Embedded Model Error Representation

Khachik Sargsyan, Xun Huan, Habib Najm

Sandia National Laboratories, Livermore, CA

LLNL SciDAC4 Prop Meeting
January 30, 2017

Sandia National Laboratories

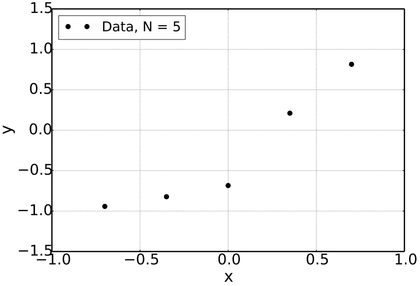
Further acknowledgements:
R. Ghanem(USC), J. Bender(LLNL), Y. Marzouk(MIT), C. Feng(MIT),
M. Eldred (SNL), C. Safta (SNL), B. Debusschere (SNL).

Main target

Model error = deviation from ‘truth’, or from a higher-fidelity model

- Represent and estimate the error associated with
  - Simplifying assumptions, parameterizations
  - Mathematical formulation, theoretical framework
  - Numerical discretization
- ...will be useful for
  - Model validation
  - Model comparison
  - Scientific discovery and model improvement
  - Reliable computational predictions
- Inverse modeling context
  - Given experimental or higher-fidelity model data, estimate the model error

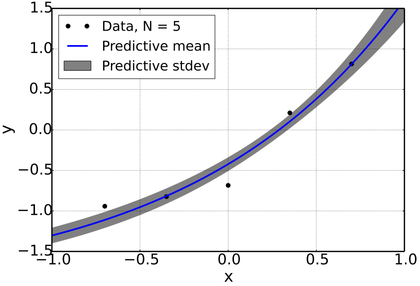
Motivation



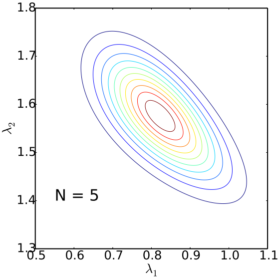
Model-data fit

- Given noisy data – Gaussian noise
- $y = g_{\text{true}}(x) + \epsilon$

Motivation



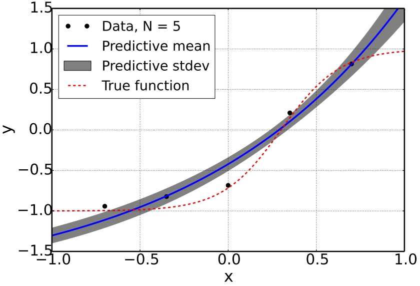
Model-data fit



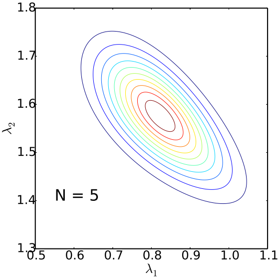
Posterior on parameters

- Employ Bayesian inference to fit an exponential model – $y_{\rm m} = f(x, \lambda)$
- Discrepancy between data and prediction presumed exclusively due to *i.i.d.* Gaussian data noise – $y = f(x, \lambda) + \epsilon_{\rm d}$
- Plotted:
  - Posterior density on the parameters
  - Preditive mean and standard deviation

Motivation



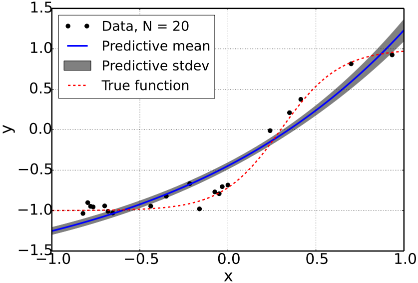
Model-data fit



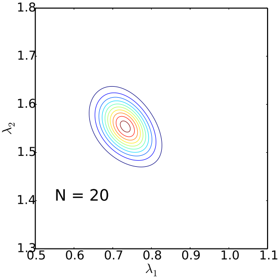
Posterior on parameters

- Employ Bayesian inference to fit an exponential model – $y_{\rm m} = f(x, \lambda)$
- Discrepancy between data and prediction presumed exclusively due to *i.i.d.* Gaussian data noise – $y = f(x, \lambda) + \epsilon_{\rm d}$
- True model $g(x)$ – dashed-red – differs from fit model $f(x, \lambda)$
- Actual discrepancy includes both data and model errors

Motivation



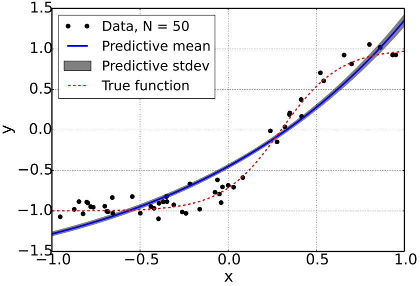
Model-data fit



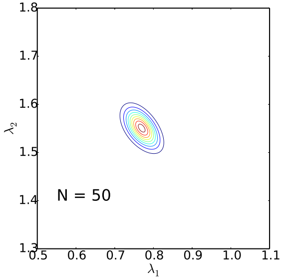
Posterior on parameters

- Increasing number of data points decreases posterior and predictive uncertainty
- We are increasingly sure about predictions based on the *wrong* model

Motivation



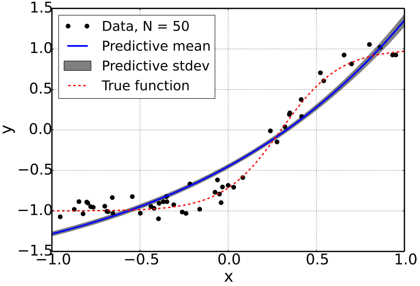
Model-data fit



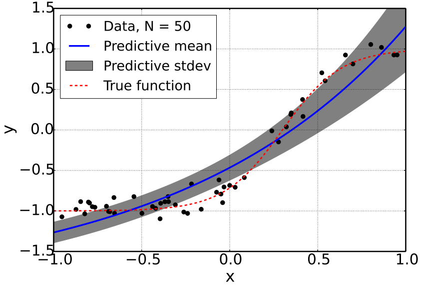
Posterior on parameters

- Increasing number of data points decreases posterior and predictive uncertainty
- We are increasingly sure about predictions based on the *wrong* model

Motivation



Model-data fit



What we want

- If the model has structural uncertainty, more data leads to biased and overconfident results
- We want to quantify model-vs-truth discrepancy in a rigorous and systematic way
    - Cannot ignore model error

Explicit model discrepancy: issues for physical models

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

- Explicit additive statistical model for model error $\delta(x)$ [Kennedy-O'Hagan, 2001]
- Potential violation of physical constraints
- Disambiguation of model error $\delta(x_i)$ and data error ϵ_i
- Calibration of model error on measured observable does not impact the quality of model predictions on other QoIs
- Physical scientists are unlikely to augment their model with a statistical model error term on select outputs
  - Calibrated predictive model: $f(x;\lambda) + \delta(x)$ or $f(x;\lambda)$?
- Problem is highlighted in model-to-model calibration ($\epsilon_i = 0$)
  - no a priori knowledge of the statistical structure of $\delta(x)$

Model error embedding: key idea

Ideally, modelers want predictive *errorbars*:
inserting randomness on the outputs has issues, so...

- Cast input parameters λ as a random variable Λ

  Black-box $$y_i = f(x_i; \Lambda) + \epsilon_i$$

- Generalize parameter forms,

  Random field $$y_i = f(x_i; \Lambda(x_i)) + \epsilon_i$$

- More generally, explore additional parameterizations,

  Extra 'physics' $$y_i = \tilde{f}(x_i; \lambda, \Theta) + \epsilon_i$$

Model error embedding: key idea

Ideally, modelers want predictive *errorbars*:
inserting randomness on the outputs has issues, so...

- Cast input parameters λ as a random variable Λ

  Black-box

  $$y_i = f(x_i; \Lambda) + \epsilon_i$$

  ... even simpler, *additive correction term*

  $$y_i = f(x_i; \lambda + \delta(\alpha, \xi)) + \epsilon_i$$

- Infer the parameters of the correction (α) with the model parameters (λ)
- Stochastic dimension ξ corresponds to model error

Model error embedding: key idea

Augment input parameters λ with additive term $\delta(\alpha, \xi)$

$$y_i = f(x_i; \lambda) + \delta(x_i) + \epsilon_i \longrightarrow y_i = f(x_i; \lambda + \delta(\alpha, \xi)) + \epsilon_i$$

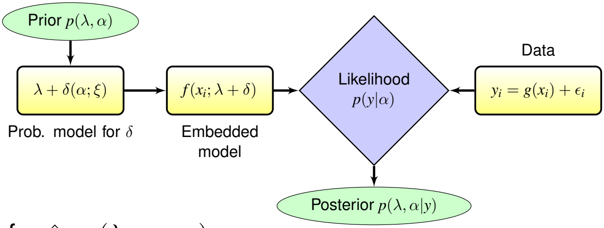
- Embed model error in specific submodel phenomenology
  - a modified transport or constitutive law
  - a modified formulation for a material property
- Allows placement of model error term in locations where key modeling assumptions and approximations are made
  - as a correction or high-order term
  - as a possible alternate phenomenology
- Naturally preserves model structure and physical constraints

Model error embedding – Bayesian density estimation

$$y_i = f(x_i; \lambda + \delta) + \epsilon_i$$

- Parametrize embedded random variable δ:
  - PDF form $\pi_\delta(\cdot;\alpha)$
  - Polynomial Chaos (PC): $\delta = \sum_k \alpha_k \Psi_k(\xi)$
  - Multivariate Normal (MVN): $\begin{cases} \delta_1 = \alpha_{11}\xi_1 \\ \delta_2 = \alpha_{21}\xi_1 + \alpha_{22}\xi_2 \\ \vdots \\ \delta_d = \alpha_{d1}\xi_1 + \alpha_{d2}\xi_2 + \cdots + \alpha_{dd}\xi_d \end{cases}$
- Inverse modeling context
  - Parameter estimation of $\lambda \Rightarrow$ parameter estimation of $\tilde{\alpha} = (\lambda, \alpha)$
  - Bayesian formulation $\underbrace{p(\tilde{\alpha}|y)}_{\text{Posterior}} \propto \underbrace{L_y(\tilde{\alpha})}_{\text{Likelihood}} \underbrace{p(\tilde{\alpha})}_{\text{Prior}}$

Model error embedding – likelihood options



- Infer $\hat{\alpha} = (\lambda, \alpha, \sigma_{\mathcal{D}})$
- Data generation model; to aid likelihood $p(y|\hat{\alpha})$ construction

$$
\begin{aligned}
y_i &= f(x_i, \lambda + \delta(\alpha, \xi)) + \epsilon_i = \\
&= f\left(x_i, \lambda + \sum_k \alpha_k \Psi_k(\xi_1, \ldots, \xi_d)\right) + \sigma_{\mathcal{D}} \xi_{d+i} = \\
\text{[NISP]} &\approx \sum_k f_{ik}(\alpha) \Psi_k(\xi_1, \ldots, \xi_d) + \sigma_{\mathcal{D}} \xi_{d+i} = h_i(\hat{\xi}; \hat{\alpha})
\end{aligned}
$$

- Full PC germ $\hat{\xi} = (\underbrace{\xi_1, \ldots, \xi_d}_{\text{Model error}}, \underbrace{\xi_{d+1}, \ldots, \xi_{d+N}}_{\text{Data noise}})$

Model error embedding – likelihood options

- Data generation model; to aid likelihood $p(y|\hat{\alpha})$ construction

$$
\begin{aligned}
y_i &= f(x_i, \lambda + \delta(\alpha, \xi)) + \epsilon_i = \\
&= f\left(x_i, \lambda + \sum_k \alpha_k \Psi_k(\xi_1, \ldots, \xi_d)\right) + \sigma_{\mathcal{D}} \xi_{d+i} = \\
\text{[NISP]} &\approx \sum_k f_{ik}(\alpha) \Psi_k(\xi_1, \ldots, \xi_d) + \sigma_{\mathcal{D}} \xi_{d+i} = h_i(\hat{\xi}; \hat{\alpha})
\end{aligned}
$$

- Full Likelihood: $L(\hat{\alpha}) = p(y|\hat{\alpha}) = p(y_1, \ldots, y_N|\hat{\alpha}) = \pi(y)$
    - Degenerate if no data noise
    - Requires multivariate KDE or high-d integration
    - Gaussian approximation: $L(\hat{\alpha}) \propto \exp\left(-\frac{1}{2}(y - \mu(\hat{\alpha}))^T \Sigma^{-1}(\hat{\alpha})(y - \mu(\hat{\alpha}))\right)$
    - Non-intrusive spectral projection (NISP) relieves the expense and provides easy access to mean $\mu(\alpha)$ and covariance $\Sigma(\hat{\alpha})$

Model error embedding – likelihood options

- Data generation model; to aid likelihood $p(y|\hat{\alpha})$ construction

$$
\begin{aligned}
y_i &= f(x_i, \lambda + \delta(\alpha, \xi)) + \epsilon_i = \\
&= f\left(x_i, \lambda + \sum_k \alpha_k \Psi_k(\xi_1, \dots, \xi_d)\right) + \sigma_{\mathcal{D}} \xi_{d+i} = \\
\text{[NISP]} &\approx \sum_k f_{ik}(\alpha) \Psi_k(\xi_1, \dots, \xi_d) + \sigma_{\mathcal{D}} \xi_{d+i} = h_i(\hat{\xi}; \hat{\alpha})
\end{aligned}
$$

- Marginalized Likelihood:
  $L(\hat{\alpha}) = p(y|\hat{\alpha}) \approx \prod_{i=1}^N p(y_i|\hat{\alpha}) = \prod_{i=1}^N \pi(y_i)$
  - Requires univariate KDE
  - Neglects built-in correlations
  - Gaussian approximation:
    $L(\hat{\alpha}) \propto \exp\left(-\frac{1}{2} \sum_{i=1}^N \Sigma_{ii}^{-1}(\hat{\alpha})(y_i - \mu_i(\hat{\alpha}))^2\right)$

Model error embedding – likelihood options

- Data generation model; to aid likelihood $p(y|\hat{\alpha})$ construction

$$
\begin{aligned}
y_i &= f(x_i, \lambda + \delta(\alpha, \xi)) + \epsilon_i = \\
&= f\left(x_i, \lambda + \sum_k \alpha_k \Psi_k(\xi_1, \ldots, \xi_d)\right) + \sigma_{\mathcal{D}} \xi_{d+i} = \\
\text{[NISP]} &\approx \sum_k f_{ik}(\alpha) \Psi_k(\xi_1, \ldots, \xi_d) + \sigma_{\mathcal{D}} \xi_{d+i} = h_i(\hat{\xi}; \hat{\alpha})
\end{aligned}
$$

- Approximate Bayesian Computation (ABC): $L(\hat{\alpha}) = \frac{1}{\epsilon} K\left(\frac{\rho(\mathcal{S}_{\mathcal{M}}, \mathcal{S}_{\mathcal{D}})}{\epsilon}\right)$
  - Mean of $f(x_i; \lambda + \delta)$ is “centered” on the data
  - The width of the distribution of $f(x_i; \lambda + \delta)$ is consistent with the spread of the data around the nominal model prediction

$$
L(\hat{\alpha}) \propto \exp\left(-\frac{1}{2\epsilon^2} \sum_{i=1}^{N} \left[(\mu_i(\hat{\alpha}) - y_i)^2 + (\sqrt{\Sigma_{ii}(\hat{\alpha})} - \gamma|\mu_i(\hat{\alpha}) - y_i|)^2\right]\right)
$$

Model Error – Predictions

$$f\left(x;\lambda+\sum_k \alpha_k\Psi_k(\xi_{1:d})\right)=\sum_k f_k(x;\alpha)\Psi_k(\xi_{1:d})$$

- Non-intrusive spectral projection (NISP) will allow
  - Posterior/pushed-forward 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

$$\mathbb{E}[y(x)]=\mathbb{E}_\alpha[\mu(x;\alpha)]$$

- Decomposition of predictive variance

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

Model Error – Predictions at data locations

$$f\,(x_i;\lambda+\sum_k \alpha_k\Psi_k(\xi_{1:d}))+\sigma_D\xi_{i+d}=\sum_k f_k(x_i;\alpha)\Psi_k(\xi_{1:d})+\sigma_D\xi_{i+d}$$

- Non-intrusive spectral projection (NISP) will allow
  - Likelihood computation
  - Easy access to first two moments:

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

- Predictive mean

$$\mathbb{E}[y(x_i)]=\mathbb{E}_\alpha[\mu(x_i;\alpha)]$$

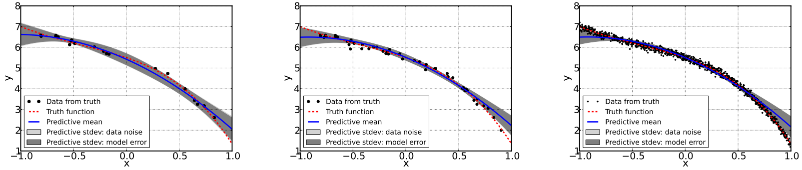
- Decomposition of predictive variance

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

More data leads to ‘leftover’ model error

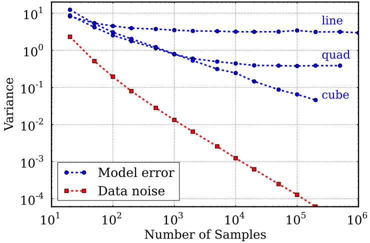
Calibrating a quadratic $f(x) = \lambda_0 + \lambda_1 x + \lambda_2 x^2$
w.r.t. ‘truth’ $g(x) = 6 + x^2 - 0.5(x+1)^{3.5}$ measured with noise $\sigma = 0.1$.

$N = 20$ $N = 50$ $N = 1000$



Summary of features:

- Well-defined model-to-model calibration
- Model-driven discrepancy correlations
- Respects physical constraints
- Disambiguates model and data errors
- Calibrated predictions of multiple QoIs



Challenges

- Density estimation is more challenging than parameter estimation
    - Inverse problem is ill-posed or intractable

  $\Rightarrow$ Employ approximate or empirical likelihoods

- Potentially a high-dimensional Bayesian problem
    - Full posterior may be inaccessible...

  $\Rightarrow$ Resort to optimization algorithms in no-noise case
    - ... or hard to sample from

  $\Rightarrow$ Adaptive MCMC algorithms, Likelihood-informed subspaces

- Sparse data or expensive high-fidelity simulations
    - With low information content, calibration may struggle

  $\Rightarrow$ More informative priors/regularization

Summary

- Represent, quantify and propagate physical model errors
- Parameter estimation $\Rightarrow$ density estimation
- Bayesian machinery to find parameters of the PDFs
- Approximate/empirical likelihoods impose constraints of interest
- Differentiates from data noise; allows model-to-model calibration
- Implemented in UQTk (www.sandia.gov/UQToolkit)
- Applied to chemical kinetics, transport modeling, LES computations...

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

- Optimal design for maximum information
- Optimal embedding
- Bayesian problem still hard; MCMC, priors, ...
- Hierarchical Bayesian viewpoint
- More intrusive embedding; problem specific

Applications

Model-to-model calibration: ignition model

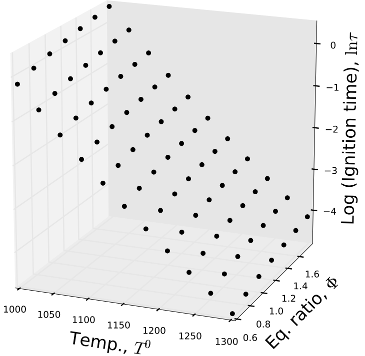
- Homogeneous ignition, methane-air mixture
- Single-step global reaction model calibrated against a detailed chemical kinetic model

- Data: ignition time; range of initial T & equivalence ratio
- Single-step model:

$$\mathrm{CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O}$$

$$\begin{aligned} \mathfrak{R} &= [\mathrm{CH_4}][\mathrm{O_2}]k_f \\ k_f &= A\exp(-E/R^oT) \end{aligned}$$

- $(\ln A, E) = \sum_k \boldsymbol{\alpha}_k \Psi_k(\xi)$

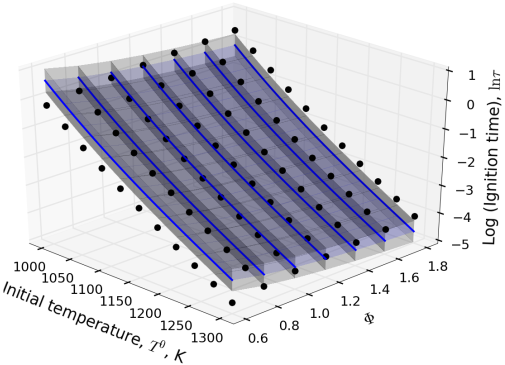


Model-to-model calibration: ignition model

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



K. Sargsyan, H.N. Najm, and R. Ghanem
"On the Statistical Calibration of Physical Models"
Int. J. Chem. Kin., 47(4): 246-276, 2015

TransCom3 Experiment of CO_2 Flux Inversion

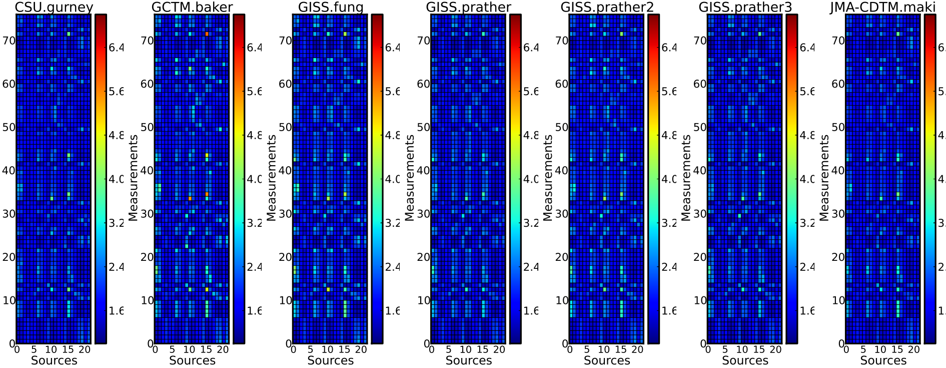
[Gurney *et al.*, Tellus B, 2003]

- Observations $\mathbf{d}$ at $N = 77$ sites around the world
- Inverse problem: find fluxes $\mathbf{s}$ at $M = 22$ locations
- Linearized 'response' model $\mathbf{R}$, such that $\mathbf{d} \approx \mathbf{R}\mathbf{s}$

$$\mathbf{d} = \mathbf{R}\mathbf{s} + \epsilon_{\mathbf{d}}$$

- Model $\mathbf{R}$ is never perfect thus contaminating the inversion
- The inferred values of $\mathbf{s}$ compensate for model deficiencies
- $\epsilon_{\mathbf{d}}$ is meant to capture data errors, but is 'entangled' with model errors

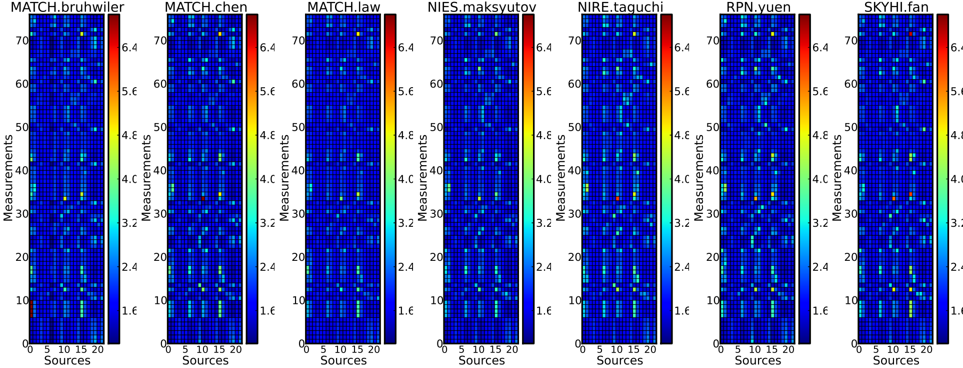
Consider 14 different response models $\mathbf{R}$



Infer fluxes $\mathbf{s}$, given measurements $\mathbf{d}$ to satisfy $\mathbf{d} \approx \mathbf{R}\mathbf{s}$

- Conventional additive Gaussian error (least-squares): $\mathbf{d} = \mathbf{R}\mathbf{s} + \xi$
- Embed probabilistic model for fluxes $\boldsymbol{s}$: $\mathbf{d} = \mathbf{R}(\mu_\mathbf{s} + \mathbf{C}_\mathbf{s}\xi)$

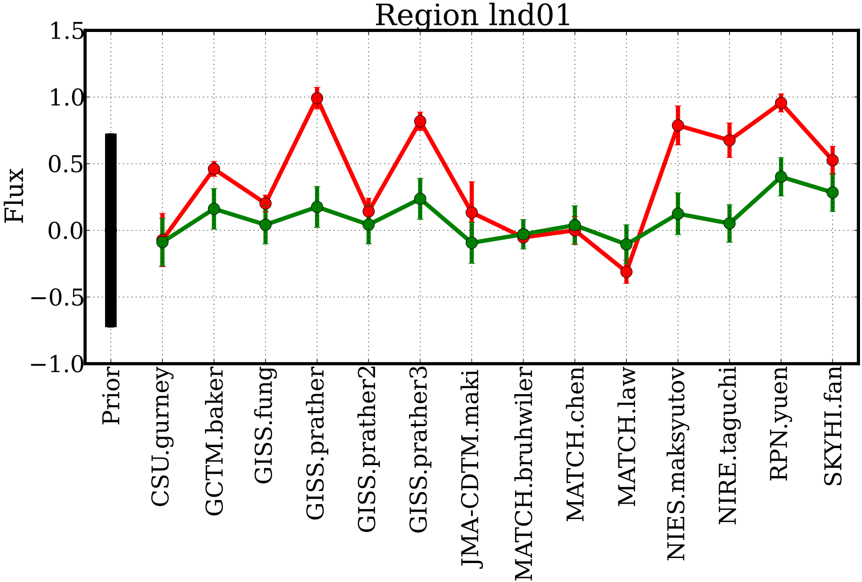
Consider 14 different response models $\mathbf{R}$



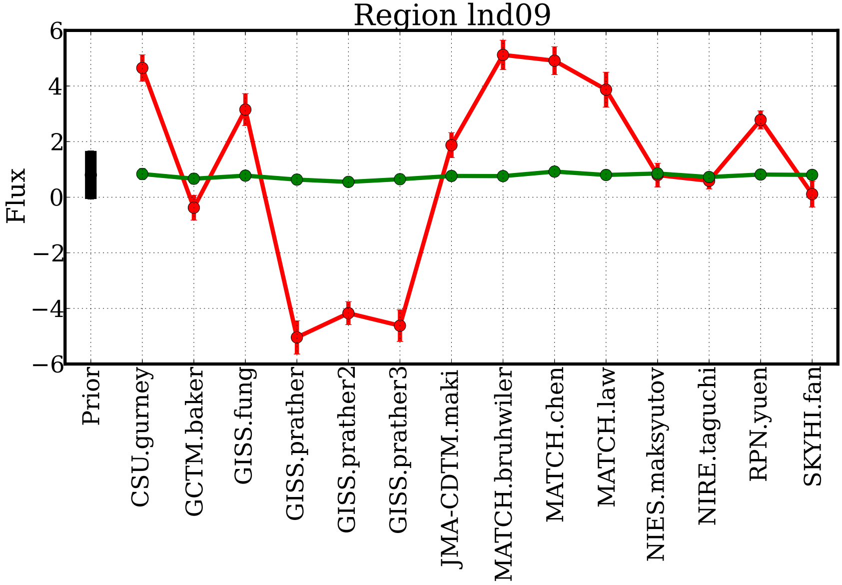
Infer fluxes $\mathbf{s}$, given measurements $\mathbf{d}$ to satisfy $\mathbf{d} \approx \mathbf{R}\mathbf{s}$

- Conventional additive Gaussian error (least-squares): $\mathbf{d} = \mathbf{R}\mathbf{s} + \xi$
- Embed probabilistic model for fluxes s: $\mathbf{d} = \mathbf{R}(\mu_\mathbf{s} + \mathbf{C}_\mathbf{s}\xi)$

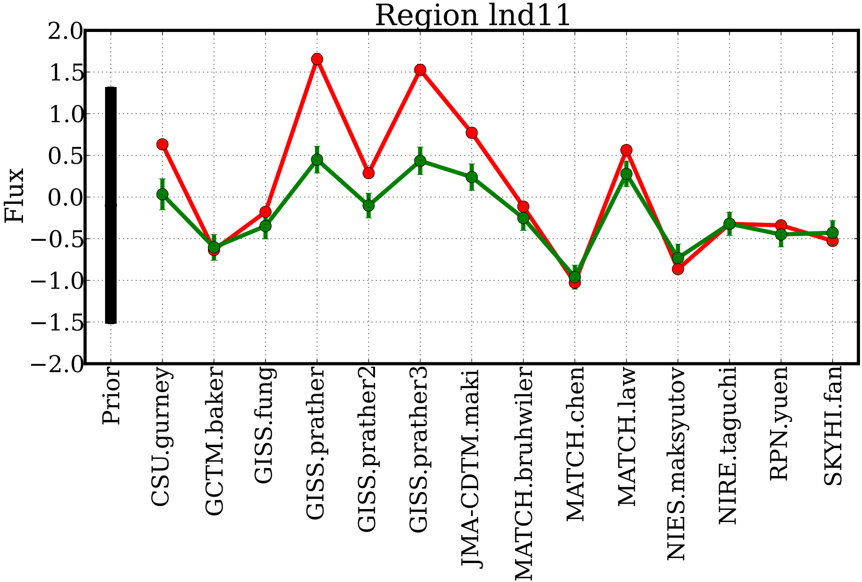
Inferred fluxes show less variability across models



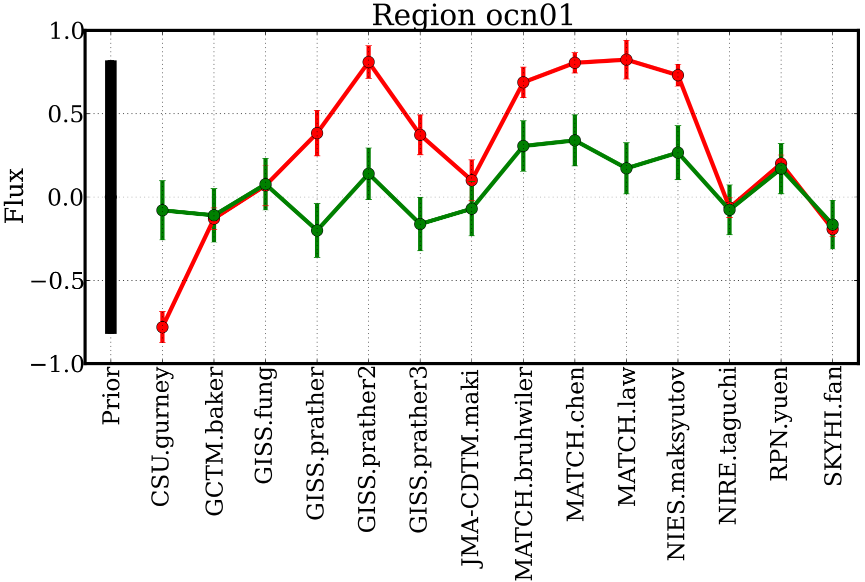
Inferred fluxes show less variability across models



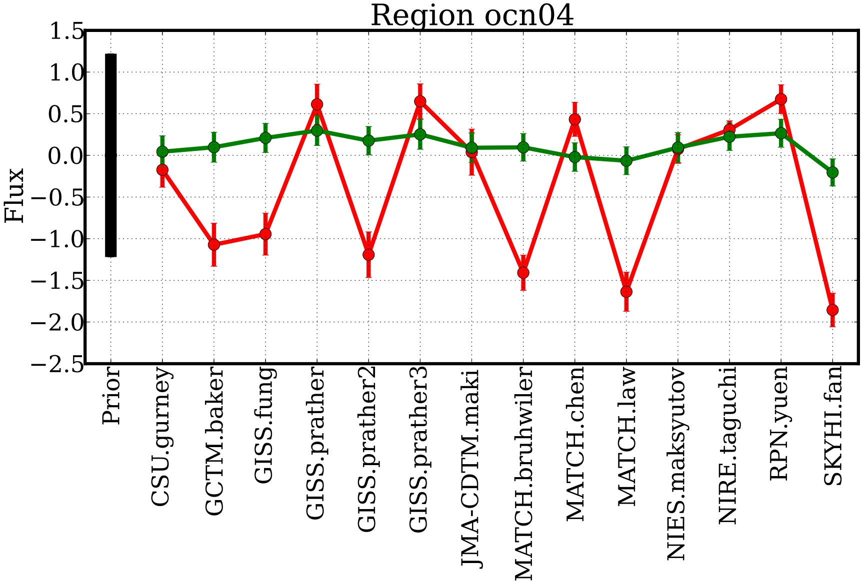
Inferred fluxes show less variability across models



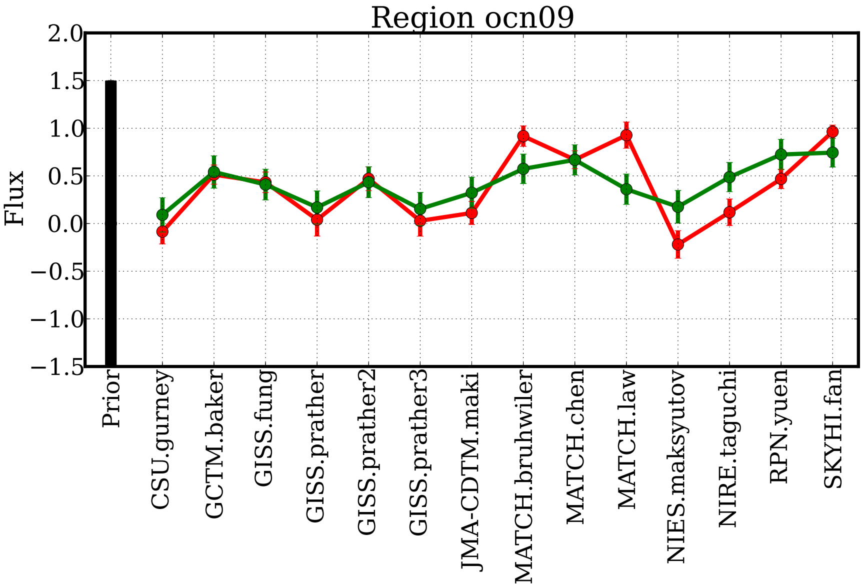
Inferred fluxes show less variability across models



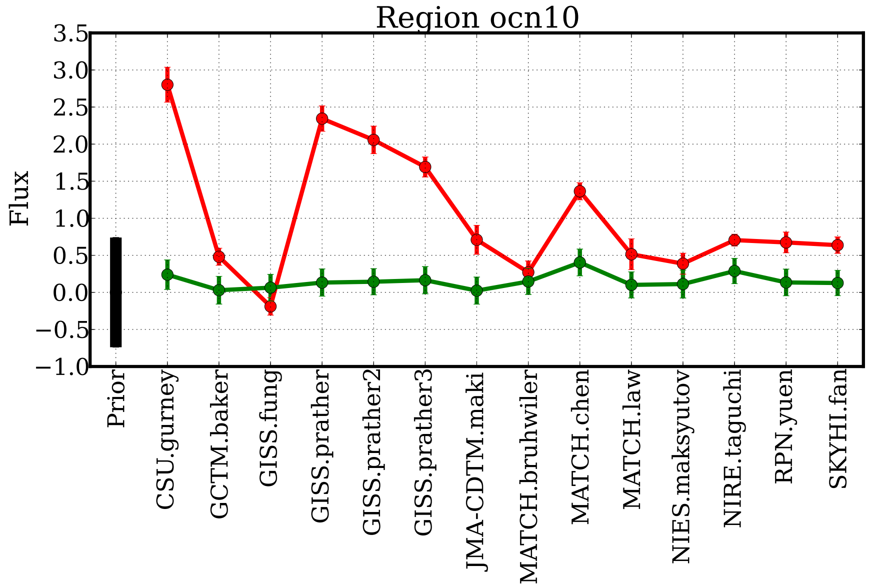
Inferred fluxes show less variability across models



Inferred fluxes show less variability across models



Inferred fluxes show less variability across models



LES computation in Scramjet engine: static-vs-dynamic SGS model calibration

Calibrate with Turbulent Kinetic Energy (TKE) data, predict both TKE and Pressure