SAND2019-12616PE

Uncertainty Quantification with Model Error

Habib N. Najm

Sandia National Laboratories
Livermore, CA, USA
hnnajm@sandia.gov

Seminar

Chinese Academy of Sciences
Beijing, China
Oct 23, 2019

Acknowledgement

K. Sargsyan, C. Safta,
M. Khalil, X. Huan, M. Eldred, G. Geraci, T. Casey, J. Oefelein,
G. Lacaze, Z. Vane, L. Hakim
— Sandia National Laboratories, CA

Y.M. Marzouk — Mass. Inst. of Tech., Cambridge, MA

This work was supported by:

- DOE Office of Basic Energy Sciences, Div. of Chem. Sci., Geosci., & Biosci.
- DOE Office of Advanced Scientific Computing Research (ASCR)
- DOE ASCR Scientific Discovery through Advanced Computing (SciDAC) program
- DARPA

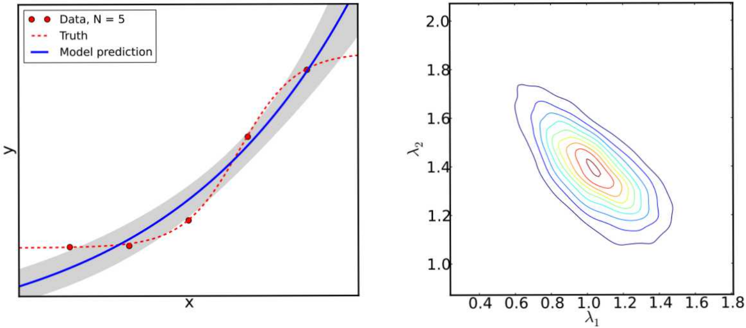
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Outline

Motivation

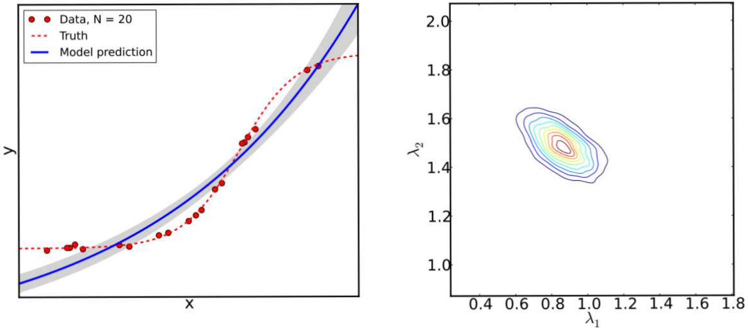
- All models are wrong in principle
- Models of physical systems rely on
  - Presumed theoretical framework
  - Mathematical formulation
- Practical models of complex physical systems rely on
  - Simplifying assumptions
  - Numerical discretization of governing equations
  - Computational software & hardware
- model error is frequently non-negligible
- Estimating model error is useful for
  - model comparison & validation
  - model improvement & scientific discovery
  - reliable computational predictions

Challenges with Model Calibration due to Model Error



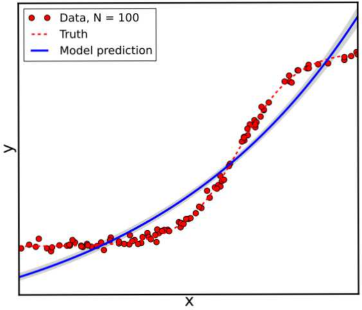
- Conventional parameter estimation context: $y_{\text{data}} = f(x, \lambda) + \epsilon_{\text{d}}$
- Additional data results in reduced parameteric posterior uncertainty
- One gets more confident about predictions with the wrong model
- Predictive uncertainty in calibrated model has no utility for prediction
- Ignoring model error leads to irrelevant predictive errors

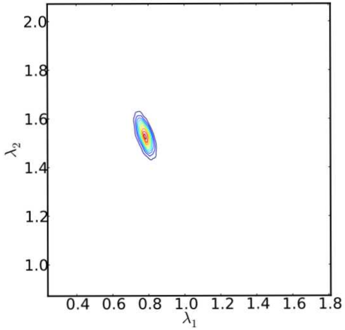
Challenges with Model Calibration due to Model Error



- Conventional parameter estimation context: $y_{\text{data}} = f(x, \lambda) + \epsilon_{\text{d}}$
- Additional data results in reduced parameteric posterior uncertainty
- One gets more confident about predictions with the wrong model
- Predictive uncertainty in calibrated model has no utility for prediction
- Ignoring model error leads to irrelevant predictive errors

Challenges with Model Calibration due to Model Error





- Conventional parameter estimation context: $y_{\text{data}} = f(x, \lambda) + \epsilon_{\text{d}}$
- Additional data results in reduced parameteric posterior uncertainty
- One gets more confident about predictions with the wrong model
- Predictive uncertainty in calibrated model has no utility for prediction
- Ignoring model error leads to irrelevant predictive errors

Statistical modeling of model error

Error framework:

Measurements: $y_{\text{data}} = y_{\text{truth}} + \epsilon_d$

Model predictions: $y_{\text{truth}} = y_{\text{model}} + \epsilon_m$

Thus: $y_{\text{data}} = y_{\text{model}} + \epsilon_m + \epsilon_d$

Error modeling – example

Model: $y_{\text{model}} = f(x, \lambda)$

Data Error: $\epsilon_d \sim \mathsf{N}(0, \sigma^2)$

Model Error: $\epsilon_m \sim \mathsf{GP}(\mu(x), C(x, x'))$

Model calibration:

Estimate model parameters λ along with those of ϵ_m, ϵ_d

Kennedy & O'Hagan 2001; Bayarri *et al.* 2002

Challenges – Physical Models

- Arbitrary choice of statistical model (*e.g.* GP) spatial structure does not take the physical model into acct
  - Potential violation of implicit constraints in physical models
  - *e.g.* incompressible flow: $\nabla \cdot v = 0$
- Difficulty in disambiguation of model & data error
- Calibration of model error on measured observable does not impact quality of other model predictions

Key idea - Targeted model error embedding

- Embed model error in specific submodel phenomenology (Berliner 2003)
  - a modified transport or constitutive law
  - a modified formulation for a material property
- Pros:
  - Allows placement of model error term in locations where key modeling assumptions and approximations are made
    - as a correction or high-order term
    - as a possible alternate phenomenology
  - explore if it can explain discrepancy on observable
  - naturally preserves model structure and associated constraints
- Cons:
  - complex likelihood $p(y|\lambda)$ for general nonlinear $f(x, \lambda, \epsilon_m)$

Consider a simple no-data-noise setting

- Calibration of a (simple) model against a complex model
- Let the complex model be presumed to represent the truth
- In this context, the data has no noise
- Discrepancy between model and data is all due to model error

$$y_{\text{data}} = y_{\text{truth}} = y_{\text{complex\_model}} = y_{\text{model}} + \epsilon_m$$

- $\epsilon_m = y_{\text{data}} - y_{\text{model}}$ is a deterministic quantity
- The only information as to the quality of the calibrated uncertain model, *e.g.* via a posterior predictive check, is in a unique ϵ_m for any x

model-to-model calibration

Model: $y = f(x, \lambda, \phi(\epsilon_m))$

- Random variable ϕ in augmented model components carries model error

Data: $D = \{(x_i, y_{\text{data},i}), i = 1, \dots, N\}$

- Goal:
  - Establish λ, $p(\phi)$ such that the likelihood of the data is high, based on the posterior predictive $p(y|D)$
- This puts us in a density estimation framework for ϕ:
  - The utility of additional data is to improve the specification of λ, and $p(\phi)$

Present Context

Embed ϵ_m in λ

- In other words: $\lambda \leftarrow \lambda + \epsilon_m$
- Model: $y = f(x, \lambda)$ with $\lambda : \Omega \to \mathbb{R}^M$
- Density estimation problem for $p(\lambda)$
- λ : a random field $\lambda(x, \omega)$, or a random variable $\lambda(\omega)$
  - focus on the latter
- Let the random variable λ be parameterized by α
  - For example, define λ as a polynomial chaos expansion

$$\lambda = \sum_{k=0}^{P} \alpha_k \Psi_k(\xi)$$

- Parameter estimation problem for $\alpha = (\alpha_0, \cdots, \alpha_P)$
- Bayesian setting
  - Prior $\pi(\alpha)$
  - Likelihood $L(\alpha) = p(D|\alpha)$

Full Likelihood

$$L(\alpha) = p(D|\alpha) = \pi_f(y_{\text{data},1}, \dots, y_{\text{data},N}|\alpha)$$

where:

$\pi_f(\cdot|\alpha)$: N-variate density of the random variable $(f_1, \dots, f_N)$ with $f_i = f(x_i, \lambda(\xi;\alpha))$

Problem: $\pi_f(\cdot)$ is degenerate in general when $N > M$

- Consider a case with $M = 1$, $\lambda \sim N(\mu, \sigma^2)$, and $f = \lambda$
- Let $N = 2$, hence $(f_1, f_2) = (\lambda, \lambda)$ for any λ sample
- With $f_1 = f_2 = \lambda$, (f_1, f_2) are dependent and $\pi_f(\cdot|\mu,\sigma)$ is non-zero only along the line $f_2 = f_1$
- $\pi_f(y_{\text{data},1}, y_{\text{data},2}|\mu,\sigma)$ is non-zero only along the line $y_{\text{data},2} = y_{\text{data},1}$
  $\Rightarrow$ potentially can ameliorate singularity with a smoothing nugget

Marginalized Likelihood

$$L(\alpha) = p(D|\alpha) = \prod_{i=1}^{N} \pi_{f_i}(y_{\text{data},i}|\alpha)$$

where

$\pi_{f_i}(\cdot, \alpha)$ is the univariate density of the RV $f_i = f(x_i, \lambda(\alpha))$

Problem: the likelihood has multiple singularities corresponding to α values leading to vanishing marginal variances at each x_i

- Gaussian example: Let $f_i \sim \text{N}(\mu_i(\alpha), \sigma_i(\alpha)^2)$, then

$$L(\alpha) = \frac{1}{(2\pi)^{N/2}} \prod_{i=1}^{N} \frac{1}{\sigma_i(\alpha)} \exp\left(\frac{(\mu_i(\alpha) - y_{\text{data},i})^2}{2\sigma_i(\alpha)^2} \right)$$

- Multiple singularities, $\sigma_i(\alpha) = 0$, $i = 1, \ldots, N$
- Posterior maximization always finds one of these singularities, fitting one point perfectly, while misfitting the rest

$\Rightarrow$ can potentially be controlled via priors on α

Approximate Bayesian Computation (ABC)

Employ a kernel density as a pseudo-likelihood to enforce select constraints:

Uncertain prediction $p(y|D)$ is centered on the data

- With $\mu_i(\alpha) = \mathsf{E}_\xi[f(x_i, \lambda(\xi;\alpha))]$:

$$\text{minimize} \parallel \mu_i(\alpha) - y_{\text{data},i} \parallel$$

The width of the distribution $p(y|D)$ is consistent with the spread of the data around the nominal model prediction

- With $\sigma_i^2(\alpha) = \mathsf{V}_\xi[f(x_i, \lambda(\xi,\alpha))]$:

$$\text{minimize} \parallel \sigma_i(\alpha) - \gamma|\mu_i(\alpha) - y_{\text{data},i}| \parallel$$

- γ is a factor that specifies the desired match between σ_i and the discrepancy $|\mu_i(\alpha) - y_{\text{data},i}|$, on average

ABC Likelihood

With $\rho(\mathcal{S})$ being a metric of the statistic $\mathcal{S}$, use the kernel function as an ABC likelihood:

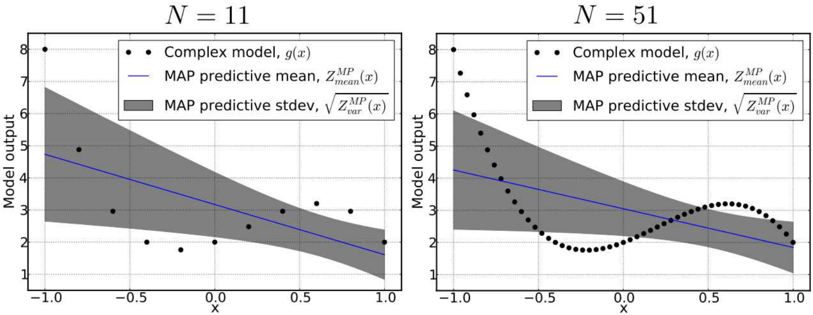
$$L_{\mathsf{ABC}}(\alpha) = \frac{1}{\epsilon} K\left(\frac{\rho(\mathcal{S})}{\epsilon}\right)$$

where ϵ controls the severity of the consistency control

Propose the Gaussian kernel density:

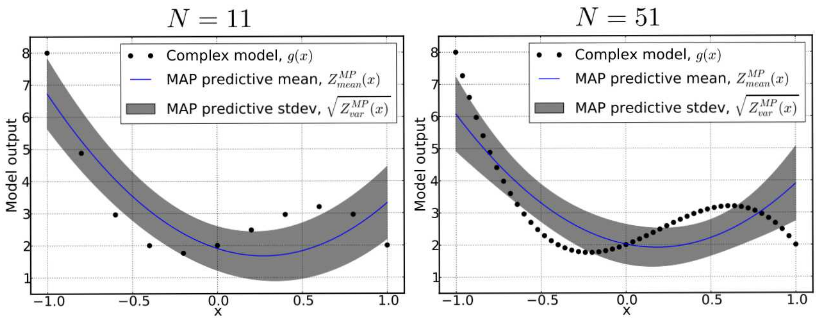
$$L_{\epsilon}(\alpha) = \frac{1}{\epsilon\sqrt{2\pi}} \prod_{i=1}^{N} \exp\left(-\frac{(\mu_i(\alpha) - y_{\mathsf{d},i})^2 + (\sigma_i(\alpha) - \gamma|\mu_i(\alpha) - y_{\mathsf{d},i}|)^2}{2\epsilon^2}\right)$$

Test problem – Cubic data fit by a line – ABC



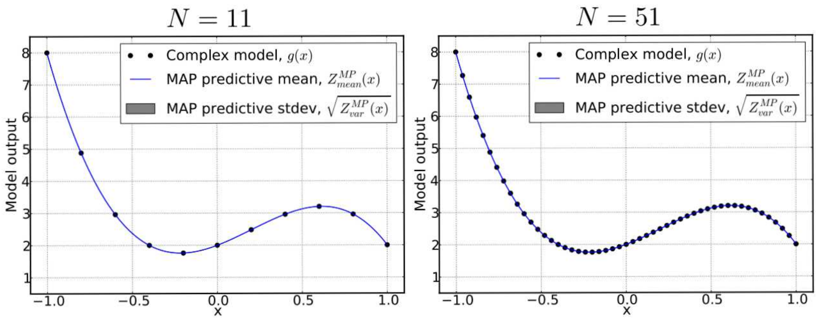
- MAP predictive (MP) mean centered on data
- MP standard deviation captures range of discrepancy
- Increasing number of data points has a small effect on both MP mean and stdev

Test problem – Cubic data fit by a quadratic – ABC



- Quadratic has better fit to the data
- Smaller MP stdev consistent with smaller discrepancy

Test problem – Cubic data fit by a cubic – ABC



- Cubic has perfect fit to the data
- Negligible MP stdev consistent with negligible discrepancy

Chemistry problem – ABC

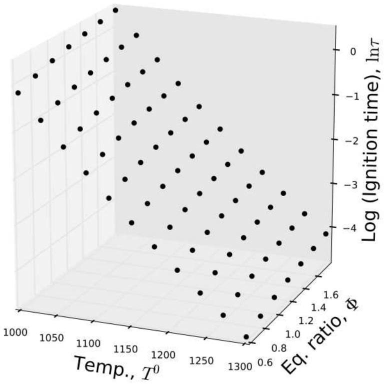
- Homogeneous ignition, methane-air mixture
- Single-step global reaction model calibrated against a detailed chemical kinetic model – ODE system
- Data: ignition time; range of initial T & equivalence ratio
- Single-step model:

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$$

$$\mathfrak{R} = [CH_4][O_2]k$$

$$k = A\,\mathrm{exp}(-E/R^oT)$$

$$\lambda = \begin{bmatrix} \ln A \\ E \end{bmatrix} = \sum_{k=0}^{P} \boldsymbol{\alpha}_k \Psi_k(\xi)$$



Constant Pressure Ignition – Problem Structure

- N species, M reactions, rate parameter vector λ
- State vector $u = (X_1, \dots, X_N, T)$ – mole fractions, temperature
- ODE system

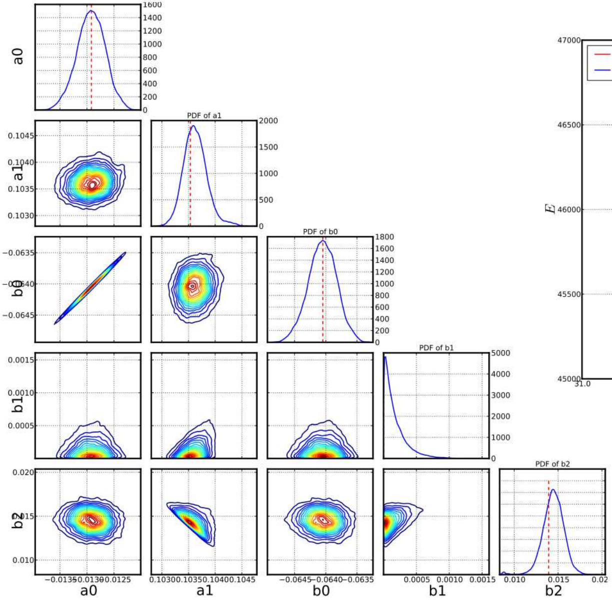
$$\begin{aligned} \frac{\mathrm{d}u_i(t;\lambda)}{\mathrm{d}t} &= w_i(u;\lambda), \quad i = 1, \dots, N \\ u(0) &= u_0 \end{aligned}$$

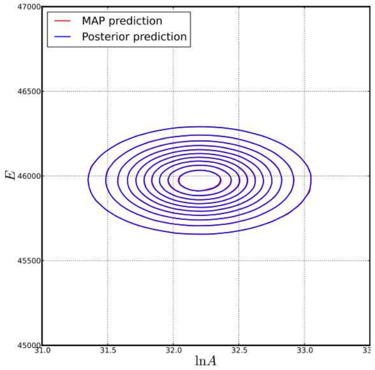
- Observable: ignition time $\tau_{\text{ign}}(u_0, \lambda) = t\,|_{T(t;u_0,\lambda)=T_{\text{ign}}}$
- Challenge, for any proposed λ, computing $\tau_{\text{ign}}(u_0, \lambda)$ is expensive
  - Large stiff ODE system for complex fuels
- Polynomial chaos formulation allows construction of a surrogate

$$\tau_{\text{ign}}(u_0, \lambda(\xi;\alpha)) = f(u_0, \xi; \alpha) = \sum_{k=0}^{P} f_k(u_0;\alpha)\Psi_k(\xi)$$

- Surrogate replaces the forward model in the Likelihood function

Posterior on α – Posterior Predictive on $(\ln A, E)$



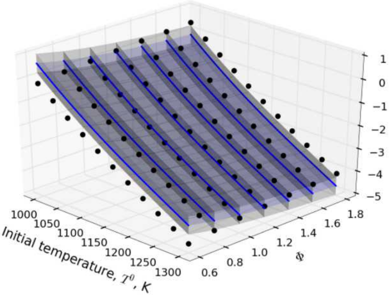


Quality of Uncertain Calibrated Model Predictions

Calibrated uncertain fit model is consistent with the detailed-model data.

Over the range of (T^0, Φ):

- MAP predictive mean ignition-time is centered on the data
- MAP predictive stdv is consistent with the scatter of the data



Consider a noisy-data setting

- Calibration of a model $y_m = f(x, \lambda)$ against noisy data
- Synthetic noisy data is generated from a "truth" model + Gaussian noise
- Discrepancy between fit model prediction and data is due to both model error & data noise

$$y = y_{\text{data}} = y_{\text{truth}} + \epsilon = f(x, \lambda) + \epsilon$$

- Modeling strategy:
  - Model λ as a random vector, represented with PC
  - Represent the noise similarly using PC
  - Estimate all PC coefficients using Bayesian inference

Model Error formulation – noisy data

$$y = f(x, \lambda) + \epsilon$$

Let $\epsilon \sim N(0, \sigma^2)$. With N *i.i.d.* data points we have

$$y_i = f(x_i, \lambda) + \epsilon_i, \quad i = 1, \dots, N$$

For Hermite-Gaussian PC:

$$\begin{aligned}
\lambda &= \sum_{k=0}^{P} \alpha_k \Psi_k(\xi_1, \cdots, \xi_d), \qquad \alpha \equiv (\alpha_0, \cdots, \alpha_P) \\
f(x, \lambda) &= \sum_{k=0}^{P} f_k(x, \alpha) \Psi_k(\xi_1, \cdots, \xi_d) \\
y_i &= \sum_{k=0}^{P} f_k(x_i, \alpha) \Psi_k(\xi_1, \cdots, \xi_d) + \sigma \xi_{d+i}
\end{aligned}$$

Augmented PC germ $\xi = (\underbrace{\xi_1, \cdots, \xi_d}_{\epsilon_m}, \underbrace{\xi_{d+1}, \cdots, \xi_{d+N}}_{\epsilon_d})$

Model Error Estimation – noisy data

Inverse problem:

- Given:
  - data:

$$D = \{(x_i, y_i)\}_{i=1}^N$$

  - data model:

$$y_i = \underbrace{\sum_k f_k(x_i, \alpha) \Psi_k(\xi_1, \cdots, \xi_d)}_{y_{\text{model}}(\epsilon_{\text{m}})} + \underbrace{\sigma\, \xi_{d+i}}_{\epsilon_{\text{d}}}, \qquad i = 1, \ldots, N$$

- Estimate parameters (α, σ)

Bayesian context:

- posterior: $p(\alpha, \sigma | D)$
- options: Full Bayesian likelihood; Marginalized; ABC
- All are viable here in principle, as the data noise introduces regularity
- We illustrate the case with a Marginalized Gaussian approximation

Calibrated Uncertain Model Posterior Predictive

- Calibrated data model : $y_i = f(x_i; \lambda(\xi; \alpha)) + \sigma\xi_{d+i}$
- Full posterior on α, σ : $\alpha, \sigma \sim p(\alpha, \sigma | D)$
- Marginal posteriors: $\alpha \sim p(\alpha|D)$, $\sigma \sim p(\sigma|D)$
- Posterior Predictive (PP):

$$p(y|D) = \int p(y|\alpha,\sigma)p(\alpha,\sigma|D)\mathrm{d}\alpha\mathrm{d}\sigma = \mathbb{E}_{\alpha,\sigma}[p(y|\alpha,\sigma)]$$

- PP Mean :

$$\mathbb{E}_{\mathsf{PP}}[y] = \mathbb{E}_{\alpha}[\mathbb{E}_{\xi}[f]]$$

- PP Variance:

$$\mathbb{V}_{\mathsf{PP}}[y] = \underbrace{\mathbb{E}_{\alpha}[\ \mathbb{V}_{\xi}[\ f\ ]\ ]}_{\text{model error}} + \underbrace{\mathbb{E}_{\sigma}[\ \sigma^2\ ] + \mathbb{V}_{\alpha}[\ \mathbb{E}_{\xi}[\ f\ ]\ ]}_{\text{data noise}}$$

Calibrated Uncertain Model Predictions

- Calibrated model : $y = f(x; \lambda(\xi; \alpha))$
- Marginal posterior on α : $\alpha \sim p(\alpha|D)$
- Pushed forward posterior (PFP):

$$p(f|D) = \int p(f|\alpha)p(\alpha|D)\mathrm{d}\alpha = \mathbb{E}_\alpha[p(f|\alpha)]$$

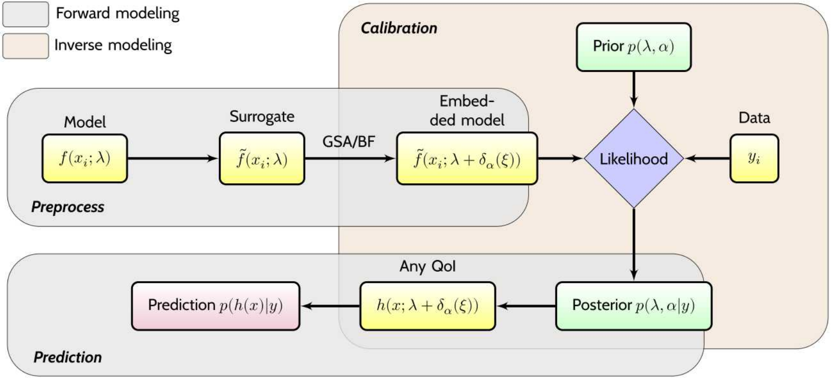
- PFP Mean :

$$\mathbb{E}_{\text{PFP}}[f] = \mathbb{E}_\alpha[\mathbb{E}_\xi[f]]$$

- PFP Variance:

$$\mathbb{V}_{\text{PFP}}[f] = \underbrace{\mathbb{E}_\alpha[\ \mathbb{V}_\xi[\ f\ ]\ ]}_{\text{model error}} + \underbrace{\mathbb{V}_\alpha[\ \mathbb{E}_\xi[\ f\ ]\ ]}_{\text{data noise}}$$

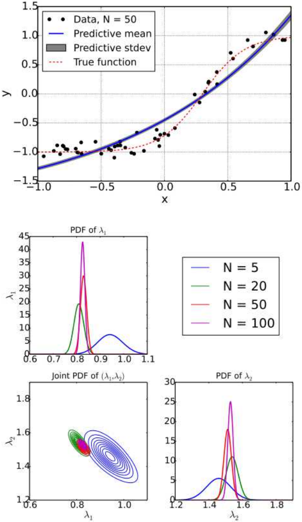
Model error embedding – workflow

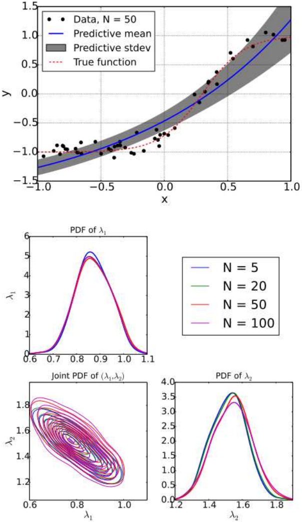


- Predictive uncertainty decomposition: Total Variance =

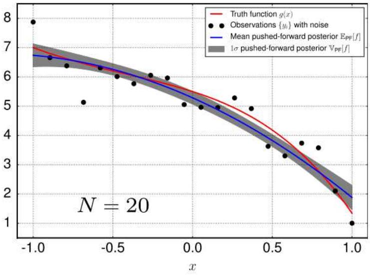
  Posterior uncertainty + Data noise + Model error + Surrogate error

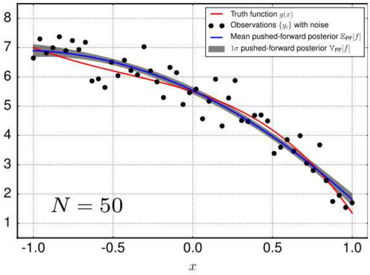
.. back to toy example



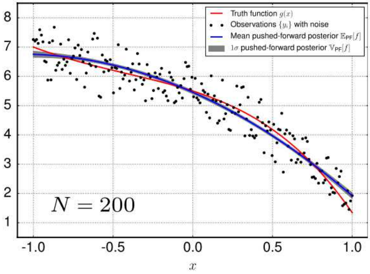


Quadratic-fit – Classical Bayesian likelihood

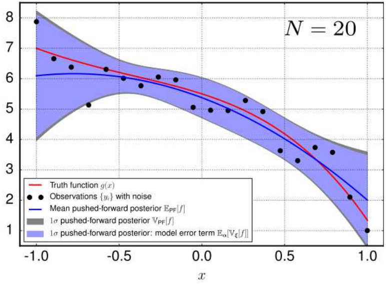


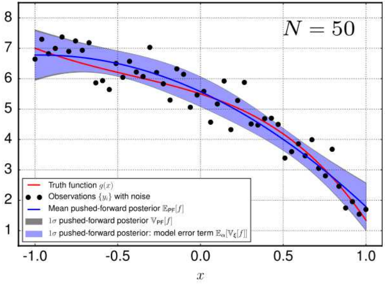


- With additional data, predictive uncertainty around the wrong model is indefinitely reducible
- Predictive uncertainty not indicative of discrepancy from truth

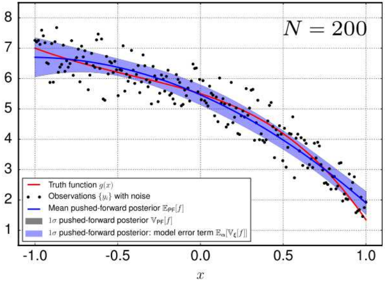


Quadratic-fit – ModErr – MargGauss

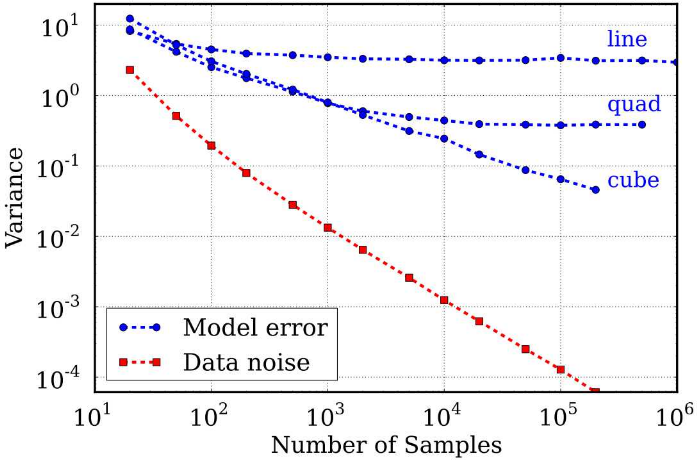




- With additional data, predictive uncertainty due to data noise is reducible
- Predictive uncertainty due to model error is not reducible



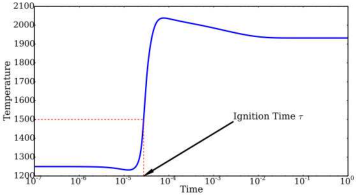
Model Error – Fit with Different Models



Ignition time in chemical kinetics

- Two-step global reaction model calibrated against shock tube experimental data
- Operating conditions: pressure P, initial temperature T_0 & equivalence ratio ϕ

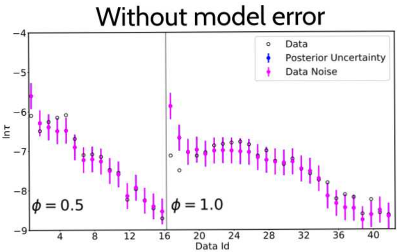
$$\begin{aligned} C_{12}H_{26} + \tfrac{25}{2}O_2 &\overset{k_1}{\rightarrow} 12CO + 13H_2O \\ CO + \tfrac{1}{2}O_2 &\underset{k_{2b}}{\overset{k_{2f}}{\rightleftharpoons}} CO_2. \end{aligned}$$

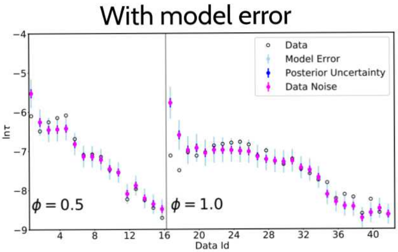


$$k_1 = Ae^{\left(-\frac{E}{RT}\right)}[C_{12}H_{26}]^{0.25}[O_2]^{1.25}$$

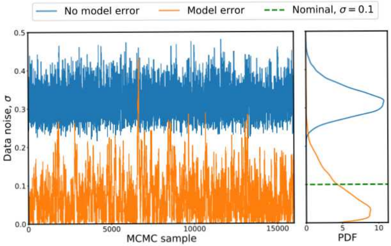
- Data: log(ignition time)
- Embedding $(\ln A, E) = \sum_k \boldsymbol{\alpha}_k \Psi_k(\xi)$

Ignition time in chemical kinetics





- Model error disambiguated from data error
- Data error correctly captured
- Meaningful extrapolative predictions



LES subgrid static-vs-dynamic – Jet-in-crossflow

Large Eddy Simulation (LES) subgrid model fidelity

- Dynamic: subgrid parameters variable in space/time, g_i
- Static : subgrid parameters constant in space/time, $f_i(\lambda)$

Target: Calibrate a static model against a dynamic model

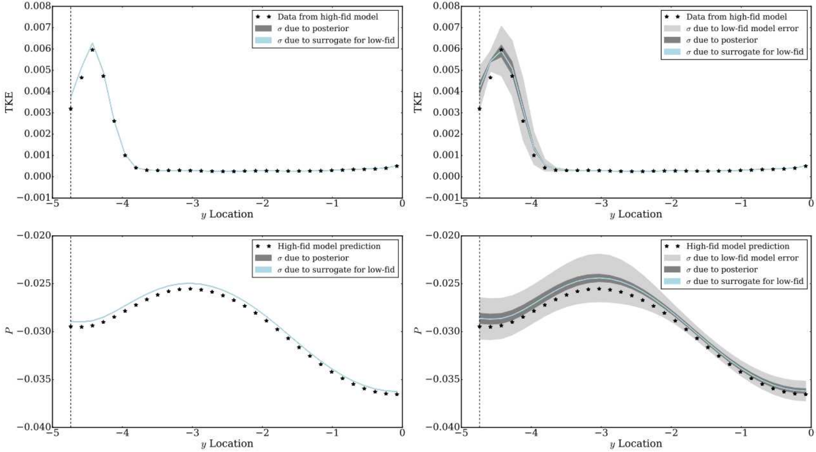
- Fit parameters $\lambda = (C_R, Pr_t^{-1}, Sc_t^{-1})$ of static model $f(\lambda)$ to data from dynamic model simulations, accounting for model error

Static model surrogate uses $4^3 = 64$ simulations of $f(\lambda)$

- Legendre polynomial expansion surrogate of 3-rd order
  - Account for surrogate error: *i.i.d.* zero-bias Gaussian noise
- Global sensitivity analysis: impact of $C_R \gg$ that of Pr_t^{-1} & Sc_t^{-1}
  - Selected only C_R for model error embedding

$$\mathbb{V}_{\text{PFP}}[f] = \underbrace{\mathbb{E}_\alpha[\ \mathbb{V}_\xi[\,f\,]\ ]}_{\text{model error}} + \underbrace{\mathbb{V}_\alpha[\ \mathbb{E}_\xi[\,f\,]\ ]}_{\alpha\ \text{posterior}} + \underbrace{\mathbb{E}_{\sigma_S}[\,\sigma_S^2\,]}_{\text{surrogate error}}$$

Calibrate with TKE data; Predict both TKE and Pressure



No model error

With model error

LES subgrid 2D-vs-3D – Jet-in-crossflow

Target: Calibrate a 2D LES model against a 3D model

- Fit parameters $\lambda = (C_R, Pr_t^{-1}, Sc_t^{-1}, I_i, I_r, L_i)$ of 2D model to data from 3D model simulations, accounting for model error
- Parameters:
  - C_R : Smagorinsky constant
  - Pr_t : Turbulent Prandtl number, and Schmidt number: Sc_t
  - I_i : Turb. intensity (inflow air) horizontal component
  - I_r : Turb. intensity (inflow air) ratio: vertical/horizontal
  - L_i : Length scale of most energetic eddies

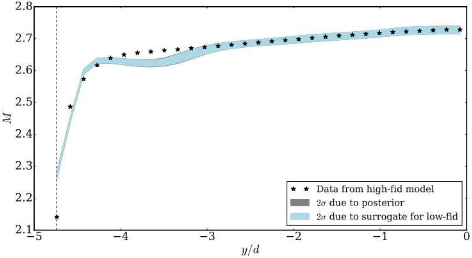
2D model surrogate construction

- Account for surrogate error: *i.i.d.* zero-bias Gaussian noise
- Global sensitivity analysis
  - Selected one parameter (I_i) for model error embedding
- Calibrate 2D model with observable: Mach no. $M(y)$ at a given x
- Predict both $M(y)$ and pressure $P(y)$, and compare to 3D model

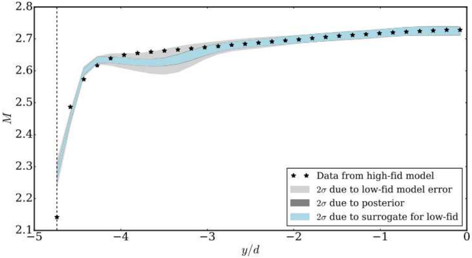
Model Error – Mach Number spanwise-average

- Model error contribution captures the discrepancy

No model error



Embedded model error

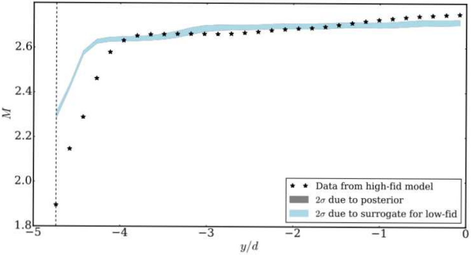


$$\sigma_i^2 = \underbrace{\mathbb{E}_{\tilde{\lambda}}\left[\sigma_i^2(\tilde{\lambda})\right]}_{\text{Model error}} + \underbrace{\mathbb{V}_{\tilde{\lambda}}\left[\mu_i(\tilde{\lambda})\right]}_{\text{Posterior uncertainty}} + \underbrace{(\sigma_i^{LOO})^2}_{\text{Surrogate error}}$$

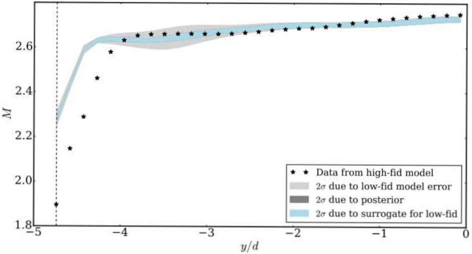
Model Error – Mach Number centerline

- Model error contribution extends as much as prior & model allow

No model error



Embedded model error



$$\sigma_i^2 = \underbrace{\mathbb{E}_{\tilde{\lambda}}\left[\sigma_i^2(\tilde{\lambda})\right]}_{\text{Model error}} + \underbrace{\mathbb{V}_{\tilde{\lambda}}\left[\mu_i(\tilde{\lambda})\right]}_{\text{Posterior uncertainty}} + \underbrace{(\sigma_i^{LOO})^2}_{\text{Surrogate error}}$$

Closure

- Presented a strategy for dealing with model error
  - targeted at physical models
- Density estimation framework – $y = f(x; \lambda(\xi; \alpha))$
- Uncertain predictions with the calibrated model include uncertainty due to both model-error and data-noise
- Results suggest disambiguation of the two components
- Demonstrations in chemical ignition and LES of jet-in-crossflow
  - Including accounting for PC surrogate error
- Limitation of model-error embedding: when no variation of the chosen parameter in the simple model could reproduce results of the detailed model
  - Expand parameter prior range(s)
  - Consider other parameters
  - Propose a modification in the model