Mixed aleatoric-epistemic uncertainty for jet engine applications

11th Dresden Probabilistic Workshop

10/10/2019 - Giulia Antinori

Proprietary Notice

This document contains proprietary information of the MTU Aero Engines AG group companies. The document and its contents shall not be copied or disclosed to any third party or used for any purpose other than that for which it is provided, without the prior written agreement of MTU Aero Engines AG.

The way uncertainty is modeled affects the robustness evaluation

In the engineering uncertainty (aleatoric or epistemic) means

- Incomplete data
- Experience based information
- Model error and imprecision
- Random variability

Assumption

How would the robustness evaluation change if the lack of knowledge would be modeled?

Aleatoric versus epistemic uncertainty

Aleatoric Uncertainty

- **Randomness** related to the physical quantities or the system
- **Irreducible**
- **Data and probabilistic information** available

Epistemic Uncertainty

- **Incomplete knowledge** and/or lack of useful data
- **Reducible** (e.g. by gathering additional data, by refining models)
- **Interval information**, max-min, imprecise models

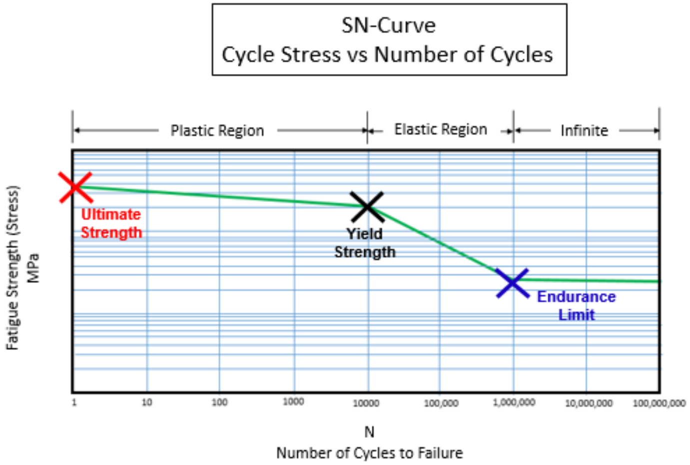
Different methods/approaches for different uncertainty types

Aleatoric	Epistemic	Mixed
Monte Carlo	Interval Analysis	Evidence Theory
Stochastic Collocation Methods	Bayesian Second Order Probability	Double Loop Approach

Example 1: from deterministic life models to possibilistic approach

Pure epistemic uncertainty quantification

- **Life tests** play a central role in the design of mechanical systems subjected to **cyclic loadings**
- **Experimental setup** is characterized by
  - ✓ Data becoming gradually available
  - ✓ **High variability** in the data



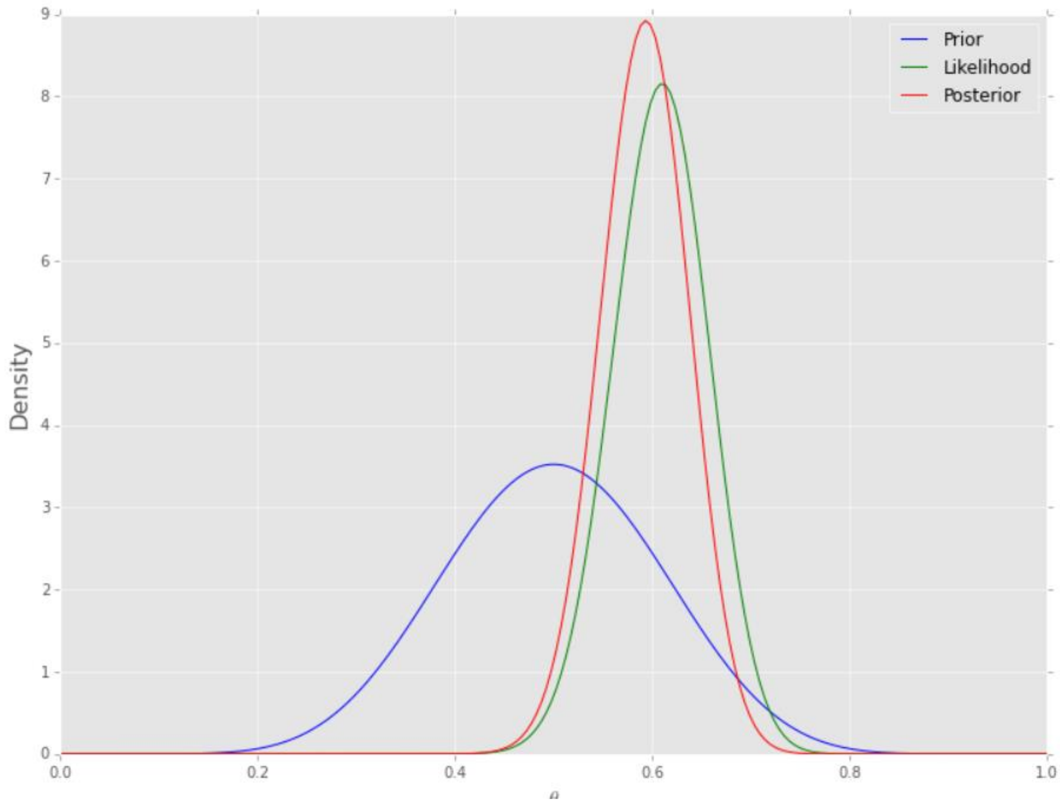
- The usual approach to estimate the life model parameters is **maximum likelihood estimation**
- No information on model parameter variability
- The **lack of knowledge** on model parameter can be addressed using **Bayesian methods**

Short overview on Bayesian probability theory

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- **Frequentist interpretation:** Bayes theorem is a **tool** to calculate the **probability of an event**, given that **another non-independent event has occured**
- **Bayesian interpretation:** probability is interpreted as a **measure of "degree of belief "**. Bayes rule links the degree of belief in an event A before and after accounting for evidence.
  - $P(A)$: the **prior**, the initial degree of belief in event A, before any data observation.
  - $P(A|B)$: the **posterior**, the degree of belief in A after accounting for the evidence B
  - $P(B|A)$: the **likelihood**, describes the compatibility of the evidence with the event A.

Graphical interpretation of Bayes theorem

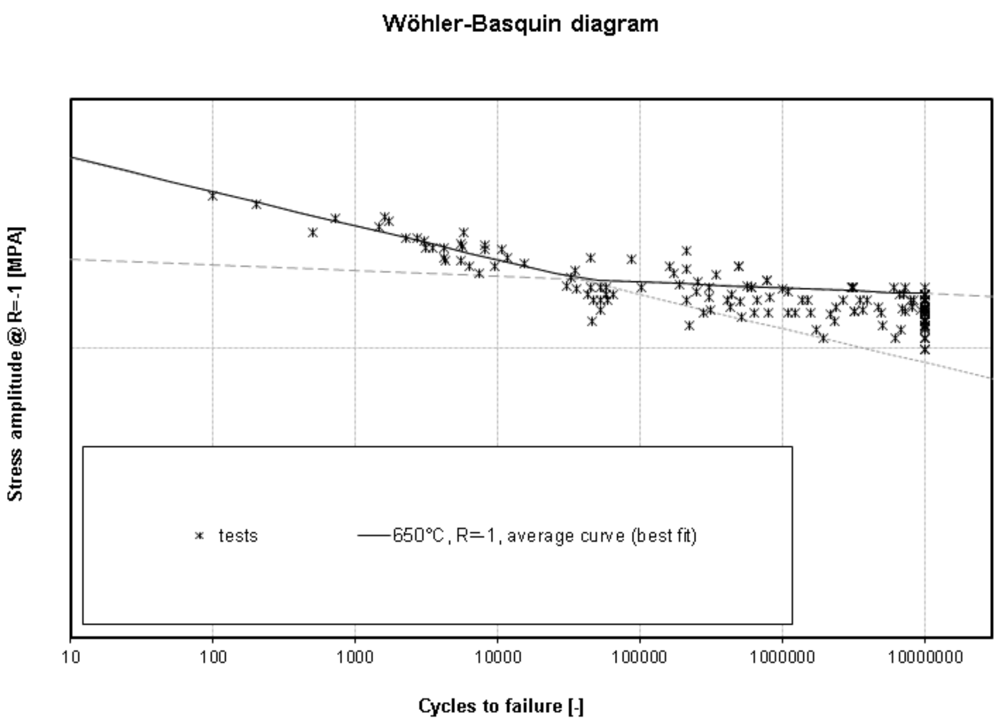


Computational Statistics with Python, Duke University lecture notes

Application: estimate parameter for fatigue life model

$$\underbrace{\sigma_{amp} \cdot \left(\frac{1-R}{2}\right)^{m-1}}_{\sigma_a{}^{walker}} = \left\{\left[A \cdot \left(2N_f\right)^b\right]^r + \left[C \cdot \left(2N_f\right)^d\right]^r\right\}^{1/r}$$

- Experimental data provided $\sigma_a{}^{walk,exp}$ and $N_f{}^{exp}$.
- Dataset including runouts.
- Parameters to be estimated: A, b, C, d, r



The uncertainty in the parameters can be modeled through priors

- **Uninformative** and **independent priors** for the parameters

$$A \sim U(0, 50000), C \sim U(0, 50000), d \sim U(-1, -0.01), r \sim U(0, 50)$$

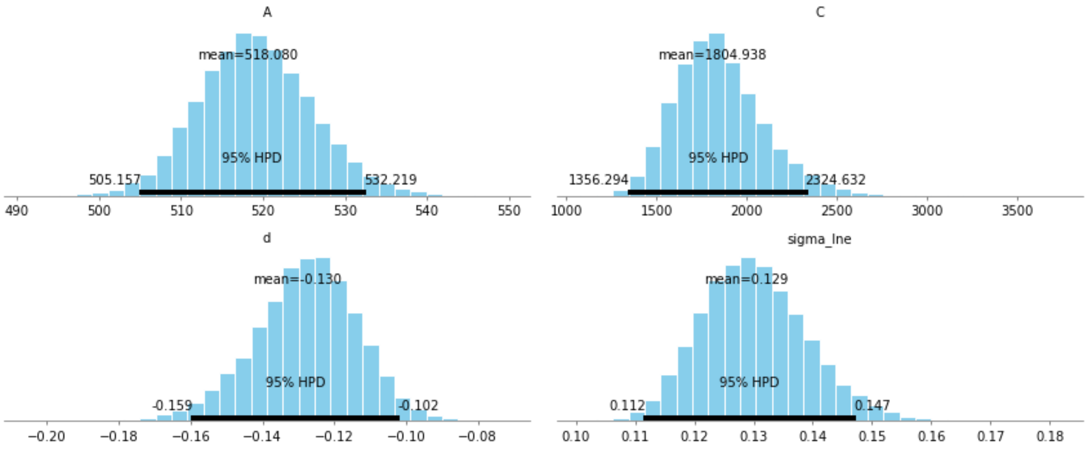
- **Uninformative prior** for the model error: $\sigma \sim U(0, \infty)$

- Available data: 105 Failures, 62 Runouts

- Bayesian calibration with Metropolis-Hastings algorithm

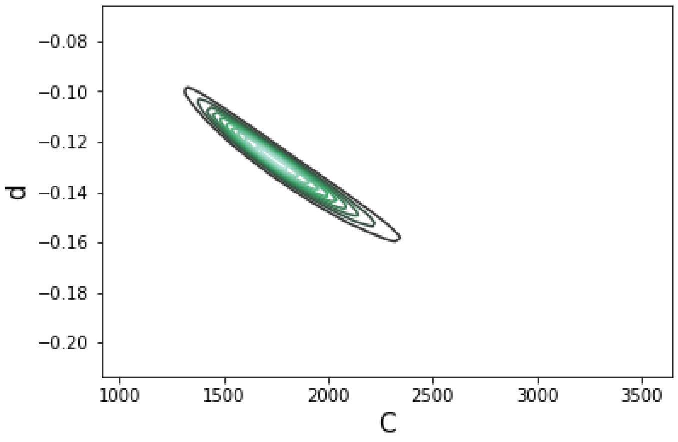
Bayesian calibration provides variability information on the parameters

Marginal Posterior Histograms



Correlations

	A	C	d	sigma_lne
A	1.000	-0.041	0.054	0.247
C	-0.041	1.000	-0.984	0.034
d	0.054	-0.984	1.000	-0.022
sigma_lne	0.247	0.034	-0.022	1.000



Add on in modeling lack of knowledge:

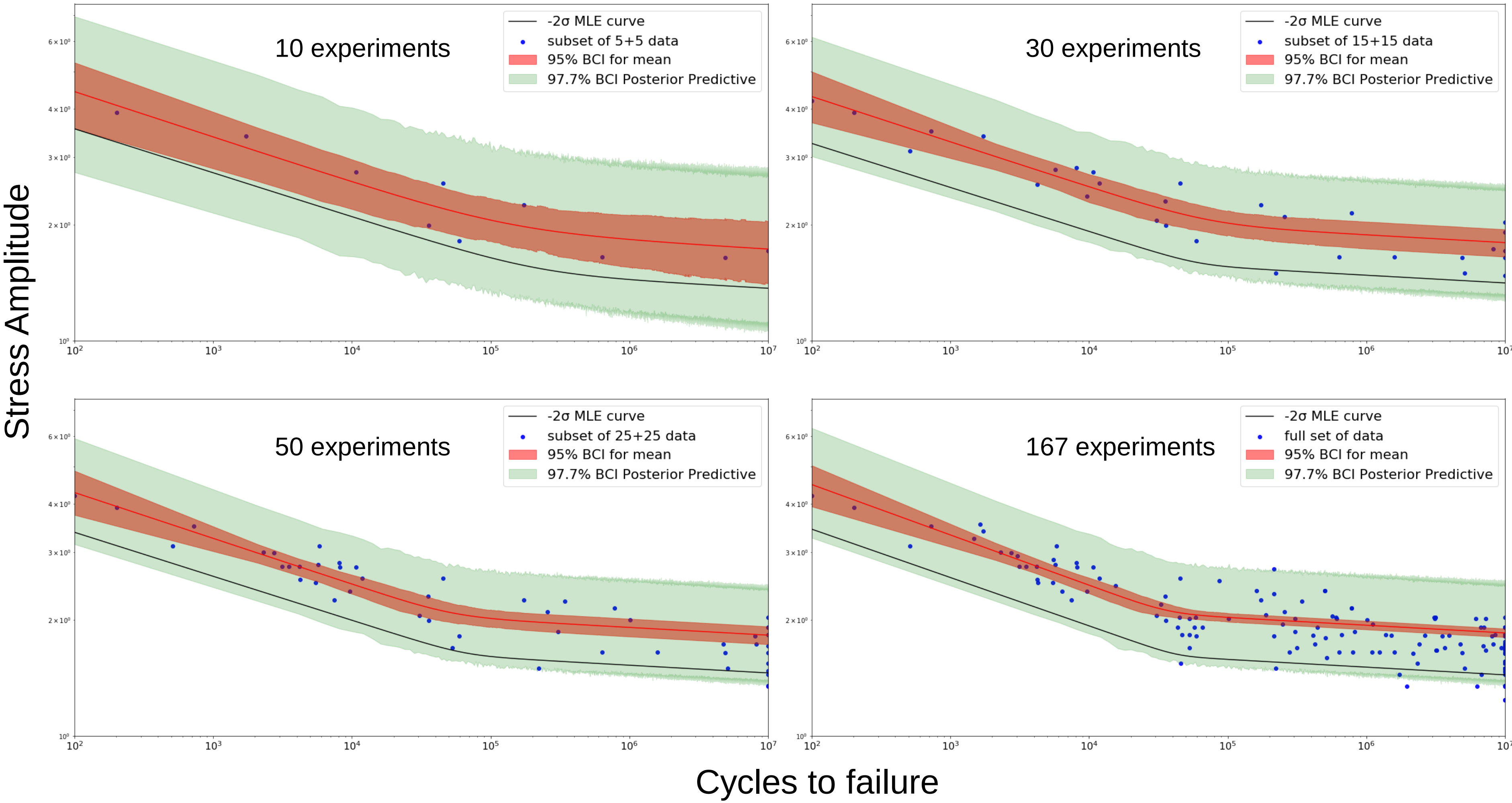
- More statistical information on parameter distribution
- Correlation information

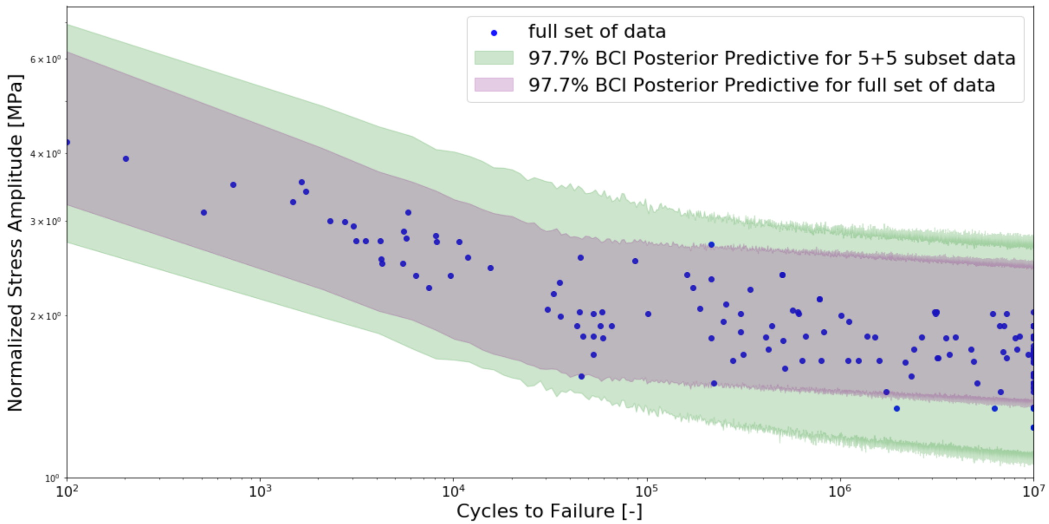
Does the number of data points affect the variance prediction?

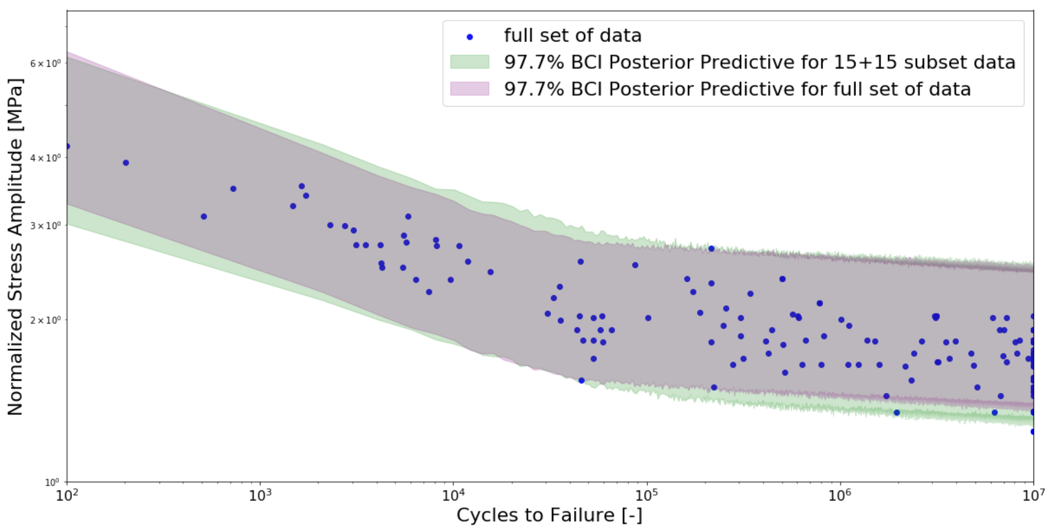
The original subset of data is subdivided in:

- ✓ 5 LCF and 5 HCF test data
- ✓ 15 LCF and 15 HCF test data
- ✓ 25 LCF and 25 HCF test data

Prediction of variance can be done by calculating the Bayesian Posterior Prediction Interval.


10 experiments
-2σ MLE curve
subset of 5+5 data
95% BCI for mean
97.7% BCI Posterior Predictive
30 experiments
-2σ MLE curve
subset of 15+15 data
95% BCI for mean
97.7% BCI Posterior Predictive
50 experiments
-2σ MLE curve
subset of 25+25 data
95% BCI for mean
97.7% BCI Posterior Predictive
167 experiments
-2σ MLE curve
full set of data
95% BCI for mean
97.7% BCI Posterior Predictive
Stress Amplitude
Cycles to failure





- 10 points are not sufficient to estimate the entire variability in the data set
- **At least 30 points should be provided.**

Conclusion

- Doing Bayesian analysis for fatigue data provides additional information on the probability distribution of model coefficients

- Using flat uniform prior distributions should be avoided and weakly informative priors should be used instead.

- The more runouts included in the dataset, the less informative the data is.

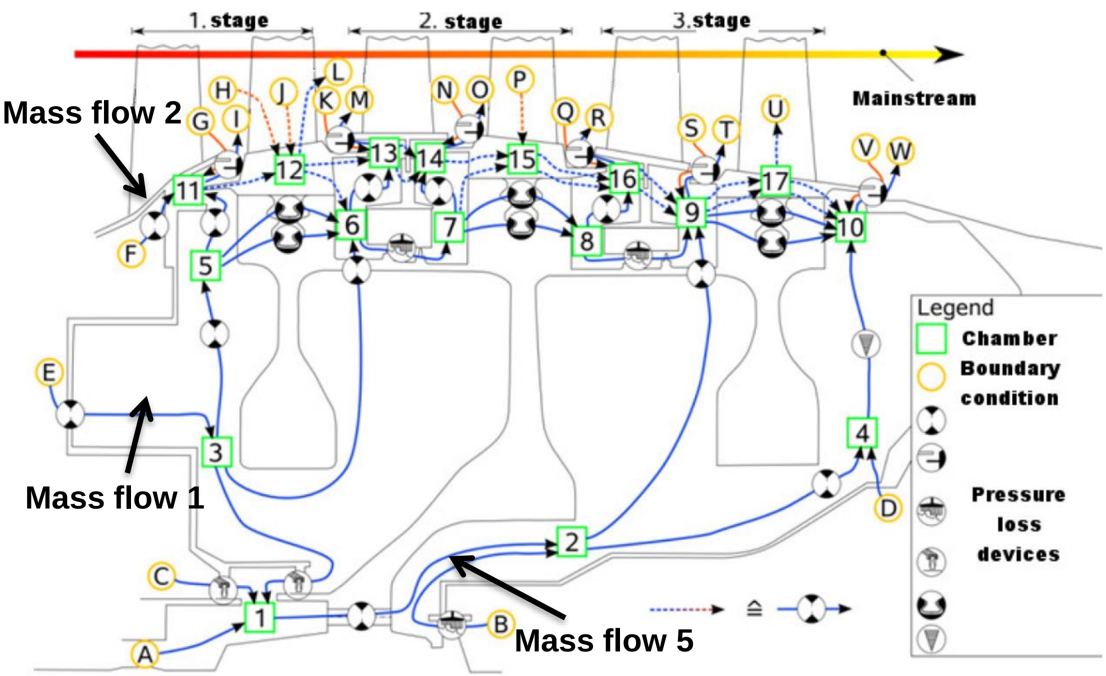
- Bayesian posterior predictive interval more conservative than -2σ MLE curve (without confidence interval)

Example 2: Manufacturing tolerance modeling in the Secondary Air System (SAS)

Mixed aleatoric - epistemic uncertainty quantification

- Model **input parameter uncertainty** the main focus.
- Only intervals available within which the true parameter value lies
- Late design phase → performing measurements is practically possible
- Focus on robustness

Jet engine secondary air system (SAS) is responsible for the cooling of the internal parts of the engine.



Input:

- Boundary conditions dependent on **performance parameters (pressure, temperature).**
- **Geometries** of the different flow elements.

Output:

- Pressure and temperature at each chamber (in green).
- Mass flow rate in each flow line (in blue).

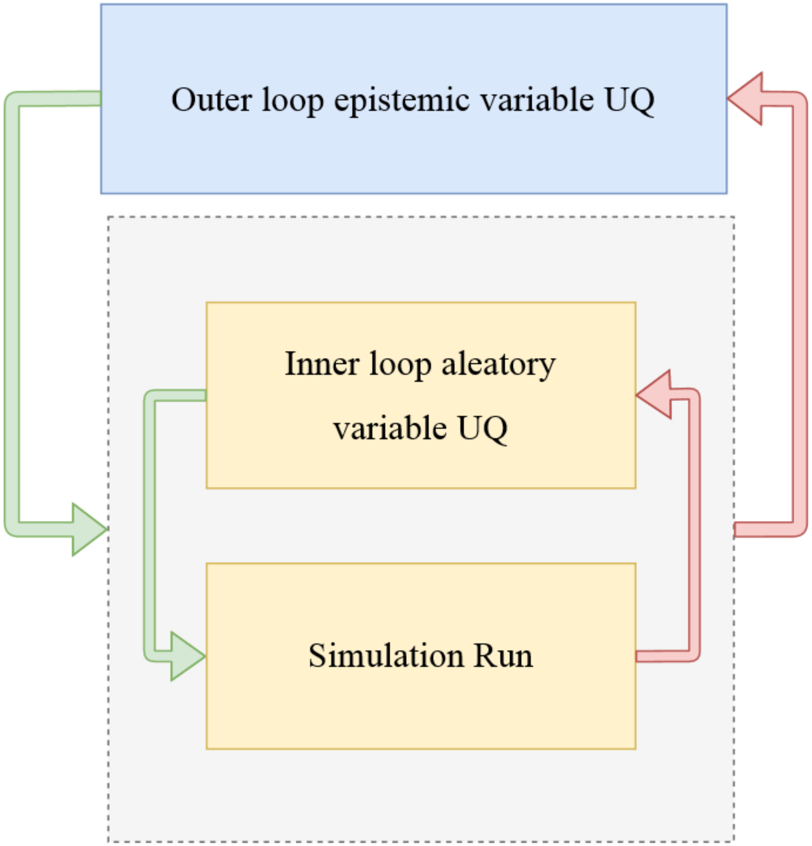
SAS uncertainties are mixed aleatoric and epistemic

Input Parameters	Uncertainty Source	Available Information	Categorization
Performance parameters	Ambient conditions Engine-to-engine variations Deterioration	Expert Information Legacy engines	Aleatory
Geometric variables (except rotor-stator gaps)	Manufacturing tolerances	Technical Drawings	Epistemic
Rotor-stator gaps	Manufacturing tolerances Engine operation	Technical Drawings -	Aleatory

12 aleatoric uncertain variables + 51 epistemic uncertain variables

The double loop approach conceptually separate epistemic and aleatoric uncertainty

$$Y = g(X), \textbf{with } \boldsymbol{X} = [\boldsymbol{X_a}, \boldsymbol{X_e}]$$



Dakota user's manual, Sandia National Laboratories.

$\boldsymbol{X_e}$ → interval information

Method:

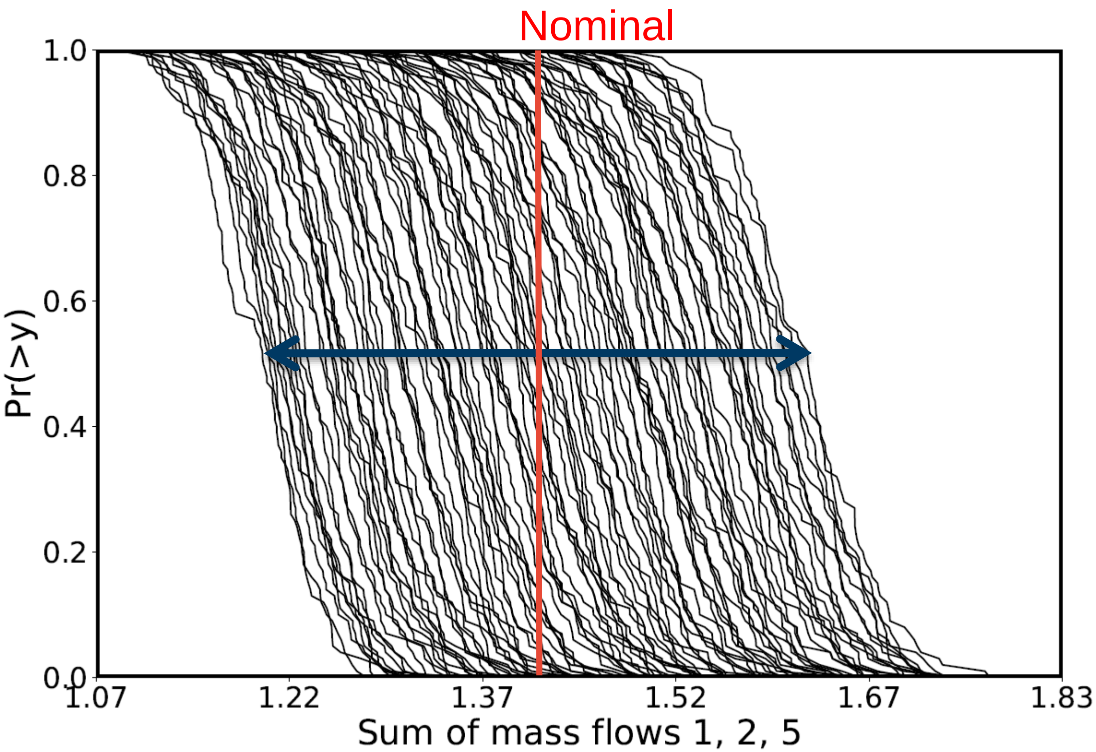
- IVP (interval value probability)

$\boldsymbol{X_a}$ → probability distribution

Method:

- Monte Carlo
- Stochastic collocation method

Aleatoric uncertainty = Shape of CCDF
Epistemic uncertainty = Interval width



- Narrowing tolerance on the geometric features of SAS means a reduction of interval width
- The **total amount of cooling air** has **less variability**
- The **tolerance** can be changed to produce an **improvement** with respect to **nominal**
- There is a **remaining variability** due to boundary conditions variation and labyrinth gap variation **which is irreducible**

Conclusion

- Parameter and model uncertainty should be modeled according to the available information
- Lack of knowledge can also be modeled through epistemic uncertainty
- **Bayesian methods for calibration** provides **additional information** useful for quantifying the **variability, e.g. in life models**
- Mixed aleatoric and epistemic uncertainty methods **separate the uncertainties both conceptually and computationally** and help obtain separately the influence of each uncertainty type

Thank you for your attention

Software

- **PyMC3**: Python library for Bayesian Statistical Modeling focusing on advanced Markov Chain Monte Carlo (MCMC) algorithms

  - Metropolis Hastings Algorithm (MH):
    - main oldest MCMC algorithm

  - No-U-Turn Sampler (NUTS):
    - State of the art MCMC algorithm which uses first order gradient information of the log-posterior density.
    - Handles only continuous parameters. Suitable in our case where we have an analytical likelihood function.