

# Bayesian Multilevel Model Calibration for Inversion of “Perfect” Data in the Presence of Epistemic and Aleatory Uncertainty

Joseph B. Nagel & Bruno Sudret

Chair of Risk, Safety & Uncertainty Quantification

MascotNum 2014, April 23-25, 2014



# Outline

- 1 Motivation: NASA UQ Challenge
- 2 Bayesian Multilevel Modeling
- 3 NASA UQ Challenge
- 4 Partial Data Augmentation

# Motivating Example

## NASA Langley Multidisciplinary UQ Challenge

- Originating from a realistic application from aerospace engineering
- Widely discipline-independent “blackbox” problem formulation

### Five UQ problems

- **Uncertainty characterization**
- Sensitivity analysis
- Uncertainty propagation
- Extreme-case analysis
- Robust design



L. G. Crespo, S. P. Kenny, and D. P. Giesy. “The NASA Langley Multidisciplinary Uncertainty Quantification Challenge”. In: [16th AIAA Non-Deterministic Approaches Conference \(SciTech 2014\)](#), National Harbor, Maryland (USA). January 13-17, 2014

# Uncertainty Characterization Subproblem

## Generic problem statement

With responses of a given forward model, learn about unknown forward model inputs, that are subject to uncertainty and variability

- Estimating forward model parameters
- Identifying the distribution of variable inputs
- Under additional nuisance variables with known distribution

## Sources of information

- $2 \times 25 = 50$  “perfect” observations
- Computational “blackbox” forward model
- “Prior” uncertainty model of its inputs

## Bayesian interpretation & solution

Bayesian calibration of a suitably defined multilevel model

Numerical solution by means of Markov chain Monte Carlo

# Section Outline

- 1 Motivation: NASA UQ Challenge
- 2 Bayesian Multilevel Modeling
  - Introduction
  - Statistical Data Model
  - "Perfect" Data Model
  - Posterior Fidelity
- 3 NASA UQ Challenge
- 4 Partial Data Augmentation

# Bayesian Multilevel Modeling: Definition

## Definition

A **hierarchical** or **multilevel model** is an “assembly of multiple submodels that are hierarchically interlinked”

## Hierarchy

- Probabilistic relations
- Deterministic mappings

## Submodels

- 1 **Forward model** (physics involved)
- 2 **Prior model** (input uncertainty)
- 3 **Residual model** (prediction error)

# Forward Model: Physical Model

## Forward model (physical model)

- Mathematical representation of a system under investigation
- Deterministic map: parameters  $\mapsto$  data

$$\mathcal{M}: (\mathbf{m}, \mathbf{x}, \boldsymbol{\zeta}, \mathbf{d}) \mapsto \tilde{\mathbf{y}} = \mathcal{M}(\mathbf{m}, \mathbf{x}, \boldsymbol{\zeta}, \mathbf{d})$$

## Input uncertainty (w.r.t. experiments $i = 1, \dots, n$ )

- Experimental conditions:  $\mathbf{d}$   $\rightarrow \mathbf{d}_i$
- Fixed yet unknown model parameters:  $\mathbf{m}$   $\rightarrow \mathbf{m}$
- Inputs with well-known variability:  $\boldsymbol{\zeta}$   $\rightarrow \boldsymbol{\zeta}_i$
- Inputs with unknown variability:  $\mathbf{x}$   $\rightarrow \mathbf{x}_i$

# Prior Model: Uncertainty & Variability

## Unknown model parameter

epistemic uncertainty:  $M \sim \pi_M(\mathbf{m})$  (parametric prior)

## Well-known input variability

aleatory uncertainty:  $\left\{ \begin{array}{l} \mathbf{Z}_i \sim f_Z(\zeta_i; \boldsymbol{\theta}_Z) \\ \text{with known } \boldsymbol{\theta}_Z \end{array} \right\}$  (structural prior)

## Unknown input variability

aleatory uncertainty:  $(\mathbf{X}_i | \boldsymbol{\theta}_X) \sim f_{\mathbf{X} | \boldsymbol{\theta}_X}(\mathbf{x}_i | \boldsymbol{\theta}_X)$  (structural prior)

epistemic uncertainty:  $\boldsymbol{\Theta}_X \sim \pi_{\boldsymbol{\Theta}_X}(\boldsymbol{\theta}_X)$  (parametric prior)



# Residual Model: Statistical Data Model

## Residual model

- Observations are in some way "imperfect" (e.g. due to measurement errors, numerical approximations and forward model inadequacies)
- Representation as  $\mathbf{y}_i = \tilde{\mathbf{y}}_i + \boldsymbol{\varepsilon}_i$  with  $\boldsymbol{\varepsilon}_i$  a realization of  $\mathbf{E}_i \sim f_{\mathbf{E}_i}(\boldsymbol{\varepsilon}_i)$

### "Imperfect" data

Data are viewed as realizations  $\mathbf{y}_i$  of random variables  $(\mathbf{Y}_i | \mathbf{m}, \mathbf{x}_i, \boldsymbol{\zeta}_i)$  distributed according to  $f(\mathbf{y}_i | \mathbf{m}, \mathbf{x}_i, \boldsymbol{\zeta}_i) = f_{\mathbf{E}_i}(\mathbf{y}_i - \mathcal{M}(\mathbf{m}, \mathbf{x}_i, \boldsymbol{\zeta}_i, \mathbf{d}_i))$

# Bayesian Multilevel Model

## Bayesian Multilevel Model

$$(Y_i | m, x_i, \zeta_i) \sim f_{E_i}(y_i - \mathcal{M}(m, x_i, \zeta_i, d_i))$$

$$M \sim \pi_M(m)$$

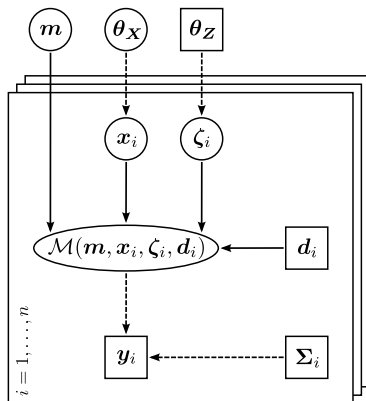
$$Z_i \sim f_Z(\zeta_i; \theta_Z)$$

$$(X_i | \theta_X) \sim f_{X|\Theta_X}(x_i | \theta_X)$$

$$\Theta_X \sim \pi_{\Theta_X}(\theta_X)$$

- Unless denoted otherwise, random variables are (conditionally) independent
- This defines a joint probability model of the entirety of random quantities

## Directed Acyclic Graph



# Bayesian Multilevel Model Calibration: Joint Inference

## Unique approach to inference

Formulate a joint probability density of all unknowns  $(\mathbf{m}, \langle \mathbf{x}_i \rangle, \langle \zeta_i \rangle, \boldsymbol{\theta}_X)$  conditioned on all knowns  $(\langle \mathbf{y}_i \rangle, \boldsymbol{\theta}_Z)$ . Then integrate out nuisance.

**Joint prior**  $\langle \mathbf{q}_i \rangle \equiv \langle \mathbf{q}_i \rangle_{1 \leq i \leq n} = (\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n)$

$$\begin{aligned} & \pi(\mathbf{m}, \langle \mathbf{x}_i \rangle, \langle \zeta_i \rangle, \boldsymbol{\theta}_X) \\ &= \left( \prod_{i=1}^n f_{X|\Theta_X}(\mathbf{x}_i | \boldsymbol{\theta}_X) \right) \left( \prod_{i=1}^n f_Z(\zeta_i; \boldsymbol{\theta}_Z) \right) \pi_{\Theta_X}(\boldsymbol{\theta}_X) \pi_M(\mathbf{m}) \end{aligned}$$

**Joint posterior**

$$\begin{aligned} & \pi(\mathbf{m}, \langle \mathbf{x}_i \rangle, \langle \zeta_i \rangle, \boldsymbol{\theta}_X | \langle \mathbf{y}_i \rangle) \\ & \propto \left( \prod_{i=1}^n f_{E_i}(\mathbf{y}_i - \mathcal{M}(\mathbf{m}, \mathbf{x}_i, \zeta_i, \mathbf{d}_i)) \right) \pi(\mathbf{m}, \langle \mathbf{x}_i \rangle, \langle \zeta_i \rangle, \boldsymbol{\theta}_X) \end{aligned}$$

# "Perfect" Data Model

## "Perfect" data

Data, i.e. forward model responses  $\tilde{y}_i = \mathcal{M}(\mathbf{m}, \mathbf{x}_i, \zeta_i, \mathbf{d}_i)$ , are viewed as realizations of random variables  $(\tilde{Y}_i | \mathbf{m}, \theta_X)$  with  $f_{\tilde{Y}_i}(\tilde{y}_i | \mathbf{m}, \theta_X)$

## Bayesian model

$$\begin{aligned}(\tilde{Y}_i | \mathbf{m}, \theta_X) &\sim f_{\tilde{Y}_i}(\tilde{y}_i | \mathbf{m}, \theta_X) \\(M, \Theta_X) &\sim \pi(\mathbf{m}, \theta_X) = \pi_M(\mathbf{m}) \pi_{\Theta_X}(\theta_X)\end{aligned}$$

## Likelihood function

$$\mathcal{L}(\langle \tilde{\mathbf{y}}_i \rangle | \mathbf{m}, \theta_X) = \prod_{i=1}^n f_{\tilde{Y}_i}(\tilde{y}_i | \mathbf{m}, \theta_X)$$

## Posterior density

$$\pi(\mathbf{m}, \theta_X | \langle \tilde{\mathbf{y}}_i \rangle) \propto \mathcal{L}(\langle \tilde{\mathbf{y}}_i \rangle | \mathbf{m}, \theta_X) \pi(\mathbf{m}, \theta_X)$$

# Forward Uncertainty Propagation: Push-Forward

Formulate a likelihood function:  $\mathcal{L}_{\text{data}}(\text{QoI}) = \mathcal{P}(\text{data}|\text{QoI})$

- Given the forward model  $\mathcal{M}: (\mathbf{m}, \mathbf{x}, \boldsymbol{\zeta}, \mathbf{d}) \mapsto \tilde{\mathbf{y}} = \mathcal{M}(\mathbf{m}, \mathbf{x}, \boldsymbol{\zeta}, \mathbf{d})$
- For fixed  $(\mathbf{m}, \mathbf{d}_i)$  consider  $\mathcal{M}_{\mathbf{m}, \mathbf{d}_i}: (\mathbf{x}, \boldsymbol{\zeta}) \mapsto \tilde{\mathbf{y}}_i = \mathcal{M}(\mathbf{m}, \mathbf{x}, \boldsymbol{\zeta}, \mathbf{d}_i)$
- Propagate **aleatory** input uncertainties  $(\mathbf{X}_i | \boldsymbol{\theta}_X) \sim f_{X|\Theta_X}(\mathbf{x}_i | \boldsymbol{\theta}_X)$  and  $\mathbf{Z}_i \sim f_Z(\boldsymbol{\zeta}_i; \boldsymbol{\theta}_Z)$  as determined by hyperparameters  $(\boldsymbol{\theta}_X, \boldsymbol{\theta}_Z)$

$$\left. \begin{array}{l} (\mathbf{X}_i | \boldsymbol{\theta}_X) \sim f_{X|\Theta_X}(\mathbf{x}_i | \boldsymbol{\theta}_X) \\ \mathbf{Z}_i \sim f_Z(\boldsymbol{\zeta}_i; \boldsymbol{\theta}_Z) \end{array} \right\} \tilde{\mathbf{Y}}_i = \mathcal{M}(\mathbf{m}, \mathbf{X}_i, \mathbf{Z}_i, \mathbf{d}_i)$$

$$\rightsquigarrow (\tilde{\mathbf{Y}}_i | \mathbf{m}, \boldsymbol{\theta}_X) \sim f_{\tilde{\mathbf{Y}}_i}(\tilde{\mathbf{y}}_i | \mathbf{m}, \boldsymbol{\theta}_X)$$

Likelihood function:  $\mathcal{L}(\langle \tilde{\mathbf{y}}_i \rangle | \mathbf{m}, \boldsymbol{\theta}_X) = \prod_{i=1}^n f_{\tilde{\mathbf{Y}}_i}(\tilde{\mathbf{y}}_i | \mathbf{m}, \boldsymbol{\theta}_X)$

# Likelihood Estimation

## KDE-based evaluation of the likelihood

$$\text{Sample } \left\{ \begin{array}{l} \mathbf{x}^{(k)} \sim f_{X|\Theta_X}(\mathbf{x}^{(k)}|\boldsymbol{\theta}_X) \\ \zeta^{(k)} \sim f_Z(\zeta^{(k)}; \boldsymbol{\theta}_Z) \end{array} \right\} \text{ for } k = 1, \dots, K$$

$$\text{Compute } \tilde{\mathbf{v}}_i^{(k)} = \mathcal{M}(\mathbf{m}, \mathbf{x}^{(k)}, \zeta^{(k)}, \mathbf{d}_i)$$


$$\hat{\mathcal{L}}_{\text{KS}}(\langle \tilde{\mathbf{y}}_i \rangle | \mathbf{m}, \boldsymbol{\theta}_X) = \prod_{i=1}^n \left( \frac{1}{K} \sum_{k=1}^K \mathcal{K}_H(\tilde{\mathbf{y}}_i - \tilde{\mathbf{v}}_i^{(k)}) \right)$$

- Free algorithmic parameters have to be tuned: number of samples  $K$  and the kernel bandwidth  $H \rightarrow$  **"optimal" parameter tuning?**
- However, the likelihood is estimated in every MCMC iteration
- The optimal tuning of free parameters has to be assessed with respect to the **posterior fidelity**

# MCMC & Likelihood Estimation

## Bayesian inference by MCMC

Construct a MC whose invariant distribution equals the posterior

- Metropolis-Hastings correction:  $\alpha = \min \left( 1, \frac{\pi_1(\mathbf{q}^{(*)}) P(\mathbf{q}^{(t)} | \mathbf{q}^{(*)})}{\pi_1(\mathbf{q}^{(t)}) P(\mathbf{q}^{(*)} | \mathbf{q}^{(t)})} \right)$
- **Statistical likelihood estimations induce modifications on the level of the Markov chain transition kernel**
- This may alter the equilibrium distribution of the Markov chain 

## Key concept: posterior fidelity

We define **posterior fidelity** as the degree as to which the induced equilibrium distribution is in congruence with the true posterior

- Posterior fidelity depends on a complex interplay between the estimator  $\hat{\mathcal{L}}$ , the true posterior  $\pi_1$  and the proposal distribution  $P$

# Posterior Fidelity & Parameter Tuning

- We do not have a means to define “optimal” parameter tunings
- Setting  $K$  and  $H$  is trading-off the fidelity of the posterior against the ease of its computation (**fidelity vs. feasibility**)
- An “appropriate” decision is made when the likelihood ratio can be estimated “sufficiently accurate”

$$\text{likelihood ratio: } \frac{\hat{\mathcal{L}}(\mathbf{q}^{(*)})}{\hat{\mathcal{L}}(\mathbf{q}^{(t)})} \approx \frac{\mathcal{L}(\mathbf{q}^{(*)})}{\mathcal{L}(\mathbf{q}^{(t)})}$$

- High posterior fidelity is achieved if “appropriate” decisions are being frequently made over the course of the Markov chain

→ **Heuristics and plausibility checks**



# Section Outline

- 1 Motivation: NASA UQ Challenge
- 2 Bayesian Multilevel Modeling
- 3 NASA UQ Challenge**
  - Prior Uncertainty Model
  - Likelihood Estimation
  - Results
- 4 Partial Data Augmentation

# The NASA Langley UQ Challenge

## Problem statement

Given forward model outputs, learn about unknown forward model inputs

### Available information

- The forward model  $\mathcal{M} \equiv h_1$  is distributed as MATLAB p-code file
- Data  $\langle \tilde{y}_i \rangle_{1 \leq i \leq 50}$  with  $\tilde{y}_i = h_1(p_{1,i}, p_2, p_{3,i}, p_{4,i}, p_{5,i})$
- Information about the forward model inputs  $(p_{1,i}, p_2, p_{3,i}, p_{4,i}, p_{5,i})$

### Types of forward model inputs

- Unknown forward model parameter to be estimated:  $p_2$
- Inputs whose variability is of inferential interest:  $p_1, p_4, p_5$
- Additional nuisance variables with known distribution:  $p_3$

## Category I & II: Aleatory & Epistemic Uncertainty

### Fixed yet unknown forward model parameter: $p_2$

- The model parameter  $m \equiv p_2$  is an unknown constant
- Its value is known to be contained in an epistemic interval  $\Delta = [0, 1]$
- **Parametric prior:**  $P_2 \sim \pi_2(p_2) = \mathcal{U}(p_2|0, 1)$

### Forward model input with well-known variability: $p_3$

- Experiment-specific realizations  $\zeta_i \equiv p_{3,i}$  for  $i = 1, \dots, n$  were sampled from a uniform population distribution  $\mathcal{U}(p_{3,i}|a_3, b_3)$
- Its hyperparameters  $(a_3, b_3) = (0, 1)$  are well-known
- **Structural prior:**  $(P_{3,i}|\theta_3) \sim f_3(p_{3,i}|\theta_3) = \mathcal{U}(p_{3,i}|0, 1)$

## Category III: Mixed Uncertainty: Correlated Gaussian

**Variable forward model inputs:**  $p_4, p_5$

- Possibly correlated Gaussian variables:  $(p_4, p_5)$

$$((P_{4,i}, P_{5,i}) | \theta_{45}) \sim f_{45}((p_{4,i}, p_{5,i}) | \theta_{45}) = \mathcal{N}((p_{4,i}, p_{5,i}) | \mu_{45}, \Sigma_{45})$$

$$\mu_{45} = \begin{pmatrix} \mu_4 \\ \mu_5 \end{pmatrix}, \quad \Sigma_{45} = \begin{pmatrix} \sigma_4^2 & \rho_{45} \sigma_4 \sigma_5 \\ \rho_{45} \sigma_4 \sigma_5 & \sigma_5^2 \end{pmatrix}$$

- Hyperparameters  $\theta_{45} \equiv (\mu_4, \sigma_4^2, \mu_5, \sigma_5^2, \rho_{45})$
- It is known that  $-5 \leq \mu_j \leq 5$ ,  $1/400 \leq \sigma_j^2 \leq 4$  and  $|\rho_{45}| \leq 1$

$$\left. \begin{array}{l} \pi(\mu_j) = \mathcal{U}(-5, 5), \\ \pi(\sigma_j^2) = \mathcal{U}(1/400, 4), \\ \pi(\rho_{45}) = \mathcal{U}(-1, 1), \end{array} \right\} \pi_{45}(\theta_{45}) = \left( \prod_{j=4}^5 \pi(\mu_j) \pi(\sigma_j^2) \right) \pi(\rho_{45})$$

## Category III: Mixed Uncertainty: Unimodal Beta

### Variable forward model inputs: $p_1$

- Variables following a unimodal beta distribution:  $p_1$

$$(P_{1,i} | \boldsymbol{\theta}_1) \sim f_1(p_{1,i} | \boldsymbol{\theta}_1) = \text{Beta}(p_{1,i} | \mu_1, \sigma_1^2)$$

- Statistical  $\boldsymbol{\theta}_1 \equiv (\mu_1, \sigma_1^2)$  or shape hyperparameters  $\boldsymbol{\theta}_1 \equiv (\alpha_1, \beta_1)$
- It is requested that  $3/5 \leq \mu_1 \leq 4/5$  and  $1/50 \leq \sigma_1^2 \leq 1/25$
- Moreover the unimodality translates into  $\alpha_1, \beta_1 > 1$
- Therefore one can state a uniform hyperprior on  $\boldsymbol{\theta}_1 \equiv (\mu_1, \sigma_1^2)$

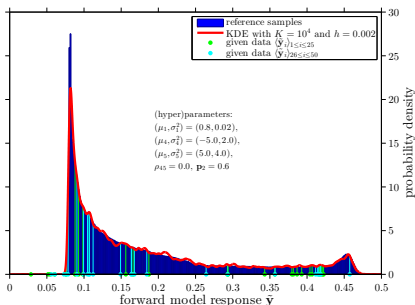
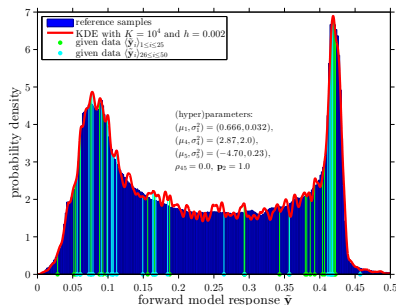
$$\boldsymbol{\Theta}_1 \sim \pi_1(\boldsymbol{\theta}_1) = \mathcal{U}(\boldsymbol{\theta}_1 | \mathcal{S}), \text{ with}$$

$$\mathcal{S} = \{(\mu_1, \sigma_1^2) \in \mathbb{R}^2 \mid 3/5 \leq \mu_1 \leq 4/5, 1/50 \leq \sigma_1^2 \leq 1/25, \alpha_1, \beta_1 > 1\}$$

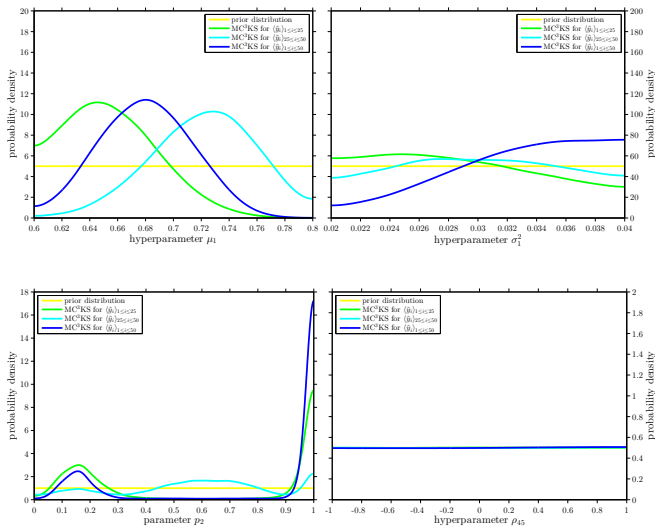
# Estimation of the Target Density $f_{\tilde{Y}_i}(\tilde{y}_i | p_2, \theta_1, \theta_{45})$

## Tuning of free algorithmic parameters

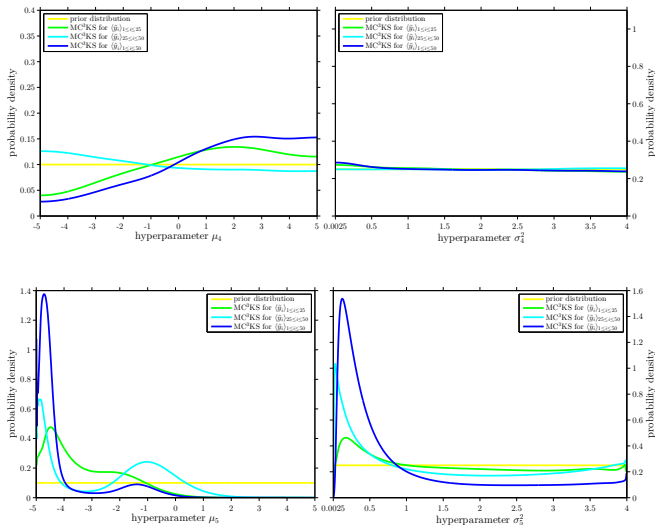
- Resource limitations bound the number of forward model runs  $K$
- Cascade of runs with  $K = 10^4$  and decreasing bandwidths  $h$
- Observations: Initial shrinkage, eventual breakdown, in between the distribution is rather stable with respect to changes in  $h$



# Marginal Posteriors of $\mu_1$ , $\sigma_1^2$ , $p_2$ and $\rho_{45}$

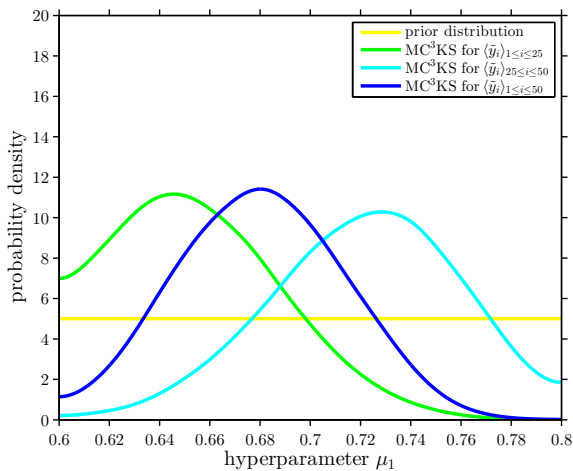


# Marginal Posteriors of $\mu_4$ , $\sigma_4^2$ , $\mu_5$ and $\sigma_5^2$

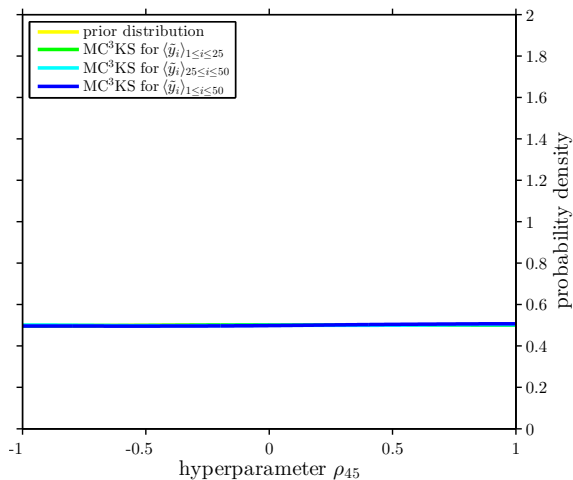




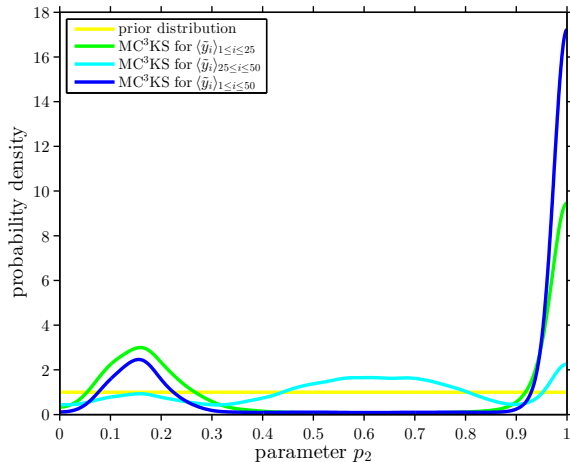
# Posterior Marginal of $\mu_1$



## Posterior Marginal of $\rho_{45}$



# Posterior Marginal of $p_2$



# Section Outline

- 1 Motivation: NASA UQ Challenge
- 2 Bayesian Multilevel Modeling
- 3 NASA UQ Challenge
- 4 Partial Data Augmentation**
  - Data Augmentation
  - Likelihood Estimation
  - Results

# Partial Data Augmentation

## Data Augmentation

- Traditionally data augmentation is a powerful MCMC technique that aims at improving numerical efficiency of posterior sampling
- Here we will try to exploit partial data augmentation for gaining posterior fidelity
- **Partial data augmentation:** explicitly infer  $\langle p_{1,i} \rangle$  as auxiliary variables, then integrate them out as nuisance
- Presume that  $\pi(\langle p_{1,i} \rangle, p_2, \theta_1, \theta_{45} | \langle \tilde{y}_i \rangle)$  is “easier” to sample than sampling from  $\pi(p_2, \theta_1, \theta_{45} | \langle \tilde{y}_i \rangle)$
- Partial data augmentation trades off dimensionality and fidelity

# Partial Data Augmentation

## Parametric & structural prior

$$\pi(\langle p_{1,i} \rangle, p_2, \theta_1, \theta_{45}) = \left( \prod_{i=1}^n f_1(p_{1,i} | \theta_1) \right) \pi_1(\theta_1) \pi_2(p_2) \pi_{45}(\theta_{45})$$

## Augmented likelihood

$$\mathcal{L}(\langle \tilde{y}_i \rangle | \langle p_{1,i} \rangle, p_2, \theta_1, \theta_{45}) = \prod_{i=1}^n f(\tilde{y}_i | p_{1,i}, p_2, \theta_{45})$$

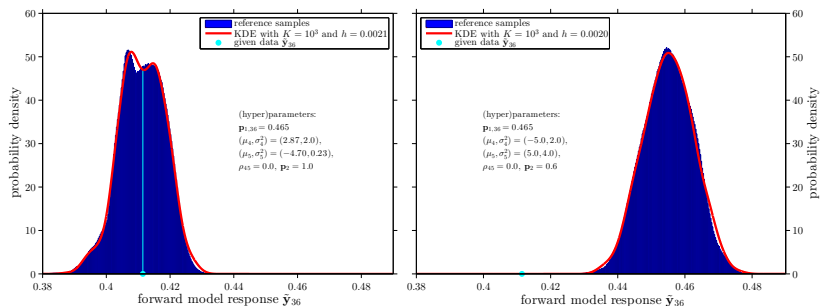
## Augmented posterior

$$\begin{aligned} \pi(\langle p_{1,i} \rangle, p_2, \theta_1, \theta_{45} | \langle \tilde{y}_i \rangle) \\ \propto \mathcal{L}(\langle \tilde{y}_i \rangle | \langle p_{1,i} \rangle, p_2, \theta_{45}) \pi(\langle p_{1,i} \rangle, p_2, \theta_1, \theta_{45}). \end{aligned}$$

## Integrate out nuisance

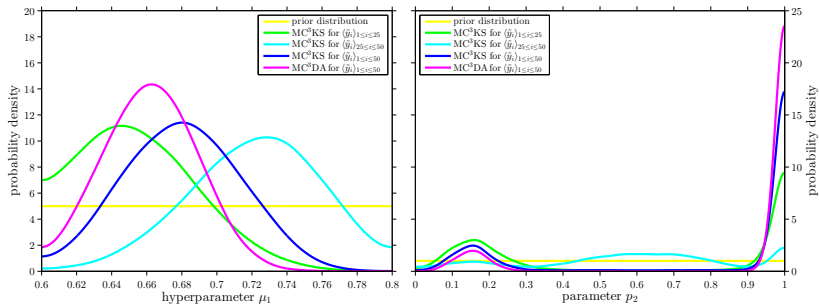
$$\pi(p_2, \theta_1, \theta_{45} | \langle \tilde{y}_i \rangle) = \int \cdots \int_{\mathcal{D}_{p_1}^n} \pi(\langle p_{1,i} \rangle, p_2, \theta_1, \theta_{45} | \langle \tilde{y}_i \rangle) d\langle p_{1,i} \rangle$$

# Estimation of the Target Density $f(\tilde{y}_i | \mathbf{p}_{1,i}, p_2, \theta_{45})$



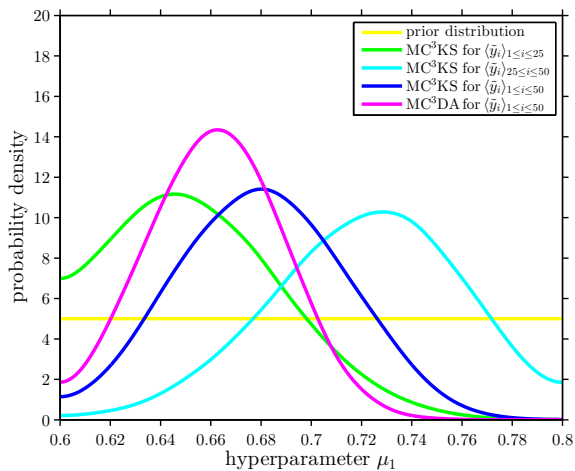
- The target density is “easier” to estimate and resembles a Gaussian
- Automatic bandwidth selection via Silverman’s rule of thumb
- Allows for more adequate likelihood estimations and therefore promises a boost in posterior fidelity

# Marginal Posteriors of $\mu_1$ and $p_2$

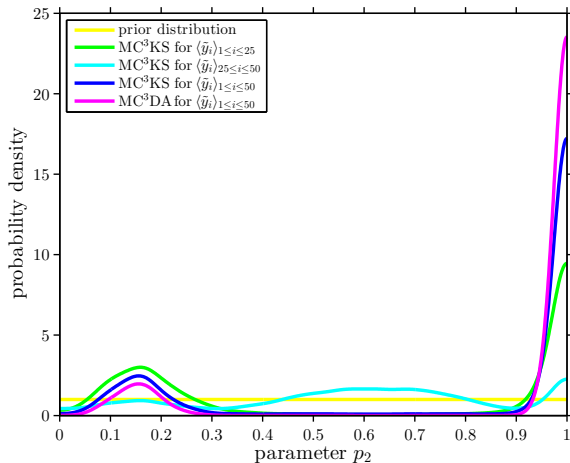




# Posterior Marginal of $\mu_1$



# Posterior Marginal of $p_2$



# Conclusions

## Conclusions

- Bayesian multilevel modeling establishes a convenient framework for solving complex inverse problems under uncertainty
- “Perfect” data model had to be devised (intractable likelihood)
- Numerical complexity has been demonstrated: Multilevel model calibration more complex than parameter estimation
- Motivates metamodeling (PCE,...) when involving computationally more expensive forward models (FEM,...)
- Posterior fidelity has been introduced as an important concept
- Data augmentation has been used to gain posterior fidelity

