Density Estimation Framework for Model Error Quantification

Khachik Sargsyan, Xun Huan, Jason Bender*, Habib N. Najm
Sandia National Laboratories, Livermore, CA, USA
currently at* **Lawrence Livermore National Laboratory, Livermore, CA, USA

ISBA World Meeting 2016

Email: ksargsy@sandia.gov

Goal: Model Error Quantification

Develop statistical framework for model error representation, quantification and propagation for physical models.

- Represent and estimate the error associated with
  - Simplifying assumptions, parameterizations
  - Mathematical formulation, theoretical framework
  - Numerical discretization
- Inverse modeling context $\quad y_i = f(x_i;\lambda) + \epsilon_i$
- Given data, calibrate for λ, accounting for model error
- Model error is deviation from 'truth'

Truth $g(x) \neq f(x;\lambda)$ Model

Model Error Challenge

Additive model discrepancy (Kennedy-O'Hagan, 2001)

$$y_i = \underbrace{f(x_i;\lambda) + \delta(x_i)}_{\text{truth } g(x_i)} + \epsilon_i$$

has challenges for physical models:

- Strong priors required on $\delta(x)$ to satisfy physical constraints
- Disambiguation of model error $\delta(x_i)$ and data error ϵ_i
- QoI-specific calibration; no extrapolation to other QoIs
- Correlation structure in $\delta(x)$ should ideally be informed by the model $f(x;\lambda)$

Idea: Embedded Model Error

Embed uncertainty within the model

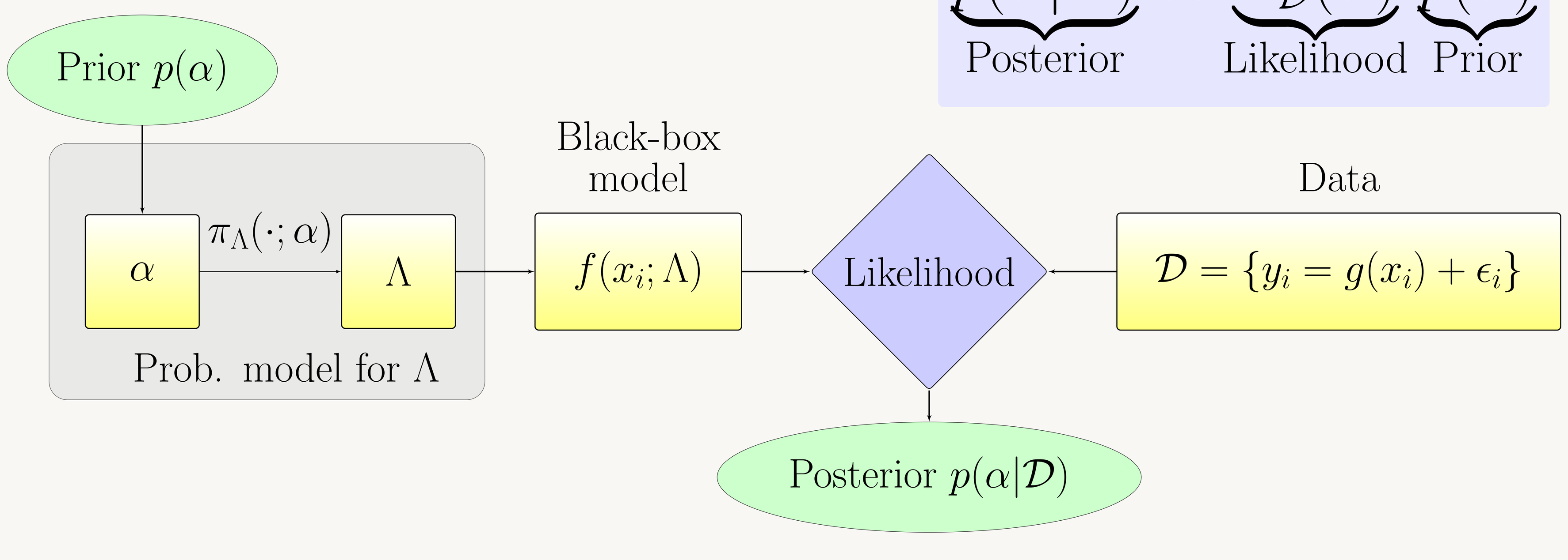
$y_i = f(x_i;\Lambda) + \epsilon_i$	$y_i = f(x_i;\Lambda(x_i)) + \epsilon_i$	$y_i = \tilde{f}(x_i;\lambda,\Theta) + \epsilon_i$
Black-box	*Random field*	*Extra 'physics'*

- Embed model error in specific submodel phenomenology
- Allows targeted model error placement
- Naturally preserves model structure and physical constraints
- Disambiguate model and data errors

- Cast input parameters λ as a random variable Λ
- Parameter estimation of λ turns into PDF estimation of Λ
- Parameterize PDF form $\pi_\Lambda(\cdot;\alpha)$
  - *e.g.* Polynomial Chaos $\Lambda = \sum_{k=0}^{K} \alpha_k \Psi_k(\xi)$
- Back to parameter estimation, now for $\alpha = (\alpha_0, \ldots, \alpha_K)$

Bayesian Density Estimation

$$\underbrace{p(\alpha|\mathcal{D})}_{\text{Posterior}} \propto \underbrace{L_{\mathcal{D}}(\alpha)}_{\text{Likelihood}} \underbrace{p(\alpha)}_{\text{Prior}}$$



K. Sargsyan, H. Najm, and R. Ghanem, "On the Statistical Calibration of Physical Models". *International Journal for Chemical Kinetics*, 47(4): pp 246–276, 2015.

Likelihood Construction

- Full Likelihood: $L(\alpha) = p(y|\alpha) = p(y_1,\ldots,y_N|\alpha) = \pi(y)$
- Marginal Apprx: $L(\alpha) \approx \prod_{i=1}^N p(y_i|\alpha) = \prod_{i=1}^N \pi(y_i)$
- Approximate Bayesian Computation: $L(\alpha) = \frac{1}{\epsilon} K\left(\frac{\rho(S_M,S_{\mathcal{D}})}{\epsilon}\right)$
- Gaussian Apprx: $L(\alpha) \propto e^{-\frac{1}{2}(y-\mu(\alpha))^T\Sigma^{-1}(\alpha)(y-\mu(\alpha))}$

Forward Prediction

$$f(x;\Lambda) = f\left(x;\sum_k \alpha_k\Psi_k(\xi)\right) \overset{NISP}{=} \sum_k f_k(x;\alpha)\Psi_k(\xi)$$

- Non-intrusive spectral projection (NISP) employed for
  - Likelihood computation and posterior predictions
  - Easy access to first two moments:

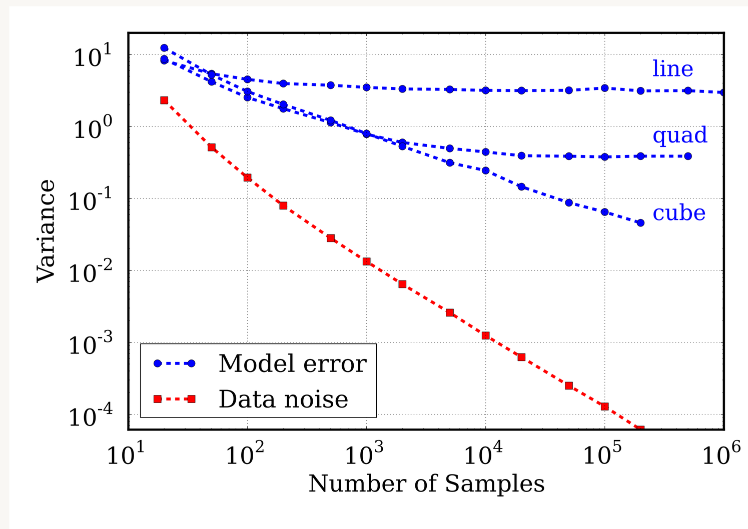
$$\mu(x;\alpha) = f_0(x;\alpha), \qquad \sigma^2(x;\alpha) = \sum_{k>0} f_k^2(x;\alpha)||\Psi_k||^2$$

- Predictive mean $\quad \mathbb{E}[y(x)] = \mathbb{E}_\alpha[\mu(x;\alpha)]$
- Decomposition of predictive variance

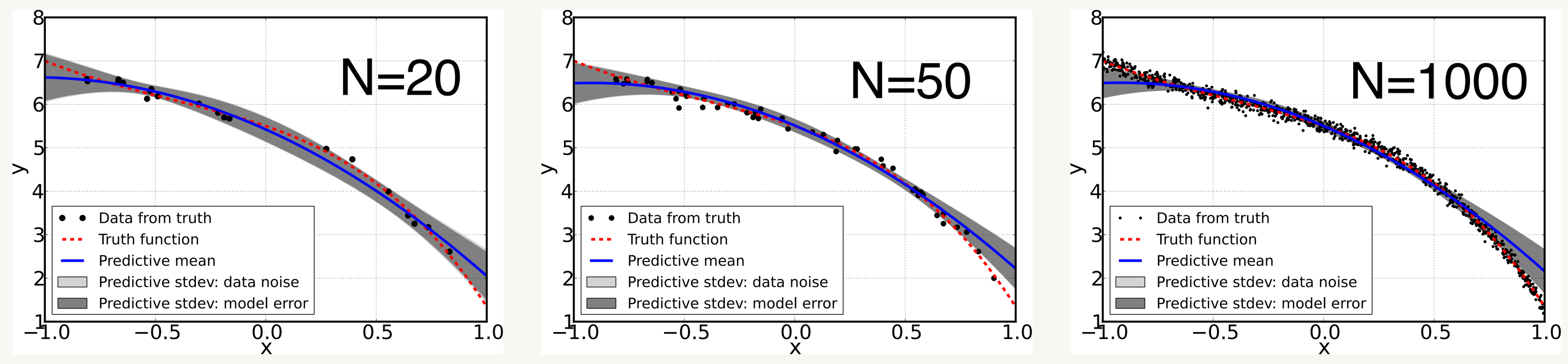
$$\mathbb{V}[y(x)] = \underbrace{\mathbb{E}_\alpha[\sigma^2(x;\alpha)]}_{\text{Model error}} + \underbrace{\mathbb{V}_\alpha[\mu(x;\alpha)] + \sigma_d^2}_{\text{Poserior/Data error}}$$

Demonstration

Calibrating linear, quadratic and cubic models w.r.t. 'truth'
$g(x) = 6 + x^2 - 0.5(x+1)^{3.5}$
measured with noise $\sigma = 0.1$



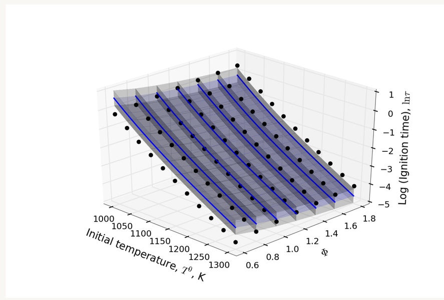
Left-over model error with increased data amount



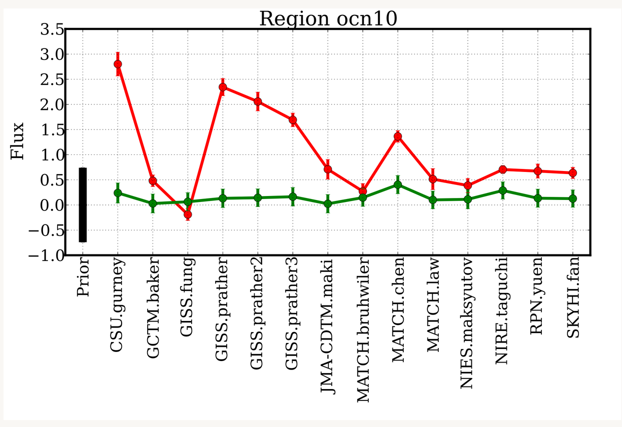
Applications

(1) Chemical ignition of methane: Model-to-model calibration

- Single-step global reaction model calibrated against a detailed chemical kinetic model



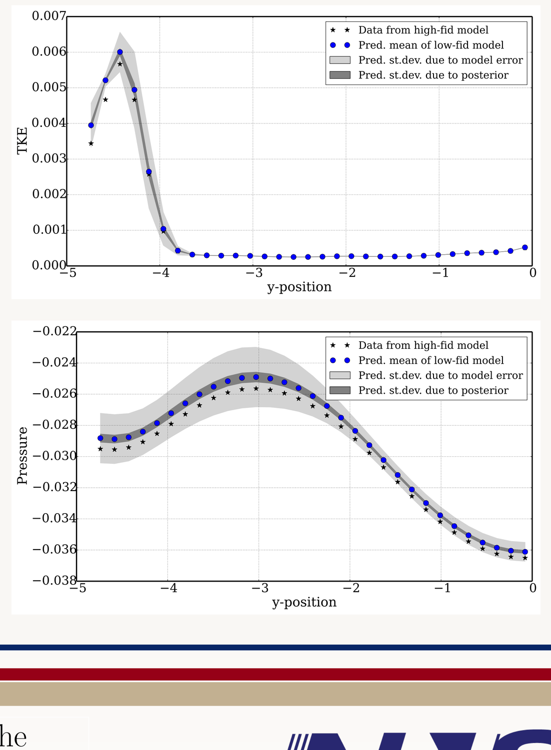
(2) Atmospheric transport: Multi-model calibration



- CO_2 measurements at 77 sites
  - Find fluxes at 22 locations
- Without model error, only 'effective' fluxes are computed

(3) Large eddy simulation of turbulent flow: Extrapolate to other QoIs

- Calibrate static subgrid model vs data from dynamic model
  - Use simulation data of turbulent kinetic energy (TKE), predict both TKE and Pressure



Support for this work was provided through the Scientific Discovery through Advanced Computing (SciDAC) program funded by the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.