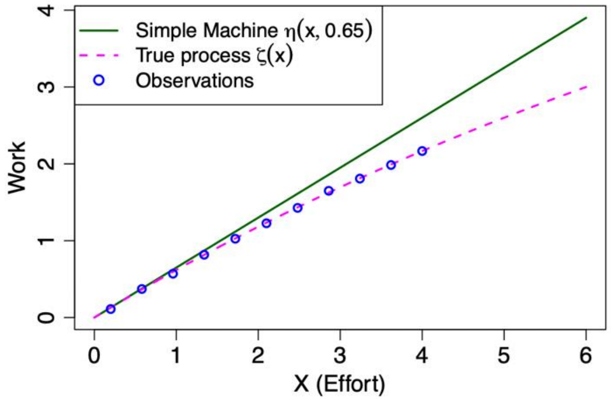


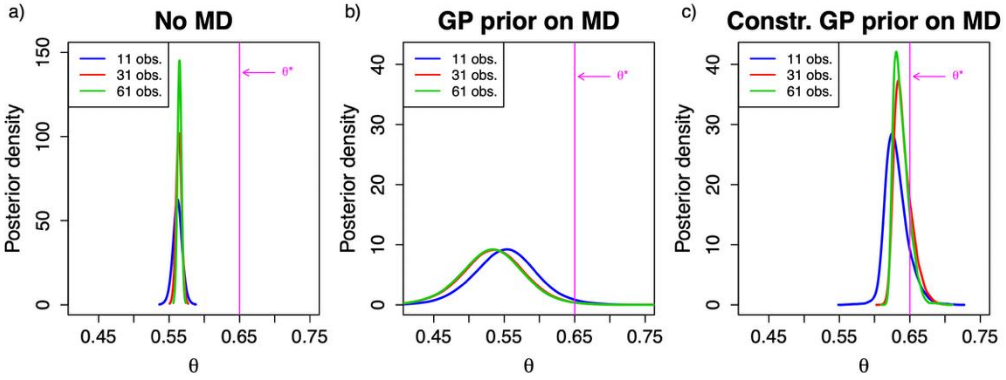
[Ramsay] https://www.osti.gov/servlets/purl/4618401
[UQ] *AIP Conf. Proc.* 3066, 480010 (2024)

Los Alamos NATIONAL LABORATORY

Model discrepancy can bias physics parameter inference (and prior information is still important!)

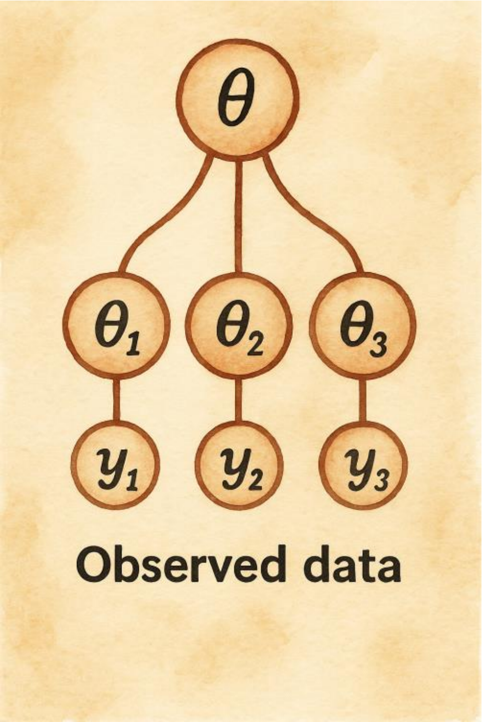
- Consider a machine that becomes less efficient with increased effort due to friction, heating, etc.
- If we naively choose a model where output and effort are linearly related, then we have introduced model discrepancy.
- If we incorporate constraints about the nature of the discrepancy (it goes to zero at zero effort and is always negative), we see better agreement in the physics parameter inference.

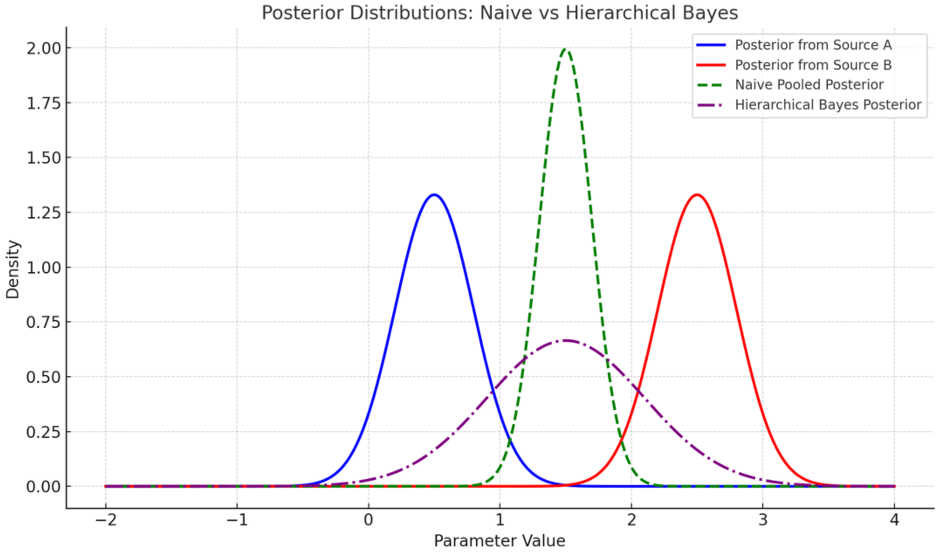




Inverse Problems **30** 114007 (2014)

Systematic biases in data have the same effect as model discrepancy: one option for both is Hierarchical Bayes





A Hierarchical Bayesian approach might assume that distinct distributions over the physics parameters can be inferred for each piece of data, but that these separate distributions are related via some parent distribution.

Go forth and UQ responsibly!

- Uncertainty quantification is not just about adding error bars — it's about asking what you don't know and why.
- Every UQ method makes assumptions — know them, question them, and check that they match your data and model.
- UQ is a toolbox, not a hammer or a magic wand

It's dangerous to go alone… take this!