$\mathcal{O}n$ Model Error and Statistical Calibration of Physical Models

H. Najm[1], K. Sargsyan[1], J. Bender[1], and R. Ghanem[2]

[1] Sandia National Laboratories
Livermore, CA
[2] University of Southern California
Los Angeles, CA

Acknowledgement

C. Safta,
– Sandia National Laboratories, CA

Y.M. Marzouk – Mass. Inst. of Tech., Cambridge, MA, USA

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: $y_{\mathsf{data}} = f(x, \lambda) + \epsilon_{\mathsf{d}}$
- 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: $y_{\mathsf{data}} = f(x, \lambda) + \epsilon_{\mathsf{d}}$
- 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

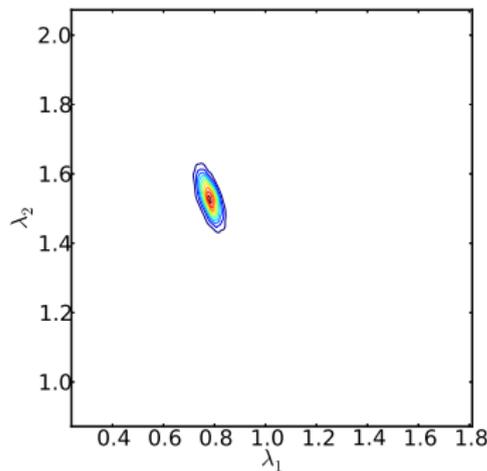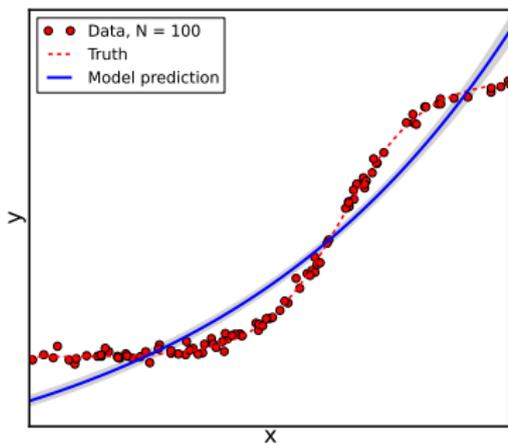# Challenges with Model Calibration due to Model Error

- Conventional parameter estimation context: $y_{\mathsf{data}} = f(x, \lambda) + \epsilon_{\mathsf{d}}$
- 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

- Physical scientists are unlikely to augment their model with a statistical model error term on select outputs

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

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

- $\epsilon_m = y_{\mathsf{data}} - y_{\mathsf{model}}$ 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: $\quad y = f(x, \lambda, \phi(\epsilon_m))$

- Random variable ϕ in augmented model components carries model error

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

- 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

- Model: $\quad y = f(x, \lambda)$ 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 $L(\alpha) = p(D|\alpha)$

Polynomial Chaos Expansion (PCE)

- Given a *germ* $\boldsymbol{\xi}(\omega) = \{\xi_1, \cdots, \xi_n\}$ - a set of *i.i.d.* RVs
  - where $p(\boldsymbol{\xi})$ is uniquely determined by its moments

Any RV in $L^2(\Omega, \mathfrak{S}(\boldsymbol{\xi}), P)$ can be written as a PCE:

$$u(\omega) = f(\boldsymbol{\xi}) = \sum_{k=0}^{\infty} u_k \Psi_k(\boldsymbol{\xi}(\omega))$$

- u_k are mode strengths
- $\Psi_k()$ are functions orthogonal w.r.t. $p(\boldsymbol{\xi})$

Orthogonal basis examples:

- Hermite polynomials with Gaussian germ
- Legendre polynomials with Uniform germ, ...
- Global versus Local PC methods
  - Adaptive domain decomposition of the support of $\boldsymbol{\xi}$

Full Likelihood

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

where:

$\pi_f(\cdot|\alpha)$: N-variate density of the random variable $(f_1, \ldots, f_N)$
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 $f_2 = f_1$

- $\pi_f(y_{\mathsf{data},1}, y_{\mathsf{data},2}|\mu, \sigma)$ is non-zero only along the line $y_{\mathsf{data},2} = y_{\mathsf{data},1}$

    $\Rightarrow$ potentially can ameliorate singularity with a smoothing nugget

Marginalized Likelihood

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

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 $f_i \sim \mathsf{N}(\mu_i(\alpha), \sigma_i(\alpha)^2)$, then

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

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

    minimize $\| \mu_i(\alpha) - y_{\mathsf{data},i} \|$

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

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

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

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

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

Likelihood construction – variants/approxmiations

- Full Likelihood
$$L(\alpha) = p(D|\alpha) = p(y_{\text{data},1}, \ldots, y_{\text{data},N}|\alpha)$$

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

- Approximate Bayesian Computation:
  Seek to satisfy the constraints:
  - $p(y|D)$ is "centered" on the data
  - The width of the distribution $p(y|D)$ is "consistent" with the spread of the data around the nominal model prediction

Prediction

$$
\begin{aligned}
p(\alpha|D) &= p(\alpha_0, \ldots, \alpha_P|D) \\
\lambda(\xi; \alpha) &= \sum_{k=0}^{P} \alpha_k \Psi_k(\xi) \\
Z(x, \xi; \alpha) &= f(x, \lambda(\xi; \alpha))
\end{aligned}
$$

MAP predictive RV $\qquad Z^{\mathsf{MP}}(x) = Z(x; \alpha_{\mathsf{MAP}})$

MAP predictive mean $\qquad Z^{\mathsf{MP}}_{\mathsf{mean}}(x) = \mathbb{E}_{\xi}[Z|\alpha_{\mathsf{MAP}}]$

MAP predictive variance $\qquad Z^{\mathsf{MP}}_{\mathsf{var}}(x) = \mathbb{V}_{\xi}[Z|\alpha_{\mathsf{MAP}}]$

Test problem – Cubic data fit by a line – ABC
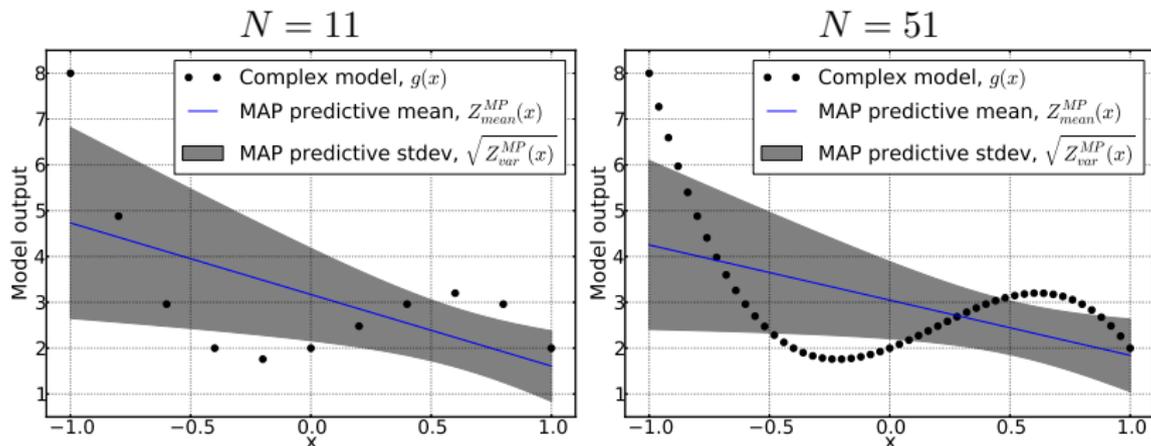
- MAP predictive (MP) mean centered on data
- MP standard deviation captures range of discrepancy
- Increasing number of data points has a small effect on both MP mean and stdev

Test problem – Posterior density on α

- Cubic data, line-fit
- Joint posterior on two elements of α

- Uncertainty in α is decreased by
  - Increasing N
  - Decreasing ϵ
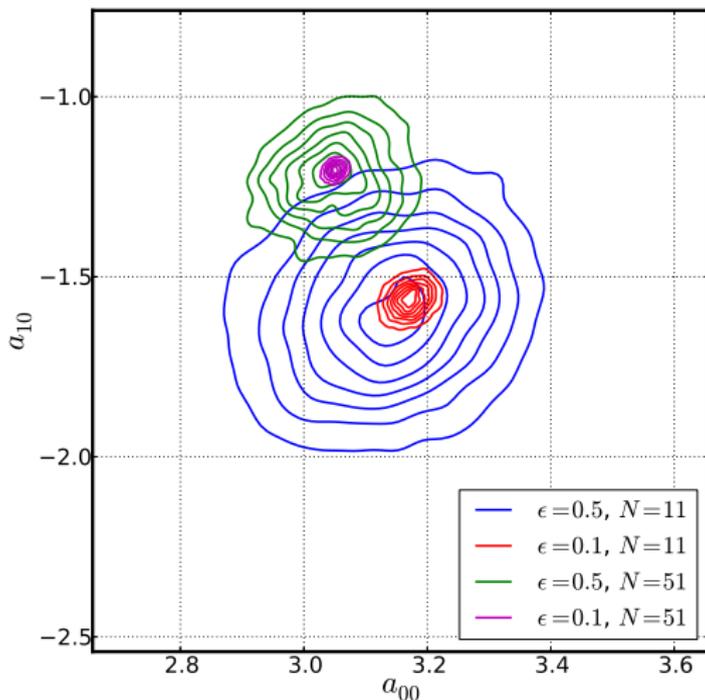
Test problem – Cubic data fit by a quadratic – ABC
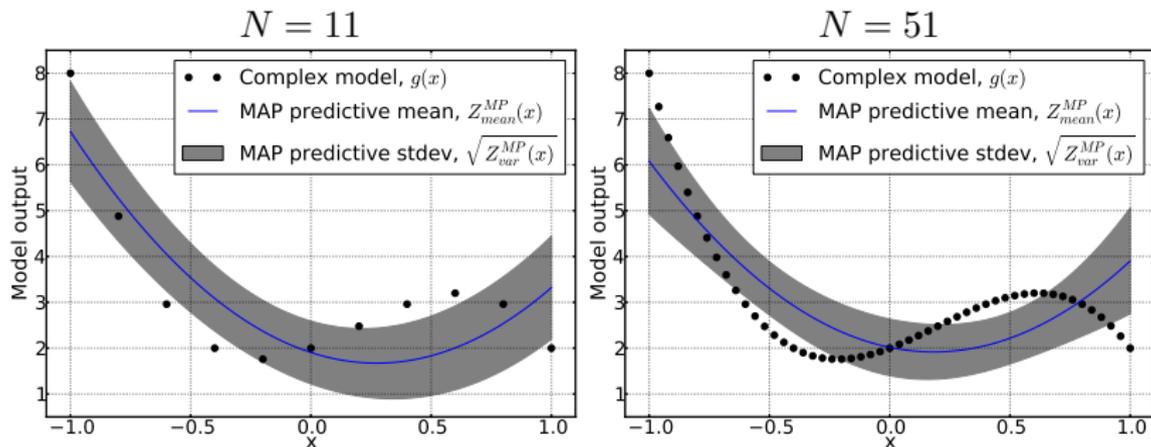
- Quadratic has better fit to the data
- Smaller MP stdev consistent with smaller discrepancy

Test problem – Cubic data fit by a cubic – ABC
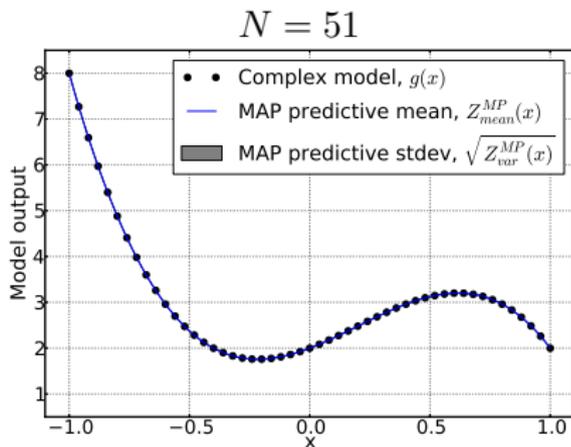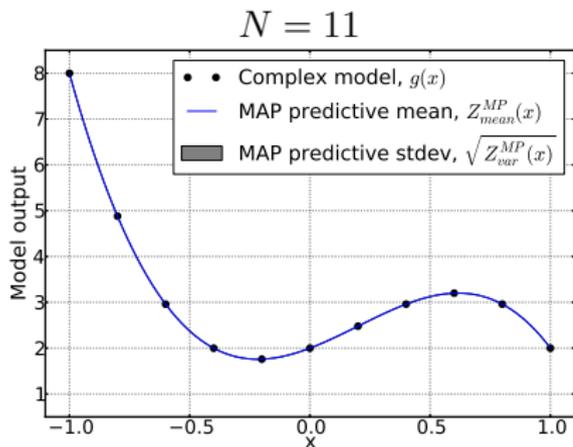
- Cubic has perfect fit to the data
- Negligible MP stdev consistent with negligible discrepancy

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:
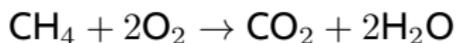
$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$$
$$\mathfrak{R} = [CH_4][O_2]k$$
$$k = A\exp(-E/R^oT)$$

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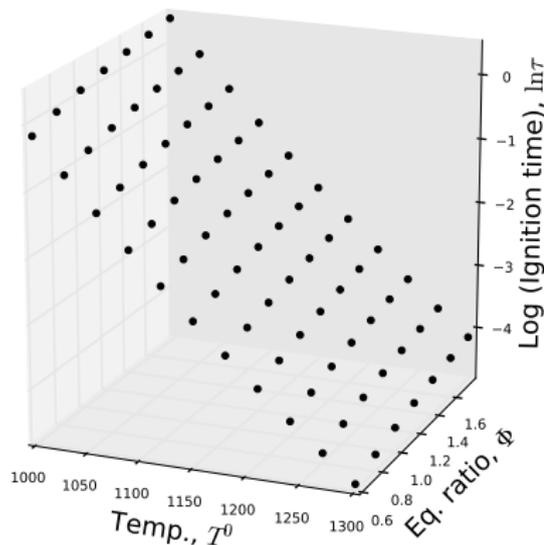
Constant Pressure Ignition – Problem Structure

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

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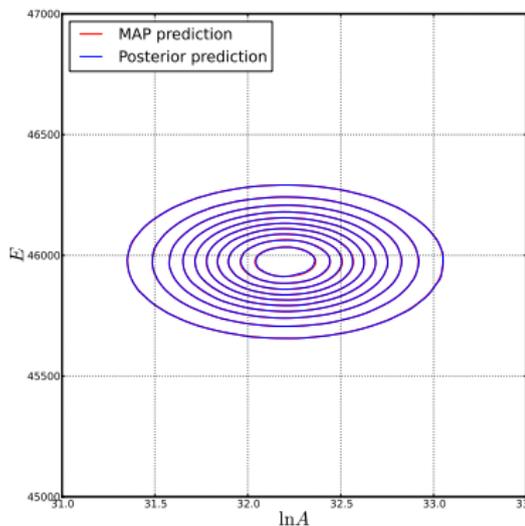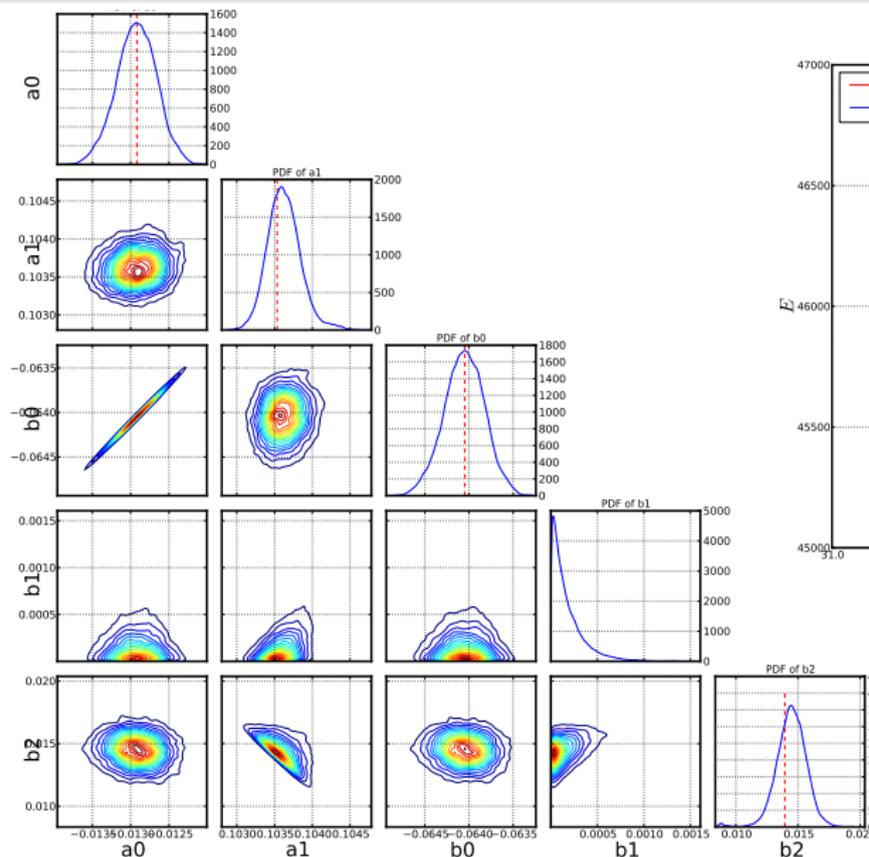u(0) &= u_0
\end{aligned}
$$

- Observable: ignition time $\quad \tau_{\text{ign}}(u_0, \lambda) = t \,|_{T(t; u_0, \lambda) = T_{\text{ign}}}$
- 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

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

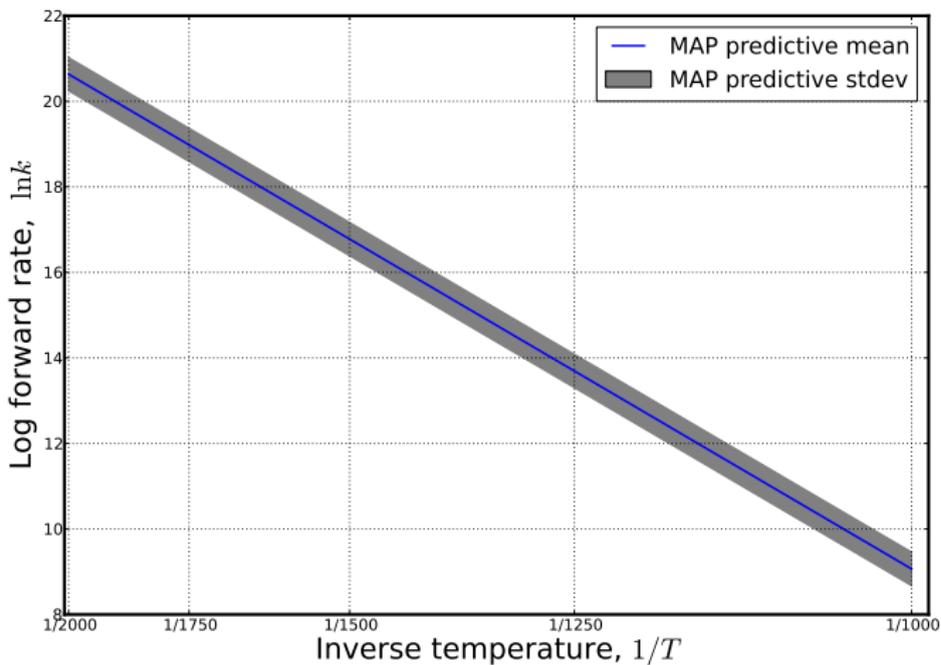- Surrogate replaces the forward model in the Likelihood function

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

Posterior predictive distribution on $\ln k$
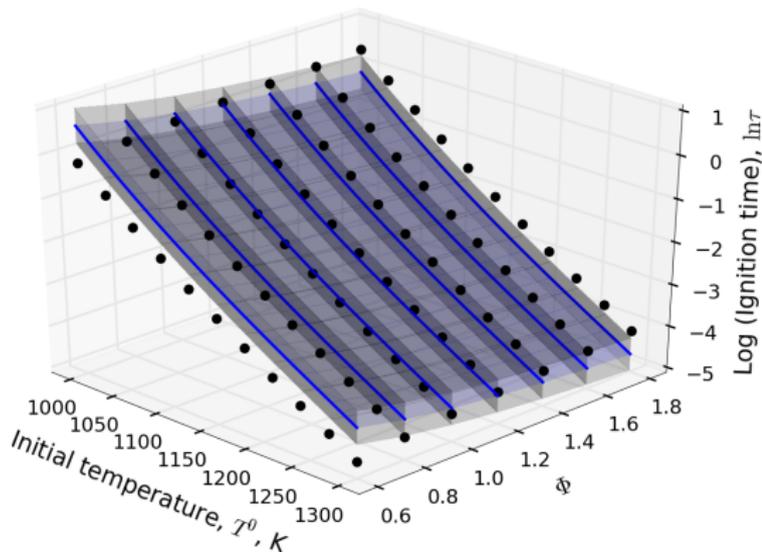
$$\mathsf{E}_\alpha[p(\ln k|D)]$$

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

n-dodecane ignition – 1
with L. Hakim, M. Khalil, J. Oefelein, and G. Lacaze, SNL

- Homogeneous ignition, n-dodecane/air mixture
- Two-step global reaction model calibrated against a "detailed" model
  - Reference chemical kinetic model: 255 species, 2289 reactions
    Narayanaswamy *et al.* 2014
- Data: ignition time; range of initial T & equivalence ratio Φ
- Two-step model:

$$C_{12}H_{26} + \tfrac{25}{2}O_2 \longrightarrow 12CO + 13H_2O$$
$$CO + \tfrac{1}{2}O_2 = CO_2$$
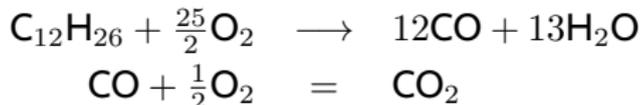
$$\mathfrak{R}_1 = Ae^{\frac{-E}{RT}}[C_{12}H_{26}]^{0.25}[O_2]^{1.25}$$

$$\mathfrak{R}_{2f} = 3.98 \cdot 10^{14} e^{\frac{-4 \cdot 10^4}{RT}}[CO][H_2O]^{0.5}[O_2]^{0.25}$$

$$\mathfrak{R}_{2b} = 5 \cdot 10^8 e^{\frac{-4 \cdot 10^4}{RT}}[CO_2]$$

- $E = \lambda_0$, $\ln A = \lambda_1 + \lambda_2 e^{\lambda_3 \Phi} + \lambda_4 \tanh((\lambda_5 + \lambda_6 \Phi)T_0 + \lambda_7$
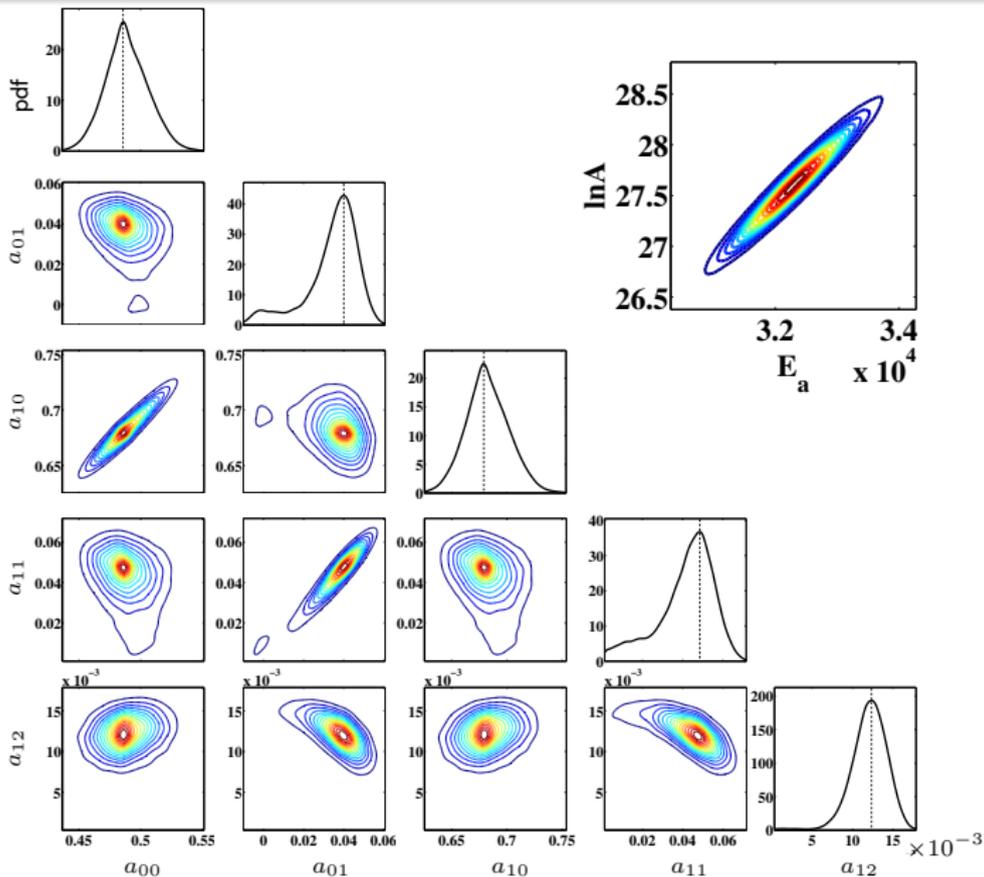
n-dodecane ignition – 2

- $(E,\ \ln A) = (\lambda_0,\ \lambda_1 + \lambda_2 e^{\lambda_3 \Phi} + \lambda_4 \tanh((\lambda_5 + \lambda_6 \Phi)T_0 + \lambda_7)$
- Original parameter vector $\lambda = (\lambda_0, \ldots, \lambda_7)$
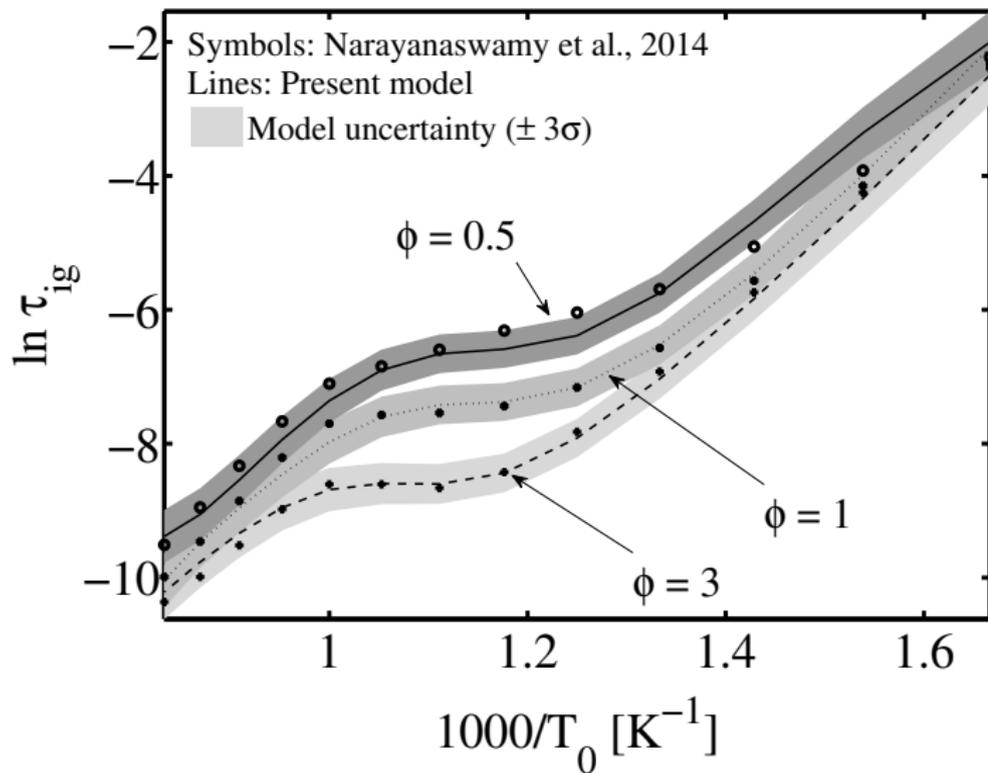- Embed model error in (λ_0, λ_1)
- PCE model:

$$
\begin{aligned}
\lambda_0 &= a_{00} + a_{01}\xi_1 \\
\lambda_1 &= a_{10} + a_{11}\xi_1 + a_{12}\xi_2 \\
\lambda_2 &= a_{20} \\
&\vdots \\
\lambda_7 &= a_{70}
\end{aligned}
$$

- ABC targets parameters $(a_{00}, a_{01}, a_{10}, a_{11}, a_{12}, a_{20}, \ldots, a_{70})$

n-dodecane ignition – 3

n-dodecane ignition – 4

Symbols: Narayanaswamy et al., 2014
Lines: Present model
Model uncertainty ($\pm 3\sigma$)

$\phi = 0.5$

$\phi = 1$

$\phi = 3$

$\ln \tau_{ig}$

$1000/T_0$ [K^{-1}]

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

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

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 : $\quad \alpha, \sigma \sim p(\alpha, \sigma | D)$

- Marginal posteriors: $\quad \alpha \sim p(\alpha | D), \sigma \sim p(\sigma | D)$

- Posterior Predictive (PP):

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

- 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 : $\quad \alpha \sim p(\alpha|D)$

- Pushed forward posterior (PFP):

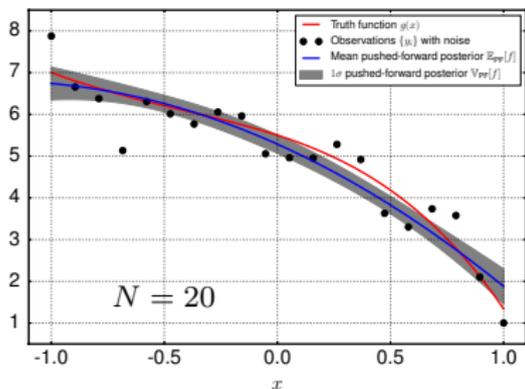$$p(f|D) = \int p(f|\alpha)p(\alpha|D)\mathbf{d}\alpha = \mathbb{E}_\alpha[p(f|\alpha)]$$

- PFP Mean :

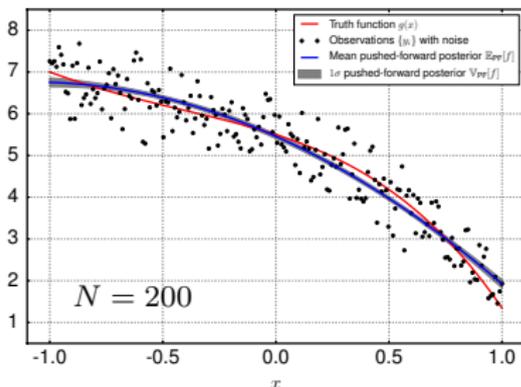$$\mathbb{E}_{\mathsf{PFP}}[f] = \mathbb{E}_\alpha[\mathbb{E}_\xi[f]]$$

- PFP Variance:

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

# Quadratic-fit – Classical Bayesian likelihood
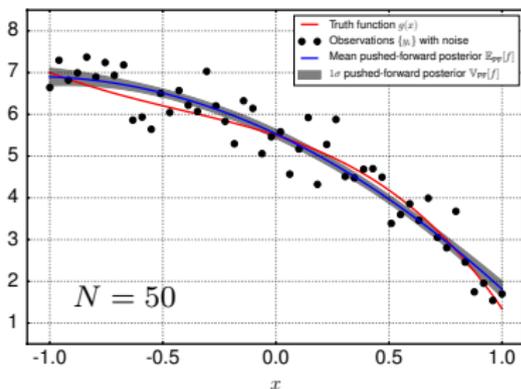
- With additional data, predictive uncertainty around the wrong model is indefinitely reducible
- Predictive uncertainty not indicative of discrepancy from truth
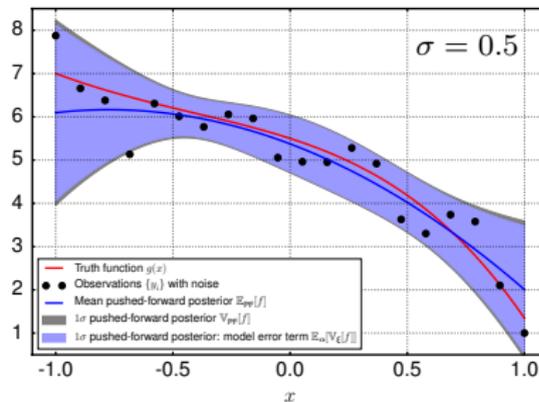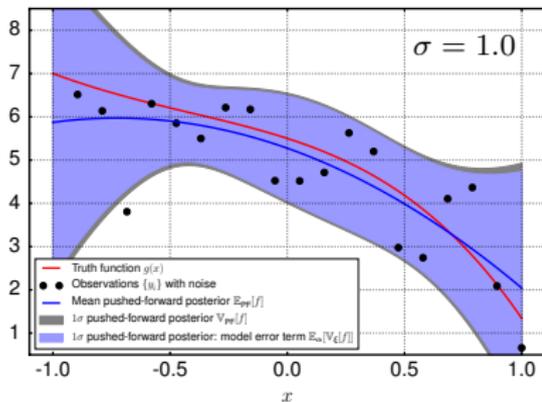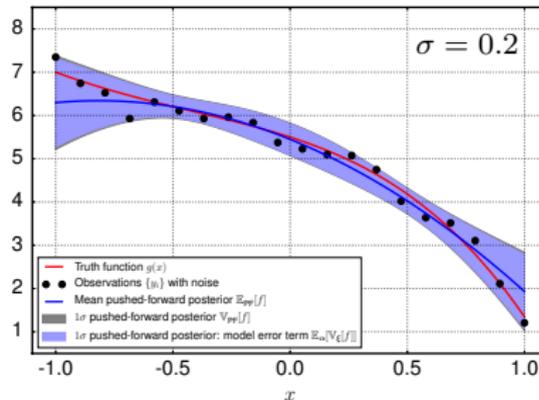
Quadratic-fit – ModErr – MargGauss
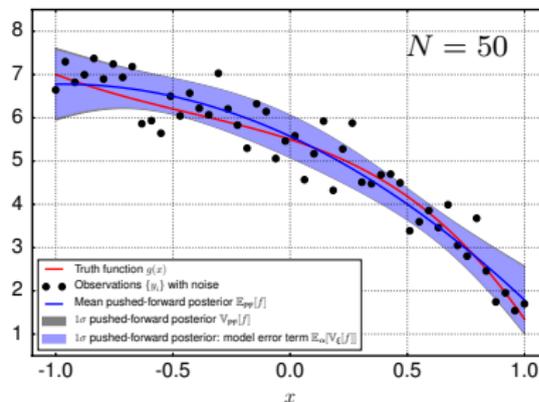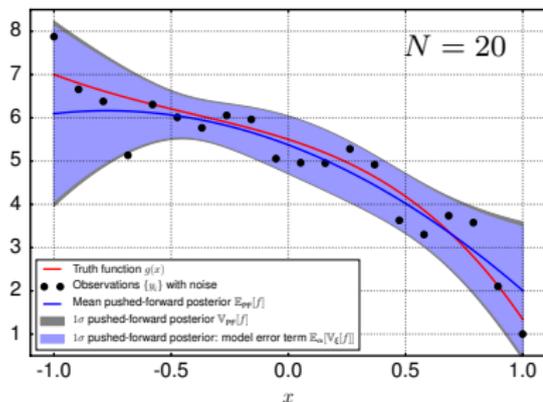
- With additional data, predictive uncertainty due to data noise is reducible

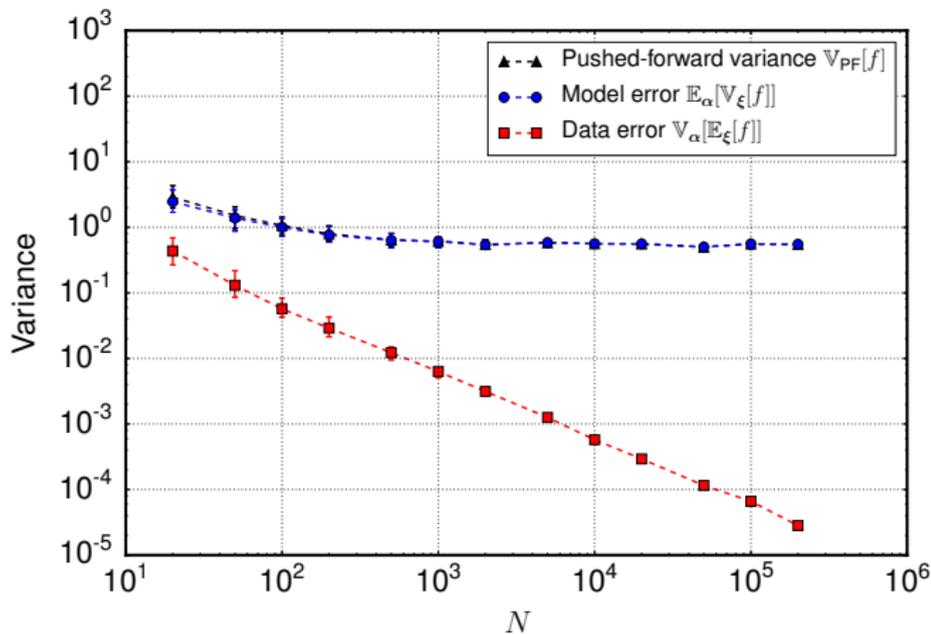- Predictive uncertainty due to model error is not reducible
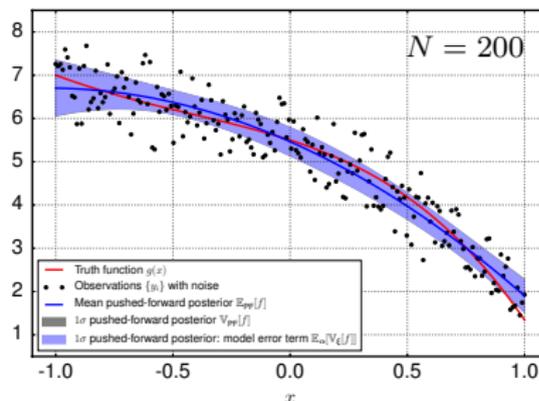
Quadratic-fit – ModErr – MargGauss

- Predictive uncertainty composed of both model-error and data-noise components
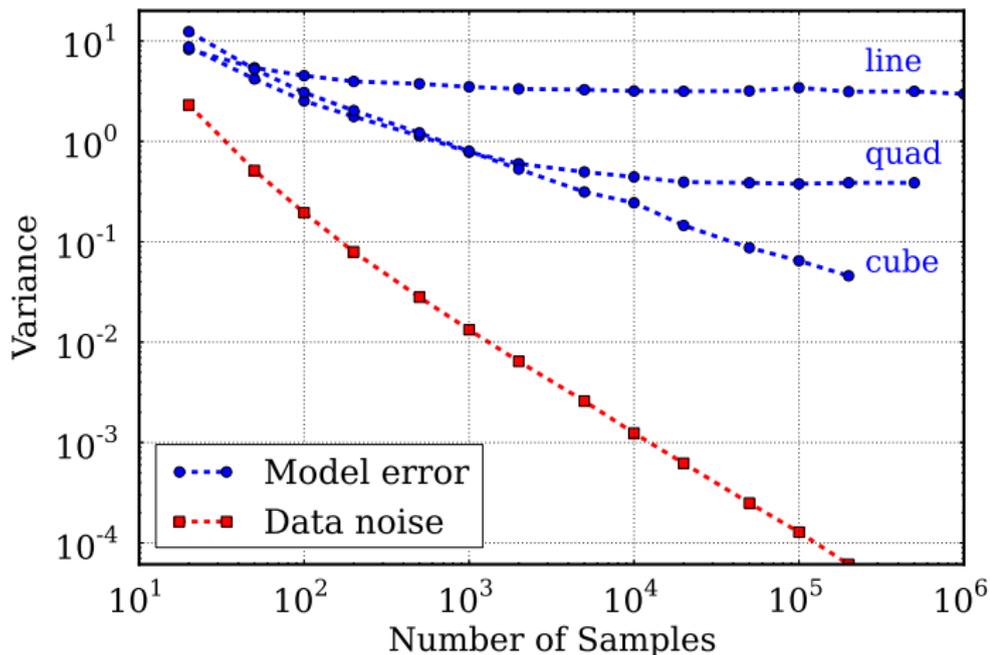- The data-noise component is reducible with lower-noise in the data

Quadratic-fit – ModErr – MargGauss

Calibrating a quadratic $f(x)$ w.r.t. $g(x) = 6 + x^2 + 0.5(x+1)^{3.5}$

Model Error – Fit with Different Models

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
- Uncertainty due to data-noise:
  - Manifested in $\mathbb{V}_\alpha[\mathbb{E}_\xi[f]]$ – Reducible with more/cleaner data
- Uncertainty due to model-error:
  - Manifested in $\mathbb{E}_\alpha[\mathbb{V}_\xi[f]]$ – Not reducible